Dissertation

Microphysics and dynamics retrievals from dual-polarization radar for very short-term forecasting

Submitted by

Renzo Bechini

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2016

Doctoral Committee:

Advisor: V. Chandrasekar

Anura Jayasumana
Paul Mielke
Juanzhen Sun

Abstract

Microphysics and dynamics retrievals from dual-polarization radar for very short-term forecasting

Nowcasting is primarily a description of the near-future forecasted atmospheric state, therefore relying heavily on observations. Besides routine meteorological observations (pressure, temperature, humidity, wind), dual-polarization weather radar provides a large amount of useful information due to the frequent-update (∼5 min) and high-resolution (∼500 m) three-dimensional sampling of the atmosphere. However, the atmospheric state variables are not readily invertible from radar remote observations, resulting in complexity in the numerical model data assimilation. This problem is normally dealt with by defining observation operators to simulate the radar variables from the model state vector.

In this work the dual-polarization radar based retrievals are developed in order to demonstrate their potential for microphysics and dynamics retrievals. In particular the analysis of radar observations in convective storms and in stratiform ice clouds revealed that specific dual-polarization signatures can be successfully related to important dynamic properties such as vertical air motions, both in convective precipitation (strong updrafts, several $\mathrm{m\,s^{-1}}$) and in stratiform precipitation (large areas of weak updrafts, tenths of $\mathrm{m\,s^{-1}}$, associated with mid-tropospheric mesoscale forcing).

Given the relevance of polarimetric signatures to dynamics retrievals, an improved hydrometeor classification method is developed based on a learn-from-data approach. In this technique, the traditional bin-based classification is replaced with a semi-supervised approach which combines cluster analysis, spatial contiguity, and statistical inference to assign the most likely class to a set of identified connected regions. The hydrometeor classification

and relevant dual-polarization signatures establish a starting point to explore new means to improve the analysis of precipitation and near-surface winds, and their subsequent nowcasting. In particular the relevance of a well-known dual-polarization feature associated with deep convection (vertical columns of differential reflectivity) is illustrated by including the microphysics and dynamics-related information into a simple method for the analysis of surface winds.

The goal of a physically consistent analysis is further pursued considering the Variational Doppler Radar Analysis System (VDRAS), an advanced four-dimensional data assimilation system based on a cloud-scale model, specifically designed for ingesting Doppler weather radar observations. The typical application using single-polarization observations from long-range S-band or C-band radars is here extended to high frequency (X-band), short range radars and dual-polarization observations. The combination of the hydrometeor classification and dual-polarization rainwater estimation allows to successfully assimilating the X-band observations, otherwise prone to relevant errors when using the reflectivity-based observation operator widely employed in numerical models. The feasibility of X-band data assimilation to contribute building a consistent analysis for nowcasting is demonstrated over the Dalls-Fort Worth test bed, where a dense network of dual-polarization X-band radars is being deployed. Eventually, a novel method for the nowcasting of precipitation and winds is built upon the VDRAS analysis, in an attempt to combine the robustness and consistency of data assimilation and the efficacy of extrapolation techniques for very short-term forecasting.

Acknowledgements

Second chances are rare in real life. This is why I'm especially grateful to Prof. V. Chandrasekar for giving me the opportunity and support to go back to the University after many years, and do research in the extraordinary environment of Colorado State University. I am thankful for the motivation, inspiration, and the many discussions held in the most diverse places (and moments) on both sides of the Atlantic Ocean. I also express my sincere gratitude to Profs. Anura Jayasumana, Paul Mielke, Branislav Notaros, and Dr. Juanzhen Sun for serving on my graduate committee. I am especially grateful to Dr. Juanzhen Sun for the support and fruitful discussions on the use of the VDRAS model for radar data assimilation (Chapter 6).

I thank all my colleagues and friends in the radar group for their support, stimulating discussions, and for the good times inside and outside the lab. I acknowledge the contribution of Dr. Luca Baldini at ISAC-CNR (Roma, Italy) for performing the electromagnetic scattering simulations in Section 3.2.2. Last but not the least, I would like to thank my family: my parents, whose encouragement for this unexpected *adventure* was just the same as when I was a teenager in high school. My wife Laura, for her unwavering faith in my talents and very special support for all those years. And our kids Sofia and Diego, who may someday look back on this and find it all funny.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

According to the Word Meteorological Organization (WMO) nowcasting is defined as a "description of current weather parameters and 0-2 hours description of forecasted weather parameters". The U.S. National Weather Service specifies zero to three hours, though up to six hours are often used in many instances. The nowcasting and the subsequent very short-term (up to 12 hours) forecasting ranges have well distinguished challenges, as compared to more traditional short-range (12-72 hours) and medium-range (72-240 hours) weather forecast, which are normally completely realized in the frame of numerical modeling.

Although nowcasting refers to the general state of the weather, particular emphasis has historically been devoted to nowcasting of precipitation. This is justified by the large impact for public safety of thunderstorms, floods and hurricanes. However, high-impact weather also includes strong winds (often associated with thunderstorms, tornadoes) and icing. In particular the nowcasting of winds has recently become the object of relevant investigation [8]. In addition to the benefits of timely warnings for potential hazards related to strong winds, the nowcasting of winds is gaining increasing interest for application in the wind energy industry, aviation and numerical dispersion modeling.

Since nowcasting is primarily a description of the current atmospheric state, observations play a fundamental role to derive an analysis resolving the small scales of the variables of interest. As a matter of fact the simplest nowcasting method can be considered the Eulerian persistence, where the future state (time $t = t_0 + dt$) of an atmospheric-related variable $\mathbf{\Psi}$ (e.g. the precipitation or the radar reflectivity) is assumed to be the same as that provided by the observations at time t_0:

$$\widehat{\mathbf{\Psi}}(t_0 + dt, \mathbf{x}) = \mathbf{\Psi}(t_0, \mathbf{x}) \tag{1}$$

Application of Eulerian persistence to the nowcasting of severe storms has obvious limitations, since rarely are thunderstorms stationary. Indeed storm cells often move with mid-tropospheric winds. Lagrangian persistence is then in general more appropriate, being likely the most effective and intuitive approach for the 0-30 minutes range forecasting of precipitation. It is intuitive, as basically relying on the same integral we all attempt to mentally solve when crossing a busy street and seizing the speed of an approaching car. Mathematically:

$$\widehat{\mathbf{\Psi}}(t_0 + dt, \mathbf{x}) = \mathbf{\Psi}(t_0, \mathbf{x} \text{ - } \mathbf{u}) \tag{2}$$

where $\mathbf{u}$ represents an estimate of the displacement (motion) vector. So the problem reduces to the calculation of the motion vectors, which is often accomplished through image processing of a sequence of past radar (or satellite) maps. The motion vectors are fixed through the forecasting period ($\frac{\partial \mathbf{u}}{\partial t} = 0$). Although in some instances the convective storms are fairly regularly advected for long time periods, in most cases the complex evolution of the storms and the interactions with the surrounding environment limit the validity of the extrapolation to no more than 20-30 minutes.

1.1. Problem statement

Numerical Weather Prediction (NWP) models resolve the larger and slower evolving scales of the atmosphere dynamics and are in general not able to beat the extrapolation-based techniques during the first 2-3 hours. Fig. 1.1 shows a schematic representation of the

forecast skill of extrapolation techniques and NWP models for increasing lead times. The theoretical limit (solid line) results from the chaotic nature of the atmosphere. The downward trend reflects the fact that in a dynamical system highly sensitive to initial conditions like the atmosphere, the unavoidable errors in the observations will grow with time into the forecast. In the late sixties Lorenz [9] predicted the finite limit of atmospheric predictability to be about two weeks.

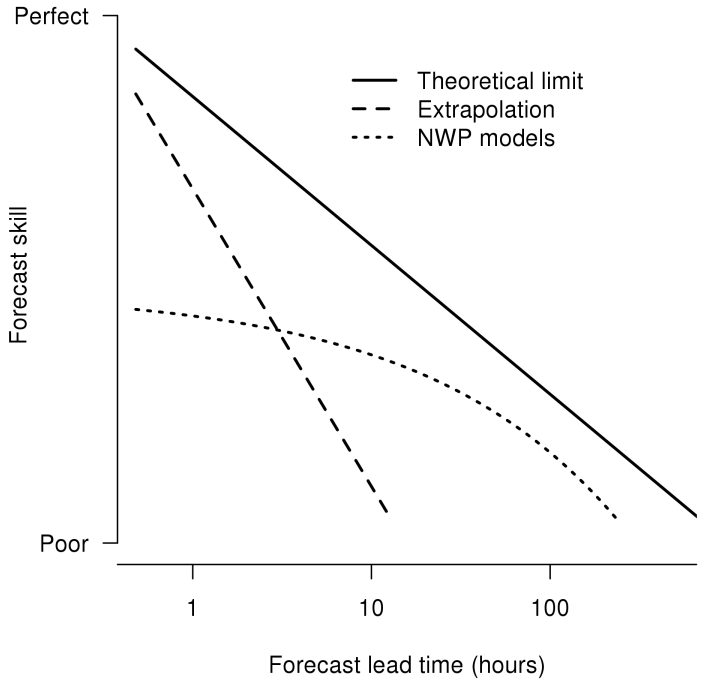


Figure 1.1: Qualitative representation of the forecast skill for extrapolation methods (dashed line) and NWP models (dotted line). The solid line indicates the theoretical limit of predictability. Adapted from [1].

The marked negative slope of the line representing extrapolation is a consequence of the lack of atmospheric physics in the method, which causes a rapid loss of the valuable initial information (observations). In the ∼0.5-3 hours forecast range the extrapolation methods have lost much of their skill, but are in general still more accurate than NWP models. The

assimilation in NWP models of dense and frequent observations as those provided by weather radar offers the possibility to improve the forecast skill, especially at very short-range.

1.2. Research objectives

The main scientific objective of this research is to explore the potential of dual-polarization observations from radars operating at different frequencies (S, C, and X band) for improving the short-term forecasting of precipitation and winds. Within this general goal, specific research objectives are devised:

(a) characterization of dual-polarization radar microphysical signatures relevant to the precipitation system dynamics;

(b) development of a hydrometeor classification algorithm applicable to both conventional long range systems (S and C band), and more attenuation-prone X-band radars;

(c) use of dual-polarization radar observations for assimilation into numerical models;

(d) development of a hybrid nowcasting technique based on data assimilation and extrapolation methods.

1.3. Organization of the dissertation

The dissertation is organized as follows. Chapter 2 introduces the radar data assimilation concepts, with a specific illustration of a popular variational assimilation scheme for weather radar observations. The overview is accompanied by a real case application with C-band and X-band data, and evaluation of the 0-3 hours forecast performance. For this and other similar stratiform precipitation events, the dual-polarization observations in the ice region of the clouds are analyzed in detail in Chapter 3, with emphasis on ice growth mechanisms and the correlation with precipitation intensity near the ground. In Chapter

4 the development of a novel hydrometeor classification scheme applicable to radars operating at attenuating frequencies is illustrated. Chapter 5 focuses on deep convection and the relevance of a specific dual-polarization signature (columns of Z_{dr}) for improving the analysis of surface winds. A more sophisticated method for the meteorological analysis of convective storms, with special emphasis on the low-level wind field, is examined in Chapter 6. In this chapter a four-dimensional data assimilation system is employed to elaborate a physically consistent analysis, based on dual-polarization observations over the Dallas-Fort Worth testbed including X-band systems. Chapter 7 focuses on a possible realization of a hybrid nowcasting system which relies on both the outcomes of data assimilation for the analysis and on a simple parametric model for the subsequent short-term forecast. Finally Chapter 8 summarizes the main results of the thesis work and illustrates the expected future developments.

CHAPTER 2

Background: data assimilation for short-term forecasting

Data assimilation is the technique by which observations are combined with a numerical model (with their respective errors) in order to provide an improved analysis of the atmospheric state at a given time. Although the origin of data assimilation for atmospheric models can be dated back to the 1950s [10], the use of weather radar data in numerical models is only a relatively new field of research.

During the last twenty years the techniques for radar data assimilation have consistently progressed, going from direct insertion of single-Doppler wind retrievals to the much more sophisticated Ensemble Kalman Filter (EnKF) initialization [11]. Most of the recent research on radar data assimilation however have focused on variational techniques [7, 8, 12, 13].

Variational data assimilation looks for an accurate analysis of the state of the atmosphere through the iterative minimization of a prescribed cost (or penalty) function. Differences between the analysis and the observations, and between the analysis and the first guess are penalized (damped) according to their perceived error. The most common implementation is the three-dimensional variational assimilation (3D-Var). In 3D-Var all observations in a given time window ($t_0 - dt$ to $t_0 + dt$, where t_0 is the analysis time) are treated if they were all taken at time t_0. Given a first guess (background) the variational minimization searches for the analysis that minimizes the penalty measuring the deviations from the observations and from the background state at t_0. Fig. 2.1 illustrates the basic steps of the 3D-Var analysis, while further discussion on the cost function and its minimization is reported in section 2.1.

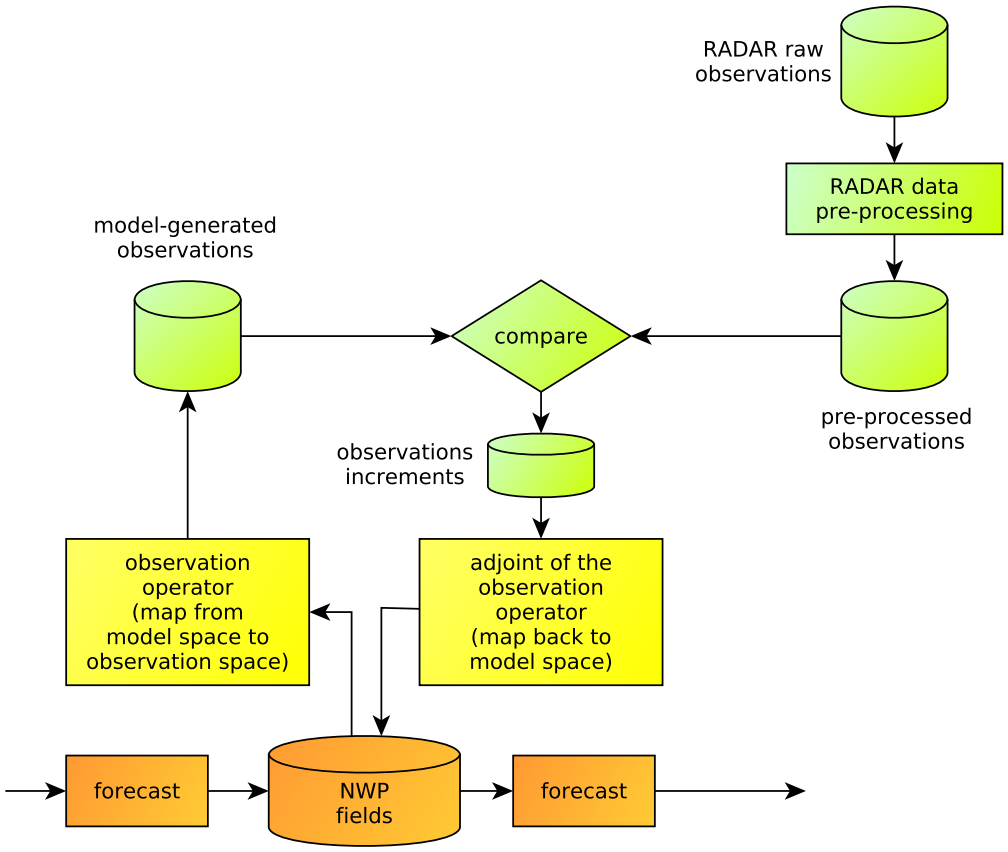


Figure 2.1: A flow diagram illustrating the 3D-Var analysis steps for radar data assimilation. Adapted from [2].

In 4D-Var (time is the fourth dimension) the model trajectory that best fits the available observations over a period of time ($t_0 - dt$ to t_0) is calculated. The minimization process is realized evaluating the gradient of a cost function with the adjoint model [10]. In this way 4D-Var searches for an optimal initial state of the model which minimizes the discrepancies between the model forecast and the observational data distributed over the assimilation window. The difference between 3D-Var and 4D-Var data assimilation is therefore the use of a numerical atmospheric model as a constraint in the latter.

The disadvantage of 4D-Var respect to 3D-Var is the increased computational cost, implied by the additional iteration of the numerical model and its adjoint over the assimilation

time window. Huang et al. [14] for example reported an increased computational time for the HIRLAM model between 7 and 30 times depending on the physics schemes. A common practical approach to avoid the computational intensive 4D-Var but still manage to take advantage of the observations at correct times is to run 3D-Var in a frequent-updating cycling mode. This is further illustrated with a real case experiment in section 2.2.

2.1. WRF 3D-Var

The Weather Research and Forecasting (WRF) model [15] is a non-hydrostatic, community-supported NWP model mainly developed by the National Center for Atmospheric Research (NCAR). In this study the Advanced Research WRF (ARW) solver with associated data assimilation package (WRF-Var) is used (version 3.4). In particular the WRF 3D-Var system [16, 17] is adopted for the assimilation of both surface and radar observations.

The current WRF 3D-Var implementation, as most variational assimilation systems, provides an analysis $\mathbf{x} = \mathbf{x}_a$ through the minimization of the prescribed cost function $J(\mathbf{x})$:

$$J(\mathbf{x}) = J_b + J_o = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - \mathbf{y}_o)^T \mathbf{O}^{-1}(\mathbf{y} - \mathbf{y}_o) \tag{3}$$

where:

$\mathbf{x}$ = analysis state vector

$\mathbf{x}_b$ = background vector

$\mathbf{B}$ = background error covariance matrix

$\mathbf{y}_o$ = observation vector

$\mathbf{y} = H(\mathbf{x})$

$\mathbf{O}$ = observation error covariance matrix

H is the (possibly non linear) forward observation operator and maps the model space

to the observation space. Eq. (3) simply states that the analysis $\mathbf{x}_a$ is the atmospheric state resulting from the minimization of its distance from the background and from the observations, weighted by their respective errors.

The background error covariance matrix $\mathbf{B}$ can be estimated from differences between the forecast and observations. However, in practice the "NMC method" [18] has been widely used to estimate the model error statistics. The method considers the difference between two different forecasts (e.g. +24 h and +12 h) valid at the same time. Specifically, the matrix $\mathbf{B}$ is obtained from:

$$\mathbf{B} \approx E\{[\mathbf{x}_f(24h) - \mathbf{x}_f(12h)][\mathbf{x}_f(24h) - \mathbf{x}_f(12h)]^T\} \tag{4}$$

where the climatological estimate in (4) is typically realized considering a period of time of at least several weeks. If n are the degrees of freedom of the model state vector $\mathbf{x}$, the calculation of the J_b term in (3) requires $\sim O(n^2)$ calculations. For a typical model domain with NX=NY=200 and NZ=50 grid points and 10 variables, n is of the order of 10^7, so the required 10^{14} calculations prevents a direct solution for the background term. In order to reduce the complexity of the problem a common solution is the introduction of the control variable transform, i.e. the definition of a control variable $\mathbf{v}$ such that $\mathbf{x}' = \mathbf{U}\mathbf{v}$, where $\mathbf{x}' = \mathbf{x} - \mathbf{x}_b$ is the analysis increment. The proper choice of the control variables ensures that the covariance between these variables is minimum, and $\mathbf{U}\mathbf{U}^T$ closely matches the full error covariance matrix $\mathbf{B}$. The matrix $\mathbf{U}$ is normally implemented with three distinct operations: $\mathbf{x}' = \mathbf{U}_p\mathbf{U}_h\mathbf{U}_v\mathbf{v}$. The horizontal ($\mathbf{U}_h$) and vertical ($\mathbf{U}_v$) transform are performed using recursive filters and empirical orthogonal function (EOF) decomposition respectively (see [16] for more details). The term $\mathbf{U}_p$ represents the physical variable transform for conversion of the control variables (stream function ψ, unbalanced part of velocity potential χ_u, unbalanced

part of temperature T_u, relative humidity q, unbalanced part of surface pressure p_{s_u}) to the model variables (u, v, T, p_s, q), taking into account the statistical balances. With this choice of control variables, the off-diagonal elements of the background error matrix are very small.

After introducing the innovation vector ($\mathbf{y}'_o = \mathbf{y}_o - H(\mathbf{x}_b)$) and the tangent linear operator $\mathbf{H}$, with elements $h_{i,j} = \dfrac{\partial H_i}{\partial x_j}$, the control variable transform allows to rewrite (3) in terms of analysis increments:

$$J(\mathbf{v}) = J_b + J_o = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}(\mathbf{y}'_o - \mathbf{HUv})^T\mathbf{O}^{-1}(\mathbf{y}'_o - \mathbf{HUv}) \tag{5}$$

The observation operator (H in eq. (3)) for the reflectivity (Z_h) and radial velocity (V_r) are [7]:

$$Z_{h,i} = 43.1 + 17.5\log(\rho\, q_{r,i}) \tag{6}$$

$$V_{r,i} = u\frac{x - x_i}{r_i} + v\frac{y - y_i}{r_i} + (w - V_t)\frac{z - z_i}{r_i} \tag{7}$$

where Z_h is expressed in dBZ and the rainwater mixing ratio q_r in $\mathrm{g\,kg^{-1}}$, r_i is the distance between a model grid point (x, y, z) and the ith radar location (x_i, y_i, z_i), V_t is the terminal fall velocity of the precipitation (expressed as a function of the rainwater mixing ratio [7]). The observation operator for the radial velocity is merely a spatial interpolation of the model wind vector components. On the other hand the reflectivity is not a model state variable and need to be calculated from the rainwater mixing ratio using eq. (6), obtained assuming a Marshall-Palmer drop size distribution.

For the minimization of the cost function in the analysis control variable space (5) the conjugate gradient method is utilized.

2.2. A 3D-Var assimilation experiment over northwestern Italy: widespread precipitation

The WRF 3D-Var assimilation system is used in the simulation of an intense and widespread precipitation event occurred in northwestern Italy at the end of April 2009. A frontal trough extending over the eastern Mediterranean Sea approached the western Alps on April 26. On April 27 the trough deepened over the central Mediterranean and a low-level pressure minimum formed in the Gulf of Genova. A moist Southerly flow established in the mid and upper troposphere (Fig. 2.2) while Easterly winds interested the lower levels. This configuration led to large rainfall accumulation both in the northern part of the region and in the southwestern sector as a consequence of the low-level Easterly flow.

With the aim of assessing the value for the assimilation of the Doppler radar observations collected by systems operating at different frequencies, the radars managed by Arpa Piemonte (Environmental Protection Agency of Piemonte) are considered. Arpa Piemonte operates a dual-polarization radar network composed of two operational C-band systems [19] and one research transportable X-band radar in North-western Italy. The measurements collected by the ARX (ARpa X-band) radar and from one of the C-band systems, the Bric della Croce radar (hereafter Bric), are used for these assimilation experiments (fig. 2.3).

The Bric radar is located at 736 m MSL on the top of the Torino hill. The X-band was deployed since 2008 at different locations for specific measurement campaigns. In 2009 (data analyzed in this chapter) the radar was deployed in Carmagnola, 235 m MSL, 16 km South of Torino. The main technical characteristics of the systems are listed in Table 4.3. The radars are dual-polarized, but the current version of WRF 3D-Var only allows assimilation of Doppler radial winds and reflectivity. The polarimetic capabilities are however exploited

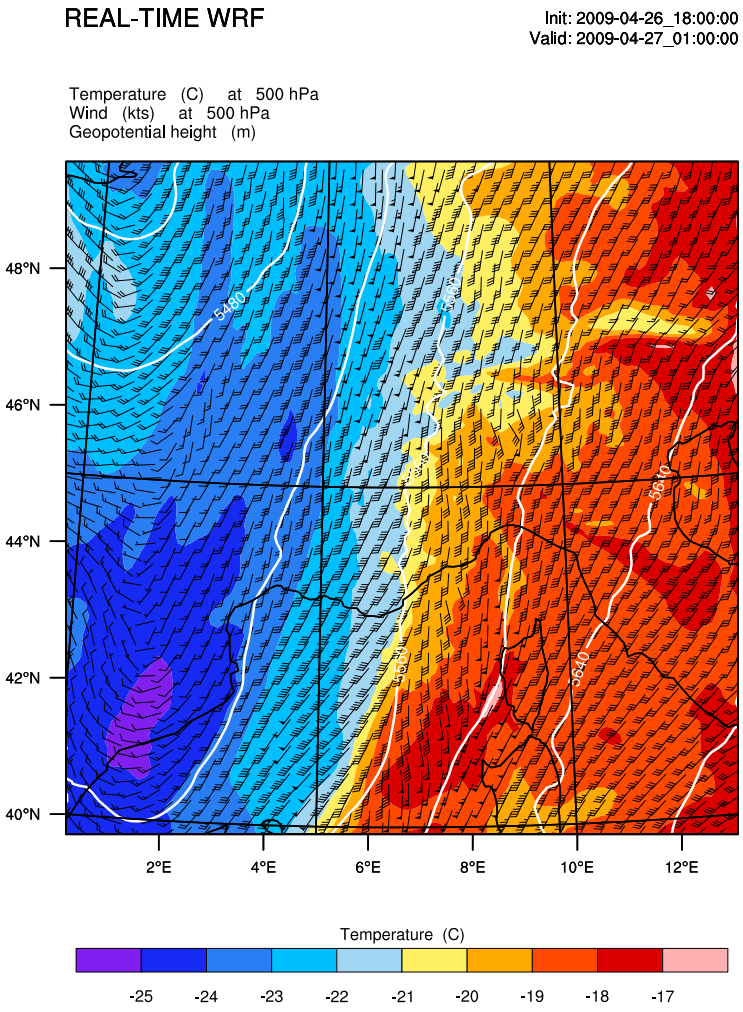


Figure 2.2: WRF control run forecast valid at 01:00 UTC over the outer domain (Fig. 2.3).

in the processing of the raw observations, in particular for clutter removal and attenuation correction.

Two nested domains are defined for the model run (Fig. 2.3), with grid resolution of 10.8 km for the outer domain and 3.6 km for the inner domain. In addition to Doppler radar, surface observations are also assimilated. While the Automatic Weather Stations (AWS) managed by Arpa Piemonte (red dots in Fig. 2.3) provide data with 30 minutes-frequency, the SYNOP messages (yellow triangles) are only available every three hours. The

Table 2.1: Radar characteristics.

	C-band *Bric della Croce*	**X-band** *ARX*
Antenna diameter	4.2 m	1.8 m
Antenna beam width	0.93°	1.29°
Antenna gain	45 dB	43 dB
Polarization type	Simultaneous HV	Simultaneous HV
Operating frequency	5.640 GHz	9.375 GHz
Transmitter peak power	250 kW	70 kW
Pulse width	0.5, 2.0 μs	0.5, 1.0, 2.0 μs
PRF	250-1200 Hz	250-2400 Hz
Range resolution	340 m	125 m
Receiver dynamic range	95 dB	95 dB
MDS	-110 dB	-110 dB
Range	170 km	50 km
Volume scan		
Number of sweeps	11	10
Elevations	-0.1° to 28.5°	4.0° to 35.0°, 90°
Pulse length	0.5 μs	0.5 μs
PRF	883/662 Hz	2400/1600 Hz
Sensitivity at 50 km	2 dBZ	4 dBZ
Scan frequency	5 minutes	5 minutes

surface observations (temperature, dew point temperature, wind direction and intensity) are therefore assimilated every three-hours, while the radar observations are assimilated every one hour.

Table 2.2 summarizes the assimilation experiments performed, with and without assimilation of radar data. The radar used in the assimilation are the Bric della Croce C-band radar, with 170 km operational range, and the X-band radar with 50 km range (Fig. 2.3). In particular, experiments using either the C-band or the X-band radar and using only Doppler radial velocity, only reflectivity or both have been considered.

Since the radar data have a much higher resolution than the model grid, the PPI measurements are averaged over Cartesian grid cells with dimension comparable with that of the model (super-observations). In this assimilation experiment two different aggregation sizes

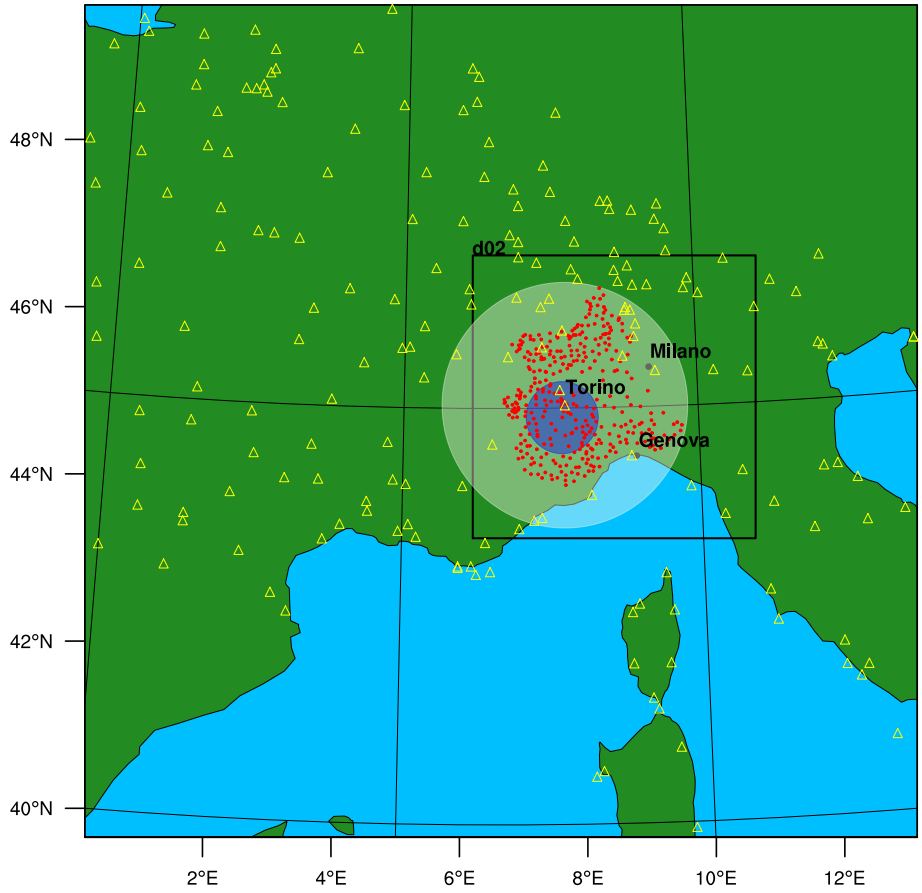


Figure 2.3: WRF two-nested domains over northern Italy with shaded disks indicating the radar range: C-band (white disk, 170 km range) and X-band (blue disk, 50 km range). The yellow triangles mark the locations of the synop stations, while the red dots the AWS regional network.

have been considered, approximately 3 km and 9 km (*radar res* in Table 2.2). The adjoint of eq. (6), i.e. conversion from Z_h to q_r, is applied before averaging the raw radar observations to avoid introducing a bias in the aggregated data. With the perturbation approaching the radar domain from south-west, the number of available observations for the assimilation increases substantially in the late evening of April 26 (fig. 2.4). The coarser aggregation size (9 km) implies a number of assimilated observations approximately one order of magnitude lower respect to the experiments with 3 km aggregation.

Table 2.2: Experiment design.

Experiment	GTS+AWS	C-band radar		X-band radar		radar res	
		Doppler	Z_h	Doppler	Z_h	3-km	9-km
SFC	X	-	-	-	-	-	-
SFC_ZV_C3	X	X	X	-	-	X	-
SFC_ZV_X3	X	-	-	X	X	X	-
SFC_ZV_CX3	X	X	X	X	X	X	-
SFC_V_C9	X	X	-	-	-	-	X
SFC_V_CX9	X	X	-	X	-	-	X
SFC_ZV_CX9	X	X	X	X	X	-	X
ZV_CX9	-	X	X	X	X	-	X

GTS and AWS data are assimilated at 3-h intervals, while radar observations at 1-h intervals.

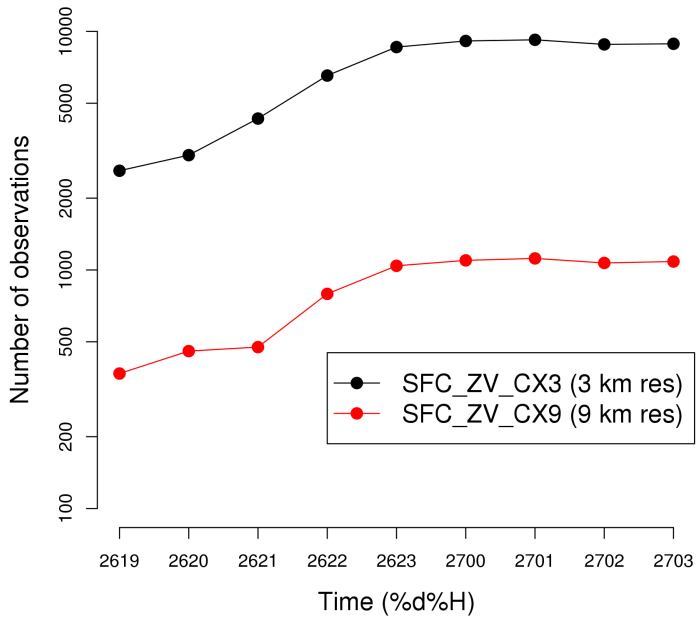


Figure 2.4: Number of observations assimilated in experiments using super-observations with 3 km resolution (SFC_ZV_CX3, black line) and 9 km resolution (SFC_ZV_CX9, red line). The y-axis is in logarithmic units.

The unambiguous velocity for the Doppler measurements is respectively 23.5 and 38.4 $\mathrm{m\,s^{-1}}$ for the C-band and the X-band radar. The Doppler data are initially processed in

the polar domain, de-aliased based on the histogram method [20] and using VAD (Velocity Azimuth Display) winds together with contiguity constraints. The data are susequently interpolated to the Cartesian space. Doppler-based filtering was not applied on the C-band radar, relying on a post-processing fuzzy logic scheme adapted from [4] to classify meteorological and non-meteorological echoes [21]. The observations identified as clutter are removed and the resulting missing values are filled in through bilinear interpolation during the subsequent super-observations processing.

Assuming that the radar observations are uncorrelated in space and time, the observation error covariance matrix $\mathbf{O}$ in (3) is assumed to be diagonal, with elements given by:

$$\epsilon(V_r) = 0.5 + \sigma(V_r)\ \mathrm{m\,s^{-1}}; \quad \epsilon(Z_h) = 2.0 + \sigma(Z_h)\ \mathrm{dB} \tag{8}$$

where σ represents the standard deviation calculated over neighboring locations and the additive constants represent the radar measurement accuracy. In the WRF 3D-Var assimilation process the observations whose innovations (obs-background) are larger than a maximum value defined as a multiple (5 is used for both V_r and Z_h for all experiments) of the observation error are rejected. The additive constants in (8) are therefore important to avoid rejection of observations whose local variability is small.

Since in the current version of the WRF 3D-Var (ver. 3.4) a simple warm-rain process is implemented to bridge water hydrometeors and other variables during reflectivity assimilation [22], the reflectivity super-observations are only assimilated below the freezing level. On the other hand the Doppler super-observations are assimilated everywhere.

Using the "NMC method" [18] as elucidated in section 2.1, the background error covariance matrix has been calculated after running the model for one month before the simulation.

The WRF model is initialized at 18:00 UTC 26 April using the 0.5° resolution GFS (Global Forecast System) analysis. The 1-h forecast valid at 19:00 UTC is used as a first-guess (background) for the assimilation of radar observations and from the resulting analysis a new forecast is started. This 1-h cycle is repeated until 03:00 UTC 27 April. At synoptic hours 21:00, 00:00, 03:00 UTC a 6-h forecast is run for subsequent validation.

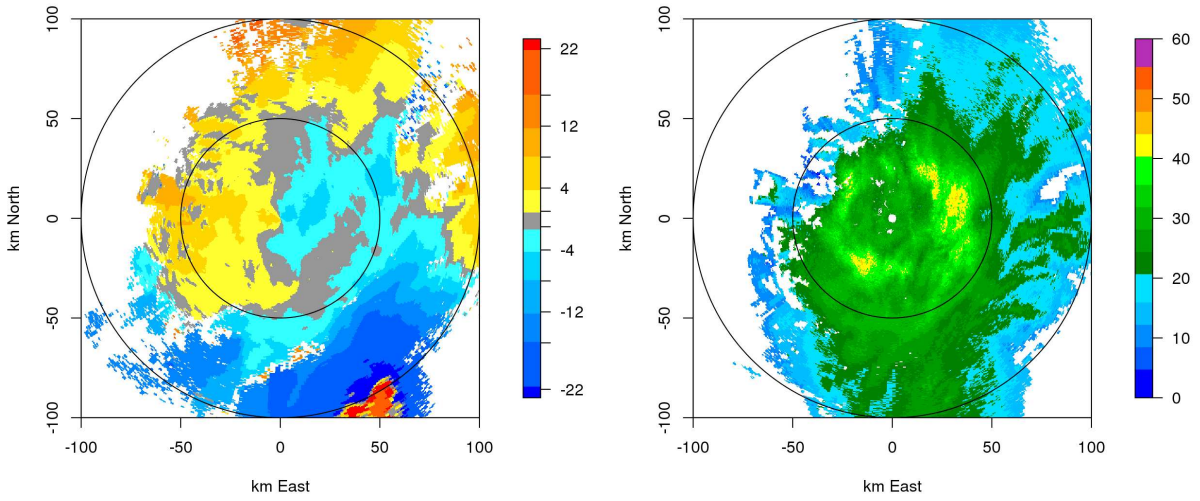
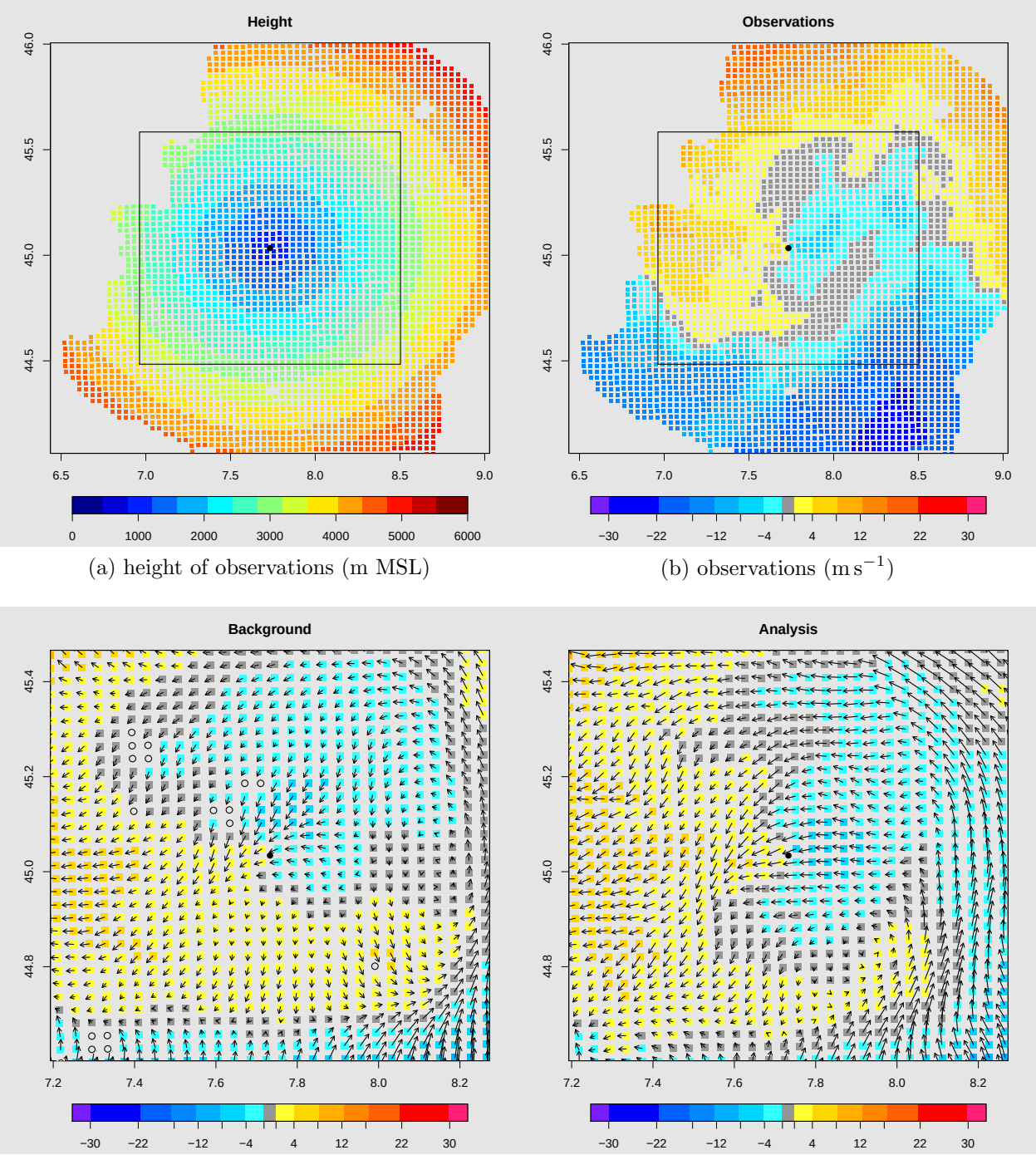


Figure 2.5: PPI of observed Doppler velocity (left) and reflectivity (right) at 03:00 UTC for C-band (2.0° elevation). Data are shown on a reduced 100 km-range.

Fig. 2.5 shows an example of the raw radar PPI of Doppler velocity and reflectivity. The Doppler field is affected by some aliased velocities (around x=+50, y=-100 km) and clutter (patches with near 0 $\mathrm{m\,s^{-1}}$ velocity in the northwestern sector). After processing including de-aliasing, clutter removal and averaging, the Doppler super-observations with 3 km-resolution look like in Fig. 2.6 (panel b)). Panel a) shows the height of the super-observations for the considered PPI, while in panel c) and d) the background and analysis wind vectors are plotted over the corresponding radial component of the velocity. The background field (here provided by a WRF 1-h forecast) is characterized by North-easterly winds at low levels, near the radar. The analysis shows how the resulting wind field is in



(a) height of observations (m MSL)

(b) observations ($\mathrm{m\,s^{-1}}$)

(c) background ($\mathrm{m\,s^{-1}}$)

(d) analysis ($\mathrm{m\,s^{-1}}$)

Figure 2.6: a) Height of the radar observations on the 2.0° elevation PPI as in Fig. 2.5. b) Doppler radial velocity super-observations (3 km resolution) after de-aliasing. c) and d) show the background and analysis radial velocities with wind vectors, on the smaller rectangular area of panel a) and b). In panel c) the open circles represent observations not assimilated due to excessive difference from the background.

better agreement with the radar observations, indicating an Easterly flow in the lowest levels, below approximately 2000 m.

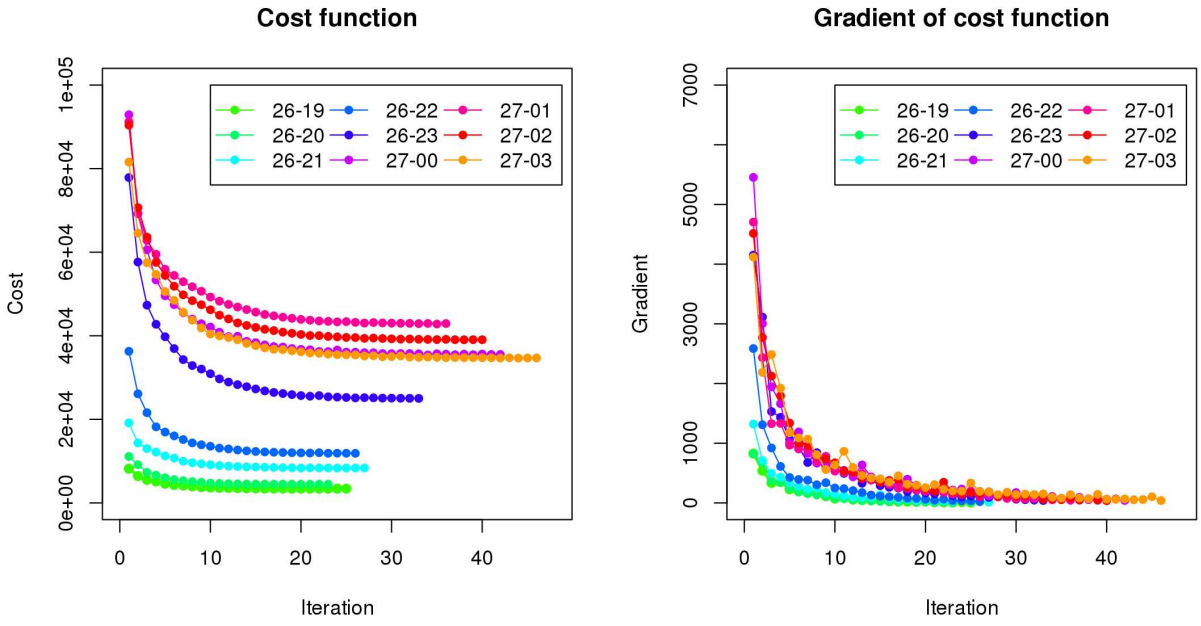


Figure 2.7: Cost function and gradient for WRF variational assimilation with 1-h time frequency (experiment SFC_ZV_CX3). The colored lines refer to different assimilation times (dd-HH in legend).

The performance of the analysis process can be monitored looking at the cost function minimization. In the WRF implementation, the iteration is stopped when the gradient of the cost function drops below a pre-determined value. Fig. 2.7 presents the cost function value (left) and its gradient (right) as the iteration progresses during the minimization. Different colors refer to the hourly analyses and the initial magnitude of the cost function is greatly determined by the number of valid available observations. After 22 UTC (blue color) the model domain is increasingly covered with radar echoes from precipitation and the cost function reduces by a factor of two or more after about 30 iterations. The gradient of the

cost function rapidly decreases during the first iterations, indicating the effectiveness of the minimization.

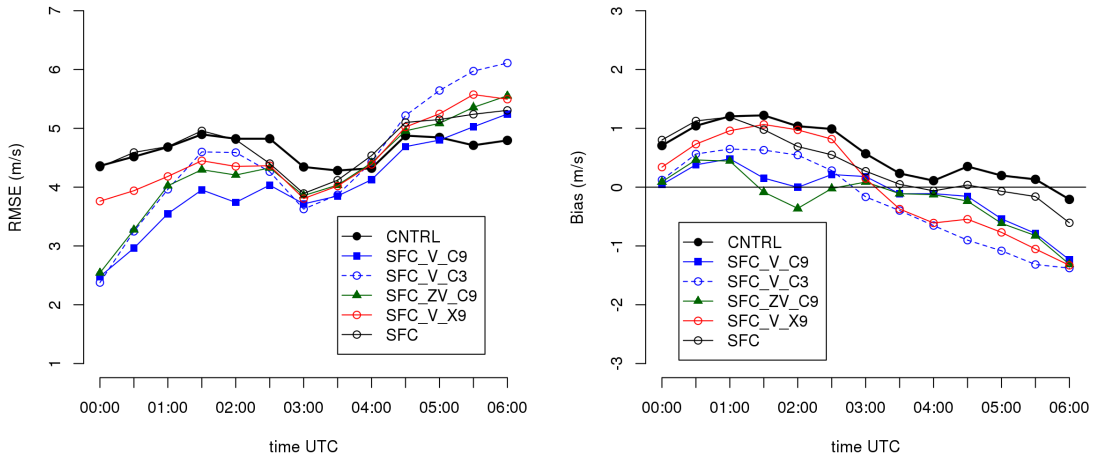
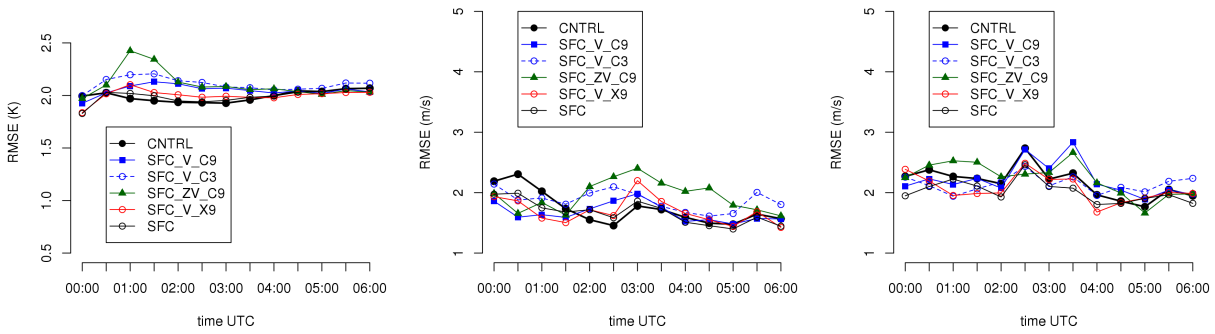


Figure 2.8: RMSE (left) and bias (right) of radial velocities from different forecast experiments verified against C-band radar observations.

2.2.1. Verification. In order to verify the impact of the assimilation a comparison with observations for the "long" 6-h forecasts starting at synoptic hours is performed. The verification is realized over the WRF inner domain (Fig. 2.3). Fig. 2.8 (left) shows the radial winds verification for the model 6-h forecast starting at 00:00 UTC 27 April, using the C-band observations as reference. The RMSE at the initial time (analysis) is reduced by about 40% with respect to the control run for all experiments with assimilation of C-band radar data (exp. SFC_V_C9, SFC_V_C3, and SFC_ZV_C9). However, the assimilation of Doppler observations with 9 km resolution showed the best impact for the forecast up to about 3 hours lead time, while consideration of higher resolution super-observations (3 km, experiment SFC_V_C3) actually led to an increase of the RMSE during the forecast. The assimilation of the X-band observations of radial winds is also beneficial for the forecast (red

line in fig. 2.8), but not as much as the C-band data, likely due to the smaller coverage (50 km) of the smaller X-band system compared to the C-band (170 km). The improvement respect to the control run gradually decreases and eventually vanishes after about 4 hours in the forecast. On the other hand the assimilation of surface observations only (exp. SFC) statistically has little effect on the upper air radial winds. The assimilation of the radar data, in particular from the C-band system, also allowed to minimize the bias (fig. 2.8, right panel). When the X-band data are used for the verification (not shown) the qualitative results are similar, although somehow noisier for the smaller extent of the radar coverage.



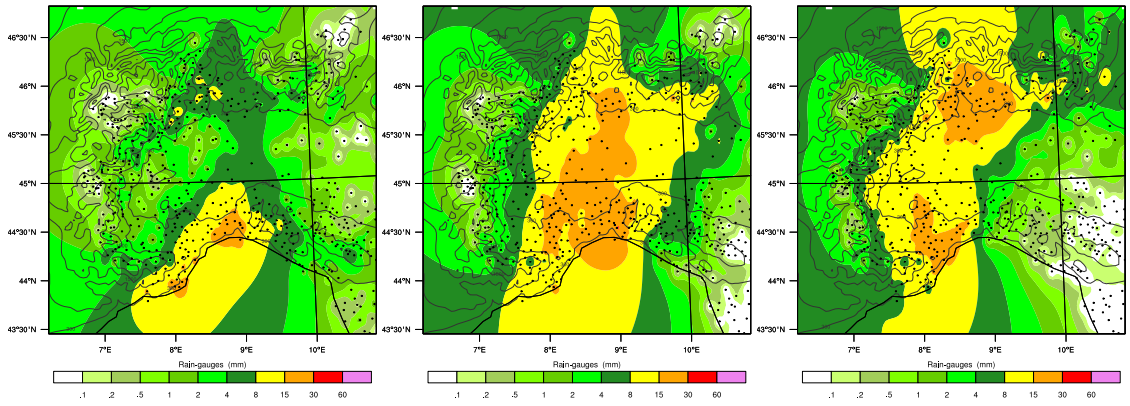
(a) T (b) u (c) v

Figure 2.9: RMSE of 2M temperature and 10M wind components (u, v) from different forecast experiments verified against surface observations.

The model forecast verification against the surface observations of temperature and winds is presented in Fig. 2.9. The temperature observations are 275 (Fig. 2.3), but only a smaller portion of the surface stations (153 observation locations) have the 10-meter wind sensor (direction and velocity).

As for the radar upper-air radial winds, at the analysis time there is a recognizable improvement for both temperature and wind, although much less relevant than in the verification with radar winds. In addition, the better performance only lasts for about 30 minutes

and 1-2 hours respectively for temperature and wind. The experiment including the assimilation of reflectivity (SFC_ZV_XC9) appears to be the most problematic when looking at the surface observations of wind and temperature. In Fig. 2.9 an increase in the RMSE is evident for temperature between 01:00 and 02:00 UTC. The wind RMSE, especially the u-component, also increases shortly after the initial time.



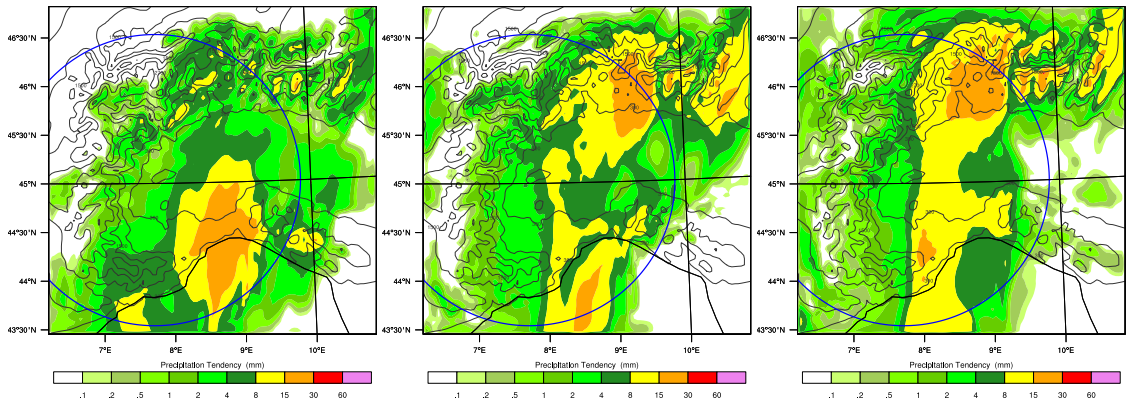
(a) Rain-gauges 21-00UTC (b) 00-03UTC (c) 03-06UTC

Figure 2.10: Observed rainfall accumulation for three consecutive 3-hourly intervals. The dots represent the locations of the rain-gauges. The overplotted contours of topography (levels 500, 1000, 1500 m) indicate the higher mountains in the western and norther sector.

For the verification of the surface precipitation a very dense network of rain-gauge observations is available (583 tipping-bucket rain-gauges, including the AWS and synoptic stations over the inner domain in Fig. 2.3, as well as additional measurements from other regional networks). Short-term forecasting of rainfall accumulation is indeed the primary motivation for the assimilation of radar observations. For the case being considered the warm-front passage helped produce continuous rain for most of the day, with 24 hour average accumulation of 76 mm over the plains, and peak values up to 140 mm. In the northern mountains, values above 200 mm were recorded. In the following discussion we focus on the period between 21:00 UTC 26 April and 06:00 UTC 27 April. The rainfall accumulation in the

3-hourly intervals 21-00, 00-03 and 03-06 UTC is analysed in detail, considering respectively the forecasts starting at 21, 00 and 03 UTC.

In Fig. 2.10 the observed rainfall accumulation is shown. The dots in the images represent the observation locations and the spatial interpolation is conducted using the Inverse Distance Weighting (IDW) method. The precipitation initially affected the coast-line, near the city of Genova (21-00 UTC), later spreaded over the Po valley (00-03 UTC) and finally concentrated near the mountains in two separate areas, one in the pre-Alpine northern region and one in South-western Apennine region.



(a) CNTRL 21-00UTC (b) 00-03UTC (c) 03-06UTC

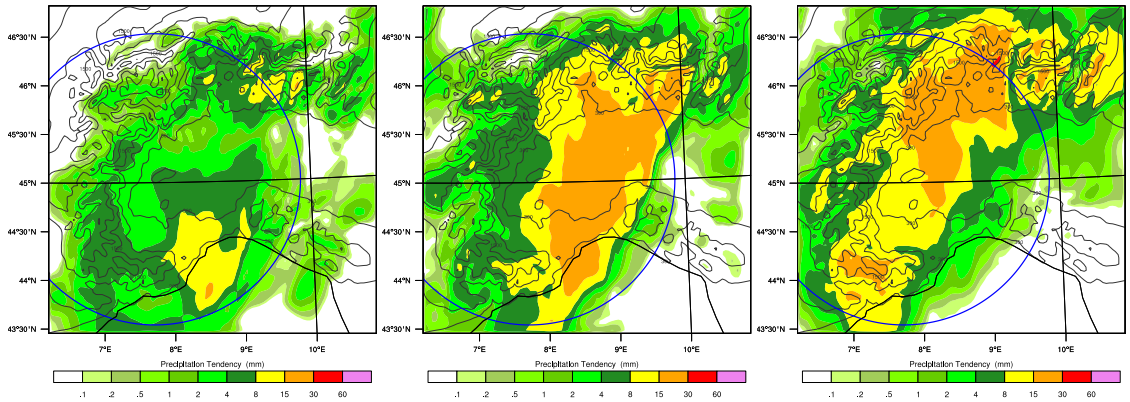
Figure 2.11: Model forecasted rainfall accumulations for the control run. The circle indicates the range of the C-band radar.

Figures 2.11 and 2.12 report the rainfall accumulation predicted by the control run and by three different experiments including assimilation of the radar observations for the same three consecutive intervals as in Fig. 2.10. To measure the skill of the precipitation forecasts two widely used statistical parameters are considered: the bias and the Threat Score (TS), also denoted Critical Succes Index (CSI), defined in Appendix A.

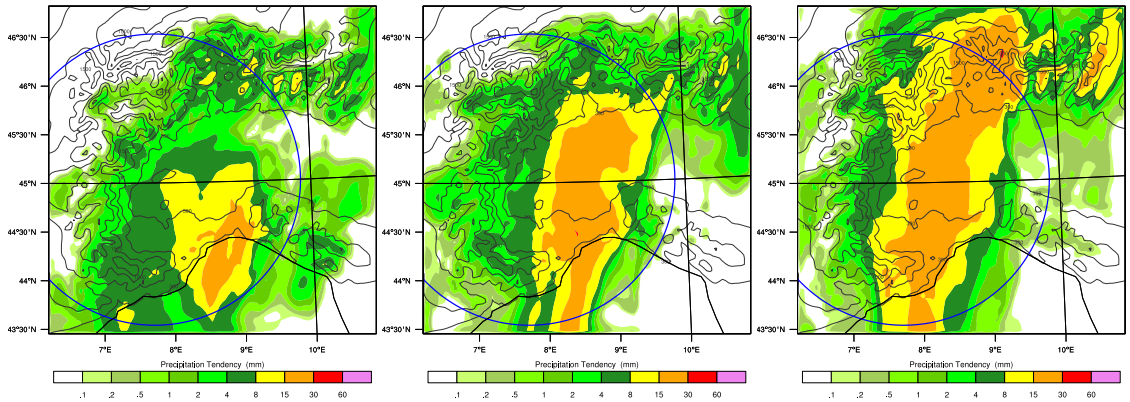
For a perfect forecast ($C = F = O$) both the bias and the TS have a value of 1. The evaluation of the bias and TS is applied to the observation sites. A summary of the precipitation forecast statistics is presented in figures 2.13, 2.14 and 2.15.

The rainfall accumulation predicted by the control run in the 21-00 UTC period (Fig. 2.11) is in good agreement with the observations, except for a general light overestimation. After midnight the control run, as a consequence of the forecasted advancement of the trough with its associated moist South-westerly flow, shows a tendency to place the major rainfall area across the central Alps, in the North-eastern part of the model domain. This is in contrast with the observations, in particular for the 00-03 UTC period when the most intense precipitation is recorded over the plains.

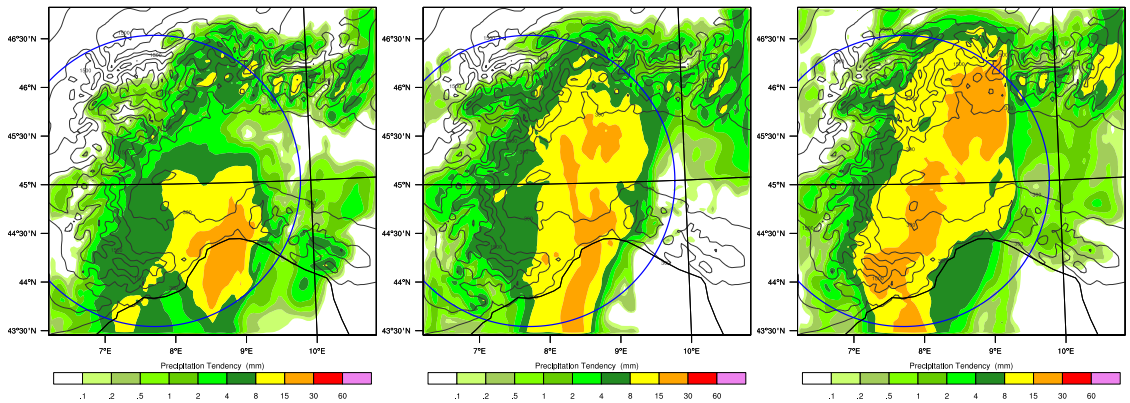
The rainfall accumulations resulting from three experiments with assimilation of radar data are reported in Fig. 2.12. All experiments include both radars but differ for the resolution and variables considered, specifically from top to bottom in the figure: assimilation of radial winds and reflectivity at 3-km resolution (panels a-c), assimilation of radial winds and reflectivity at 9-km resolution (panels d-f), assimilation of radial winds only at 9-km resolution (panels g-i). The rainfall accumulations are calculated from the +0 to +3 hours forecasts, e.g. forecast starting at 00 UTC for the 00-03 UTC accumulation. The 3-km resolution experiment (top row) shows an improved localization of the rainfall amounts with respect to the control run between 00 and 03 UTC, but an appreciable negative bias during the first period (21-00 UTC), when the predicted precipitation on the coast-line does not exceed 15 mm. For the same period a similar assimilation experiment with a coarser resolution (9-km, middle row in Fig. 2.12) actually gives a better result, with very small placement error and negligible bias. In general the 9-km resolution super-observation provides better performances in the assimilation. It is possible that the 3-km super-observations, being even



(a) 21-00UTC (b) 00-03UTC (c) 03-06UTC



(d) 21-00UTC (e) 00-03UTC (f) 03-06UTC



(g) 21-00UTC (h) 00-03UTC (i) 03-06UTC

Figure 2.12: Model forecasted rainfall accumulation with assimilation of both surface observations and radar data for three different experiments. Top row: SFC_ZV_CX3 , middle: SFC_ZV_CX9 , bottom: SFC_V_CX9. The circle indicates the range of the C-band radar.

finer than the model grid (3.6 km), may introduce some small scale noise that increases the model spin-up.

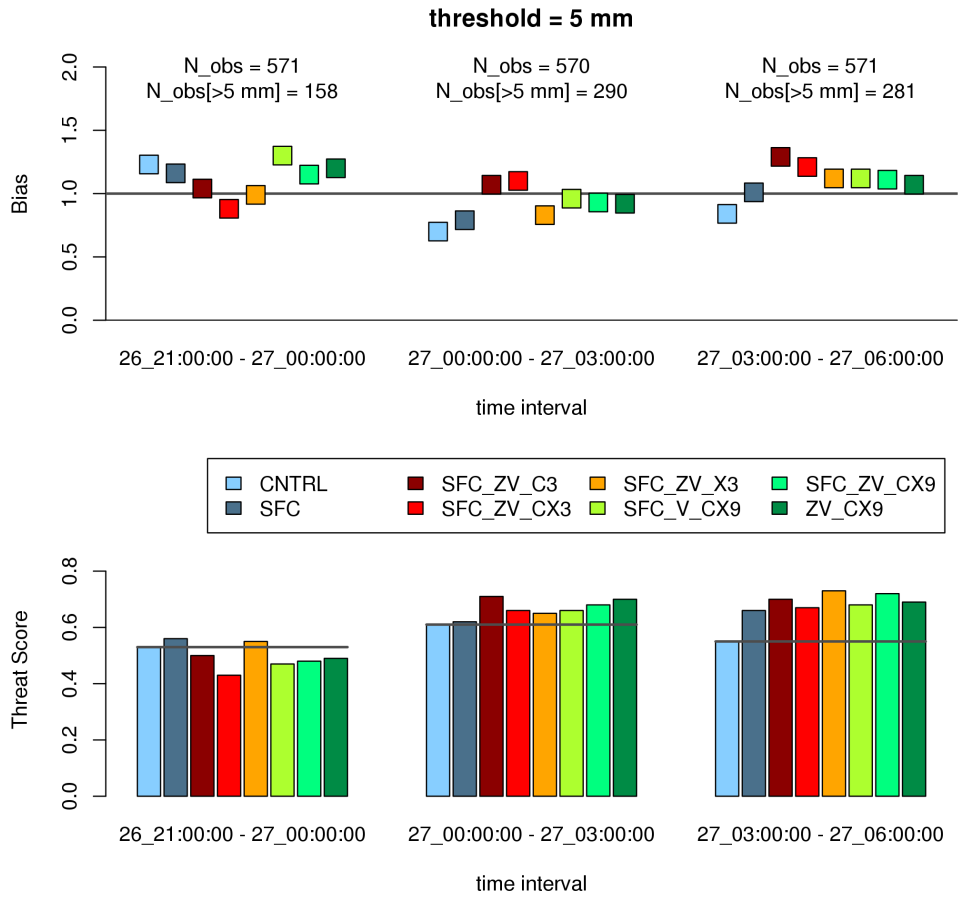


Figure 2.13: Assimilation statistics for rainfall > 5 mm (0-3 hours forecast).

The bias and TS are calculated for two different rainfall thresholds, 5 and 12 mm (figs. 2.13 and 2.14). For the 5 mm thresholds the assimilation of the radar data shows no improvement during the first time period (21-00 UTC), likely due to the relative scarcity of the observations. After 00 UTC the impact on the skill of the rainfall forecast increases with the assimilation of the radar data, in particular when the 9-km resolution super-observations are considered. However, the most remarkable impact of the radar can be observed for the 12

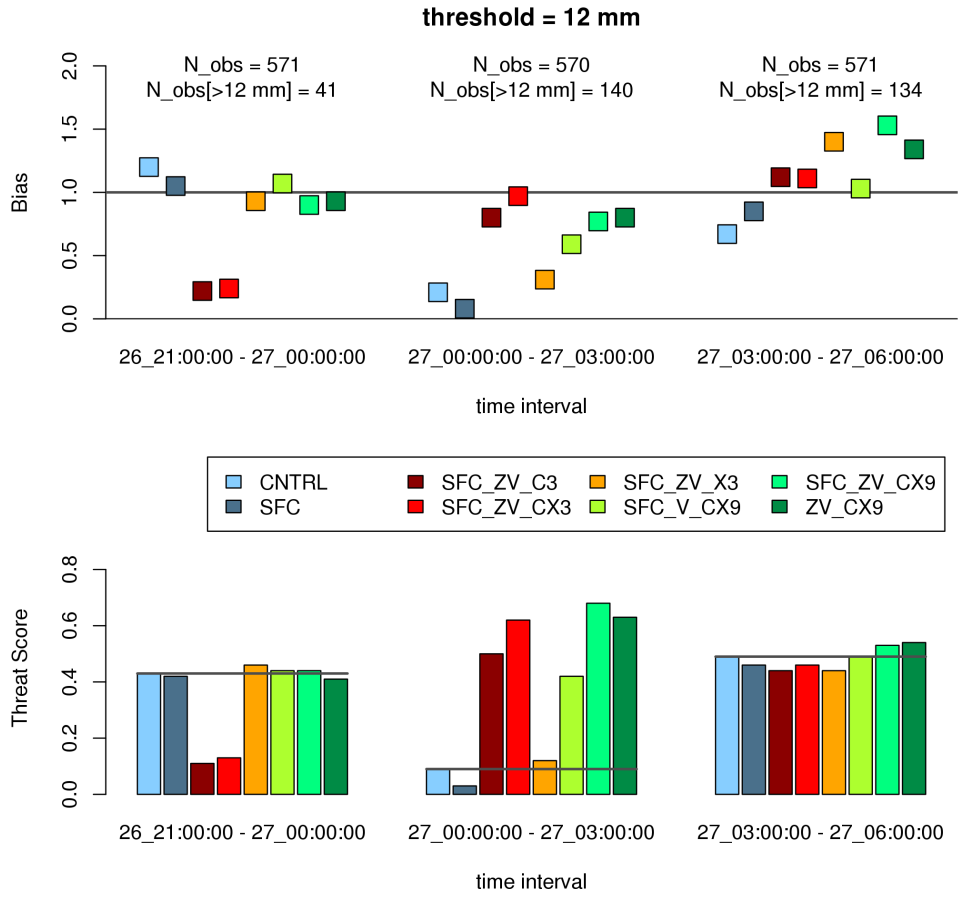


Figure 2.14: Assimilation statistics for 3-h rainfall accumulation > 12 mm (0-3 hours forecast).

mm threshold (Fig. 2.14). In particular for the 00-03 UTC interval, with maximum rainfall intensity over the plains, the Threat Score jumps from 0.09 to 0.68 for the experiments with assimilation of reflectivity and radial winds from both radars. The underestimation present in the control run is also considerably reduced, with the bias going from 0.21 to 0.77. During the same time interval the assimilation experiments including the X-band data show a perceptible improvement, even though the X-band radar range is completely within the C-band domain (Fig. 2.3). The TS for the experiments with the C-band radar at 3-km resolution (SFC_ZV_C3) is 0.50 and increases to 0.62 when the X-band data are assimilated

as well (SFC_ZV_CX3). The X-band radar is located on the plains at 235 m height, while the C-band is over a hill at 736 m height. It seems likely that the better sampling of the lowest atmospheric layers by the X-band, in addition to the complementary radial winds from a different observation point, contributes to a better description of the moist flow and the resulting rainfall field.

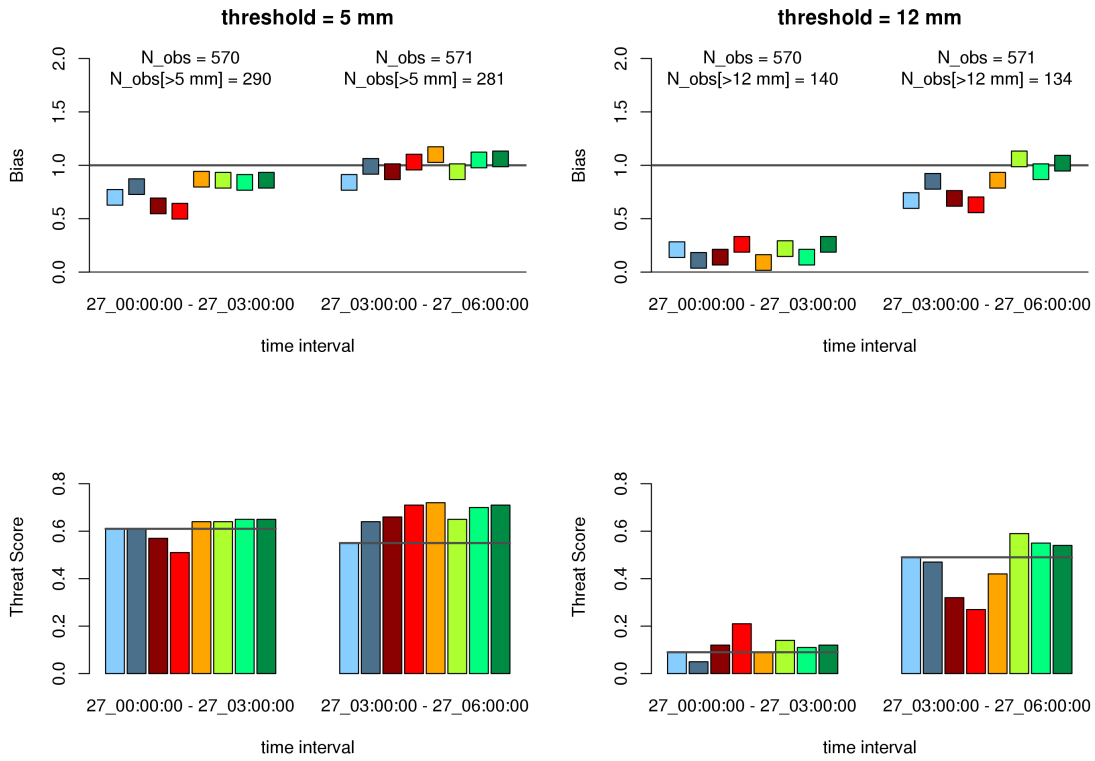


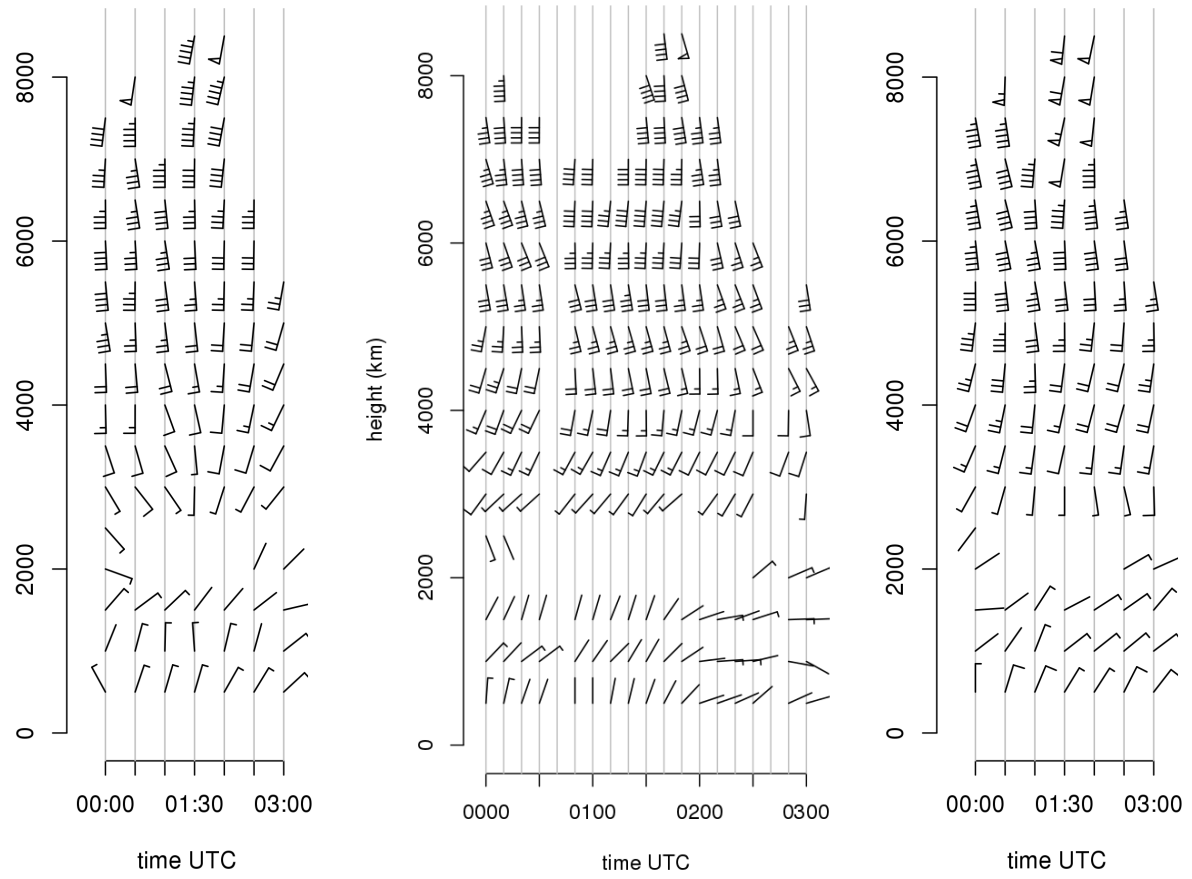
Figure 2.15: As in figure 2.13 and 2.14, same legend, but for 3 to 6 hours rainfall accumulation forecast.

Finally Fig. 2.15 presents a summary statistics for the longer range forecast (+3 to +6 hours, e.g. the rainfall in the 03-06 UTC interval is evaluated considering the forecast starting at 00 UTC). The impact of the radar observations is reduced in this case, although still noticeable both in terms of the reduced bias and increased TS. The assimilation of surface observations alone instead has in general a negligible impact on the rainfall forecast.

For the case study analysed the positive impact of the assimilation of radar observations, in particular the radial winds, has been shown. The reflectivity observations have a not well defined impact. In particular there is a small additional positive impact on the rainfall forecast (when reflectivity is assimilated together with radial winds), but a negative impact on the surface observations of temperature, indicating the complexity in dealing with the balances involving latent heat transfers. It is worth noting that Sugimoto et al. [12] in an OSSE (Observing System Simulation Experiments) have obtained an improvement not only for the rainfall fields, but also for the un-observed fields of wind, thermodynamics, and microphysics. In their OSSE however, a cloud analysis was introduced to modify the background before performing a 3DVAR minimization procedure.

2.2.2. Discussion. Given the supposed central role of the improved dynamics for the more accurate rainfall forecast in the case studied, an analysis of the un-observed upper-air horizontal winds (as opposed to the radial winds as in Fig. 2.8) may help to shed more light on the impact of the Doppler winds assimilation. In [12] a limited ability of the WRF 3D-Var sytem in retrieving the un-observed tangential wind component in a single-Doppler assimilation experiment was demonstrated. On the other hand the use of multiple radars improved significantly the retrieval of the tangential wind component.

Using a VAD (Velocity Azimuth Display) technique it is possible to use the observed radial velocity to retrieve the areal mean vertical profile of the horizontal wind above the radar. Since these retrievals were not considered in the assimilation, they can be used to verify the performance of the upper-air wind analysis and forecast. The algorithm proposed by Tabary et al. [23], based on the azimuthal derivative of the observed radial velocities, is used to compute the wind vectors presented in Fig. 2.16, panel b). In the same figure the vertical profiles predicted from the control run at 00 UTC (a) and the experiment SFC_V_C9,



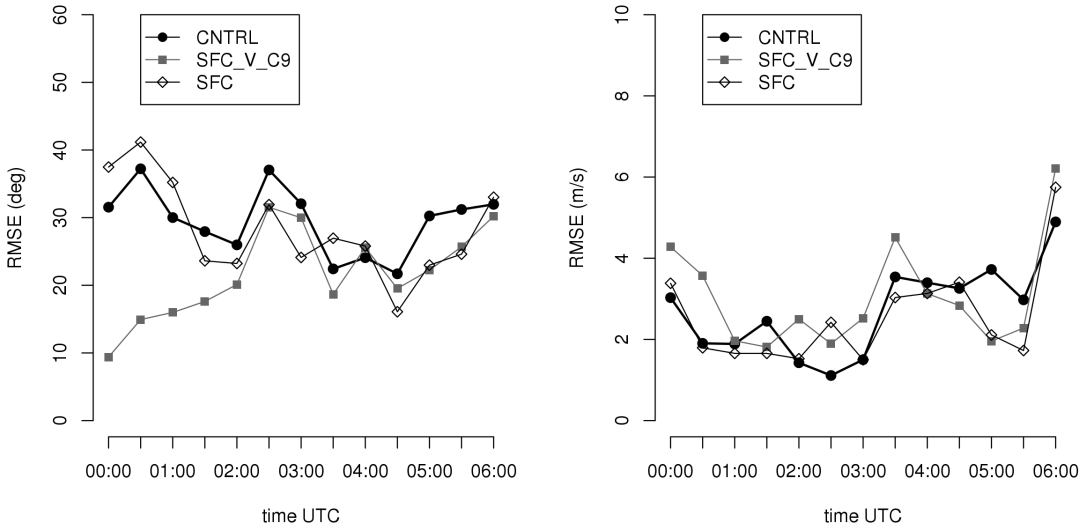
(a) CNTRL (b) VAD radar (c) SFC_V_C9

Figure 2.16: Wind profiles from the control run (a), X-band observation-based VAD (Velocity Azimuth Display) retrievals (b) and SFC_V_C9 assimilation experiment. Wind vectors are represented by wind barbs: each short barb represents 5 knots, each long barb 10 knots, pennants are 50 knot.

assimilation of radial winds from the C-band radar (c), are also plotted. The VAD in panel b) is retrieved from the X-band observations, which are not used in the assimilation experiment SFC_V_C9.

The control run basically predicts a regular clockwise wind rotation with height (wind veering) associated with the warm front at the X-band radar location (Fig. 2.16a). This monotonic wind veering contrast with the VAD retrievals, which show a stronger wind shear until approximately 3000 m (wind from the South-West as opposed to the South-easterly

flow in the control run) and then a backing of the wind (counter-clockwise rotation) up to 6000 m. At higher levels the wind veers again and sets from the South in good agreement with the control run. The radial winds assimilation run yields an analysis (00 UTC) closer to the observations, reproducing the alternate veering-backing revealed by the VAD. Most noticeably, this type of wind pattern is maintained reasonably well in the forecast for about 2-3 hours.



(a) wind direction

(b) wind speed

Figure 2.17: Verification against VAD wind profile retrievals from the X-band.

An objective verification of the wind direction and velocity is presented in Fig. 2.17. While the assimilation of surface observations alone has substantially no impact on the upper-level winds, the improvement brought by the assimilation of the radial winds is clear in the first hours of the forecast, although limitedly to the wind direction. It is very likely that the improved horizontal wind field (direction) in the mid-tropospheric levels plays a role in the more accurate rainfall forecast, also considering the possible interaction of the moist flow with the Alps, whose top heights reach 4800 m in the region. For this winter-type

events the precipitation mainly forms at mid-levels (through the cold rain Bergeron process, see discussion in Chapter 3), so it is very important to get a good wind analysis for this layer. Radar radial winds have shown to provide a clear positive impact with this respect.

The fact that the positive impact is limited to 2-3 hours in the forecast is at least partially ascribable to the lack of radar observations over most of the WRF outer domain. In fact a wind velocity of 15 $\mathrm{m\,s^{-1}}$ at mid-levels will advect a tracer for about 160 km in three hours, a distance comparable with the actual radar range.

As recognized in several studies [7, 12, 24] the assimilation of reflectivity observations presents important challenges. Latent heat transfers and non-linear observation operators (eq. 6) are among the important topics of current research in the radar data assimilation field. When the precipitation is widespread and covers a large portion of the model domain the radial winds assimilation alone in general provides a relevant improvement, while the addition of reflectivity has not always a defined positive impact. In addition reflectivity should only be assimilated below the freezing level because of the warm rain process-based partitoning of the total water mixing ratio. This is currently a serious limitation, since even during summer mid-latitudes precipitation systems the ice microphysics in the middle to upper troposphere play important roles in kinematic and thermodynamic processes. While there are on-going investigations on the inclusion of more sophisticated partitioning schemes to take into account the ice phase, it will be important to provide radar products of hydrometeor classification for both verification and direct input into future assimilation schemes.

CHAPTER 3

Polarimetric radar observations in the ice region of precipitating clouds

Polarimetric radar observations have been widely exploited in rainfall estimation since the introduction of dual-polarization radars in the seventies [25]. However the ice phase of precipitating clouds have received much less attention, partly because of the intrinsic difficulty in the direct observations of ice clouds, needed to provide in situ observations for comparison.

During the last couple of decades several studies in the literature have focused on observations of ice particles with polarimetric radars. These studies often included the analysis of in situ aircraft measurements and mainly considered the radar observations of differential reflectivity [26–29] and linear depolarization ratio [30, 31, 27, 28].

Several papers considered specific differential phase shift (K_{dp}) measurements of ice particles. K_{dp} is defined as one-half the range derivative of the two-way propagation differential phase (Φ_{dp}), which is the phase shift occurring between the horizontally- and vertically-polarized pulses along the propagation path [32]. Some of these studies attributed the K_{dp} signatures in the ice region of convective storms to the effects of cloud electrification [33–36]. Only few studies reported observations of differential phase in snow [37–41]. Namely Hendry et al. [37] reported values of $K_{\mathrm{dp}} \sim 0.4^{\circ}\,\mathrm{km}^{-1}$ in snow at S-band and values up to $1^{\circ}\,\mathrm{km}^{-1}$ at Ku-band in heavy snow. Vivekanandan et al. [38] have also shown S-band differential phase shift changes by 10 ° above the bright band in the stratiform region of a mesoscale convective system, with peak values of K_{dp} up to $0.5^{\circ}\,\mathrm{km}^{-1}$.

More recently, positive values of K_{dp} in the ice region of stratiform precipitating clouds have been shown to be linked to increased surface precipitation rates at S band [41]. These authors found regions of enhanced K_{dp} are observed near the -15°C isotherm and suggested that they are related to the growth of dendritic crystals. The observed values of K_{dp} are only few tenths of $^{\circ}\,\mathrm{km}^{-1}$ at S band, but are expected to be higher and more easily detectable at higher frequencies such as those of C and X band due to the K_{dp} scaling with frequency ([25], their eq. 7.101).

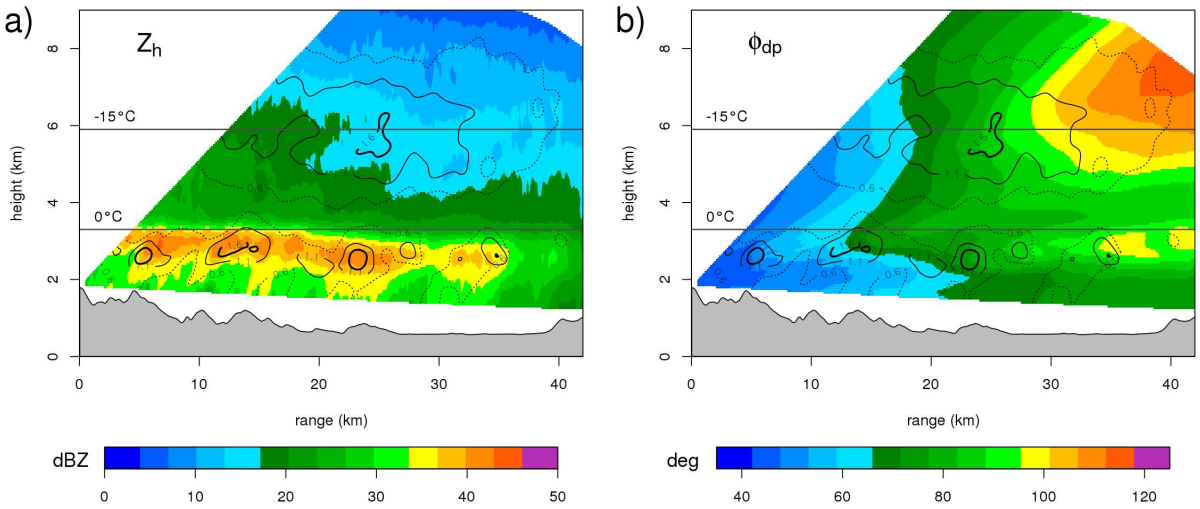


Figure 3.1: Elevation scan from Colle di Tenda on 14 August 2010 at 1743 UTC, along the 340° azimuth. a) reflectivity (Z_h), b) filtered differential phase shift (Φ_{dp}). The underlying topography is shown in gray. K_{dp} contours are overplotted in both panels: isolines at 0.6 (dotted line), 1.1 (solid line), 1.6 (thick solid line)$^{\circ}\,\mathrm{km}^{-1}$. The horizontal gray lines mark the location of the 0°C and -15°C temperature levels as inferred from the nearby radiosounding of Cuneo Levaldigi (WMO code 16113).

Elevation scans at high space resolution collected by the Arpa Piemonte transportable ARX X-band radar system during summer stratiform precipitation, for example, show Φ_{dp} increases in excess of 40° in the ice region, with K_{dp} peak values up to $\sim 2^{\circ}\,\mathrm{km}^{-1}$ in the mid-tropospheric layer at 5-7 km height (Fig. 3.1), corresponding to temperatures in the range -9 to -20°C as measured by a nearby radio sounding.

This study aims at characterizing the phenomenological behavior of K_{dp} in the ice region of generic mesoscale precipitation systems, irrespective of the phase of precipitation (liquid or solid) at the surface, at C-band and X-band radar frequencies. Differential phase shift measurements have the well-known advantages of being 1) immune to radar calibration, 2) immune to attenuation due to propagation, 3) mainly insensitive to partial beam blocking [42, 25, 43–45]. One of the potential problems with differential phase measurements is the backscattering differential phase δ at C and X band. In fact radar does not measure directly Φ_{dp} , but the total differential phase shift Ψ_{dp}, which equals the sum of the differential propagation phase and δ, i.e. $\Psi_{\mathrm{dp}} = \Phi_{\mathrm{dp}} + \delta$. Typical ice particles in the ice region of stratiform clouds around -15°C are Rayleigh scatterers at both C-band and X-band radar frequencies (Section 3). This fact makes δ negligible [46], removing a potential source of error in deriving reliable specific differential phase estimates. The primary source of data for this paper is the C-band Bric della Croce radar, close to Torino (Italy). The data from this system allowed an extensive analysis of vertical profiles of the polarimetric measurements in the northwest Italy subalpine region. Although K_{dp} is the main focus of this study, differential reflectivity (Z_{dr}) measurements from this radar are exploited to infer the dominant crystal habit. Additionally, measurements collected by the transportable ARX X-band system during a selected event are also considered. This made possible a thorough analysis of K_{dp} in the ice region with simultaneous observations at C-band and X-band radar frequencies (5.640 and 9.375 GHz respectively). Interpretation of radar measurements at the -15°C isotherm level are supported by electromagnetic scattering simulations performed at these frequencies.

3.1. Radar data and processing

The data considered are from the same radars described in chapter 2.2, Table 4.3. Φ_{dp} is filtered using the Hubbert and Bringi [47] scheme. Then the rain profiling algorithm based on Testud et al. [48] is applied to correct the reflectivity for path attenuation. Attenuation correction is performed exclusively in the rain medium (i.e. below the melting level) to avoid introducing a bias due to differential phase increases in the ice region. K_{dp} is estimated using the Wang and Chandrasekar [49] technique. This technique has been implemented in the operational post-processing chain of the Arpa radar network and proven to be suitable for the unsupervised processing of large data sets. In fact it does not require the estimation of the system differential phase (which may typically fluctuate with time as a consequence of hardware maintenance) and is insensitive to differential phase aliasing. The K_{dp} estimation requires measurements with a good Signal to Noise Ratio (SNR), to avoid too noisy phase measurements. A 5 dB threshold was used for this study, which implies that the effective sensitivity of the two radar volume scans using the short pulse (Table 4.3) is decreased from 2 dBZ (4 dBZ) at 50 km range to 7 dBZ (9 dBZ) for the C-band (X-band) radar respectively. The vertical profiles of temperature needed for the freezing level identification in the attenuation correction procedure and for the K_{dp} analysis on which this paper is focused, are taken from the 3-hourly output of the local area COSMO model (www.cosmo-model.org) run over Italy (analysis at 00 UTC and subsequent 3-hourly forecasts). In previous verification studies (COSMO Newsletter # 7, chapter 5.4) the accuracy of mid-tropospheric temperature analysis and 24-h forecast has been shown to be within 1°C. The vertical profiles are calculated considering all the radar volumes collected with a 5-minute frequency in a given time interval (typically 1 hour or 24 hours). Only sweeps at elevation angles below 15° are

used since the polarimetric signals decrease with higher view angles, due to the general hydrometeor orientation with the symmetry axis in the vertical direction. Additionally, range bins farther than 50 km from the radar are excluded to avoid excessively sparse vertical sampling. The data are then binned in 0.3 km height levels. The lowest available height is 900 m (layer 750-1050 m) for the C-band radar and 300 m (layer 150-450 m) for the X-band radar.

Differential reflectivity measurements from the C-band radar are used in the following sections, in addition to Z_h, $K_{\rm dp}$ and $\rho_{\rm HV}$, in order to support the characterization of the dominant crystal habit limitedly to stratiform precipitation systems. As shown in Wolde and Vali [27] and more recently in Williams et al. [50], strongly distinctive $Z_{\rm dr}$ signatures are associated with either dendrite crystals or plate crystals. Plate crystals, owing to their high volume fraction of ice and low axial ratio may give $Z_{\rm dr}$ values up to 6-7 dB, while $Z_{\rm dr}$ for dendrites with lower bulk density attains values in the range 1-2 dB. The required accuracy on the $Z_{\rm dr}$ measurements to distinguish between the aforementioned crystal habits is therefore not as stringent (say 0.1-0.2 dB) as for use in quantitative polarimetric rainfall estimation algorithms.

Figure 3 shows a scatterplot of the hourly average $Z_{\rm dr}$ vs. Z_h for the lowest level (900 m) of 260 vertical hourly profiles in stratiform precipitation (refer to Section 4 for the profiles processing and the stratiform/convective classification). Only profiles with the freezing level above 1.5 km are considered to ensure the measurements are in the rain medium. The $Z_{\rm dr}$ values tend to approximately 0 dB when the reflectivity is below 20 dBZ, as expected due to the mostly spherical shape of small droplets. Displayed for comparison (dashed line) is also the $Z_{\rm dr}$ vs. Z_h average relation from Bringi et al. [3]. Differential attenuation at C band may introduce a bias affecting $Z_{\rm dr}$ measurements especially in the ice region, where

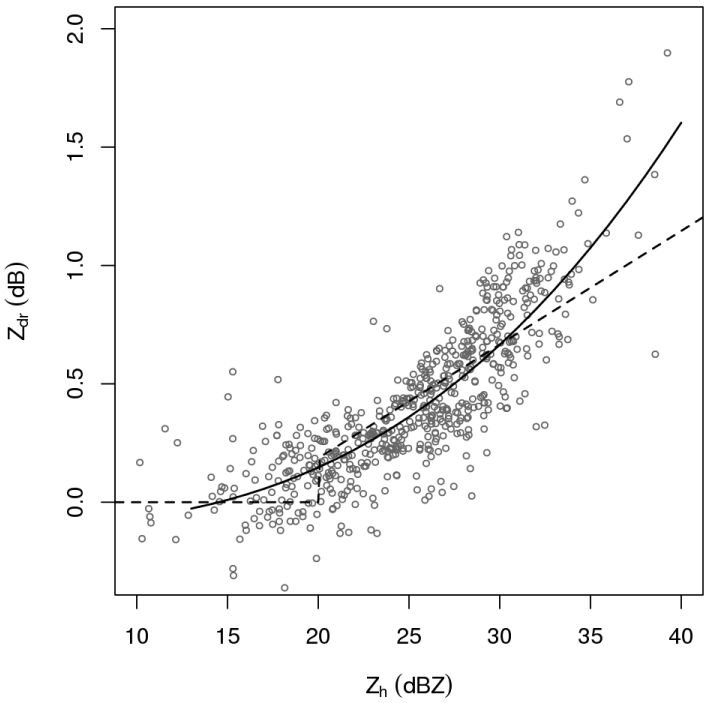


Figure 3.2: Scatterplot of hourly averages of $Z_{\rm dr}$ and Z_h in rain from the lowest height level of the 260 profiles in stratiform precipitation (Section 4). The solid line represent the nonlinear fit to the data, the dashed line the $Z_{\rm dr}\,(Z_h\,)$ relation from Bringi et al. [3].

the radar pulse is sensing the cloud after passing through the rain and the melting layer. To check the effect on the average profiles we derived statistics from large dry aggregates right above the melting layer in stratiform precipitation. Low density dry aggregates are known for the intrinsic low $Z_{\rm dr}$ [44]. The vertical profiles of $Z_{\rm dr}$ show a local minimum just above the freezing level (Fig. 14), where large aggregates are most likely to be found. The temperature corresponding to this minimum for the daily profiles in stratiform precipitation (27 events between April 2009 and December 2010, Section 4) is -2.6 $\pm$ 1°C and the associated average $Z_{\rm dr}$ is 0.23 $\pm$ 0.14 dB, varying in the range -0.06 to 0.52 dB. Although these values are slightly higher than the 0.25 dB upper limit reported in Ryzhkov and Zrnic [39] for cold snow, the limited variability over the long period analyzed is a good indication of the stability of the system and ensures the considered profiles are not significantly affected by

differential attenuation. Overall, the Z_{dr} average profiles for the stratiform precipitation cases can be considered as accurate within 0.2 dB, which is enough for the purpose of crystal habit identification.

3.2. Widespread stratiform precipitation case

In this section the same stratiform precipitation event analysed for the WRF 3D-Var assimilation experiments in section 2.2 is considered.

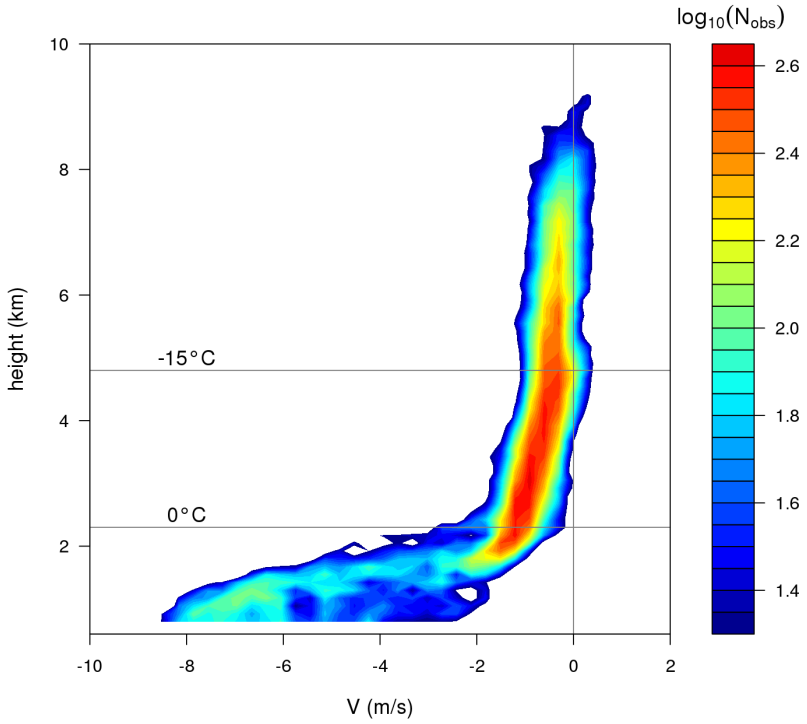


Figure 3.3: Frequency plot of daily Doppler velocity observations from vertical looking X-band radar scans (27 April 2009). The color scale represents the number of observations N_{obs} in logarithmic units, using 0.3 $\mathrm{m\,s^{-1}}$ and 0.125 km intervals for velocity and height respectively. The horizontal gray lines mark the temperature levels of 0 and -15°C, while the vertical gray line marks the 0 $\mathrm{m\,s^{-1}}$ velocity.

The X-band radar performed a scan strategy that included an acquisition at vertical incidence within each 5-min volume scan. Doppler velocity observations at vertical incidence (Fig. 3.3) allow to infer that riming was likely not relevant in the vicinity of the radar.

In fact the lack of a significant number of particles with Doppler velocity lower than -1.5 $\mathrm{m\,s^{-1}}$ (downward) is here taken as an evidence of the negligible generation of graupel by heavy riming [51]. However the capture of small supercooled cloud droplets by snow crystals can still take place when the air is saturated over water and simultaneous growth of ice particles and liquid droplets is effective [52, 53], without leading to the formation of heavily rimed crystals like graupel. The ARX vertical observations are therefore considered as suitable for investigating ice particle distributions mainly resulting from vapor deposition, light riming, and aggregation.

3.2.1. Polarimetric radar observations. Figure 3.4 shows the C-band polarimetric radar measurements collected at 0025 UTC during a PPI scan at the elevation of 7.4°. Both Z_h and Z_{dr} show the characteristic increase in the melting layer, with a corresponding low correlation coefficient ρ_{HV} in the range 0.8-0.96, deriving from the co-existence of liquid and partially frozen particles [39]. The values of ρ_{HV} in the western portion of the domain are in general lower due to the weaker radar echo (low SNR) and the Z_{dr} field is consequently noisier. At higher levels, on the East side of the radar domain, Z_{dr} shows marked positive values (up to 2.2 dB) along an azimuth sector about 135°-wide centered on the -15°C temperature level at ∼4.7 km MSL. The corresponding reflectivity varies in the range 10-25 dBZ. These values are fairly consistent with the category A microphysical regime (dendrite crystals) described in Williams et al. [50], who found the enhanced Z_{dr} layer (1-3 dB) in different case studies at temperatures between -13 and -15°C, and with corresponding reflectivity between 10 and 30 dBZ. Similar Z_{dr} signatures in this temperature range were also previously reported by Sauvageot et al. [54], Wolde and Vali [27, 28], Andric et al. [55]

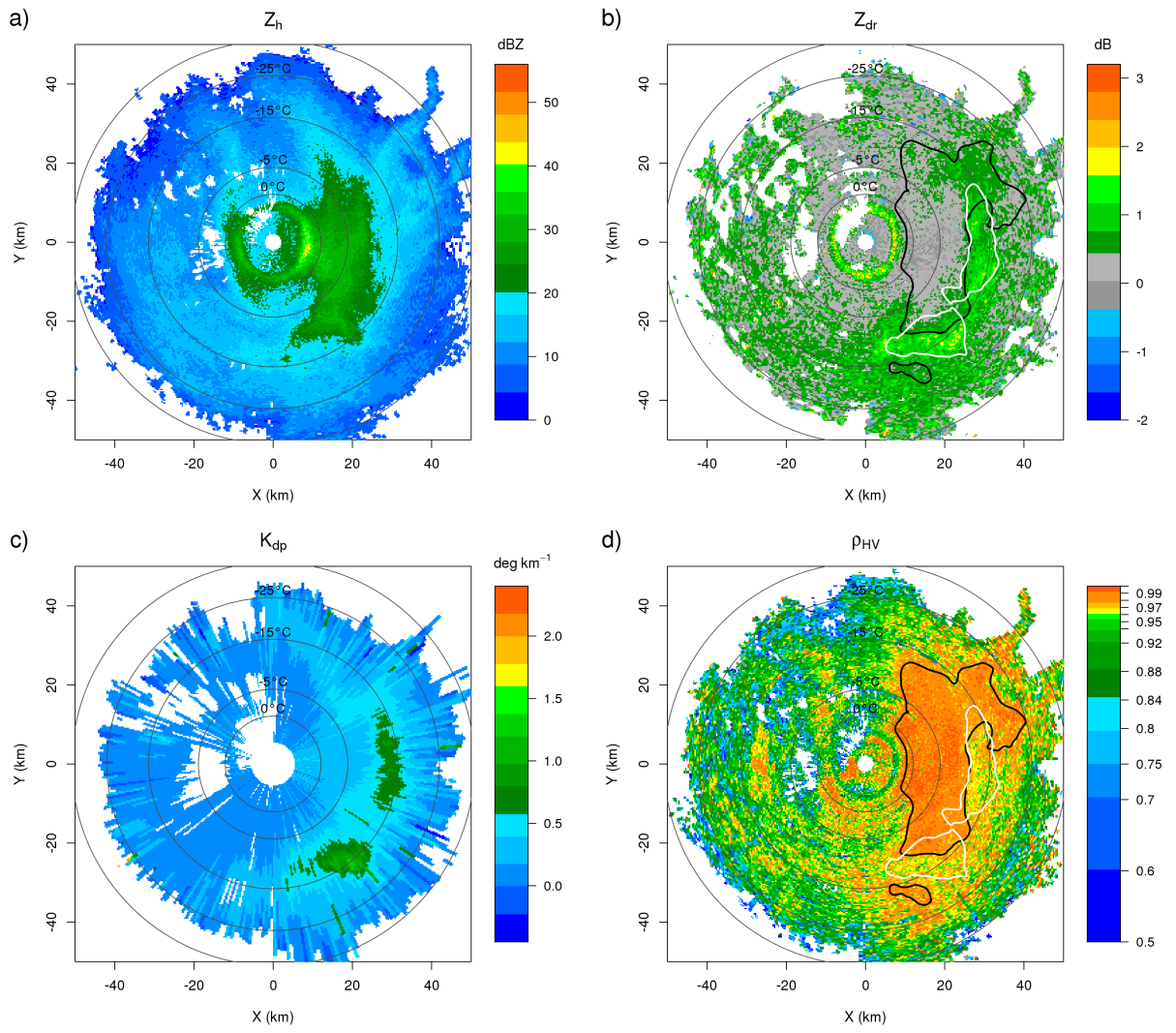


Figure 3.4: PPI at 7.3° elevation at 0025 UTC from the C-band radar. a) reflectivity Z_h, b) differential reflectivity Z_{dr}, c) specific differential phase K_{dp}, d) correlation coefficient ρ_{HV}. The overplotted circles mark several significant temperature levels from the local model analysis at 00 UTC. The thick solid black (white) line in panels b) and d) represent respectively the 0.98 ($0.5^\circ\,\mathrm{km}^{-1}$) level contour of ρ_{HV} (K_{dp}) respectively. Note the overall good spatial matching between the positive Z_{dr} and K_{dp} patterns around -15°C, in a region of relatively low correlation coefficient ($\rho_{\mathrm{HV}} < 0.98$).

and Kennedy and Rutledge [41]. In particular Wolde and Vali [27] analyzed in-situ measurements with microphysical probes and W-band airborne polarimetric observations, reporting Z_{dr} up to 2 dB for dendritic crystals within a nimbostratus cloud at -13°C.

The K_{dp} field (Fig. 3.4c) shows a similar behavior as Z_{dr} at this time, with the highest positive values around $1.3^{\circ}\,\mathrm{km}^{-1}$ at X=20 km, Y=-25 km. It is interesting to note that the positive signatures (Z_{dr} and K_{dp}) lie in a region of relatively low correlation coefficient ρ_{HV} (Fig. 3.4d). The aforementioned positive peak appears confined in an area where $\rho_{\mathrm{HV}} <$ 0.98, with higher correlation coefficient below (closer ranges). The same is observed on the tongue-shaped positive Z_{dr} signature extending to the North, around X=28 km, Y=5 km. In general the region where $Z_{\mathrm{dr}} \geq$1 dB at altitudes where the temperature is lower than -5°C is characterized by ρ_{HV} in the range 0.95-0.98. The transition from relatively low correlation and high Z_{dr} to higher correlation and lower Z_{dr} may indicate the onset of aggregation, with highly oblate crystals being replaced by larger particles with lower density and less anisotropic shapes.

One hour and a half later at 0155 UTC (Fig. 3.5), the melting layer thickened and lowered by about 200 m, from ∼1.8 to ∼1.6 km MSL, apparently as a consequence of the presence of larger aggregates (higher reflectivity above the freezing level) that require more time to melt. On the other hand the main positive Z_{dr} pattern aloft appears on average shifted to a slightly higher altitude, from ∼4.7 to ∼5.1 km MSL. The quantitative distribution of Z_{dr} , and its relation with the correlation coefficient, did not changed significantly, with the southern peak only increasing from 2.2 to 2.4 dB. The appearance of patches of Z_{dr} in the northeastern sector, between -10 and 25°C , suggests a transition to a less stratified structure of the upper cloud. Despite the certain stratiform signatures (principally the well-defined bright band), this moderate cellular structure aloft seems to indicate a degree of turbulence consistent with updraft motions stronger than those associated with the average widespread frontal lifting. In his radar study of a warm frontal region Heymsfield [56] has shown the generating cells above the warm frontal zone to be associated with a distinctly cellular

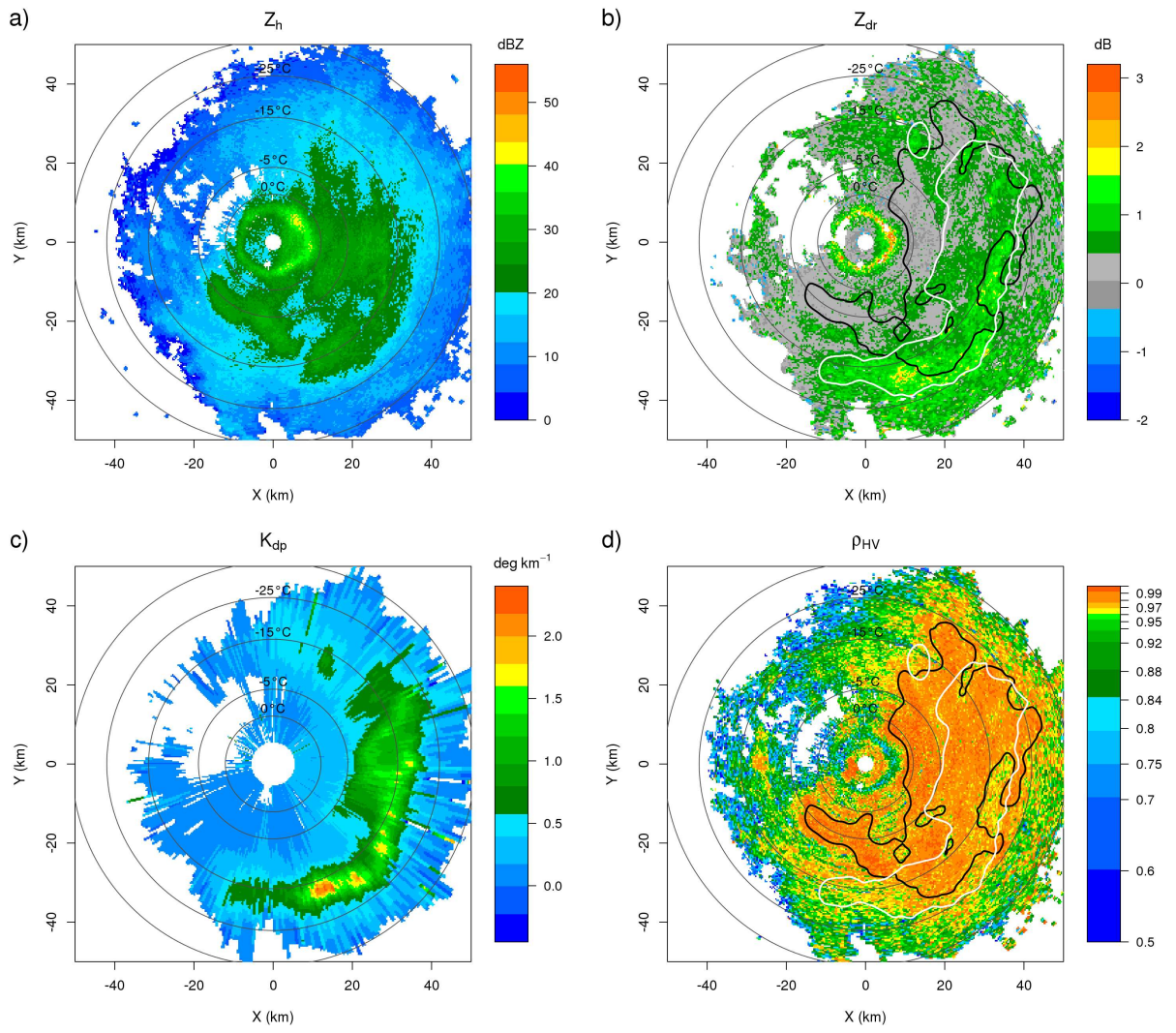


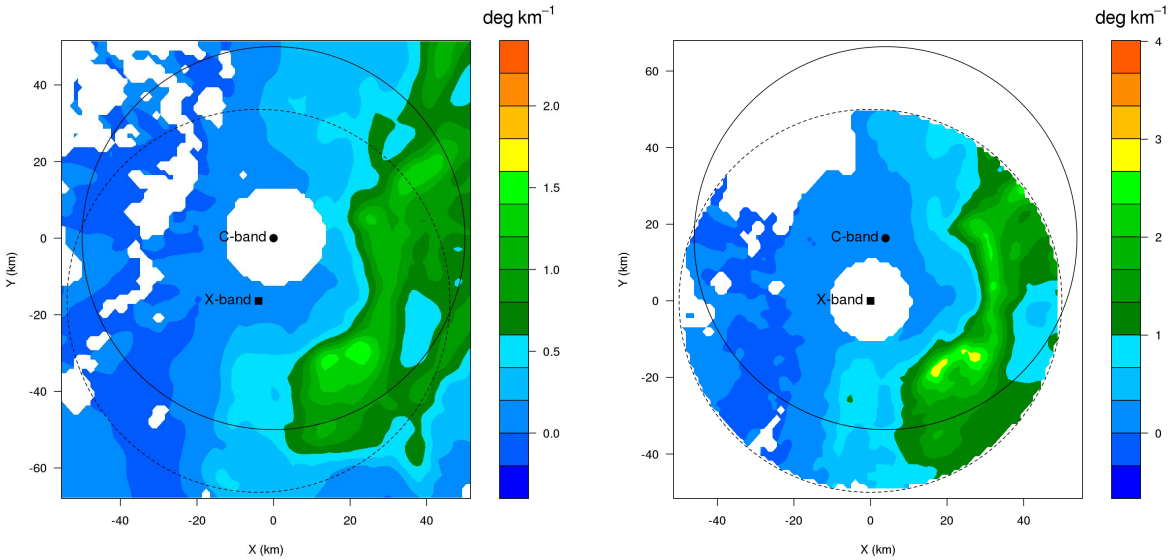
Figure 3.5: Like Fig. 3.4, but at 0155 UTC. In this case the positive $K_{\rm dp}$ pattern extends at lower altitudes, compared to $Z_{\rm dr}$, in the eastern sector (see text of Section 3.2.1).

structure and vertical velocity up to ∼0.4 $\mathrm{m\,s^{-1}}$. According to Korolev and Mazin [52] updraft velocity in the range 0.1-1 $\mathrm{m\,s^{-1}}$ is sufficient for typical ice particle concentrations and sizes to reach the water saturation. In addition, at higher temperatures (e.g. -15 < T < -5°C) the saturation over water can be reached with weaker updrafts, suggesting that the small supercooled droplets may have traveled within the updraft before reaching the altitude of the -15°C isotherm. When the saturation over water is exceeded, the new formation of supercooled droplets contributes to deplete the available water vapor and maintain the air

supersaturation level close to the saturation over water. In this water saturated environment, the most likely crystal habit is the dendritic type ice crystals [57, 58].

While the positive $Z_{\rm dr}$ values aloft did not change significantly between 0025 (Fig. 3.4) and 0155 UTC (Fig. 3.5), the $K_{\rm dp}$ field shows a substantial increase, from a peak value of $1.3^\circ\,{\rm km}^{-1}$ at 0025 UTC to $2.3^\circ\,{\rm km}^{-1}$ at 0155 UTC (southern sector). In addition, unlike the rather good spatial matching between the $Z_{\rm dr}$ and $K_{\rm dp}$ positive patterns at 0025 UTC, 90 minutes later we note how $K_{\rm dp}$ extends to lower heights, particularly in the western sector ($K_{\rm dp} > 0.5^\circ\,{\rm km}^{-1}$ down to the -5°C temperature level) in a region of low differential reflectivity and high $\rho_{\rm HV}$ (>0.98). While $Z_{\rm dr}$ is only related to the (reflectivity-weighted) axis ratio and density of the hydrometeors, $K_{\rm dp}$ is also a function of the mass of the crystals ([25], their eq. 7.101). So when aggregation starts to reshape the PSD near the -15°C level (see also comments about Fig. 16 in Section 4) both the axis ratio and density of the largest particles are reduced, likely causing the sudden decrease of the $Z_{\rm dr}$ values, but a rather smoother decrease of $K_{\rm dp}$. This behavior will be shown to be quite a common feature of the vertical profiles of polarimetric measurements in stratiform precipitation (Section 3.3).

In order to compare the radar observations at C band and X band in the ice region we computed the average values in a 1.3 km-depth layer (4.2-5.5 km MSL) centered on the -15°C temperature level. Fig. 3.6 shows the resulting average $K_{\rm dp}$ fields over the study area at 0155 UTC. There is excellent agreement between the two radar estimates, partly owing to the proximity of the two systems. The X-band $K_{\rm dp}$ estimates are higher resolution compared to the C-band estimates because of the difference in range resolution (125 m versus 340 m, Table 4.3). Hourly average vertical profiles of the polarimetric radar measurements have been calculated for both radars. In addition to the common statistical parameters (mean and standard deviation), the 20-quantiles (vigintiles) have also been calculated [59] to

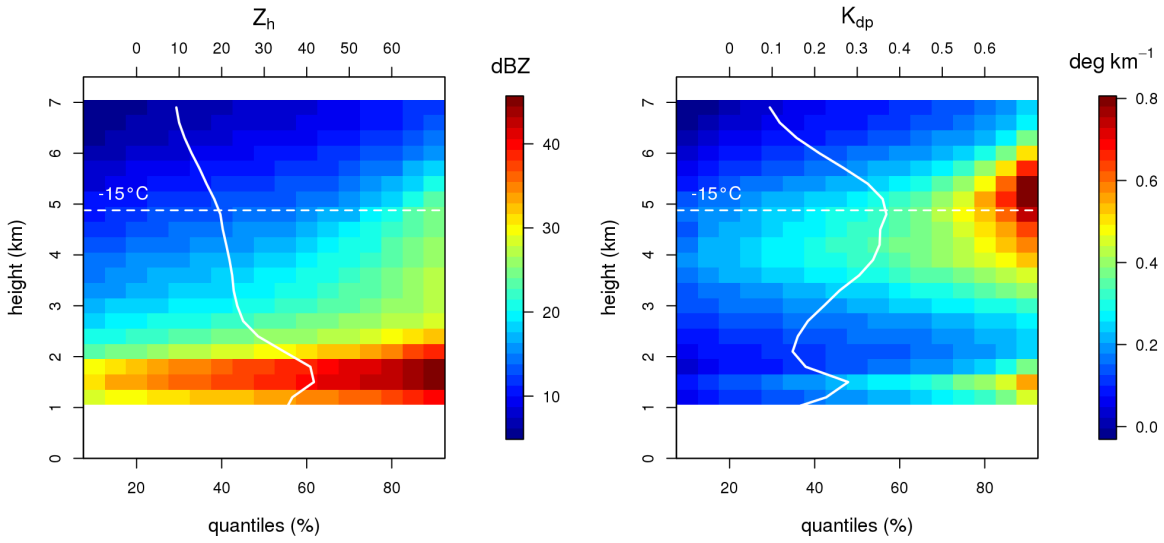


(a) C-band (b) X-band

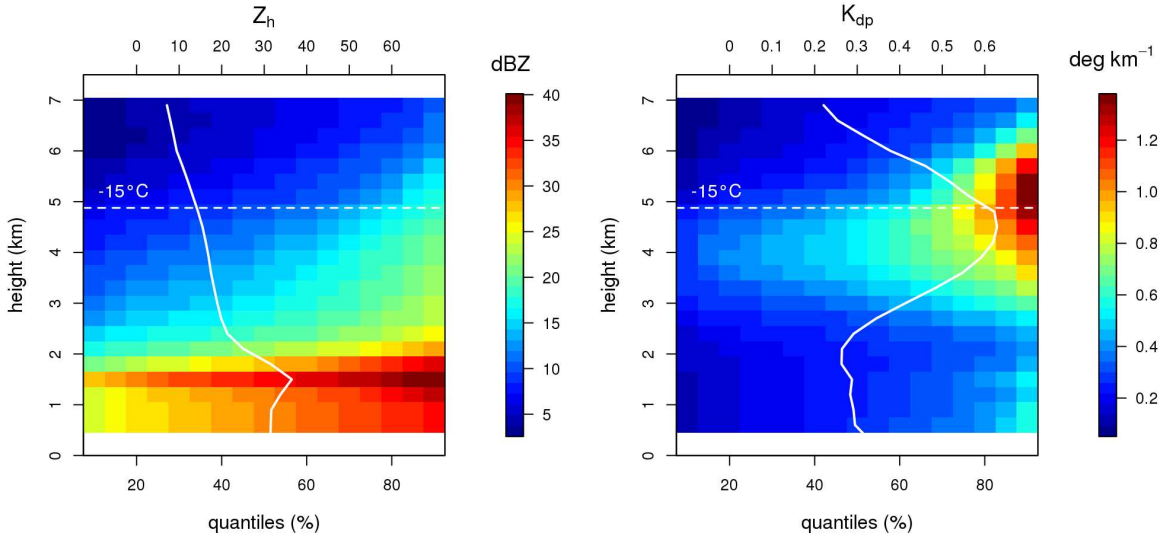
Figure 3.6: Average K_{dp} in the layer 4.2-5.5 km MSL (around -15°C), at 0155 UTC, 27 April 2009. a) C-band radar. b) X-band radar. A 50 km-range ring is overplotted for the C-band radar (solid line) and the X-band radar (dashed line) in both panels. For ease of comparison, the color palettes are scaled by the 1.67 factor, corresponding to the ratio of the operating frequencies.

better represent the distribution of radar measurements, which are in many instances highly non-Gaussian. In fact, the quantiles divide the sample in 20 subsets (5% spaced classes) containing approximately the same number of observations and represent the boundaries between the resulting equal-populated classes. This allows a more comprehensive description of the data and an effective graphical representation of the distribution, with special emphasis on extreme values. Conversely, a simple average of hourly or daily profile may hide significant polarimetric signatures resulting from a transient microphysical process.

Fig. 3.7 show the Z_h and K_{dp} quantiles distribution with height for the C-band (a) and the X-band (b) radar data, corresponding to one hour (04-05 UTC) of light to moderate precipitation. The quantiles are only plotted between 10 and 90% to exclude possible outliers from the visualization. The overlaid average K_{dp} profiles at the two frequency bands show a



(a) C-band



(b) X-band

Figure 3.7: a) Graphical representation of the distribution of reflectivity (left) and K_{dp} (right) at all height levels (y-axis, 0.3 km vertical spacing) for the 04-05 UTC data collected on 27 April 2009 by the C-band radar. Colors represent the values of the quantiles (x-axis) of Z_h (a) and K_{dp} (b). The overplotted white line in both panels represents the average profile (upper scale) and the horizontal white dashed line marks the height of the -15°C level. b) Same as a), but for the X-band radar.

peak between 4.5 and 5.0 km MSL, roughly corresponding to the -13 to -15°C temperature interval as per the 03 and 06 UTC COSMO model forecast. In general K_{dp} in the ice region at X band is observed to be about 1.7 times higher than K_{dp} at C band, in good agreement with the ratio of the operating frequencies of the two radars. Peak values of observed K_{dp} during the event are $\sim 2^{\circ}\,\mathrm{km}^{-1}$ at C band and $\sim 3.5^{\circ}\,\mathrm{km}^{-1}$ at X band, respectively. Such values are consistent with the scattering simulations reported in Kennedy and Rutledge [41]. When K_{dp} values are well above the typical measurement error (i.e. $K_{\mathrm{dp}} > 0.2^{\circ}\,\mathrm{km}^{-1}$ at C band), K_{dp} fields at the -15°C level at C and X band are highly correlated (the Pearson's correlation for the pairs of K_{dp} maps is r>0.9), indicating the robustness of the K_{dp} estimates in the ice region. Reflectivity fields instead are less correlated (0.7<r<0.9) likely due to the variability of the response of C vs. X band to a given PSD and to path attenuation uncertainties determined both in rain and melting layer.

Figure 3.8 shows the frequency plot of K_{dp} at C and X band at the -15°C level for the first 10 hours of the day, when the highest rainfall amounts were recorded by the surface gauges in the area (accumulations between 12 and 51 mm). The relation between the two frequencies appears roughly linear, although for higher K_{dp} the C-band values underestimate the corresponding X-band scaled estimates. This behavior only concerns the peak values and it is believed to be a simple consequence of the different range resolution (Table 4.3), for which the higher range resolution at X band allows to detect more intense local maxima.

3.2.2. Comparison with electromagnetic scattering simulations. T-matrix simulations of scattering at C and X band have been performed to interpret radar measurements at the -15°C isotherm level, using simplified assumptions based on those adopted by Kennedy and Rutledge (2011). The population of particles at the -15°C level is modeled as a mix of highly oblate spheroid crystals for diameter smaller than 3 mm with axis ratio varying

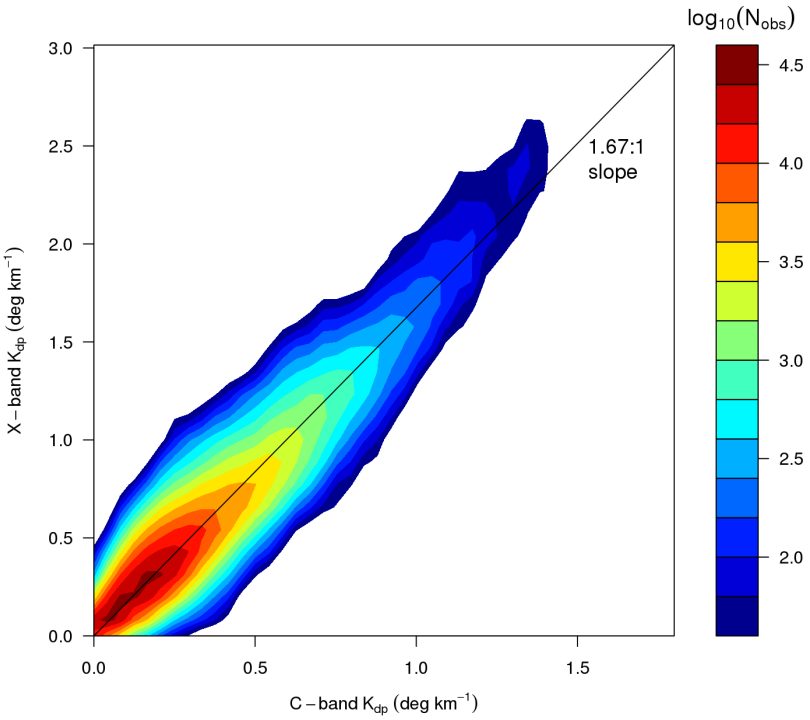


Figure 3.8: Frequency plot of $K_{\rm dp}$ at C-band (x-axis) and X-band (y-axis) for the ten hours interval 0000-0955 UTC on 27 April 2009, in the layer 4.2-5.5 km MSL (around $-15°$C). The color scale represents the number of observations $\rm N_{obs}$ in logarithmic unit, using $0.05°\,\rm km^{-1}$ intervals.

between 0.05 and 0.15 and bulk density values as reported in Heymsfield et al. (2004) and aggregates for larger diameters. Exponential PSD are assumed with N_0 varying between $50{\times}103$ to $400{\times}103$ $\rm cm^{-1}m^{-3}$ and Λ between 25 and 45 $\rm cm^{-1}$ [60, 41]. Simulations show that Z_h values are not significantly influenced by the frequency, $K_{\rm dp}$ can be scaled according to the ratio of wavelengths and, finally, resonance effects such as the differential phase shift upon backscattering are negligible at both frequencies.

Figure 3.9 shows the observed $K_{\rm dp}$ vs. Z_h dispersion at C band (panel a) and X band (b), together with the results from the scattering simulations. The observed radar variables are the averages in the 4.2-5.5 km height layer during ten consecutive hours, as in Fig. 11. Given the influence of wet radome attenuation on X-band power measurements, the reflectivity

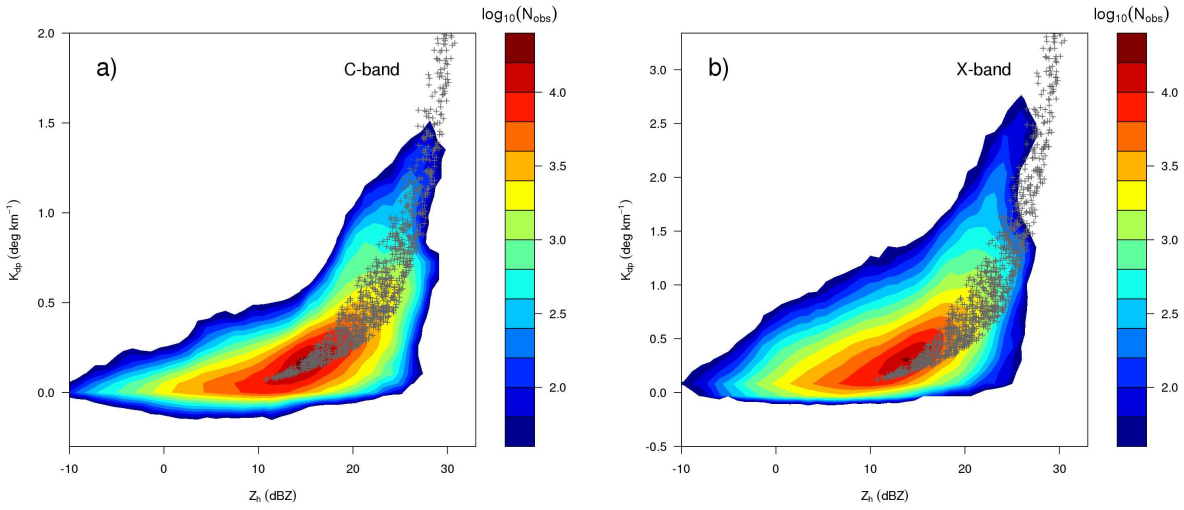


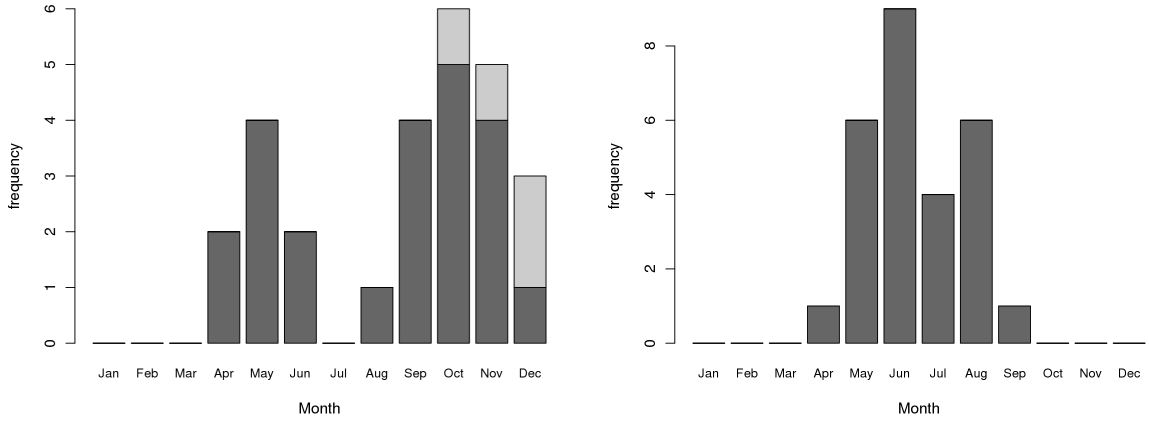
Figure 3.9: Frequency plot of K_{dp} vs. Z_h (colors) at C-band (a) and X-band (b) for the ten hours interval 0000-0955 UTC on 27 April 27 2009, in the layer 4.2-5.5 km MSL (around -15°C). The color scale represents the number of observations Nobs in logarithmic units, using 1 dB and $0.05^{\circ}\,\mathrm{km}^{-1}$ intervals for Z_h and K_{dp} respectively. Overplotted (gray crosses) are the simulated values for the PSD representing ice particles at the -15°C level.

values for each scan have been corrected based on the 1-minute rain rate measurements from the co-located optical disdrometer [61]. A fairly good agreement at C band is noted for the positive K_{dp} values. The agreement is slightly worst at X band due to the higher dispersion of the observations and a reflectivity bias of about -1.4 dB, compared to the C-band measurements. The higher dispersion of the X-band reflectivity is likely related to the increased uncertainty affecting the power measurements due to 1) residual wet radome attenuation and 2) under-estimation of the path attenuation. The average -1.4 dB bias between the X-band and C-band radar may also be a result of the excess attenuation at the higher operating frequency.

3.3. Analysis of polarimetric vertical profiles from C-band radar observations

Space-time average vertical profiles allow a statistical evaluation of the vertical structure of the atmosphere from radar measurements. Specifically the relation between measurements at different height levels can be addressed without explicitly considering the height-dependent advection, relying on the assumption that within a given observation period (∼hours) the observed cloud water content aloft will mainly precipitate within the same (sufficiently large) sampling area. More than one year of data routinely collected at C band between 2009 and 2010 by the operational radar was analyzed leading to a selection of 54 significant rainy days. This choice was objectively made based on the regional raingauge network by selecting only the days when at least 5 mm of cumulative precipitation was recorded by at least one raingauge over the area of interest, defined in this instance by the 50 km-range area around the C-band radar only. Daily and hourly profiles were calculated for all polarimetric measurements, allowing to work on a wide statistical sample and check possible differences at the two time scales. In order to perform an objective partitioning of the dataset into stratiform/convective cases, the characteristic distribution of reflectivity vertical profiles described for example in Steiner et al. [62] and Yuter and Houze [63] is exploited to define a metric for convectivity named Radar Convective Parameter (RCP), described in the Appendix B.

For the purpose of the following analyses we arbitrarily define an event to be stratiform when RCP is lower than 50th percentile and convective elsewhere. Figure 3.10 shows the histograms of the resulting monthly distribution of stratiform and convective days. A well-defined seasonal distribution arises from the automated classification, with convective events mainly concentrated in the summer months between June and August and stratiform events more frequent in autumn and spring. May is typically a transition period, with



(a) Stratiform (b) Convective

Figure 3.10: Histogram of the monthly distribution of the 54 selected events. a) Stratiform cases. The light gray bars represent the events with freezing level below 1.5 km MSL. b) Convective events. The stratiform/convective classification is based on the RCP value (Appendix B). Note the bimodal distribution for stratiform precipitation, with peaks during spring and fall. On the other hand, the convective events are concentrated during the summer months.

both stratiform and convective events. Winter is the driest season in northern Italy, the dataset analyzed being not an exception, with no significant precipitation between January and March.

3.3.1. Polarimetric vertical profiles and their relation with crystal habit.

Figure 3.11 shows the hourly vertical profiles of the four radar variables Z_h , $Z_{\rm dr}$ $K_{\rm dp}$ and $\rho_{\rm HV}$, with colors representing the RCP value. The decreasing magnitude of the correlation coefficient (Fig. 3.11d) for lower RCP values can be ascribed to the bias introduced by the conventional zero lag $\rho_{\rm HV}$ estimator at low SNR [64], more likely to occur in stratiform precipitation. The change from stratiform to convective conditions can be clearly seen by the vanishing of the bright band nose in the reflectivity profiles, fairly correlated with the increasing RCP. The melting layer causes a well-defined local enhancement Z_h , $Z_{\rm dr}$ and $K_{\rm dp}$,

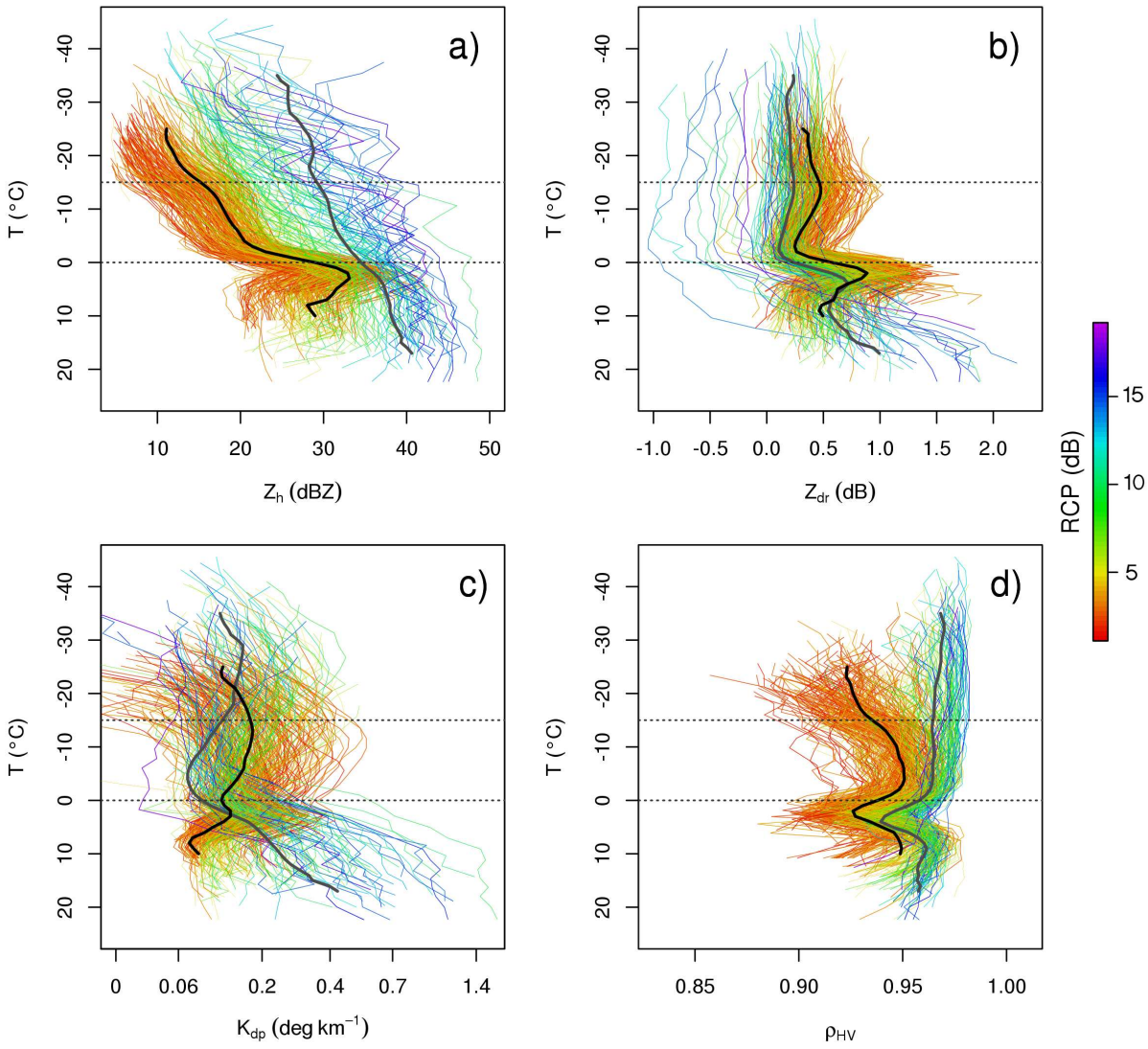


Figure 3.11: Hourly vertical profiles of C-band Z_h (a), Z_{dr} (b), K_{dp} (c) and ρ_{HV} (d) colored according to their respective RCP value. The RCP quantiles (0, 25, 50, 75, 100%) are respectively (1.1, 2.7, 3.9, 7.3, 21.7 dB). The black (gray) thick lines represent the average of the daily profiles for stratiform (convective) events. It is possible to note several Z_{dr} convective profiles (high RCP) notably affected by differential attenuation (negative values up to -1 dB). In order to highlight the variations for small values, the K_{dp} profiles are plotted on a log-axis.

occurring (with slight differences, depending on the considered measurements) just below the freezing level, evident in profiles with RCP lower than approximately 5 dB (see also the solid thick lines in Fig. 3.11, representing the average of all daily profiles). The melting layer

is also very well captured in the profile of ρ_{HV}, by virtue of its sensitivity to the diversity of orientation, shape, size and thermodynamic phase of the hydrometeors. Interestingly, the melting layer signature is well defined in Z_{dr} and ρ_{HV} also for profiles with higher RCP, i.e. with increasing convective characteristics [65].

The increase around the -15°C level on the other hand is only depicted in the Z_{dr} and K_{dp} profiles. As already noted in Section 3.2, the peak appears sharper in Z_{dr} than in K_{dp}, likely as a consequence of the different shape and mass-dependency of the two variables. The analysis of Z_{dr} reveals that the peak around -15°C in the hourly profiles for stratiform precipitation varies between 0.1 and 1.3 dB (mean values), while the 90th quantile span over 0.5-2.5 dB, without any apparent relation with the reflectivity. K_{dp} instead shows a fairly defined increase with reflectivity, which is further discussed in the next subsection. Also noteworthy is the sudden increase in ρ_{HV} below the -15°C level for stratiform precipitation (Fig. 3.11). A similar increment is also observed in the reflectivity profiles (while Z_{dr} decreases), being indicative of the transition to a warmer layer, where aggregation produces larger particles with lower density and the overall anisotropy of the medium is reduced.

The generally moderate Z_{dr} peak encountered in this study ($Z_{\mathrm{dr}} \leq 2.5$ dB for 90% of the data) suggests that values as high (6-7 dB) as reported by either Wolde and Vali [27] for high density plate crystals in the size range 0.2-1.5 mm or by Williams et al. [50] are relatively uncommon in our geographical region for the type of meteorological events considered. Actually, Wolde and Vali [27] observations of dendritic crystals with sizes up to 4-6 mm ($\sim$2 dB at -13°C) refer to nimbostratus (Ns) observations, while the highest observed Z_{dr} (up to 7 dB) refer to hexagonal plates in shallow altocumulus (Ac), which are typically not associated with surface precipitation. We note in this context that the 5 mm threshold on the surface rainfall adopted as a selection criterion for this study led to select

a number of events characterized by similar warm frontal synoptic conditions, especially relevant for their hydrological impact. While the mesoscale forcing associated with a warm front passage is expected to produce large areas of weak updrafts, the vertical velocity field is certainly modulated by smaller scale turbulence, as evidenced by the frequent insurgence of the $Z_{\rm dr}$ patched patterns shown in Fig. 3.5. Outside the weak updraft regions it is then reasonable to expect that the air saturation over water is not always reached, making the environment favorable to hexagonal plate crystals [57, 58]. However, due to the much smaller sizes of these higher-density crystals, and considering the sensitivity limitations of our cm-wavelength radar (Section 2), the fact that we did not observed in the cases analyzed a correspondingly high $Z_{\rm dr}$ cannot be considered an indication of the lack of hexagonal plate crystals in these clouds.

Korolev et al. [66] provided statistics from four measurement campaigns in stratiform clouds. Although the plate category was not included in that study (due to the insufficient resolution of the particle measuring system), the frequency of occurrence in the -10°C to -15°C temperature interval of the largest particles (>500 μm, most likely to influence radar measurements) is shown to be dominated by irregular crystals (particles having an irregular or random shape - 64%), followed by dendrites (28%). Given this general partition, it is not surprising to observe in our region the average $Z_{\rm dr}$ distribution shown in Fig. 3.11.

3.3.2. $K_{\rm dp}$ ENHANCEMENT IN THE ICE REGION. The $K_{\rm dp}$ enhancement around the -15°C level in the ice region of stratiform situations (Fig. 3.11c) appears close to a local maximum in the height derivative of the reflectivity vertical profile (Fig. 3.11a, the slope of thick black line increases towards the ground after crossing the -15°C level). According to Lo and Passarelli [60], the increased height derivative of reflectivity may be interpreted as an

indication of the transition from vapor deposition to aggregation processes. Although aggregation normally occurs at temperatures warmer than -5°C , a secondary maximum between -10 and -16°C may exist when the arms of the dendritic crystals become entangled [67].

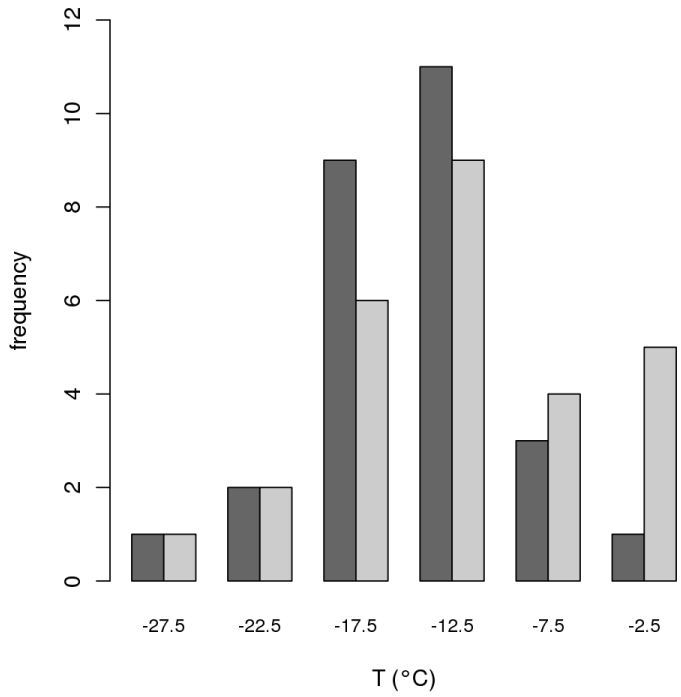


Figure 3.12: Histogram of the temperature distribution of the maximum K_{dp} within the ice region ($\mathrm{T} < 0°\mathrm{C}$) from the daily vertical profiles. Dark bars refer to the daily 90th percentile K_{dp}, while light bars to the daily average K_{dp}.

Figure 3.12 shows the temperature distribution of the maximum K_{dp} location in the ice region ($\mathrm{T} < 0°\mathrm{C}$) for the profiles classified as stratiform. The 5°C-interval histograms show both the temperature distribution of the peak values derived from the daily average K_{dp} (light bars) and from the daily 90th percentile (dark bars). The peak of the 90th percentile is well defined and centered at -14.5°C (mean value), while the average K_{dp} shows a less pronounced peak around -12°C. This is a clear indication that positive values of K_{dp}, when observed in the ice portion of the precipitation system, are preferentially associated with the dendritic

growth zone around -15°C , as already pointed out in the case study discussion of Section 3.2. The average $K_{\rm dp}$ profiles are smoothed by averaging process and a local peak may not be easily recognizable. In fact, processes like growth by vapor deposition, as well as riming by small supercooled droplets when the air is saturated over water, are enhanced by updrafts, which normally occur over a confined space-time portion of the stratiform precipitation event.

3.3.3. Statistical correlations between observations at -15°C and lower levels. Using particle trajectory calculations, Kennedy and Rutledge [41] have shown the linkage between the positive $K_{\rm dp}$ aloft and the snowfall intensification at the surface. While their conclusions were limited to snow, from a hydrological point of view it is extremely valuable to identify specific radar observations that may represent a precursor for the surface rainfall onset or intensification. We therefore exploit the large dataset of radar vertical profiles elaborated for the present study, seeking for statistical correlation between the radar observations aloft (specifically the $K_{\rm dp}$ enhancement) and near the surface.

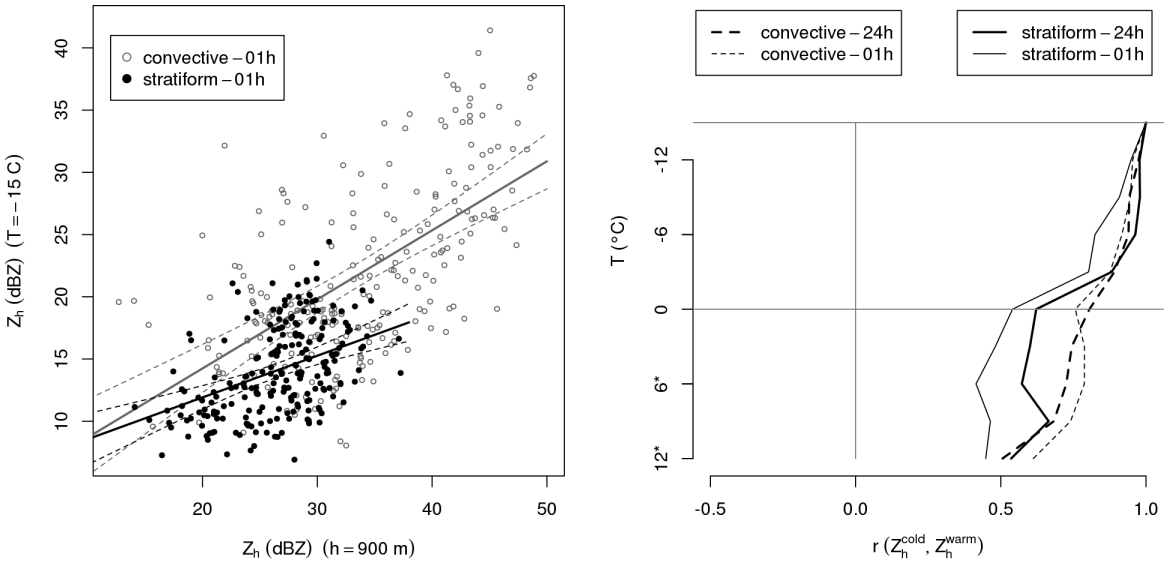
For the following analysis only events in which the freezing level was above 1.5 km MSL are considered (dark gray bars in the histogram of Fig. 3.10), in order to avoid vertical gradients associated with melting. This is done to allow meaningful comparisons with the lowest level (900 m MSL) radar averages of Z_h and $K_{\rm dp}$, which are taken as representative of the mean rainfall rate near the surface. Among the initial 54 rainy days, only four were discarded based on the above threshold on the freezing level. We denote as $r(P^{cold}, Q^{warm})$ the Pearson correlation coefficient between a radar measurement P at the upper *cold* level of -15°C (P^{cold}) and the measurement Q at warmer temperatures (Q^{warm}, where $warm$ denotes temperatures above -15°C at lower heights). Specifically, we are interested in analyzing (P^{cold}, Q^{warm}) pairs such as $(Z_{\rm h}^{cold}, Z_{\rm h}^{warm})$, $(\tilde{K}_{\rm dp}^{cold}, \tilde{K}_{\rm dp}^{warm})$, $(\tilde{K}_{\rm dp}^{cold}, Z_{\rm h}^{warm})$, where $\tilde{K}_{\rm dp} = 10\log_{10}(K_{\rm dp})$ is used to linearize the relation between Z_h and $K_{\rm dp}$ (Section 3.2). Note that

this procedure is not directly affected by the change in the dielectric factor among profiles since the correlation is calculated between two radar variables, each at a given temperature level, e.g. at -15°C (cold) and Z_h at +5°C (warm). However, some minor influence on the correlation at different levels is possible since the change in the dielectric factor affects the radar detectability of particles.

In general the microphysical processes relevant for the precipitation growth depend on the temperature, and not on the absolute height. Therefore, we want to compute statistics as a function of the air temperature (taken from the COSMO local area model) to compare vertical profiles from different meteorological events. In order to compensate for varying surface temperature, the temperature at altitudes below the melting level (T > 0°C) is normalized to the average (over all events) lowest level temperature of 12°C, according to:

$$\tilde{T} = 12 \frac{T}{T_{900m}} \,^{\circ}\mathrm{C} \tag{9}$$

Figure 3.13a shows a scatterplot of the -15°C hourly reflectivity versus the lowest level reflectivity for convective (open circles) and stratiform (filled circles) events. The overplotted regression lines with confidence intervals emphasize the significant positive correlation between the reflectivity aloft and near the surface for both stratiform and convective precipitation. Fig. 3.13b summarizes the computed correlation values for all the temperature levels, the correlation, with 3°C-spacing, for both hourly and daily profiles. In Fig. 20 the analogous $r(\tilde{K}_{\mathrm{dp}}^{cold}, \tilde{K}_{\mathrm{dp}}^{warm})$, correlation is presented. The correlation is calculated on a total of 526 hourly profiles (260 stratiform and 266 convective) and on the 50 daily profiles. Although the daily profiles form a much lower statistical sample (23 stratiform and 27 convective events), the same qualitative results are found.



(a) (b)

Figure 3.13: a) Scatterplot of hourly average Z_h aloft (-15°C) versus Z_h at the lowest level (900 m MSL) for stratiform and convective precipitation. The solid black (gray) line is the regression for stratiform (convective) cases; the dashed lines mark the 99% confidence interval. b) Correlation coefficient between Z_h aloft (-15°C) and Z_h at lower levels (3C-spacing) for both hourly and daily average profiles. The superscript (*) on the y-axis temperatures indicate normalized values (eq. 9).

The statistical significance of the correlations between the -15°C level and the lower levels subsequently reported is always ensured at the 0.5% level (α=0.005 is the significance level under which the null hypothesis of no correlation is rejected) for both hourly and daily stratiform profiles. Reflectivity shows a decreasing correlation from the initial value of unity at -15°C (correlation between the -15°C level reflectivity with itself) to approximately 0.4-0.6 at the lowest level. There is a qualitative difference between stratiform and convective cases, with lower correlation in the melting layer for the stratiform precipitation and a more linear trend for convection. This behavior, and the higher slope of the regression between Z_h aloft and near the surface (Fig. 3.13a), are explainable in terms of the lower stratification (higher mixing) for the convective cases and keeping in mind that the small scale variability

if filtered out by space-time integration in the calculation of the mean profile. No significant differences are observed between the 1-hour and the 24-hours correlation profiles. This result is indeed expected, since it forms the basis for the widely used Vertical Profile of Reflectivity (VPR) correction methods [68, 69].

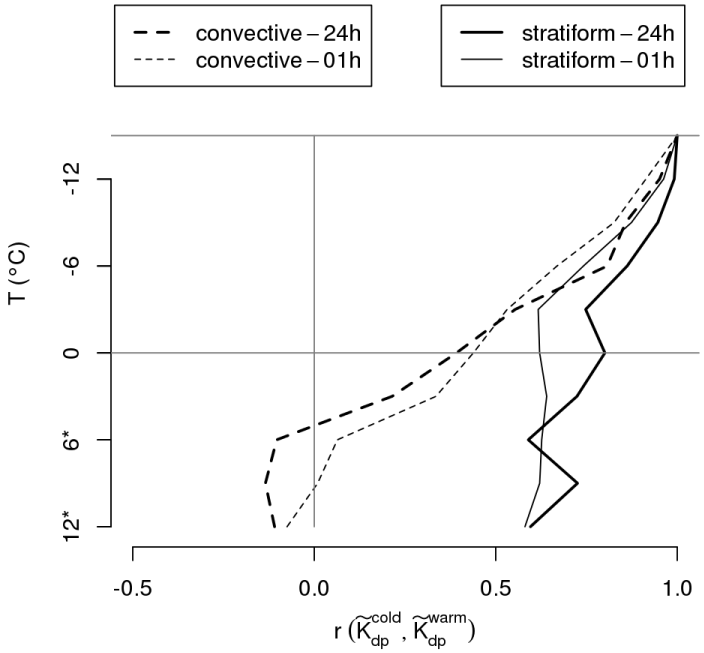
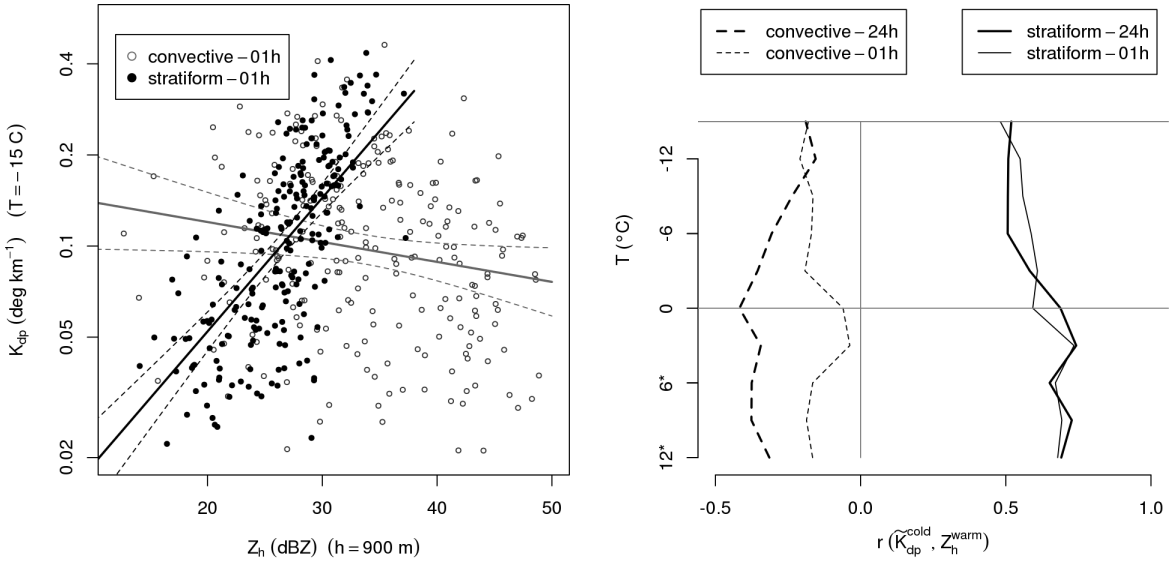


Figure 3.14: Correlation coefficient between log-transformed $K_{\rm dp}$ aloft (-15°C) and $K_{\rm dp}$ at lower levels (3°C -spacing) for both hourly and daily average profiles.

Less predictable is the $r(\tilde{K}_{\rm dp}^{cold}, \tilde{K}_{\rm dp}^{warm})$ correlation (Fig. 3.14). For convective cases the correlation decreases rapidly to approximately zero below the melting layer, indicating that the specific differential phase in the rain medium has, in general, no relation with the values observed aloft in the ice region. On the other hand, for stratiform cases, the $r(\tilde{K}_{\rm dp}^{cold}, \tilde{K}_{\rm dp}^{warm})$ correlation shows a behavior closer to the $r(Z_{\rm h}^{cold}, Z_{\rm h}^{warm})$ correlation, with moderate values (> 0.5) at all temperature levels.

For stratiform events the measured range of the hourly 90th percentile $K_{\rm dp}$ aloft over all the events is -0.02 to 0.99° km^{-1}, with 8% of the values of $K_{\rm dp} < 0.1$° km^{-1}. On the other hand, the range of hourly 90th percentile $K_{\rm dp}$ at the lowest 900 m height level is 0.04 to 0.59° km^{-1}, with a higher fraction (31%) below the 0.1° km^{-1} threshold. We notice that the correlation of $K_{\rm dp}$ aloft with the corresponding $K_{\rm dp}$ at the lower levels could be affected by the lack of sensitivity of $K_{\rm dp}$ measurements in the light rain, such as that typically occurring in light rainfall intensities of stratiform precipitation systems. For this reason it is useful to consider also the $r(\tilde{K}_{\rm dp}^{cold}, Z_{\rm h}^{warm})$ correlation, i.e. the correlation between $K_{\rm dp}$ at the -15°C temperature level and Z_h at lower altitude levels, since Z_h has no such sensitivity limitation within the 50 km range.



(a) (b)

Figure 3.15: a) Scatterplot of hourly average $K_{\rm dp}$ aloft (-15°C) versus Z_h at the lowest level (900 m MSL) for stratiform and convective precipitation. The solid black (gray) line is the regression for stratiform (convective) cases; the dashed lines mark the 99% confidence interval. b) Correlation coefficient between $K_{\rm dp}$ aloft (-15°C) and Z_h at lower levels (3°C -spacing) for both hourly and daily average profiles. The superscript (*) on the y-axis temperatures indicate normalized values (eq. 9).

A scatterplot of the hourly average $K_{\rm dp}$ at the -15°C temperature level versus Z_h at the lowest 900 m level, with regression lines for stratiform and convective events, is shown in Fig. 3.15a. The regression line over the stratiform sample shows a positive slope indicating an increase in $K_{\rm dp}$ with increasing Z_h . On the contrary, the points representing convective cases appear mostly randomly distributed with a weak negative correlation. Fig. 3.15b shows the correlation summary for all levels. For stratiform cases the correlation varies between approximately 0.5 above the freezing level and 0.7 below, for both hourly and daily profiles.

If we enforce a stricter definition of stratiform precipitation, considering only events with RCP $<$ 25th percentile, then the correlation between $K_{\rm dp}$ at -15°C and Z_h at 900 m increases to approximately 0.8 for both hourly and daily profiles, while it remains around 0.5-0.6 between Z_h at -15°C and Z_h at 900 m (not shown). This marked correlation can be seen as an indication of the ability of $K_{\rm dp}$ aloft to represent the ice water content (IWC) ([25], their eq. 7.17). Since we are applying a relevant space-time averaging in the derivation of the vertical profiles, Z_h in the rain layer is well related to the precipitation liquid water content (LWC). In fact a Marshall-Palmer exponential relation represents fairly well the drop size distribution (DSD) after enough space-time averaging [70]. For ice PSD mainly arising from water vapor deposition (plate-like crystals) the IWC has been shown to be linearly related to $K_{\rm dp}$, as opposed to the power law relation between the IWC and Z_h (Vivekanandan et al. 1994). In the absence of other microphysical processes contributing to IWC generation at lower altitudes (riming, accretion) and neglecting evaporation, the LWC below the freezing level is expected to be strictly connected to the IWC around -15°C . This is because the particle distribution undergoes modifications mostly through aggregation in the layer between -15°C and 0°C and therefore should keep its IWC essentially unchanged. The correlation

between $K_{\rm dp}$ aloft and Z_h in rain is then taken as an evidence of a plausible representation, on average, of the IWC by $K_{\rm dp}$ in this particular ice region of the stratiform clouds.

The results shown in Figs. 3.12-3.15 clearly reveal the existence of a statistically significant link (at the 0.5% significance level) between peak values of $K_{\rm dp}$ in the ice region and the precipitation near the surface for mesoscale widespread precipitation systems. No positive correlation is found for convective precipitation, for which other processes such as heavy riming may overwhelm the depositional growth of ice. A slightly negative correlation is found actually for convective events (Fig. 3.15b), more pronounced (r$\sim$-0.3), but less significant (15% significance level) for daily average profiles. This negative correlation, opposite to the stratiform case, can be attributed to the increasing role of particle growth by riming, which contributes to generate higher-density particles with lower $K_{\rm dp}$ in the ice region and larger Z_h over the whole vertical profile, resulting in higher rainfall intensity near the surface.

Unlike $K_{\rm dp}$, the reflectivity aloft was shown to be fairly correlated (r$\sim$0.5) with the lowest level reflectivity for both stratiform and convective precipitation (Fig. 3.13). The $K_{\rm dp}$ peak around the -15°C temperature level appears to be therefore a distinguished feature of stratiform precipitation, where the vapor deposition mechanism, possibly combined with condensational growth of liquid droplets in the water saturated cloud, can be considered the dominant ice-particle growth processes [71, 53].

3.4. Discussion and conclusions

Differential phase shift measurements and specific differential phase estimators have received considerable attention in recent years mainly due to their potential for hydrological applications [72, 73]. $K_{\rm dp}$-based rainfall rate estimators in particular have become increasingly popular, especially at attenuating frequencies such as C band and X band for their

insensitivity to path attenuation and the increased sensitivity for light rain measurements at these frequencies [74]. In contrast, measurements in the ice region of precipitating systems where the hydrometeors can exhibit considerable anisotropy to dual polarization radar methods have been only occasionally considered [37–41]. The present study has extended the results of Kennedy and Rutledge [41] who analyzed K_{dp} at S band in several winter storms in Colorado. For the present research a large dataset of measurements collected between 2009 and 2011 by the operational C-band radar of Bric della Croce (Torino, Italy) was processed. Additionally, a single significant stratiform event observed by the C-band radar and by the nearby deployed ARX transportable X-band system has been studied in more detail.

A first result of the present work elucidates the microphysics behind the studied signatures of K_{dp} in stratiform precipitation. The analysis of the coincident C-band and X-band measurements in the ice region demonstrated that, in agreement with electromagnetic scattering simulations, the ice particles in the region centered around -15°C are Rayleigh scatterers because K_{dp} scales with frequency. Vapor deposition, associated with riming by small supercooled droplets at water saturation [53], is the most relevant snow growth process in stratiform clouds, occurring in the region where the difference between the saturation vapor pressure over water and the saturation vapor pressure over ice is the greatest. In this region, between approximately -12 and -16°C , plate-like crystals present the most effective shape for the deposition of the ambient water vapor by virtue of the high surface-to-volume ratio [67]. The ice crystal habit is in general a function of both temperature and supersaturation (with respect to ice). In particular, the appearance of hexagonal plate crystals (high density) or dendrites (lower density) in the -10 to -20°C is governed by the relative humidity and the consequent level of supersaturation in the cloud [58]. As noted in Williams et al. [50], dendritic crystals represent the most likely crystal habit for the considered temperature

range in water saturated regions. A detailed analysis of the polarimetric radar signatures for a single case showed qualitative and quantitative agreement with previous observations of dendritic crystals [27]. The following evaluation of the polarimetric characteristics of the average vertical profiles for 27 days of stratiform precipitation allowed to extend this evidence and to deduce that dendrite crystals represent a common crystal habit in the -15°C region, for autumn and spring stratiform precipitation in the Italian subalpine region.

The presence of dendritic crystals, whose branches are prone to become entangled [67] and collect the available supercooled droplets to form rimed aggregates, is probably one plausible reason for which these signatures are so easily revealed by cm-wavelength radars [55, 41, 50], whose detection capabilities are strongly dependent on the availability of large scatterers. This leads us to remark that, although dual-polarized radar observations provide an invaluable tool for the observation of ice clouds, the crystal habits and relevant microphysical processes can only be comprehensively depicted using a wider range of observing instrumentation systems, ultimately relying on in-situ measurements.

Secondly, statistical evidence is presented that the K_{dp} signature in stratiform precipitation occurs near the -15°C level, irrespective of altitude. Being independent of the precipitation type at the ground, it appears to be a relevant feature not only in snowfall, but more generally for widespread rainfall events, which in the area of interest are most likely to occur during spring and fall. For the 27 days analyzed the height of the -15°C temperature level ranged between 3.9 and 6.0 km MSL and the surface daily average temperature between 1.3 and 16.9°C. For over 70% of the cases the maximum value of K_{dp} above the freezing level was found between -10 and -18°C. A restriction of the dataset to include only the days with the freezing level height above 1.5 km MSL (23 stratiform daily profiles, 260 hourly profiles) was applied to calculate the correlation between the radar observables (Z_h and K_{dp}) aloft

and the corresponding observations in rain at the lowest available level (900 m MSL). Since the low level average Z_h can be taken as adequately representative of the surface precipitation rate for light to moderate intensities, the correlation found between K_{dp} in ice at the -15°C level and Z_h in rain (r$\sim$0.7, Fig. 3.15) is considered especially relevant for potential hydrological applications and short term precipitation forecasting.

The skeweness of the reflectivity distribution is considered to define a convective parameter named RCP (Radar Convective Parameter, described in Appendix). More generally, the skewness of the distributions of the polarimetric variables is an important characteristic that could be further exploited to derive valuable information about the microphysical state and processes of precipitation particles. A local characterization of the distributions, e.g. by n-quantiles calculation over limited space-time domains, may also have a potential to benefit the performance of hydrometeor classification schemes [75, 4].

Fig. 3.16 finally summarizes the correlation statistics (K_{dp} aloft vs. the near-surface reflectivity) as a function of the RCP value associated with the analyzed hourly profile. In this plot the Pearson correlation r and the rank (Spearman) correlation r_S are shown for ten equal-sized classes, each bar representing the correlation calculated over approximately 50 samples. The rank correlation does not assume a linear relationship, but just a monotonic one, so there is no need to linearize the K_{dp} - Z_h relation by taking the log values of the former variable, as for the Pearson correlation. The rank correlation is in general higher than the Pearson correlation, as expected, being more robust to outliers [76]. The plot highlights the dramatic difference between stratiform and convective precipitation. In particular, the trend presents a marked step change around RCP = 4 dB, with the lowest 40% of the total sample having a rank correlation near $r_S = 0.8$. This presents clear evidence that the statistical

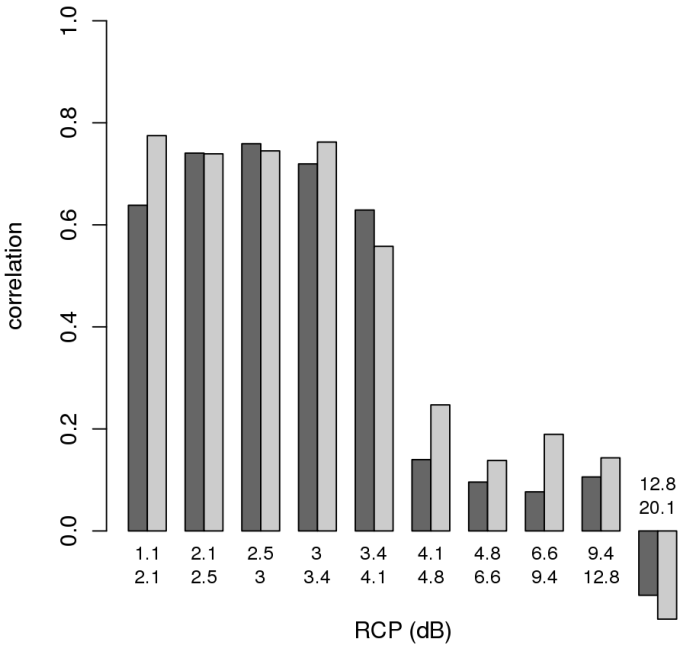


Figure 3.16: Pearson correlation (dark bars) and Spearman (rank) correlation (light bars) between hourly Kdp aloft (-15°C) and Z_h at the lowest level for ten equal-sized classes of the Radar Convective Parameter (RCP). The values on the x-axis at the base of the bars denote the left and right limits of the RCP classes.

correlation between the K_{dp} signatures associated with snow growth aloft and the surface precipitation rate is unique to stratiform precipitation.

The correlation statistics focused on events with a relatively high freezing level ($>$ 1.5 km MSL) for ease of comparison with the low level reflectivity, representative of the surface precipitation rate. While precipitation associated with purely stratiform rain has generally moderate consequences on the subalpine Italian territory, even weak snowfall events over the plains can have a major impact, especially on ground transportation and aviation. For this reason it will be fundamental to assess whether and under which circumstances the presented results may extend to snowfall episodes. At the same time, the origin of the statistical correlation between observations aloft and surface precipitation will need to be further explored. In particular time-lagged analysis between K_{dp} in the ice region and surface

precipitation (from either gauges or radar) will help to assess if there is a potential for prognostic applications.

CHAPTER 4

Hydrometeor classification

The WRF 3D-Var reflectivity assimilation (version 3.4), similarly to other radar assimilation schemes, is currently relying on a warm-rain process physics [22]. This means that only rainwater increments are considered in the minimization of the cost function. Ice-phase schemes are nowadays under development by researchers within the numerical model community. Once those schemes will become available more radar information will be usable for assimilation, providing accurate microphysical retrievals that are expected to improve numerical weather prediction. In particular, methods will need to be devised to incorporate into the assimilation schemes the information about the hydrometeor phase and type (rain, hail, snow,...) from dual-polarization radar.

Besides microphysical retrievals application for numerical models, hydrometeor classification is also relevant to:

- QPE (Quantitative Precipitation Estimation), providing a guidance for application of the most suitable semi-empirical relation to estimate the rainfall (snowfall) intensity;
- discriminate between hail and heavy rain;
- identify aircraft icing conditions.

In view of the numerical modeling anticipated progresses and the aforementioned potential applications, we devise a need for a robust radar hydrometeor classification methodology. Although classification schemes have been developed and even operationally employed over the last decade at S band, frequencies such as C band and X band present specific challenges due to non-Rayleigh particle scattering, attenuation and differential attenuation.

The radar variable employed for hydrometeor classification are listed in table 4.1, with the corresponding definitions in terms of:

- wavelength λ;
- radar cross sections at horizontal or vertical polarizations $\sigma_{h,v}$;
- dielectric factor $k_{w,i}$ (water,ice);
- complex forward scatter amplitudes at horizontal and vertical polarizations f_h, f_v;
- elements of the backscattering matrix S_{hh}, S_{vv}.

Table 4.1: Definition of the radar variables relevant for hydrometeor classification.

Radar variable	Definition	Unit
Reflectivity	$Z_{h,v} = 10\log_{10}\left[\frac{\lambda^4}{\pi^5\lvert k_{w,i}\rvert^2}\int_{D_{min}}^{D_{max}}\sigma_{h,v}(D)N(D)dD\right]$	dBZ
Differential Reflectivity	$Z_{dr} = Z_h - Z_v$	dB
Specific Differential Phase Shift	$K_{dp} = \frac{180}{\pi}\lambda Re\int_{D_{min}}^{D_{max}}[f_hD - f_vD]N(D)dD$	$^\circ\,\mathrm{km}^{-1}$
Correlation Coefficient	$\rho_{\mathrm{HV}} = \left\lvert\frac{\langle S_{vv}\cdot S^*{}_{hh}\rangle}{\langle S^2{}_{hh}\rangle^{1/2}\cdot\langle S^2{}_{vv}\rangle^{1/2}}\right\rvert$	-

D_{min} and D_{max} are the minimum and maximum diameter of the hydrometeors in the distribution $N(D)$. Asterisk is complex conjugate, and the angle brackets indicate expectation values.

The early work of Straka et al. [77] provides a basic description of the radar variables as for their characteristics and potential in hydrometeor classification. More recent studies

have extended the discriminating capabilities in specific circumstances, in particular for ice particles.

Most of the recent hydrometeor classification schemes are based on fuzzy logic [78, 75, 4, 79–81]. When the fuzzy logic algorithm is applied to noisy input variables, the resulting classification is also noisy. This is inherently due to the fact that the classification is performed on the individual radar resolution bins (hereafter simply bins) with typical size $\sim 1^\circ \times 100\,\mathrm{m}$, irrespective of the surrounding observations. For operational applications in particular it is desirable to have a smooth and clear output to ease the interpretation by the end user. An obvious way to get a smooth output is by filtering either the input variables or the final classification. In this paper we present an attempt to produce a noiseless classification by combining in a unique way:

- the quality (or strength) of the classification
- the spatial coherence and self-aggregation propensity of the observations
- basic microphysical constraints

The aim is to define an objective and completely automatic procedure to attain a result as close as possible to the classification that a human expert could perform. The items listed above are natural information sources that a radar expert is normally considering to partition the physical space in different categories based on the multiple inputs provided by the polarimetric radar observations. The adopted approach lies within the class of semi-supervised learning and is implemented here around a core cluster analysis module. In this type of learning methods, also called constraint-based, the auxiliary background knowledge or a given amount of labeled data is used to improve the partition of the data space. The technique is basically heuristic. In fact the supervised component of the analysis relies on

constraints and rules which can be ultimately considered background knowledge supported by precipitation microphysics.

The proposed method assumes that a standard bin-based classification scheme is already available and is intended to improve the initial output, through noise reduction and exploitation of the self-aggregation process inherent in cluster analysis. The ultimate goal of the method is to attain a simplified picture of the hydrometeors distribution, summarized by a limited number of connected regions in the two-dimensional observation space (either PPI or RHI).

The development of the classification methodology considerably relies on techniques commonly used in data mining and digital image processing. In particular an adaptation of the K-means clustering algorithm [82] is developed to incoporate a spatial contiguity constraint and a penalty term for the inclusion of microphysical constraints. One of the distinguishing points of the proposed methodology is that the final classification is performed over connected regions, as opposed to traditional methods which are applied to the individual range bins or Cartesian grid points. The connected regions in the radar observations domain can be derived based on digital image analysis techniques. In this work, the Connected Component Labeling (CCL) algorithm [83, 84] is employed for the identification and unique labeling of regions populated with adjacent bins assigned to the same hydrometeor class.

For the definition of the dual-polarization variables ranges and membership functions for the set of hydrometeor classes we rely on the relevant published work, in particular [77, 4, 79] for S-band, [85–87] for C-band, [80, 88, 89] for X-band.

In Section 4.2 the proposed methodology is described in detail, and in Section 4.3 application examples using data from different radars operating at frequencies ranging from

S-band to X-band are illustrated. Section 4.4 presents a statistical evaluation of the method and finally Section 7.4 summarizes the main results.

4.2. Method for noiseless hydrometeor classification

The basic building block of the classification procedure is the quite popular fuzzy logic method depicted in Fig. 4.1. The success of the fuzzy logic techniques in hydrometeor classification is likely due to the ease of implementation. All that is needed is a proper set of rules, generally implemented through analytical membership functions. The specific fuzzy logic technique adopted here is essentially based on the classification methods described in [78] and [79], with few modifications discussed in Section 4.2.1. The overall structure of the classification algorithm is represented by the diagram in Fig. 4.2. The building blocks 1-4 encompassed by a the dashed line are discussed in detail in the following sub-Sections. Hereinafter we'll refer to the output of blocks 2, 3 and 4 in Fig. 4.2 respectively as bin-based, cluster-based and region-based classification.

The input radar variables for the hydrometeor classification are: reflectivity Z_h , differential reflectivity $Z_{\rm dr}$, specific differential phase shift $K_{\rm dp}$ (the range derivative of the differential phase shift $\Phi_{\rm dp}$), cross-polar correlation coefficient $\rho_{\rm HV}$. The vertical profile of temperature T from either a nearby observed sounding or numerical model output is also considered, leading to a total of $\mathrm{N_{var}} = 5$ inputs. In addition, for the identification of non-meteorological echoes (Section 4.2.1), the spatial variance of the differential reflectivity $\sigma(Z_{\rm dr})$, the spatial variance of the total differential phase shift $\sigma(\Psi_{\rm dp})$ and the Doppler velocity V are also used. The observed total differential phase shift is defined as:

$$\Psi_{\rm dp} = \Phi_{\rm dp} + \delta_{hv} \tag{10}$$

where δ_{hv} is the differential phase shift upon backscattering [25]. To avoid over-estimating the variability of $Z_{\rm dr}$ and $\Psi_{\rm dp}$ due to physical trends in the observations (e.g. $\Psi_{\rm dp}$ increasing in heavy rain), the variances are calculated on the residuals after linear regression over five consecutive range bins.

The algorithm is designed to work on the radar observations and textures in the two-dimensional (either PPI or RHI) polar domain (radial θ, range R). The choice is in the first instance dictated by the need to limit the overall computational resources for real-time applications, avoiding interpolation of all radar variables and textures. Working in the polar domain also has the notable advantage to avoid loss of information near the radar and preserve the contiguity properties in the observation space.

The classification is partitioned in $\rm N_{cl}$=11 hydrometeor classes: *large drops* (LD), *drizzle* (DR), *rain* (RA), *heavy rain* (HR), *rain + hail* (RH), *hail* (HA), *graupel* (GR), *wet ice* (WI), *dry ice* (DI), *crystals* (CR), *dendrites* (DN). In addition to the above 11 hydrometeor classes, non hydrometeor decision classes such as *clutter* (CL) and *clear air* (CA) categories are considered in the preliminary fuzzy-logic classification (Section 4.2.1). The most notable difference between the hydrometeor classes used here respect to most previous studies is the addition of a specific class for dendrite crystals. Kennedy and Rutledge [41] for S-band radar, Bechini et al. [90] for C-band and X-band have shown that dendrite crystals have specific dual-polarization signatures. $K_{\rm dp}$ (as high as $2^\circ\,{\rm km}^{-1}$ at C-band) and the environmental temperature (ranging between approximately -10°C and -20°C) are the most relevant variables employed for the identification. Another new feature of this classification system is a macro-class (LIQUID, MIXED, SOLID, HAIL) associated to the hydrometeor classes, for use in the final step of the classification (block 4 in Fig. 4.2). The LIQUID macro-class

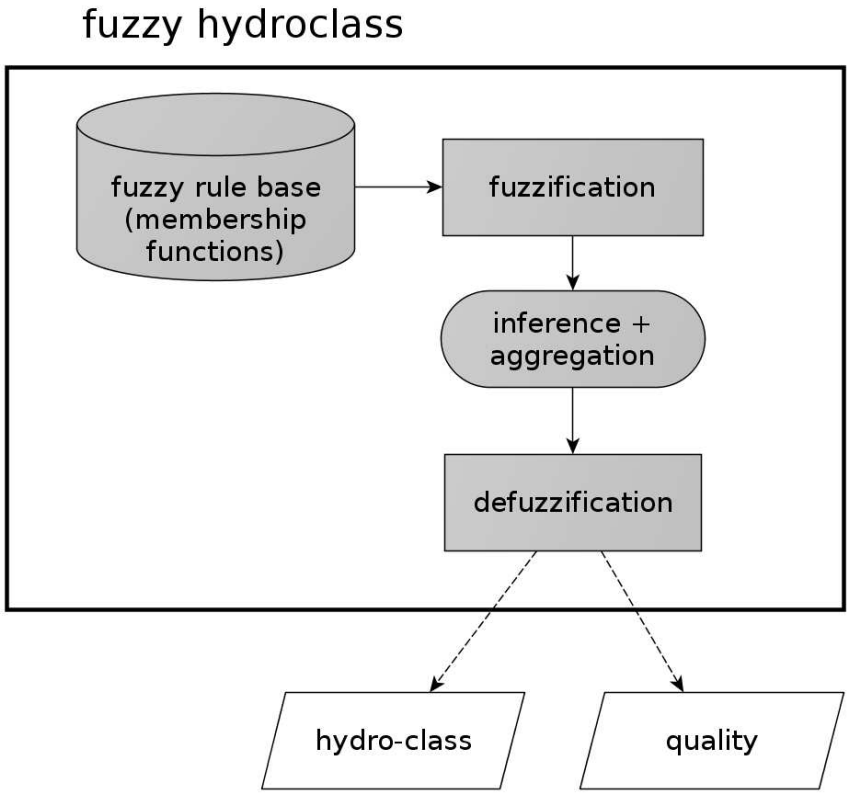


Figure 4.1: Diagram of the basic fuzzy-logic classification. Adapted from [4].

includes four hydrometeor types (LD, DR, RA and HR). The MIXED macro-class is only composed of WI, SOLID includes GR, DI, CR and DN, while HAIL includes RH and HA.

4.2.1. Preliminary processing and fuzzy-logic classification. Several processing tasks normally performed separately are here integrated in a single procedure. The reason is twofold: on one hand the total computational time is reduced, on the other hand the adopted scheme provides an environment favorable to deal with the interconnection between attenuation correction and hydrometeor classification in the development process.

The preliminary processing of the dual-polarization moments includes clutter identification, Ψ_{dp} filtering, K_{dp} calculation and attenuation correction. The fuzzy logic hydrometeor

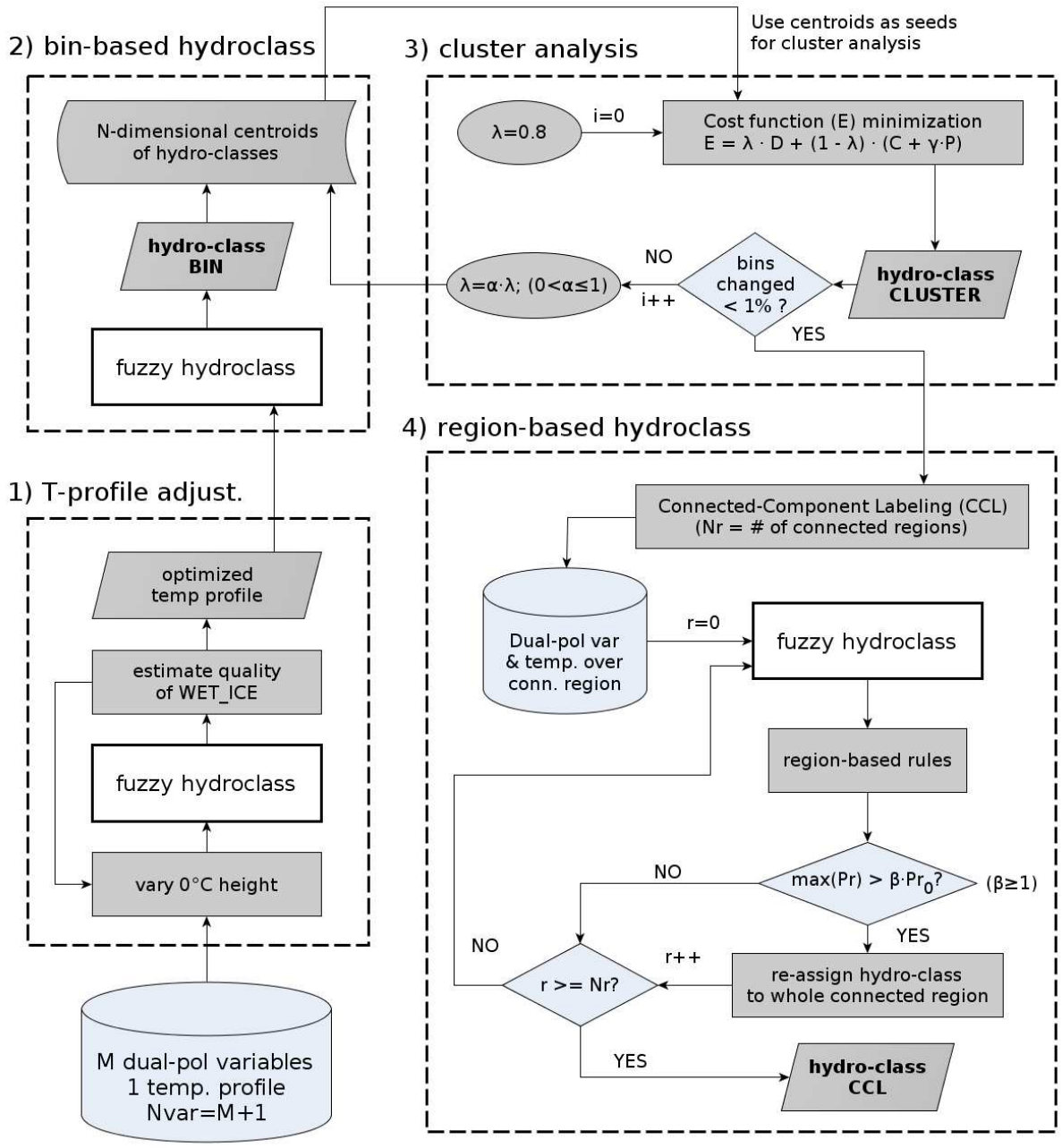


Figure 4.2: Diagram of the full four-step classification. The box labeled "fuzzy hydroclass" is represented in detail in Fig. 4.1. See text for further details.

identification algorithm follows the classical fuzzification - inference - aggregation - defuzzification flow depicted in Fig. 4.1. In addition to the class for a specific bin, the fuzzy-logic

algorithm provides in output the quality (described hereinafter) associated with the specific class assignment. This is an important point because the quality of the classification is needed in the subsequent cluster analysis (Section 4.2.3). To represent the membership functions (MBF) both beta functions [4] or trapezoidal functions [79] have been used. Park et al. [91] proposed asymmetric trapezoidal functions to better approximate the shapes of the probability distributions. Keeping a similar approach we define an asymmetric beta distribution, with values ranging from 0 to 1, as:

$$\text{MBF}(x) = \begin{cases} \dfrac{1}{1+(\frac{x-m}{a_l})^{2b_l}}; & x \leq m \\ \dfrac{1}{1+(\frac{x-m}{a_r})^{2b_r}}; & x > m \end{cases} \tag{11}$$

where x is the value of the input variable (radar or temperature), m is the center of the function, a_l/a_r the left/right side half width, and b_l/b_r is the left/right side slope. In the current implementation $b_r = b_l$ for all hydrometeor classes, while different values are used for the non-meteorological classes *clutter* and *clear air*. In fact the parameters for the *clutter* and *clear air* classes are derived by means of statistical analysis over samples of known echo type [92] and the values of b_l/b_r can be set to better match the empirical distributions.

For the hydrometeor categories the parameters m, a_l, a_r, b_l, b_r are derived for each class and radar parameter, based on relevant observational and modeling studies. Specifically we mainly rely on [77, 4, 79] for S-band, [85–87] for C-band, [80, 88, 89] for X-band.

As an example, Table 4.2 reports the parameters defining the beta function for the four dual-polarization variables at S-band. Similarly to [79] we deal with physical dependencies between radar variables for specific hydrometeor classes, by introducing a dependency between the parameters m, a_l/a_r and Z_h . This is visually represented in Fig. 4.3, where the

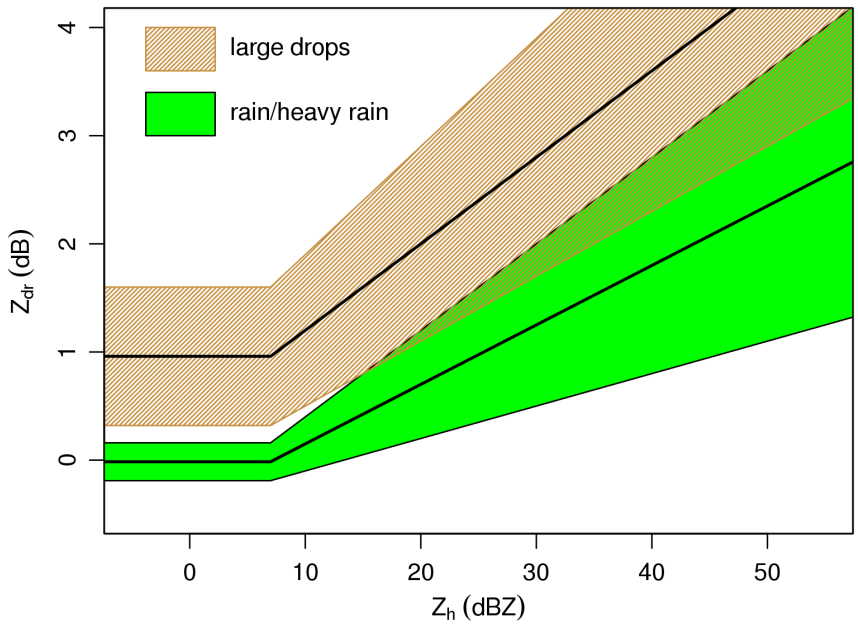


Figure 4.3: Membership functions for the input variable $Z_{\rm dr}$ and hydrometeor classes *large drops* and *rain/heavy rain*. As reported in Table 4.2 the values of m and a_l/a_r are expressed as a function of Z_h. The thick solid line represents the 1.0 central value of the beta function $m(Z_h)$, while the lower and upper boundaries of the colored areas represent the 0.5 value of $m(Z_h) - a_l(Z_h)$ and $m(Z_h) + a_r(Z_h)$ respectively ($a_l = a_r$ in this case).

MBF of $Z_{\rm dr}$ is plotted for the *large drops* and *rain/heavy rain* classes. The *large drops* class is essentially defined as the rain category, following a similar increase with Z_h but with higher $Z_{\rm dr}$. *large drops* can be found either below the freezing level, originating from the melting of large snow flakes, or within strong updrafts indicating the presence of supercooled liquid drops [93, 94].

The inference rule is given by the weighted sum of the MBF of temperature and radar parameters [79, 95]:

Table 4.2: Values of the parameters used to define the asymmetric beta function (eq. 11) for the dual-polarization radar variables at S-band. When $a_r \neq a_l$ both values are indicated, otherwise a single value is reported. For some classes the parameters m and a_l/a_r are expressed as a function of $\widetilde{Z_h}$ (where $\widetilde{Z_h} = Z_h$ $[Z_h > 7\ dBZ]$; $\widetilde{Z_h} = 7$ $[Z_h \leq 7\ dBZ]$) to account for the expected physical correlation.

	Z_h				$Z_{\rm dr}$				$\rho_{\rm HV}$				$K_{\rm dp}$			
Class	m	a_l/a_r	b_l/b_r	w	m	a_l/a_r	b_l/b_r	w	m	a_l/a_r	b_l/b_r	w	m	a_l/a_r	b_l/b_r	w
large drops	40	15.	8	1.0	0.4+0.08· $\widetilde{Z_h}$	0.5+0.02· $\widetilde{Z_h}$	2	0.8	0.99	0.03	2	1.0	0.3	0.3/1.0	2	0.5
drizzle	0	25	8	1.0	0.5	0.7	2	0.8	1.00	0.02	2	0.8	0.01	0.2	3	0.5
rain	34	14	8	1.0	-0.4+0.055· $\widetilde{Z_h}$	0.8+0.025· $\widetilde{Z_h}$	2	0.8	0.99	0.03	2	1.0	0.3	0.3/1.0	2	0.5
heavy rain	50	10	8	1.0	-0.4+0.055· $\widetilde{Z_h}$	0.8+0.025· $\widetilde{Z_h}$	2	0.8	0.99	0.05	2	1.0	0.5	0.2/10.0	2	0.8
rain + hail	65	15	4	1.0	0.5	1.5	3	1.0	0.95	0.07	2	0.8	5.0	5.0	8	1.0
hail	65	13	4	1.0	-2.0	3.0	3	1.0	1.00	0.06	2	0.8	0.0	1.0	2	1.0
graupel	42	14	5	1.0	1.2	2.0	2	0.8	1.00	0.06	2	1.0	0.8	1.5	2	0.5
wet ice	20	25	4	1.0	1.5	1.7	3	1.5	0.88	0.08	2	1.0	1.0	1.2	2	0.5
dry ice	22.5	15	5	1.0	0.2	1.0	2	0.4	1.00	0.05	2	1.0	0.2	0.3	2	1.0
crystals	0	22.5	5	1.0	1.5	3.5	2	0.5	0.98	0.06	2	1.0	0.5	0.7	2	1.0
dendrites	20	15	5	1.0	1.0	0.5/1.5	2	1.0	1.00	0.07	3	0.5	0.8	0.4/2.5	2	2.0

$$S_k = \frac{\sum_{i=1}^{N_{var}} [w_k(x_i) \cdot \mathrm{MBF_k}(x_i)]}{\sum_{i=1}^{N_{var}} w_k(x_i)} \tag{12}$$

where $w_k(x_i)$ are the weights associated with the input variable x_i and hydrometeor class k. S_k is the resulting confidence associated with the class k and the output class is the one for which S_k is maximum. The additive inference rule has the advantage to mitigate the effect of measurement errors, especially on the dual-polarization parameters, which may show unreliable values mainly due to partial beam filling, side lobes effects, low signal to noise ratio. The weights are a priori assigned based on the confidence and the discriminating capability of the temperature and radar variables for every hydrometeor type.

The weight associated with the temperature is not a priori fixed as for the radar variables, but is defined to be a parabolic function of the temperature itself, with a minimum value (1.0) at +1.25 °C . The weights then double at -14°C (+16°C) and triple at -20°C (+23°C). Lim et al. [95] proposed a hybrid scheme where the sum inference rule is used for Z_{dr} , K_{dp} , ρ_{HV} , while the product rule is used for Z_h and T. In such scheme the reflectivity and temperature are strong constraints, with the advantage of reducing most misclassifications due to overlapping dual-polarization radar MBF. But on the other hand the multiplicative membership function of T introduces very sharp and often unrealistic transitions across precipitation phase changes. The use of the parabolic weight for the temperature is intended to exploit the temperature information to minimize misclassifications in regions far from phase transitions and to rely more on radar information to discriminate near the melting layer.

A special feature of the algorithm implementation is that the *clutter* is treated exactly as an additional class and is identified (optionally, depending on the specific signal processor filtering activation) within the same fuzzy logic volume processing. This allows a reduced total computation time. More specifically, within the loop over the radials, the *clutter* is identified at radial n by imposing a threshold value on its confidence (when $S_{clutter} > 0.5$ the bin is flagged as *clutter*, irrespective of the confidence associated to the other meteorological classes). The rest of the hydrometeor classes are treated within the same loop, but considering the preceding radial $n-1$, which is already flagged for clutter. This is devised to allow possible calculation of averages and textures in range for the purpose hydrometeor classification, excluding the bins marked as clutter.

The processing of the n^{th} radial also includes filtering Ψ_{dp} using the Hubbert and Bringi [47] scheme to remove δ_{hv}, after the clutter bins are flagged. K_{dp} is calculated by a linear moving window regression in a two-step procedure: a first estimate is obtained over a large window (e.g. ± 16 range bins). In the second step the initial K_{dp} estimate is used to set the proper window width (smaller for higher K_{dp}). In this way it is possible to retain the small scale variability in heavy precipitation and at the same time limit the noise in weaker precipitation regions. More sophisticated algorithms are available for the estimation of K_{dp} , but mainly for application to rainfall estimation [49, 96, 97]. For the purpose of hydrometeor classification the current approach has the advantage of being simple to implement, robust (it provides an estimate of K_{dp} even in regions of low signal to noise ratio (SNR) where Ψ_{dp} is more noisy) and completely independent of the other polarimetric variables. However, if a K_{dp} estimate is already available from a separate processing, this can be used as input for the hydrometeor classification.

The rain profiling algorithm proposed by Testud et al. [48] is applied to correct the horizontal reflectivity for path attenuation, while differential attenuation is linearly estimated from the horizontal attenuation α_h [98]. Attenuation correction is performed exclusively in the rain medium (i.e. below the freezing level) to avoid introducing a bias due to differential phase increases in the ice region (Chapter 3). However, especially for X band, the attenuation corrected measurements are subject to increasing uncertainty with range, due to the inherent approximation of the attenuation estimate. An obvious implication is that Z_h and Z_{dr} measurements at farther ranges, when attenuation is relevant, should be given less confidence in the hydrometeor classification. This is implemented by applying an arbitrarily defined attenuation-dependent and exponentially-decaying weight to Z_h and Z_{dr} (multiplying the weight $w_k(x_i)$ in eq. 12), as a function of the estimated horizontal attenuation:

$$W_{Z_h,Z_{dr}} = e^{-(\alpha_h/\mathrm{coeff}_{Z_h,Z_{dr}})^2} \tag{13}$$

where the values of $\mathrm{coeff}_{Z_H,Z_{dr}}$ is set to 20 dBZ and 10 dB respectively for Z_h and Z_{dr}.

Before hydrometeor classification is performed, optional range averaging over the radar observations can be applied, excluding the bins flagged as *clutter*.

Based on the MBF defined in eq. 11, it is possible to define the quality (or strength) of the classification in several ways, e.g.:

- confidence of the identified particle class (max confidence among all hydrometeor classes: $\max_k(S_k)$);
- difference between the confidence of the first and second identification [81];

- difference between the confidence of the first identification and the highest confidence of a particle with a different macro-class.

After detailed evaluation the simplest method a) was chosen for exhibiting the most robust performance in the definition of the seeds for the cluster analysis (Section 4.2.3).

4.2.2. Adjustment of the freezing level height. Due to the uncertainty typically affecting the temperature vertical profile (space-time representativeness of observed soundings, model uncertainty) a preliminary step in the algorithm (block 1 in Fig. 4.2) aims at fine tuning the location of the freezing level. In the literature several approaches have been proposed, based on the dual-polarization measurements with focus on the location of the minimum of ρ_{HV} [39, 99] or considering the change in the vertical gradient of Z_{dr} [95]. Here an iterative procedure is adopted, to adjust the level of the 0°C isotherm based on the results of the bin-based hydrometeor classification itself. A direct search algorithm [100] is used to find the temperature increment (positive or negative) that maximizes the average quality of the bins identified as *wet ice*. The proposed recursive approach exploits all the dual-polarization information available, without the need of an additional specific processing.

The temperature optimization for the application examples discussed hereinafter in Section 4.3 is represented in Fig. 4.4. In order to accept the estimated temperature shift the curvature of the *wet ice* quality vs. dT curves represented in the plots is considered. A threshold on the 2$^{\mathrm{nd}}$ order derivative estimated from non-uniformly spaced data points [101] is used in the current implementation, which allowed to accept all temperature shifts corresponding to the maxima in Fig. 4.4.

Finally the whole temperature profile is shifted by the resulting increment. When the absolute value of this final temperature increment is significant (e.g. >0.5°C), the new

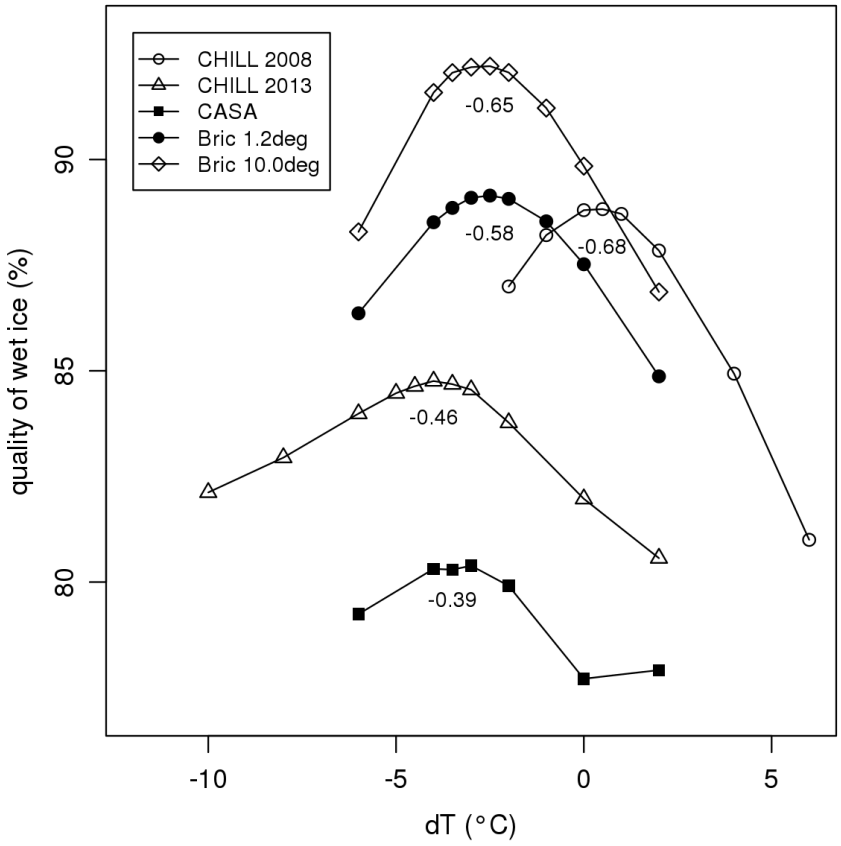


Figure 4.4: Average quality of the *wet ice* class for different temperature shifts during the maximization loop (block 1 of the diagram in Fig. 4.1). The number below each curve represents the estimated 2^{nd} order derivative (curvature) in units of % $^{\circ}\text{C}^{-2}$. All curves show values below (higher curvature) the adopted threshold of -0.3 % $^{\circ}\text{C}^{-2}$.

estimation of the freezing level height is used to re-run the attenuation correction and the bin-based hydrometeor classification. Although discontinuities may arise from one scan to the next, in a operational implementation a stable estimate should be attained through filtering the available preceding time series of the temperature shift.

The current implementation relies on the four polarimetric radar variables (Z_h, Z_{dr}, ρ_{HV} and K_{dp}) and temperature for the identification of *wet ice*, as for all the other hydrometeor classes. It is likely that the adjustement of the freezing level could be further improved by additionally considering the variance of Ψ_{dp} and the vertical derivatives of Z_{dr} specifically

for the *wet ice* category. In fact these additional parameters should improve the identification of *wet ice* and also make it less susceptible to small biases in Z_h and Z_{dr}.

4.2.3. Cluster analysis. The bin-based fuzzy-logic classification performed using the adjusted temperature profile is the initial step of the cluster analysis (block 3 in Fig. 4.2). The centroids of each hydrometeor class are calculated and used as initial values (seeds) for an analysis based on K-means clustering [82].

The K centroids (K being the number of hydrometeor classes found in the current data: $\mathrm{K} \leq \mathrm{N_{cl}}$) are defined by a vector $\boldsymbol{\mu}_k$ whose components are the quality-weighted average values of the radar variables Z_h, Z_{dr}, K_{dp} and ρ_{HV} (observation vector $\mathbf{x}$) for each hydrometeor class. The ith component of the vector $\boldsymbol{\mu}_k$ can be expressed as:

$$\mu_{k_i} = \frac{\sum_{n=1}^{{N_k}^{obs}} [S_k \cdot x_i]}{\sum_{n=1}^{{N_k}^{obs}} S_k} \tag{14}$$

where S_k is the confidence (eq. 12) associated with the class assigned by the bin-based classification to the nth bin and x_i is the ith radar variable. ${N_k}^{obs}$ is the number of observations (bins) that have been assigned to the class $k \in [1{:}K]$.

The temperature T is not included in the cluster analysis, which entirely relies on the radar observations. In fact the inclusion of T in this step may lead to an excessive vertical stratification in the resulting analysis. In addition, only the 11 hydrometeor classes are considered hereinafter, i.e. the identification of *clutter* and *clear air* is finalized in the bin-based classification step. It is also worth emphasizing that the membership functions and associated weights for the radar parameters are not used in the clustering process, which is

designed to extract additional information from the observations with a "learn from data" approach.

K-means clustering is used to partition the data space, by iteratively computing the affinity between a hydrometeor assignment and each of the K already existing clusters. Lakshmanan [102] proposed a segmentation technique for the radar reflectivity using textures. Applying a K-means clustering approach to the vector $\mathbf{x}$ of dual-polarization variables, under smoothness assumption a cost function E_k can be defined for every bin in the polar domain:

$$E_k = \lambda \cdot D_k + (1 - \lambda) \cdot (C_k + \gamma \cdot P_k) \tag{15}$$

where k indicates the hydrometeor class. D_k is the Mahalanobis distance [103] between the cluster vector $\boldsymbol{\mu}_k$ for the class k and the observation vector $\mathbf{x}$ for the current bin, defined as:

$$D_k = \sqrt{(\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)} \tag{16}$$

where Σ is the covariance matrix of the observations and the superscript T indicates the transposed vector. Since the covariance matrix is in general positive definite, the matrix Σ can be first decomposed into into a product of a lower triangular matrix and its transpose using Cholesky decomposition and then inverted. The Mahalanobis distance is widely used in cluster analysis and provides an objective method for the normalization of the observation vector, whose components (Z_h , Z_{dr} , K_{dp} , ρ_{HV}) represent different variables with different physical units (dBZ, dB, $^\circ\,\mathrm{km}^{-1}$, unitless). The use of the Mahalanobis distance for vector normalization allows to overcome one of the typical limitations of K-means clustering, i.e. the tendency to produce spherical clusters arising from the adoption of the Euclidean metric.

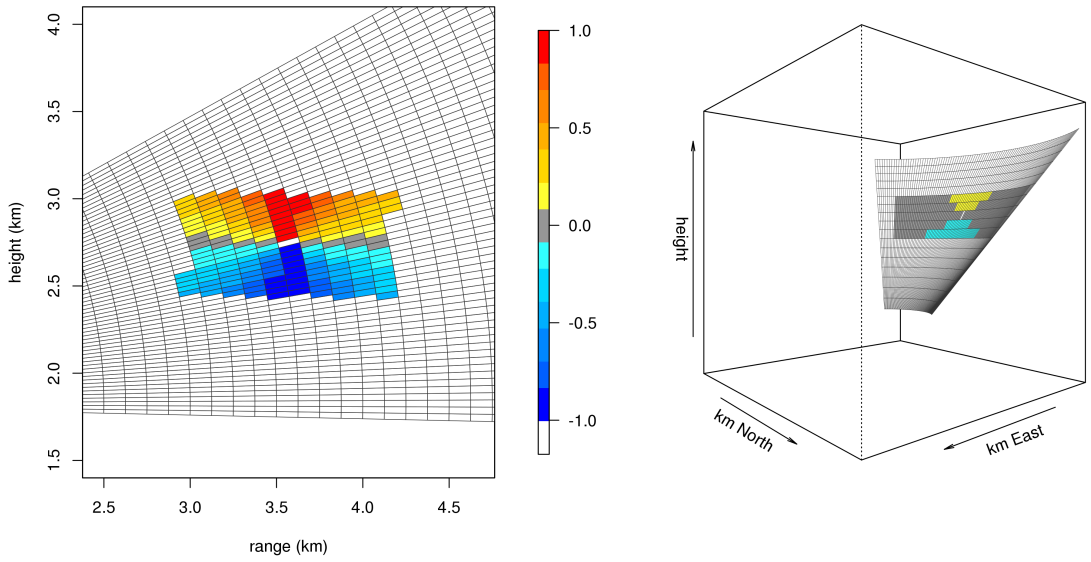


Figure 4.5: Geometry of the contiguity window for an elevation scan (a) and an azimuth scan (b). The colors represent the $cos(\phi)$ term in eq. 18 (penalty term).

C_k is a measure of contiguity given by the number of neighboring bins with a hydrometeor class different from the candidate class k for the current bin:

$$C_k = \frac{\sum_{t=-N_\theta}^{N_\theta} \sum_{r=-N_R}^{N_R} (1 - \delta(J_{t,r} - k))}{N_{tot}} \tag{17}$$

$J_{t,r}$ is the class of the neighboring bin in the polar array, N_R and N_θ are the number of bins in range and azimuth (or elevation) respectively and N_{tot} the number of valid observations:

$$N_{tot} \leq (2\,N_\theta + 1)(2\,N_R + 1).$$

N_R and N_θ are set according to a desired width and height of the contiguity window, typically of the order of ~ 1 km $\cdot$ 1 km or less. For the higher elevations in vertical scans (RHI) it is necessary to cut the polar contiguity window to keep a uniform geometry in the Cartesian projection, as shown in Fig. 4.5.

P_k is a penalty term introduced to provide some weak physical constraints to the cluster analysis. In the actual configuration the constraints only involve the relative position in the vertical of different hydrometeors. For example it is more likely to find *hail* rather than *rain* above a candidate *rain* + *hail* bin. The checkerboard diagram in Fig. 4.6 illustrates the implemented pairwise constraints. The columns indicate the unlikely hydrometeors (black boxes) above a given candidate k (top labels), while the rows indicate the unlikely hydrometeors below a given candidate (left labels). Those penalty terms add up after proper weighting according to the relative position of the neighboring bin respect to the candidate bin. The penalty term P_k can then be expressed as:

$$P_k = \frac{\sum_{t=-N_\theta}^{N_\theta} \sum_{r=-N_R}^{N_R} cos(\phi) \cdot (B_p[k, J_{t,r}] - B_p[J_{t,r}, k])}{N_{tot}} \tag{18}$$

where B_p represents the Boolean penalty matrix in Fig. 4.6 (black boxes are ones and white boxes are zeroes), with k representing the column index and $J_{t,r}$ the row index. ϕ is the angle between the vertical and the line connecting the candidate bin k and the neighboring bin. The polar bins in Fig. 4.5 are colored according to the weighting term $cos(\phi)$. A good performance has been achieved over many different cases with the multiplying factor γ in eq. 15 set to a value of the order of 10. The penalty term is clearly more relevant for RHI scans, where the observations are distributed on a vertical plane rather than on a low elevation PPI conical surface (Fig. 4.5).

Fig. 4.7 gives a picture of the typical convergence rate for two values of λ (λ=0.2 and λ=0.9). With low values of λ (more weight to contiguity) the convergence is in general faster. Decreasing the value of λ with increasing iteration n was found to be a good option

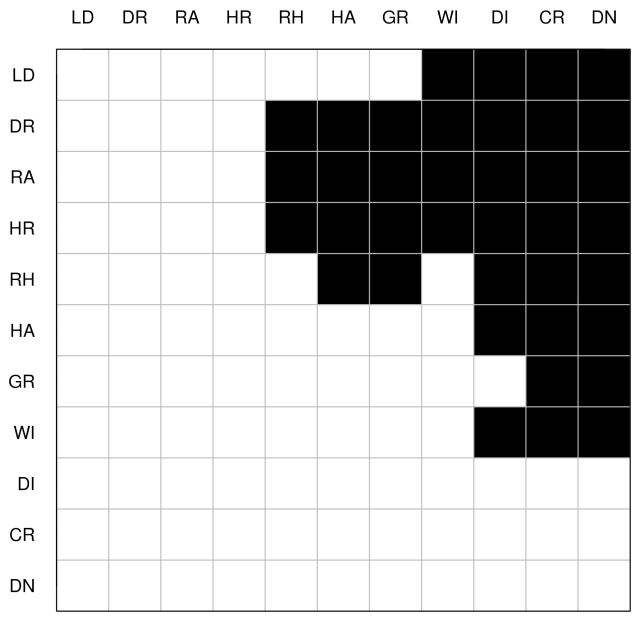


Figure 4.6: Checkerboard representation of the Boolean penalty matrix B_p in equation 18. The x and y axis labels represent the hydrometeor classes: *large drops* (LD), *drizzle* (DR), *rain* (RA), *heavy rain* (HR), *rain + hail* (RH), *hail* (HA), *graupel* (GR), *wet ice* (WI), *dry ice* (DI), *crystals* (CR), *dendrites* (DN). Black boxes in a given column indicate hydrometeors to be penalized above a candidate (top labels), while black boxes in a row indicate hydrometeors to be penalized below a candidate (left labels).

to provide a robust classification result in few iterations:

$$\lambda^n = \alpha \cdot \lambda^{n-1}; \;\; 0 < \alpha \leq 1 \tag{19}$$

In this way during the first iterations the K clusters may change significantly by self-adapting to the observations, while the last iterations provide the desired spatial smoothing. The iteration is stopped when the number of bins changed in the classification array falls below a given threshold (1% of the total number of valid bins).

4.2.4. Region-based classification. Depending on the value of the parameters in the cost function (eq. 15), especially when high values of λ are used, the K-means cluster

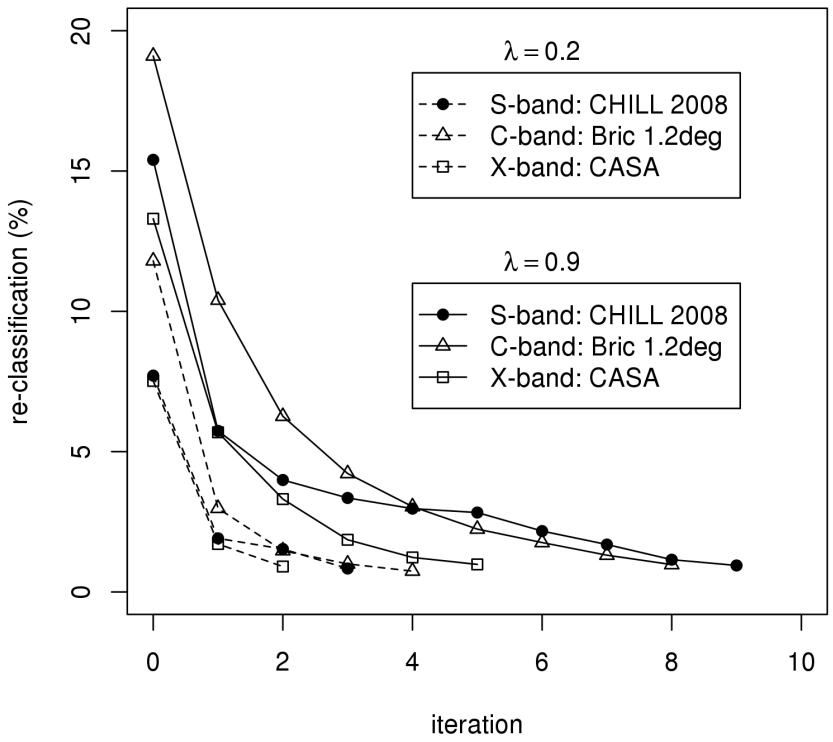


Figure 4.7: Number of bin classification changes at each iteration of the cluster analysis, for three different cases (Section 4.3) and two fixed values of λ (α=1 in eq. 19).

analysis may cause some cluster to drift too much from the initial values. As a consequence the original hydrometeor class may become no more appropriate for a given region. The formulation of the cost function with increasing emphasis on the contiguity component (eq. 19) yields a virtually noiseless classification map. On this map it is practical to identify a limited number of connected regions (N_r) by running a Connected-Component Labeling (CCL) algorithm [83, 84]. The implemented CCL algorithm works by checking for 8-connectivity (four horizontal and vertical neighbors, and four diagonal neighbors) in binary input images. One hydrometeor class is processed at a time, by setting all the pixels marked with the current class to one and the pixels with a different class assignement to zero. From this input binary

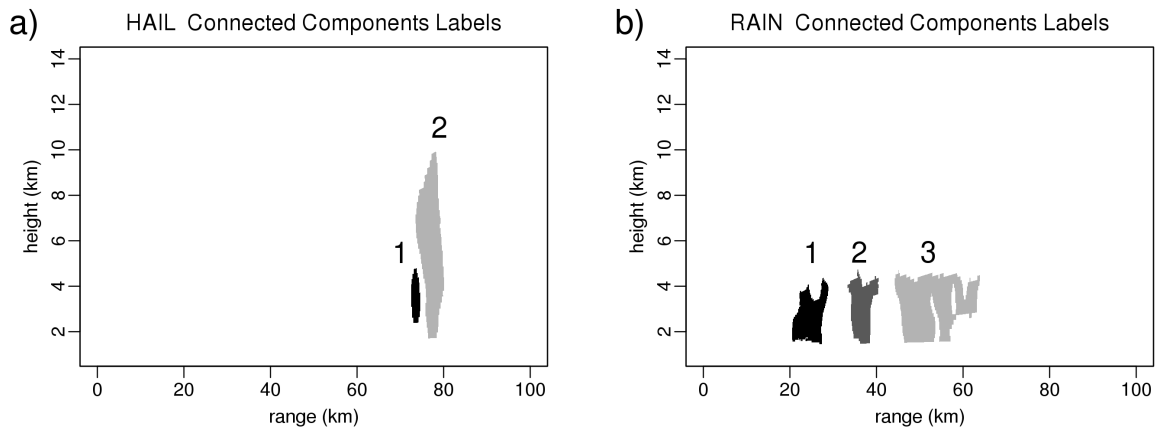


Figure 4.8: Illustration of the Connected Component Labeling (CCL) for the case study reported in Fig. 4.10. Two connected regions identified as *hail* (panel a)) and three regions of *rain* are identified and sequentially labeled (overplotted numbers).

image all the connected regions are sequentially labeled. An example illustrating the CCL results for two hydrometeor classes is reported in Fig. 4.8, where the labels identifying the connected regions are represented by different gray levels. In this example, referring to the case study analysed hereafter in Section 4.3.1, two separate connected regions of *hail* and three regions of *rain* are identified. These plots elucidate one of the main features of the proposed approach, i.e. the aim to deal with homogeneous and spatially connected regions instead of individual range bins.

The fuzzy logic hydrometeor classification can now be applied on the set of connected regions in a similar way it is applied to the individual bins. It is in principle only necessary to identify a proper parameter (typically the mean or the median) representing the underlying statistical distribution for every region. Alternatively, a direct comparison between the Probability Density Function (PDF) of the univariate observations and the associated membership function can be performed. This latter approach is illustrated in Fig. 4.9. In

practice the value of $\mathrm{MBF_k}(x_i)$ in eq. 12 is replaced by an affinity measure defined as:

$$\mathrm{AFF_k}(x_i) = 1 - \frac{A_k^{out}}{A_k^{tot}}(x_i) \tag{20}$$

where A_k^{out} is the portion of the area under the PDF (dashed line in Fig. 4.9) for variable x_i, lying outside the domain defined by the membership function for the class k (hatched area in Fig. 4.9). A_k^{tot} is the total integral of the PDF, obtained from the observations using a Gaussian kernel smoothing [104]. The $\mathrm{AFF_k}(x_i)$ parameter, encompassing the 0-1 range just like the membership function in eq. 11, provides a measure of the conformity of the observed distribution to the theoretical expectation. For most cases this definition provides a similar estimate as the one simply (and quickly) obtained using the average or median value of the distribution. But in some instances, e.g. highly non-normal distributions or distributions with large tails exceeding both sides of the membership function, it gives a more robust indication about the affinity of the region with a given hydrometeor class. The classification then proceeds as for the bin-based application described in Section 4.2.1.

The resulting region-based classification may either confirm the original class of the connected region (which means that the cluster analysis did not drift excessively) or propose a new assignment. At this point we can further exploit the availability of a statistical sample for every connected region (as opposed to the single value bin) to introduce some simple physically-based rules. These rules rely on the 0, 25, 50, 75 and 100% quantiles of the temperature distributions and are implemented as if *statements* within the loop over the hydrometeor classes, for a given connected region $nreg$. With k being the index of the candidate class:

- If macro-class(k) is $LIQUID$

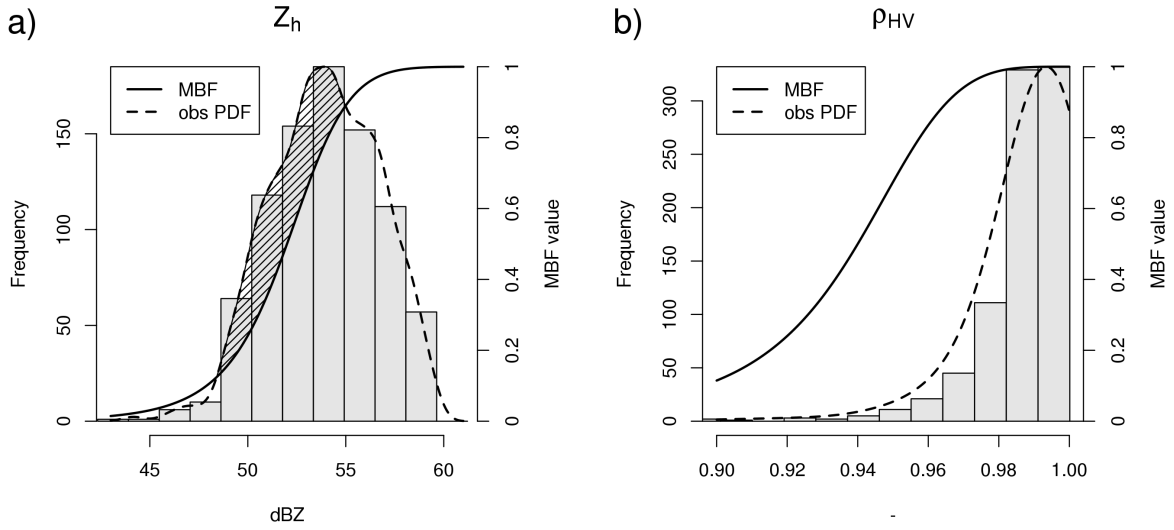


Figure 4.9: Observed distribution (histograms) of the reflectivity (a) and the correlation coefficient (b) within the connected region 2 classified as *hail* in Fig. 4.8a. The solid line represents the corresponding membership function (right y-axis), while the dashed line is the kernel density estimated PDF of the observations.

  - if $T^{75}_{nreg} < T^{25}_{wet\ ice}$ then skip k
  - if $T^{0}_{nreg} <$ -12°C then skip k

- If macro-class(k) is $SOLID$
  - if $T^{25}_{nreg} > T^{75}_{wet\ ice}$ then skip k
  - if $T^{100}_{nreg} >$ +12°C then skip k

The superscript of T represents the percentile of the temperature. $T^{25}_{wet\ ice}$ and $T^{75}_{wet\ ice}$ are respectively the 25 and 75% quantiles over all the regions identified as *wet ice*. Therefore some of the above conditions are only applicable if at least one *wet ice* region is present in the classification map. The scope of the rules is to avoid some clearly unreliable situations, e.g. when a *rain* region is entirely above the melting layer. This kind of situation is not likely to occur but still possible, since the temperature information is not included in the cluster

analysis. Other common features, in convective systems, are regions of liquid hydrometeors extending from near the surface to above the ambient freezing level within an updraft. These regions will not be affected by the adopted rules and will be preserved.

The region is finally reassigned only if the quality of the new assignment is higher than the quality of the original class by a given relative threshold (parameter β in block 4 of the diagram in Fig. 4.2). The higher the value of β, the higher the confidence posed in the self-adapting attitude of the data. Lower values of β are more suitable for an operational conservative approach. In the following application examples a value of β=4% is used. This means that the new proposed assignment must have a confidence at least 4% higher than the confidence associated with the original class to be accepted.

The region-based classification is not performed if a region is too scarcely populated (<50 bins) because a reliable PDF may not be estimated. The bins of these regions are assigned after the loop in block 4 (Fig. 4.2) is completed, based on the mode of the surrounding bins.

Although at the moment only few very simple rules are adopted, it is clear that working with connected regions considerably expands the potential of the classification. The properties of the spatially-connected statistical samples (density function, orientation, vertical/horizontal extension, etc.) can be further exploited to introduce more physically-based constraints.

4.3. Application examples

In order to test the method, dual-polarization data from different radar systems operating at S, C and X band are considered (Table 4.3). In particular, the following examples are based on measurements collected by the S-band CHILL radar in Colorado [105], the C-band Bric della Croce radar in Italy and the CASA IP1 X-band system in Cyril, Oklahoma [106].

The classification may vary significantly depending on the choice of the parameters λ, α and γ in eq. 15 and 19. In the following examples, except where explicitly noted, the initial value of λ is set to 0.8, α=0.75 and γ=12.0. The λ parameter in particular determines the adopted approach to the cluster analysis. High values of λ allow the clusters to vary significantly from the first guess classification. The input variables may assume values departing from the initial definition of the classes through the membership functions. Low values of λ on the other hand produce a classification relying more on spatial contiguity. In this case the distribution of the input variables within the regions may widen, although the average values will not change significantly. With the set of chosen values it is found that in general less than 10% of the regions are reassigned in the final step of the classification (fuzzy logic applied to the connected regions, as depicted by block 4 in Fig. 4.2). This is considered an acceptable compromise between a heavy data-adaptive approach (high values of λ) and an analysis based on spatial contiguity.

4.3.1. S-band, deep convection (Colorado). Fig. 4.10 shows a RHI scan through a deep convective storm in Colorado observed by the CHILL radar on 22 June 2013 at 23:06 UTC. The radar operated in Alternate Transmit Simultaneous Receive (ATSR) transmission, with 128 integrated pulses and a radial resolution of 0.25° . The most close and recent temperature profile available for real-time processing is from the Denver 12 UTC sounding. From this profile the height of the freezing level is located around 4650 m MSL. The temperature optimization loop (block 1 in Fig. 4.2) provides a considerable temperature shift of -3.75°C , as inferred from the average quality of the *wet snow* bins in Fig. 4.4. Although the analysed scene is mainly convective, the more stratiform region closer to the radar provides enough data for a fairly robust adjustement of the freezing level height (re-located around 4000 m MSL), as can be inferred by visual inspection of the differential reflectivity field between 20

Table 4.3: Main characteristics of the three radars used for the application and evaluation of the method. SHV stands for Simultaneously transmitting H and V polarizations, while ATSR is for Alternate Transmit Simultaneous Receive mode. The Pulse Repetition Frequency (PRF) of the C-band radar is staggered with ratio 3/2.

	S-band CHILL	**C-band** Bric	**X-band** CASA IP1
Location	Greeley, CO	Torino, Italy	Cyril, OK
Altitude (m)	1432	736	424
Antenna diameter (m)	8.5	4.2	1.5
Antenna beam width (°)	1.0	0.93	1.8
Antenna gain (dB)	43	45	37
Polarization type	ATSR & SHV	SHV	SHV
Operating frequency (GHz)	2.725	5.640	9.403
Transmitter peak power (kW)	800	250	25
Pulse width (μs)	1.0	0.5	0.4
PRF (Hz)	1000	882/588	2000
Range resolution (km)	0.150	0.340	0.096
Range (km)	140	170	40

and 60 km range. In particular the separation line between solid and liquid/mixed phase particles in Fig. 4.10c follows quite well the boundary between marked positive and weakly positive $Z_{\rm dr}$. The bin-based classification after the temperature adjustement is represented in Fig. 4.10e, while Fig. 4.10f shows the region-based final classification using a different pair of coefficients (λ=1.0, α=0.95). Both region-based classifications (panels b and f) show a considerable improvement respect to the first-guess bin-based map (panel e) in terms of noise reduction, but the choice of the λ and α coefficients determines some difference in the output results. As previously mentioned the classification with λ=0.8 and α=0.75 is a compromise between a conservative approach (low λ), which would basically act like a spatial filter, and a more data-adaptive approach (high λ). In this case the final classification is basically the output of the cluster analysis (block 3 in Fig. 4.1), since only a couple of small

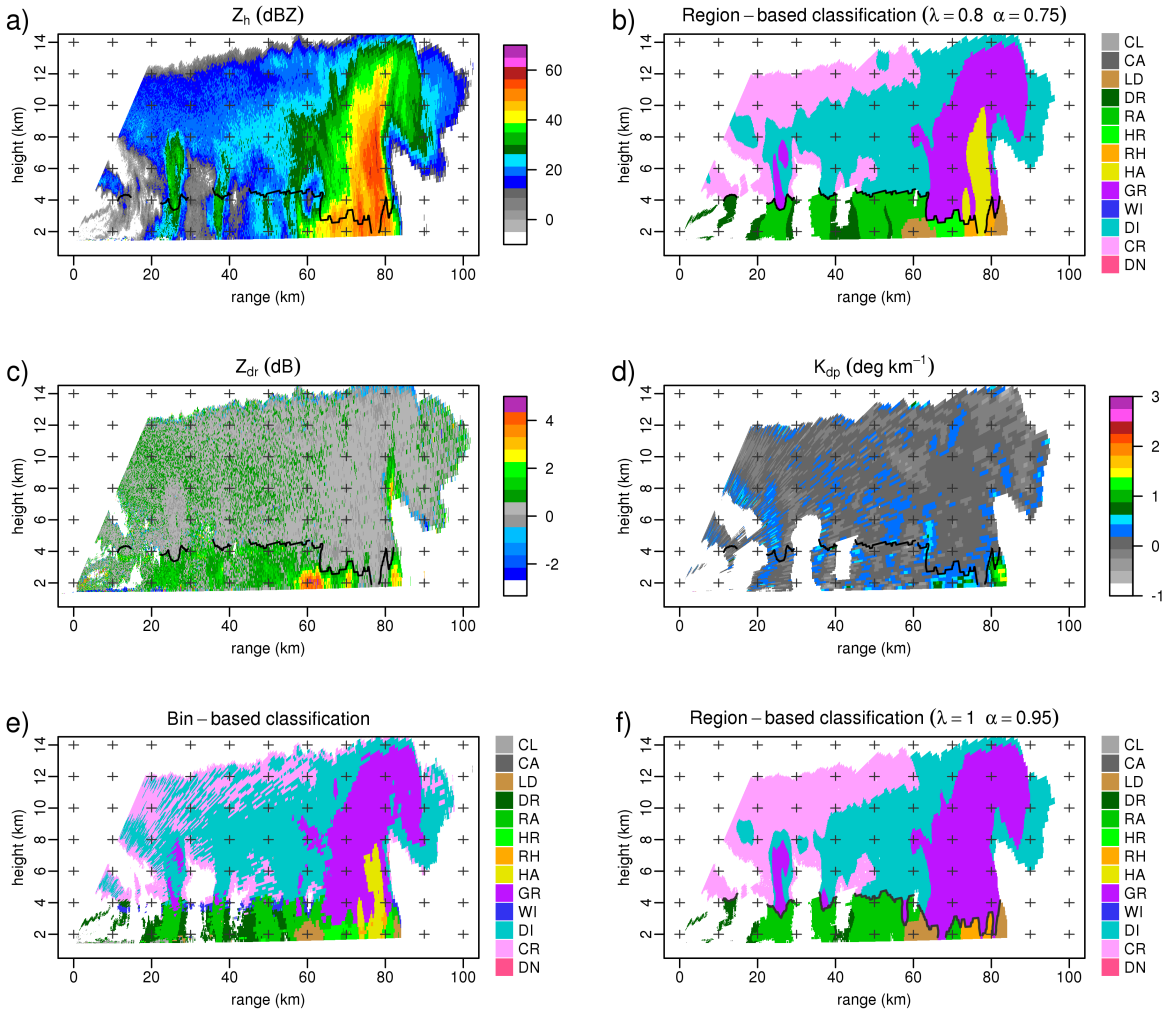


Figure 4.10: RHI along the 135° azimuth on 22 June 2013 at 23:06 UTC, from CHILL radar. Reflectivity (a) and final region-based hydrometeor classification with λ=0.8 and α=0.75 (b). The classes in the legend are: *clutter* (CL), *clear air* (CA), *large drops* (LD), *drizzle* (DR), *rain* (RA), *heavy rain* (HR), *rain + hail* (RH), *hail* (HA), *graupel* (GR), *wet ice* (WI), *dry ice* (DI), *crystals* (CR), *dendrites* (DN). Z_{dr} (c), K_{dp} (d), the bin-based hydrometeor classification after temperature adjustement (e) and the region-based classification using λ=1.0 and α=0.95 (f). The overplotted solid black line in panels a) to d) marks the separation between solid and liquid/mixed phase particles in the final classification of panel b).

wet ice connected region (composed of less than 100 polar bins) have been reassigned in the final step (block 4).

The use of high values of λ through the K-means cluster iterations (λ=1.0, α=0.95) on the other hand may significantly modify the characteristcs of the clusters. Several regions

showing a major evolution in their extension and properties have been reassigned based on the fuzzy logic classification of the connected regions. Specifically five regions (three of which composed by more than 10^3 polar bins) have been reassigned in the final step of the classification, out of 36 total connected regions. Some of the large *drizzle* regions in Fig. 4.10f were previously classified as *wet ice* after the cluster analysis. These regions have grown considerably during the K-means cluster iterations, reaching an extension from the freezing level down to the surface, which caused the subsequent automatic reassignment to a more plausible class. This specific behavior can be seen as a lack of a well spatially defined signature for the *wet ice* (although still sufficient for the freezing level adjustment). A moderate enhancement in Z_{dr} is apparent, while both Z_h and ρ_{HV} do not present the expected spatial trend in a typical well defined melting layer.

The most relevant difference respect to the classification in Fig. 4.10b is the lack of the *hail* region, replaced in the lowest levels by a slightly enlarged *rain hail* region. The *rain hail* is distinguished from the overhanging *graupel* mainly for the lower correlation coefficient (and higher temperature), while K_{dp} reaches values up to $1^\circ\,\mathrm{km}^{-1}$ in *rain* between 65 and 75 km range but $K_{\mathrm{dp}} \sim 0^\circ\,\mathrm{km}^{-1}$ between 75 and 80 km near the surface and above (Fig. 4.10d). The *hail* area aloft grown during the cluster analysis at the expenses of the surrounding *graupel*, leading to the inclusion of excessively low reflectivity values to be still acceptable for a *hail* classification. The median reflectivity lowered from 54.6 dBZ in the bin-based classification to 50.9 dBZ, leading to a final reassignment to the *graupel* class. The distinction between *hail* and *graupel* in this case merely lies on the absolute value of the reflectivity, since all the other polarimetric variables show rather weak signatures. This causes the *hail* area to grow within the *graupel* bigger region until an equilibrium is reached when the two regions have similar extensions. This behavior can be viewed as a limitation

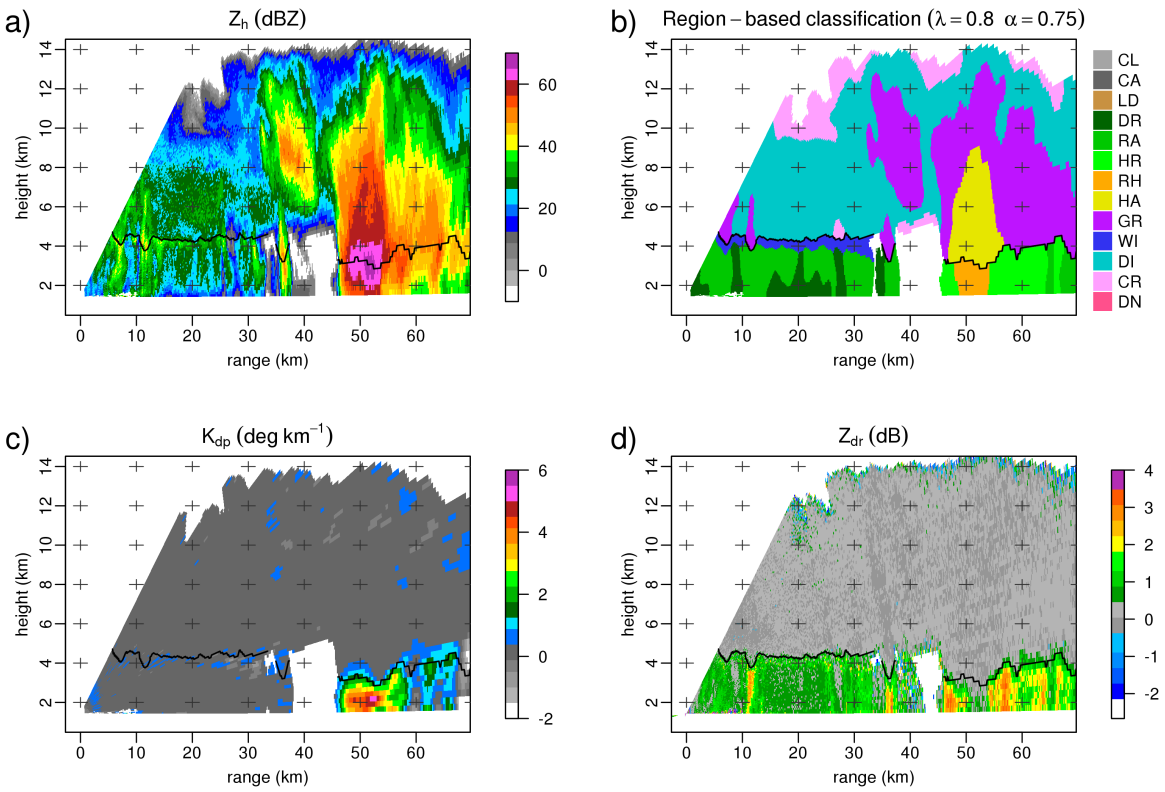


Figure 4.11: RHI along the 45° azimuth on 2 July 2008 at 22:57 UTC, from CHILL radar. Reflectivity (a), region-based hydrometeor classification (b), K_{dp} (c) and Z_{dr} (d).

of the method, but it is actually stressing the low discriminating power between *hail* and *graupel* based on the available observations in this specific case. An easy workaround in this type of situation may be a later reassignment of a *graupel* sub-region solely based on the reflectivity value.

Fig. 4.11 illustrates another example of a deep convective storm observed in July 2008 by the CHILL radar. As for the previous case, these dual-polarization measurements are also collected in ATSR mode, but with a larger number of integrated pulses (256) and a radial resolution of 0.2° resulting in high quality moments and very little noise. This can be considered a text case, with very well defined dual-polarization signatures. In particular the K_{dp} signal is very strong and helps to designate precisely the boundary between the *hail*

and *rain + hail* regions. The *rain + hail* region is characterized by the high Z_h and $K_{\rm dp}$, but lowering $Z_{\rm dr}$ in particular between 50 and 55 km range. In the stratiform region on the other hand $K_{\rm dp}$ provides very little information for the region-based classification, while $Z_{\rm dr}$ is fundamental for the definition of the transition between *dry ice*, *wet ice* and *rain*.

The two examples presented elucidate the most noteworthy characteristc of this method. The region-based classification is empowered by a controllable degree of self-adaptation to the most prominent spatial polarimetric signatures. This essentially means that the procedure itself does not need any specific tuning for a particular radar or operating frequency. At S band the classification is especially sensitive to Z_h and $Z_{\rm dr}$, while at attenuating frequencies like C and X band $K_{\rm dp}$ is automatically gaining importance owing to the frequency scaling and the consequent better definition of spatial patterns for several hydrometeors.

4.3.2. C-BAND, STRATIFORM PRECIPITATION (ITALY). Observations collected in PPI mode with 1.0° angular resolution by the Italian C-band dual-polarization operational radar located in Torino are discussed in this section. The case presented refers to a warm frontal stratiform precipitation system illustrated in detail in Chapter 3. The freezing level optimization loop estimates a -2.25°C temperature shift (Fig. 4.4), lowering the 0°C altitude from approximately 2.3 km (Milano Linate sounding at 00 UTC) to about 2.0 km MSL. Fig. 4.12 shows the PPI at 1.2° elevation of the radar dual-polarization observations and the resulting hydrometeor classification. The radar is located on a hill and surrounded by complex orography. The dual-polarization measurements are subject to side-lobe contamination which enhances the measurement noise and determines low $\rho_{\rm HV}$ values over some parts of the domain. Hydrometeor classification in stratiform widespread precipitation is in general an easy task when considering significant elevation angles, owing to the sharp transitions along the range on the conical surface. For low elevations however, the designation of the melting

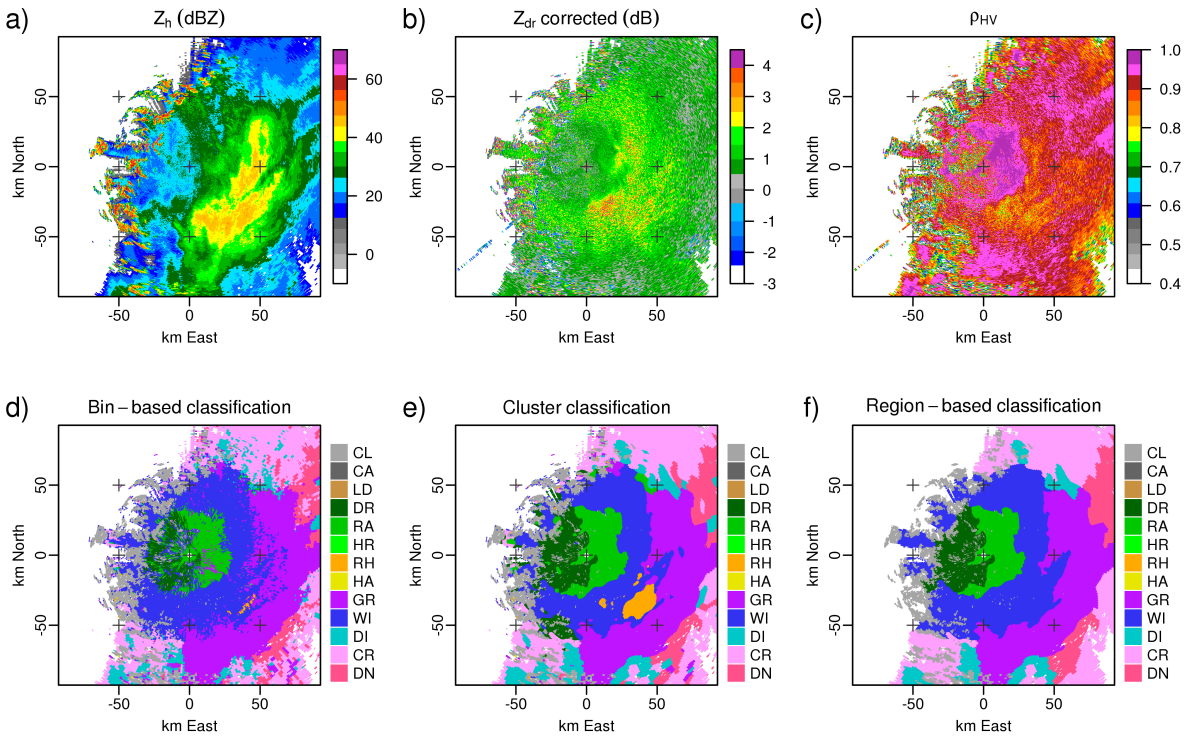


Figure 4.12: PPI at 1.2° elevation from the C-band Bric radar, on 27 April 2009 at 01:30UTC. observed Z_h (a), attenuation corrected Z_{dr} (b), ρ_{HV} (c), bin-based hydrometeor classification (d), cluster-based classification (e) and final region-based classification (f).

layer boundaries is not straightforward. In Fig. 4.12d the bin-based classification shows a noisy map, with ambiguous results especially in the low reflectivity area in the western sector. The low correlation and noisy Z_{dr} caused by side lobe effects produced many clearly wrong *wet ice* identifications close to the surface. The final classification in contrast shows a definite improvement, mainly in terms of overall noise reduction and very well separation between liquid (*drizzle* and *rain*) and mixed (*wet ice*) precipitation. The lower boundary of the melting layer appears reasonably well defined with good enough spatial resolution.

In the bin-based classification (Fig. 4.12d) 157 bins are suspiciously identified as *rain hail*, due to the relatively high attenuation corrected reflectivity (the 25, 50 and 75% percentile are respectively 50.1, 51.1, 52.5 dBZ) and the low correlation coefficient (0.817, 0.870, 0.898).

From this initial classification an enlarged region has grown during the cluster analysis (Fig. 4.12e), encompassing about 1200 bins with (25, 50 and 75%) percentiles of Z_h and ρ_{HV} respectively (46.4, 48.0, 49.6 dBZ) and (0.839, 0.874, 0.902). This expanded *rain hail* region (and few other smaller) is eventually reassigned to *wet snow* in the final step, because of the low affinity measure $\mathrm{AFF_k}(x_i) = 0.49$ for the intensity reduced Z_h distribution. The high K_{dp} values up to 1.8 $^\circ\,\mathrm{km}^{-1}$ (not shown) well above the melting layer are responsible for the two regions of dendrites crystals adjacent (and at higher altitude) to the extended *graupel* area. This classification appears fairly consistent with the existence of a mid-level weak widespread updraft associated with light riming above the melting layer (*graupel*) and sustaining the crystal growth by vapor deposition around the -15°C temperature region (*dendrites*), as discussed in detail in Chapter 3.

4.3.3. X-band, convective precipitation (Oklahoma). Fig. 4.13 shows the RHI observations (0.25° radial resolution) and hydrometeor classification from the CASA IP1 test-bed data in Oklahoma on 20 May 2011 at 04:14 UTC. The temperature profile is taken from Norman (OK) sounding observations at 00 UTC, reporting the height of the freezing level around 4.5 km. The 0°C level is then lowered to about 4.1 km MSL after the temperature optimization loop, providing an estimate of -3.75°C shift (Fig. 4.4).

The X-band observations are corrected for attenuation and differential attenuation. Fig. 4.13a and 4.13b show respectively the observed and the attenuation-corrected reflectivity. Attenuation up to 17 dB for Z_h and 2.2 dB for Z_{dr} is estimated. Given these high attenuation estimates, the weights associated with Z_h and Z_{dr} are reduced according to eq. 13. This explains how the *wet ice* region identification is possible, despite negative Z_{dr} values beyond the *rain hail* core. The presence of the *rain hail* region may actually be responsible for an underestimation of the differential attenuation, stressing the ultimate importance

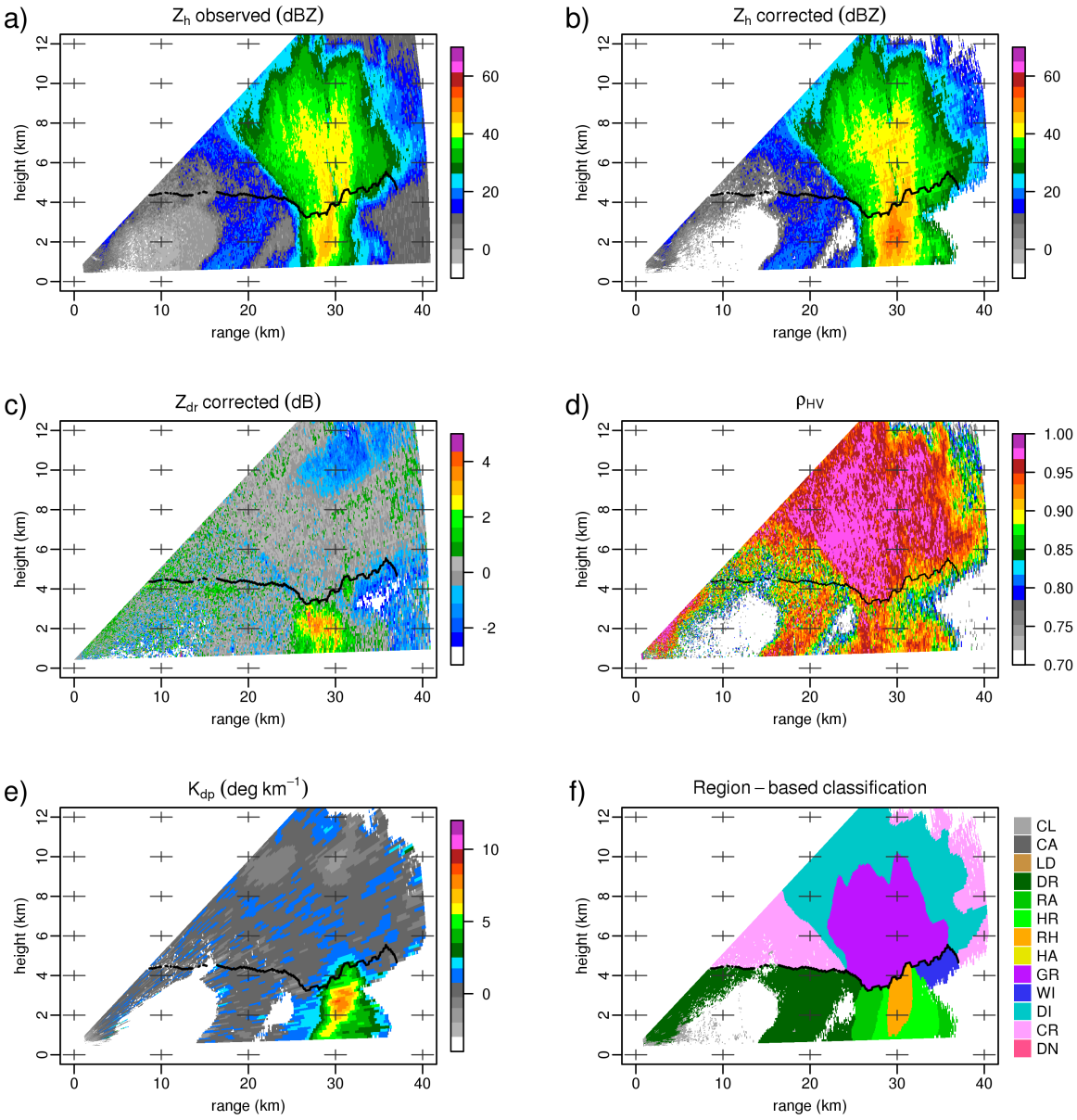


Figure 4.13: CASA IP1 RHIs of dual-polarization variables: observed Z_h (a), attenuation corrected Z_h (b), Z_{dr} (c), K_{dp} (d), ρ_{HV} (e) and final hydrometeor classification (f) on 20 May 2011 at 04:14 UTC, along the azimuth 101.6° .

of coupling the hydrometeor classification and attenuation correction for short wavelength radars. The classification appears in general reasonable with few well distinguished regions. It is especially apparent the relevance of the K_{dp} observations for delineating the modulated boundary between liquid and frozen particles.

The polarimetric signatures of K_{dp} and Z_{dr} at approximately 10 km height and 30 km range (negative values) point to the existence of vertically oriented ice crystals, associated with electrical activity inside the storm [33–35]. This specific hydrometeor class is not considered in our scheme, and the more generic definition adopted for *crystals* implies a relatively low reflectivity ($\leq$22 dBZ), which eventually causes this upper portion of the storm to lie within the *dry ice* region.

4.4. Evaluation of the method

A comprehensive direct validation of hydrometeor classification algorithms from radar is in general not possible due to the scarcity of in situ measurements. Even when in situ observations are available, the comparison is challenging considering the largely different sampling volumes of radar and cloud probes, as well as specific issues with airborne cloud particle sampling [107–109]. In this work we rely on an internal evaluation approach, considering the application examples illustrated in the previous section. In addition to the obvious noise reduction in the final classification, in the following sub-sections we focus on a metric for the evaluation of clustering algorithms and on the sensitivity to noise and bias in the input observations.

4.4.1. Clustering evaluation with Davies-Bouldin (DB) index.

Davies and Bouldin [110] introduced a metric for the internal evaluation of clustering algorithms, with the useful property of being independent on both the number of clusters analyzed and the method used to produce the data partition. The internal evaluation scheme provides a measure of the goodness of the cluster partitioning, based on quantities inherent to the employed dataset. The definition of the DB index relies on the ratio of a measure of the

scatter within the cluster (intra-cluster distance) to a measure of the separation between the clusters (inter-cluster distance):

$$\mathrm{DB} = \frac{1}{K}\sum_{k=1}^{K} R_k \tag{21}$$

where

$$R_k = \max_{k \neq j}(R_{k,j}); \quad R_{k,j} = \frac{\overline{d}_k + \overline{d}_j}{d_{k,j}} \tag{22}$$

In eq. 22 K is the number of clusters, $\overline{d}_k$ and $\overline{d}_j$ represent the dispersions (average distance between each point in the cluster and the centroid of the cluster) of clusters k and j, while $d_{k,j}$ is the distance bewteen the centroids of clusters k and j, using the same metric defined in eq. 16. It is clear that a lower DB value is associated with a better partitioning of the data, indicating that the hydrometeor classes are well separated based on the input dual-polarization measurements.

Rather than the absolute value of the DB index, we consider the relative variation of the index between different classification schemes. This allows in particular to assess the improvement achieved using the proposed methodology, respect to the traditional bin-based classification. Table 4.4 reports the DB index obtained for the cases described in Section 4.3 (with a supplemental higher elevation for the Bric case) in the three steps of the classification, after the adjustement for the temperature profile (blocks 2, 3 and 4 in Fig. 4.2). In addition, the results from two completely unsupervised K-means clustering are reported. The first one ($K-\mathrm{means}^{\mathrm{bin}}$) is obtained using the seeds from the bin-based classification (with the same number of classes). The second K-means clustering ($K-\mathrm{means}^{\mathrm{opt}}$) is performed using the

optimal number of classes (with random seeds), determined as the one which minimizes the DB index. In this case several experiments with independent initializations revealed that the effect of the random seeds on the final partions is quite limited, i.e. the variability in the number of output classes is in general ≤ 1 and the standard deviation of the DB index is ≤ 0.02.

For the cases considered, with the exception of the CASA X-band RHI, the cluster analysis produced an improvement in the partition of the hydrometeor classes, respect to the first-guess bin-based classification. This may sound tautological given the purpose of the clustering algorithm. But in fact, considering the specific minimization process including the additional contiguity and penalty terms (eq. 15), a reduction of the DB index is not automatically guaranteed. This result indicates that the method has the ability to conjugate a reduction in the output classification noise with a tighter affinity within the same class and a better separation of the hydrometeor classes. In most cases the DB index reduction is achieved with a concurrent reduction of the number of the hydrometeor classes, i.e. some hydrometeor type poorly represented in the bin-based classification have been suppressed in the cluster analysis.

During the last step of the classification process (region-based hydroclass in Fig. 4.2) the number of classes can be further reduced following the reassignment of one or more connected regions. This is the case for example for the CHILL 2013 RHI with λ=0.8 and α=0.75, where a couple of residual *wet ice* regions are converted to *rain* and *drizzle*, and the DB index of the final region-based classification is further decreased. For the same case, but with λ=1.0 and α=0.95, on the other hand the reassignments of some connected regions (in particular the *hail* region to *graupel*) reflects in an increase of the DB index. In fact while the high weight of the distance component in the clustering analysis (λ=1.0) is

responsible for a better clustering (DB=1.83 as compared to DB=2.24 for the same case with λ=0.8), the excessive deviation from the reference membership functions and the subsequent reassignment as discussed in Section 4.3.1 implies a deterioration of the clustering tightness, but still a better performance than the bin-based classification (DB=3.96).

The unsupervised K-means clustering using the seeds from the bin-based classification ($K-\text{means}^{\text{bin}}$ in Table 4.4) predictably shows a better performance in terms of the DB index, respect to the region-based classification. The DB index for the optimized K-means clustering ($K-\text{means}^{\text{opt}}$) is the lowest by definition, but may imply a consistent change in the number of hydrometeor classes. In fact the number of represented classes in the analysis is much lower respect to the final region-based classification for the CHILL 2013 and CASA cases, while it is higher for the Bric PPIs.

The K-means clustering classifications using the initial classes from the bin-based classification are shown in Fig. 4.14 for the CHILL 2008 and the CASA case of Fig. 4.11 and 4.13 respectively. The limitation of these analyses is clear, with several hydrometeor classes deviating too much from the original definition. Both the *hail* and *rain hail* regions have grown excessively in the CHILL case, including much lower reflectivity values. In addition, the lack of a temperature constraint in this process has led to the wrong *dry ice* classification of many bins below the melting layer and *drizzle* in the highest portion of the cloud (low reflectivity, high correlation and near zero K_{dp} and Z_{dr}). In the CASA case (Fig. 4.14d) the weak reflectivity region closer to the radar, due to the lack of any well defined polarimetric signature (and low ρ_{HV} likely arising from non-uniform beam-filling), resulted in a noisy classification from the surface up to about 7 km MSL. A fairly coherent region in the upper portion of the storm is identified and marked as *hail*, whereas this region is most likely to contain vertical ice crystals originating from cloud electrification processes.

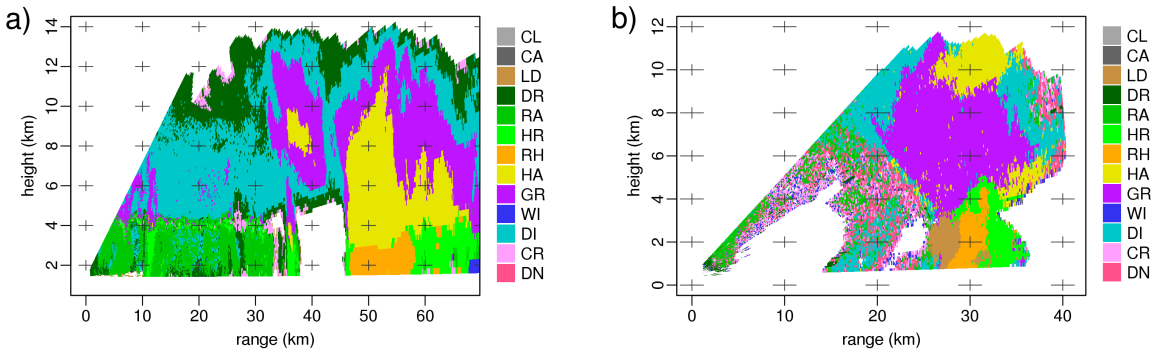


Figure 4.14: K-means unsupervised classification ($K-\text{means}^{\text{bin}}$ in Table 4.4) for a) CHILL 2008 case (Fig. 4.11) and b) CASA case (Fig. 4.13).

Looking back at the cluster analysis, the increased DB value for the CASA RHI (DB=3.44 in Table 4.4, compared to DB=3.09 for the bin-based classification) seems actually ascribable to the influence of the weak reflectivity region closer to the radar. In fact if the DB index is calculated only for the main storm, by simply a applying a reflectivity threshold of 25 dBZ, the resulting DB values are 2.30 (11 classes), 1.84 (9) and 1.66 (7) respectively for the bin-based, cluster and region-based classifications. These values present a similar qualitative behavior as in the other cases, showing a significant improvement in the final classification.

This case emphasizes the importance of the supervised component in the classification process. The membership functions, variables weighting, temperature optimization are all important components to derive a reliable first-guess. Furthermore the contiguity/penalty terms in the cluster analysis and the subsequent possible reassignment of the connected regions have a fundamental role in controlling the aggregation process. The effort pursued with the proposed method is ultimately to balance the data-adaptive strength of the clustering technique with the conventional fuzzy-logic, physically-based approach. In fact a pure unsupervised clustering may give better partitioning results in terms of DB index, but at the cost of results not physically meaningful (Fig. 4.14).

Table 4.4: Davies-Bouldin internal evaluation index (DBI) for the classification examples in Fig. 4.10, 4.11, 4.12, 4.13. The first three columns report the values of the index for the intermediate (bin-based and cluster analysis) and final (region-based) classification results. In brackets the number of distinct hydrometeor classes is indicated. The last two columns report the DBI value for the unsupervised K-means clustering, with the same number of classes as resulting from the bin-based classification ($K-means^{bin}$) and with the optimal number of classes ($K-means^{opt}$).

	Bin	*Cluster*	*Region*	$K-\text{means}^{\text{bin}}$	$K-\text{means}^{\text{opt}}$
CHILL 2013	3.96 (11)	2.24 (10)	2.14 (9)	1.21 (11)	1.02 (5)
CHILL 2013(*)	3.96 (11)	1.83 (10)	2.34 (8)	-	-
CHILL 2008	2.00 (10)	1.43 (10)	1.42 (9)	1.11 (10)	1.01 (9)
CASA	3.09 (11)	3.44 (9)	3.89 (8)	1.25 (11)	1.11 (3)
Bric 1.2°	2.67 (11)	2.45 (7)	2.51 (7)	1.37 (11)	1.27 (15)
Bric 10.0°	2.02 (7)	1.61 (6)	1.63 (6)	1.23 (7)	1.14 (14)

All classifications using λ=0.8 and α=0.75 except (*) which is obtained using λ=1.0 and α=0.95 (Fig. 4.10f).

Finally, the very limited reduction of the DB index for the C-band 1.2° PPI scan in Table 4.4 may indicate a difficulty to improve the first-guess hydrometeor classification, in terms of clusters tightness, for low elevations radar scans. Data at low elevations are sometimes lacking well defined polarimetric signatures, especially in vertically stratified clouds where the radar resolution volume may encompass particles in different thermodynamic phases. Running the method on a higher elevation (10°) from the same volume actually produces a better result (last line in Table 4.4), in agreement with the RHI scans. The cluster analysis in fact achieves a more significant decrease of the DB index in this case (from ∼2.0 to ∼1.6).

4.4.2. Sensitivity to noise and bias. In this section we consider the sensitivity to input noise and bias, with the aim to assess the impact of the cluster analysis in mitigating the effect of perturbed inputs. For this purpose we compare the bin-based classification with

the cluster-based classification (output of block 3 in Fig. 4.2), after properly altering the original radar measurements described in Section 4.3.

The sensitivity to additional noise in the input variables Z_h and Z_{dr} is shown in Fig. 4.15, while Fig. 4.16 reports the corresponding results for ρ_{HV} and K_{dp}. The white noise is generated using a Gaussian distribution with increasing standard deviation (x-axis). The extent of the noise considered is clearly higher than the normal levels found in real observations, but is used here to emphasize the fundamental qualitative trend. The lines in the plots represent the number of polar bins that changed class, compared to the reference classification using the original un-perturbed observations. The colored areas between the two lines (bin-based and cluster-based classifications) highlight the improvement (light blue color) or worsening (orange color) of the proposed method in terms of robustness to the noise. The fact that the classification using perturbed inputs remains closer to the original one is taken as an indication of its robustness, especially for real-time applications.

With few exceptions the classification resulting from the proposed method shows a lower sensitivity to the noise in the radar observations, compared to the bin-based classification. The most notable improvement is obtained for the ρ_{HV} and K_{dp} radar measurements. The specific asymptotic behavior in the ρ_{HV} plots (Fig. 4.16, panels *a*) to *e*)) seems due to the fact that although the measurable range of ρ_{HV} is 0 to 1, the actual range for meteorological targets is about one order of magnitude shorter, i.e. $\sim$0.85 to 1.0. Consequently the contribution of ρ_{HV} to the particle classification becomes practically irrelevant for additional white noise with standard deviation in excess of 0.1.

The plots referring to the 1.2° PPI from the Bric radar (all four radar variables: panels *d*) and *i*) in Fig. 4.15 and 4.16) and the 2013 RHI from CHILL (limitedly to Z_h and Z_{dr}: panels *b*) and *g*) in Fig. 4.15) show a small impact for the new classification method. As

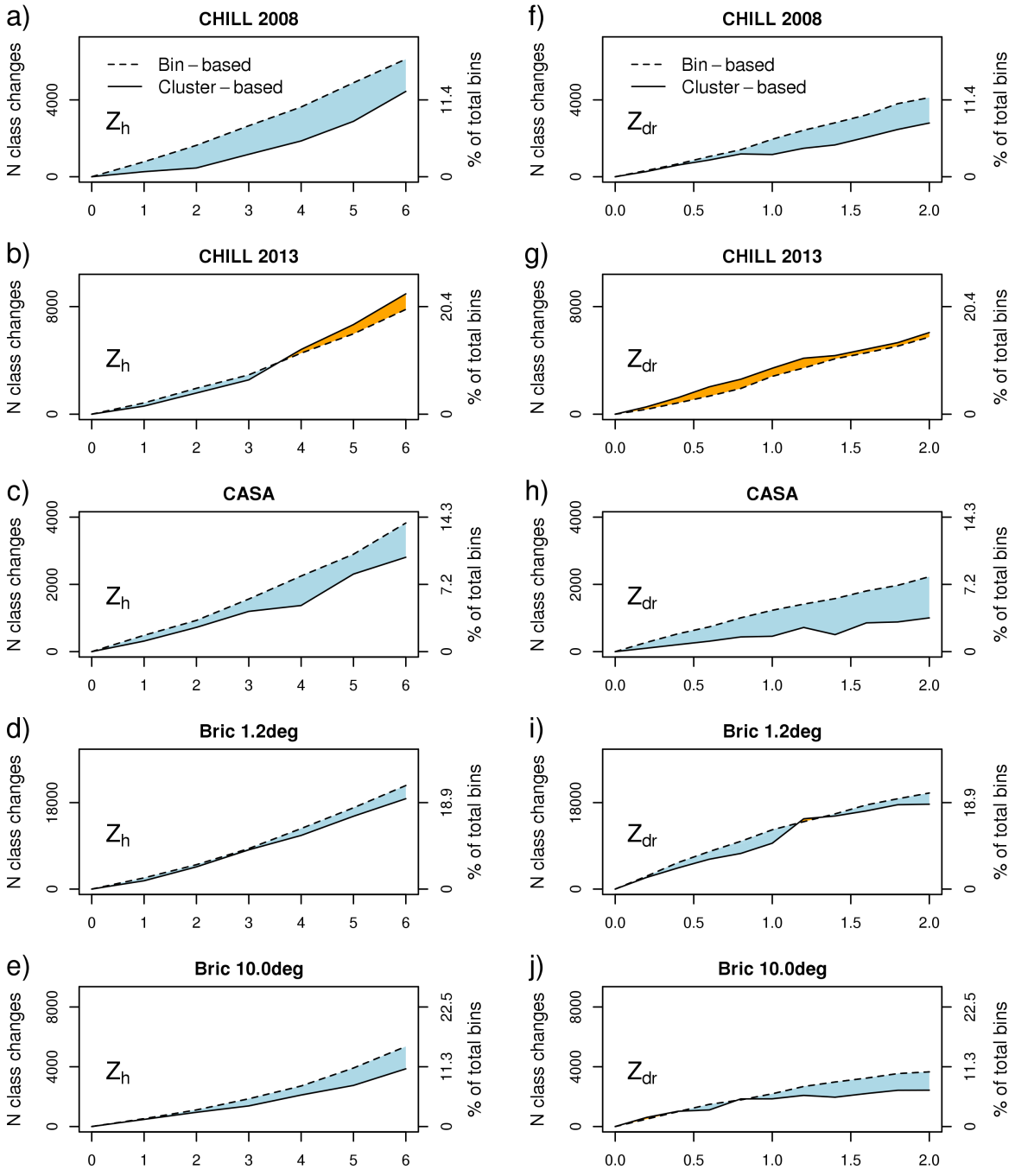


Figure 4.15: Sensitivity to an artificial Gaussian white noise in the input observations Z_h (panels a) to e)) and $Z_{\rm dr}$ (panels f) to j)) for the cases analysed in Section 4.3. The plots report the number of polar bins that changed class with respect to the reference hydrometeor map without added noise, for the bin-based and the cluster-based classifications (output of block 3 in Fig. 4.2, with λ=0.8, α=0.75). In addition, the right y-axis reports the fraction (in %) of bins changed respect the total number of valid bins. In order to facilitate the interpretation, the area between the two lines is colored in light blue when the cluster-based classification produced less class changes, in orange otherwise.

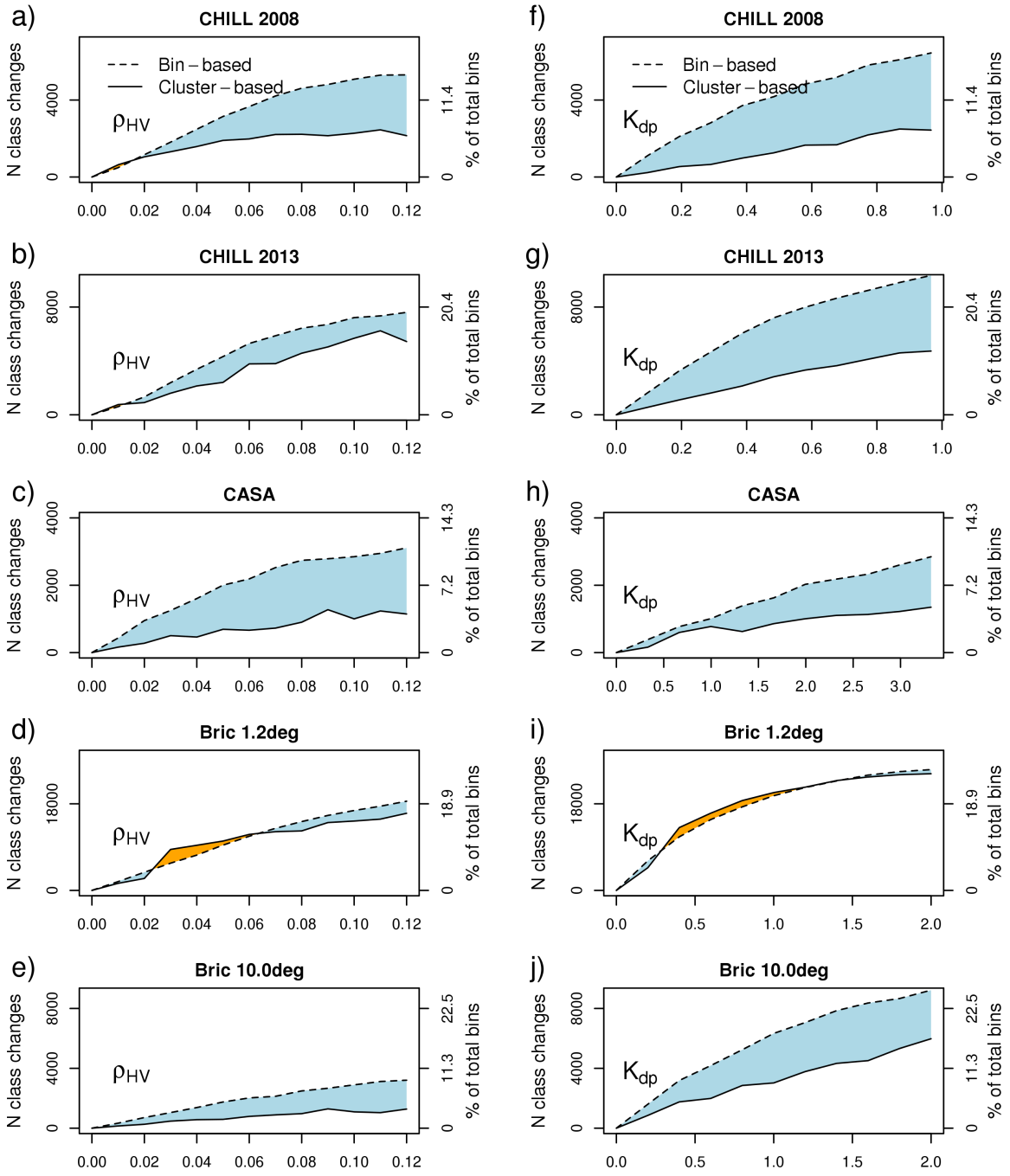


Figure 4.16: As in Fig. 4.15, but for the input observations ρ_{HV} (panels a) to e)) and K_{dp} (panels f) to j)). The K_{dp} noise introduced is scaled with the radar frequency.

observed in Section 4.4.1 the beneficial effect of partition clustering is quite limited for the low elevation Bric scan and this may explain the neutral impact in terms of sensitivity to

additional noise in the radar observations. For the CHILL 2013 case on the other hand, there is a lower sensitivity to noise in ρ_{HV} and K_{dp}, while for Z_h and Z_{dr} no significant differences are noted. A more detailed analysis shows that this specific behavior may be ascribed to the generation of several small regions of *wet ice* in the stratiform portion of the storm, when the noise in Z_h or Z_{dr} is increased. This splitting into multiple small regions around the melting layer has the effect of vanishing the advantage of the cluster analysis in relation to noise sensitivity. The same negative effect is not seen in the CHILL 2008 case, which is also showing a relevant stratiform region. The main difference between the two cases is that in the 2008 case (Fig. 4.11) the melting layer signature is well marked in few variables, i.e. Z_{dr}, ρ_{HV} (not shown), and weakly in Z_h and K_{dp}, while in the 2013 case (Fig. 4.10) only Z_{dr} shows the characteristic increase due to the transition between snow and rain. As a consequence, when the noise in Z_{dr} is increased, the cluster analysis has no other observations to rely on in order to preserve the original consolidated regions.

The sensitivity to additional bias in the input power variables Z_h and Z_{dr} is also considered and represented in Fig. 4.17. While the effect of additional reflectivity bias (Fig. 4.17, panels *a*) to *e*)) is rather ambiguous, the sensitivity to the Z_{dr} bias (Fig. 4.17, panels *f*) to *j*)) is significantly lower for the cluster-based classification, compared to the bin-based classification. This is interpreted as one of the beneficial effects of the data-adaptive aggregation process during the cluster analysis. The unsupervised K-means clustering would actually be completely insensitive to systematic biases in the observations. The proposed semi-supervised procedure is however able to maintain a certain degree of insensitivity. The key point is the adaptive aggregation of the observations during the cluster analysis, eventually attaining a set of homogeneous connected regions. This is possible of course when the classification relies on multiple input variables, so that if one variable is biased the

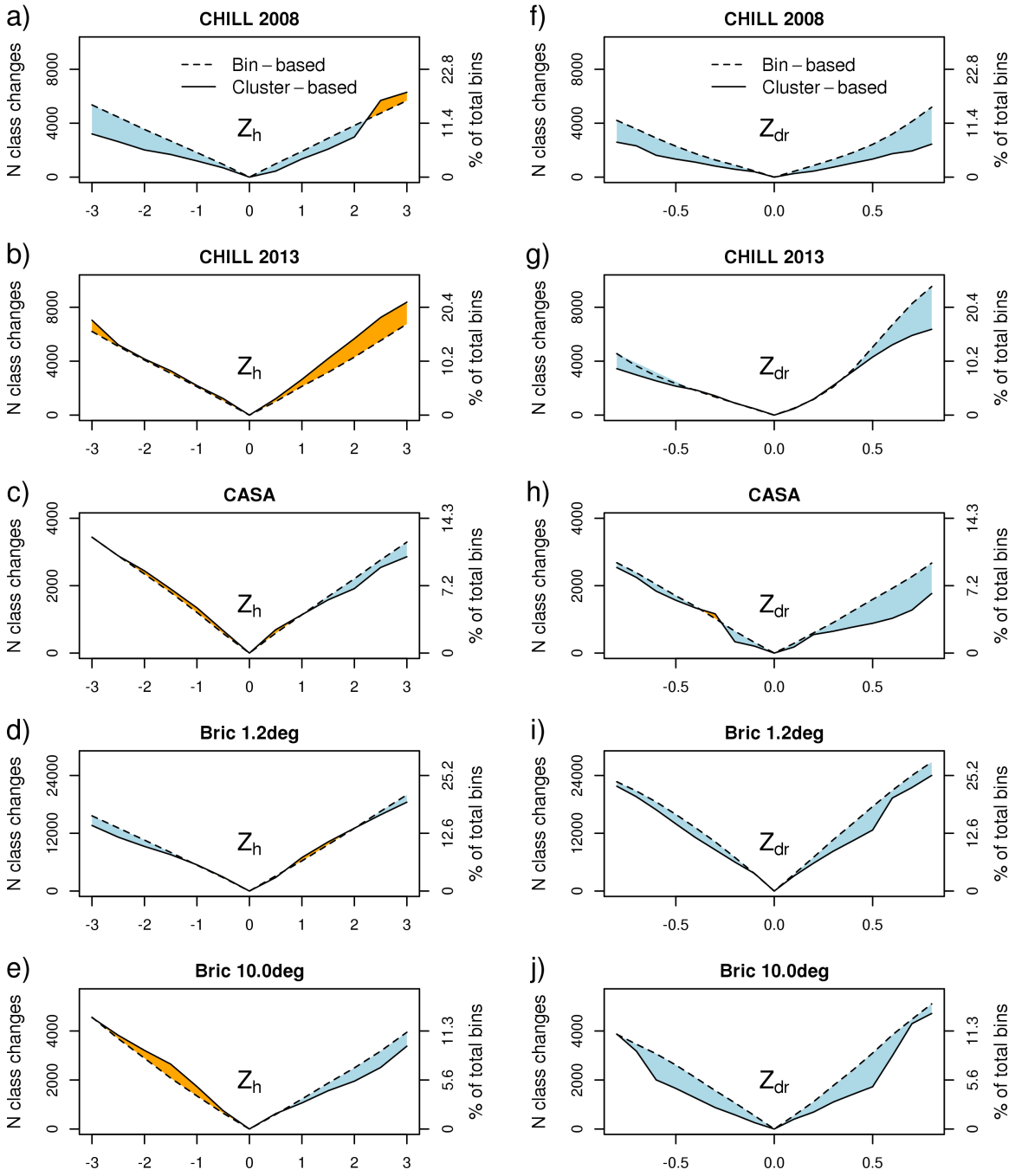


Figure 4.17: Similar to Fig. 4.15, but here the sensitivity to an artificial bias is shown, for the input observations Z_h (panels a) to e)) and $Z_{\rm dr}$ (panels f) to j)).

other variables will still drive a specific hydrometeor identification and involve a degree of self-adaptation to the biased observations in the cluster analysis.

The reflectivity does not show a clear improvement in the sensitivity to the additional bias. This behavior may be explained considering the peculiarity of the reflectivity measurements, as opposed to the other dual-polarization variables considered. The correlation coefficient ρ_{HV} and the differential variables Z_{dr} and K_{dp} are often characterized by abrupt changes corresponding to transitions between different hydrometeor types. This natural segregation behavior is exploited in the cluster analysis. Reflectivity by contrast is an absolute power measurement which does not depend on the shape of the hydrometeors, so it most often behaves as a continuum involving only gradual quantitative transitions. For this reason it appears to be less suited for a data-adaptive approach in a clustering analysis. As a matter of fact a key limitation of the pure K-means clustering is the tendendy to produce clusters of approximately similar size, as was shown in the CHILL 2013 example (Fig. 4.10) using λ=1.0. In that case the *hail* region expanded excessively mainly because of the continuous decrease of the reflectivity outwards from the storm core. The current method already limits this issue through the use of the cost function involving the additional contiguity and penalty terms ($\lambda \leq 1$, $\alpha << 1$).

4.4.3. Implementation and performance evaluation. The procedure is coded in C language for its efficiency and portability. The implementation is modular, in the sense that various parts of the procedure can be activated by editing a text configuration file. In fact, depending on the specific radar data being used, some of the processing may result redundant, e.g. in some cases the input radar data may already contain a K_{dp} estimate. All the radar-specific parameters (radar constant, operating frequency, etc.), membership functions (one table for each class and frequency band) and algorithm-specific parameters (λ, α, γ, extension of the contiguity window, etc.) are defined in external configuration files.

This provides a considerable flexibility for testing and adaptation to new radars/wavelengths, without need to edit and recompile the source code.

Table 4.5 presents a summary of the algorithm performance for the cases analysed, running on a 2010 laptop hosting a Intel *i*7 CPU M 620 @2.67GHz. The total wall-clock time is clearly dependent on the number of range-bins processed, as reported in the first row. While the RHI scans took about 10-12 s to complete the classification procedure, the processing of the low elevation PPI needed almost twice that time to complete. The 10° elevation PPI on the other hand only took ∼8 s. One of the most time-consuming modules, i.e. the temperature optimization loop, does not need to run at every radar scan. In fact the height of the freezing level is not expected to vary significantly at the time scale of the scan update frequency. In an operational implementation this module may only be executed e.g. every 30 minutes and only for selected elevations from a 3D polar volume. Overall, the performance appears compatible with real-time operations.

4.5. Discussion and conclusions

This study has two major goals:

- exploit the spatial information content of dual-polarization radar observations for the purpose of hydrometeor classsification
- enhance the standard bin-based fuzzy logic classification by additionally considering the quality of the classification itself, the spatial coherence and the self-aggregation propensity of the radar observations

Thus the basic fuzzy logic classification method is improved with a data-adaptive approach, typical of clustering methods (unsupervised component). Additional supervised

Table 4.5: Summary of the detailed performance of the hydrometeor classification procedure for the different cases analysed. The total number of valid range bins processed is indicated in the first row. The partial times are provided for the modules depicted in Fig. 4.2, with item 2) including $\Psi_{\rm dp}$ filtering and the calculation of $K_{\rm dp}$. For items 1) and 3) the number of iterations performed is indicated in brackets. The I/O time represents the time spent for reading and writing the radar data files.

	CHILL 2008	CHILL 2013	CASA	Bric 1.2°	Bric 10.0°
N valid data bins	35,188	40,046	28,173	95,368	35,512
			Wall-clock time (s)		
1) T-profile adjust. (N_{iter})	3.2 (8)	5.7 (11)	2.6 (8)	9.2 (9)	3.5 (9)
2) Bin-based hydroclass	0.7	0.8	1.1	2.4	0.9
3) Cluster analysis (N_{iter})	5.2 (4)	6.3 (5)	7.6 (4)	7.2 (6)	2.9 (5)
4) Region-based hydroclass	0.1	0.2	0.1	0.5	0.3
I/O	0.2	0.3	0.1	0.2	0.2
TOTAL	9.5	13.3	11.5	19.5	7.8

components are provided in the form of pairwise constraints (penalty term) and elementary physical rules (temperature limits for the connected regions).

The method is implemented in a four-steps procedure (Fig. 4.2), including optimization of the temperature profile (step 1), bin-based fuzzy logic classification (2), cluster analysis (3) and fuzzy logic classification applied to the connected regions (4). The key features of the proposed method are a degree of self-adaptability and the shift from classical bin-based classification schemes to a region-based classification.

The method has been tested on few cases in different meteorological situations and using data from radars operating at different frequencies. These preliminary application examples show promising results, especially for RHI scans. Continuous scanning in elevation clearly helps to detect the vertical variations in the dual-polarization observations, with a resulting higher segregation across a region encompassing a range of different microphysical species.

The classification also worked fairly well for the analyzed PPI scan in widespread precipitation, likely due to the large amount of available data across the transition regions. However more tests considering a wider range of precipitation events is needed to asses the potential for real-time application to volume scans resulting from a collection of PPI at fixed elevation angles.

An analysis of the performace showed that the method is suitable for operational applications. In addition, for real-time applications the recent history of clusters description should be considered to improve their statistical representativeness, providing a classification both robust and stable over time. To this end, the use of specific RHI interleaved with volume scans (some radars, e.g. during CASA IP1, actually perform this type of complex strategy) will be investigated to better train the clusters of the hydrometeor classes.

The sensitivity to noise and bias in the input variables has been shown for five cases (three RHIs and two PPIs). The semi-supervised approach seems able to maintain a certain degree of insensitivity to a measurement bias in Z_{dr}. For Z_h however there is no defined improvement respect to the bin-based method. In addition the method showed a reduced sensitivity to noisy observations. In fact, the addition of a white Gaussian noise to the four radar variables (Z_h, Z_{dr}, K_{dp} and ρ_{HV}) resulted in less changes in the classification resulting from the cluster analysis (compared to the classification with original measurements) when the proposed method is used. This is certainly a noticeable outcome of the method, which is able to provide a clean and appealing classification also for operational applications. It is clear that for a generalization of these preliminary findings the analysis of a larger and comprehensive dataset is needed.

Going from a bin-based to a region-based classification has several advantages. The regions are identified from the semi-supervised analysis, which is able to reduce the intra-cluster distance and increase the inter-cluster distance (lower Davies-Bouldin index). This allows a more clear separation between the different hydrometeors. The final fuzzy logic classification applied to the connected regions is very robust, being based on a statistical sample rather than on single point measurements. In addition, the classification is locally adaptive, in the sense that it automatically exploits and adapt to the most marked spatial signatures in the polarimetric observations.

Having defined a classification map with (relatively) few connected regions may in the future allow the inclusion of further spatial processing, e.g. based on the polygons defining the regions, the boundaries between different regions or their orientation (e.g. horizontally elongated for melting hydrometeors, vertically-elongated for hail). In general, working with sets of contiguous bins, instead of pointwise measurements, considerably widens the potential of the classification algorithm, by allowing the inclusion of more sophisticated physical constraints based on either spatial properties or microphysical processes.

Working with regions however also brings in some issues specific to spatial analysis, in particular the modifiable areal unit problem [111]. In fact when point-based measurements are aggregated into contiguous regions, the resulting statistics is affected (biased) by the size and boundaries of the regions. The size effect, mainly controlled by the contiguity window in our procedure (Fig. 4.5), must be considered in relation with the processes one is aiming to represent. The scale of the process is generally different from the spatial resolution of the data. In cloud microphysics the scale of the process can range from sub-pixel (processes within the radar resolution bins, e.g. turbulent mixing) to much larger scales, like organized updrafts, phase changes across the melting layer, etc. The heuristic approach adopted in the

proposed methodology is clearly aimed at representing the larger scale features in the cloud, by aggregating a multitude of radar bins with similar characteristics.

In addition to the parameters specified for the definition of the membership functions and the weights associated to the input variables, the current procedure involves additional parameters in the cost function (λ, α, γ), and few other arbitrarily defined thresholds. The parameters of the cost function in particular can be set depending on the specific purpose of the classification. For example, if λ is very low (near zero) then the method simply acts like a smoother. If λ is set to a high value (close to one) on the other hand, the method could be useful for exploratory analysis and eventually for refinement of the membership functions.

On the downside, the increased degrees of freedom in this method may complicate the set-up for a specific system and also the interpretation of the final results. However, the proposed method is mainly devised to be super-imposed on an existing fuzzy-logic classification algorithm. Therefore the membership functions are assumed to be already fairly established and the set-up of the procedure should only involve the specification of the cost function and few other parameters.

CHAPTER 5

A methodology to analyze the wind field in the storm environment using dual-polarization observations

Radar provides high density observations inside the storm, but in general a poor sampling of the environment around. However, the thermodynamic characterization of the storm environment is fundamental for a comprehensive analysis, the basic building block of any forecasting model.

Conventionally the problem of incorporating the available observations into a physically-consistent analysis is addressed within the framework of data assimilation (Chapter 2). For the assimilation of radar data the typical strategy involves the definition of proper observation operators to map the model space (model variables at model grid points) into the radar observation space (reflectivity, Doppler velocity, dual-polarization parameters at sampling locations). For the assimilation techniques based on statistical estimation theory (optimal interpolation, variational assimilation, ensemble Kalman filter) the resulting analysis aims at the minimization in a least squares sense of the distance between the observations and the model background [10], given a proper definition of the model and observations covariances. However, this global minimization may not preserve important local features inherent in the radar data. In fact the assimilation of radar observations for convective storms, as opposed to stratiform systems (Chapter 2), presents relevant challenges. In particular, serious issues may arise if the background (a model state used as first guess) has no convection whereas the radar indicates precipitation, or vice versa. For very short-term forecasting of convective precipitation and winds, it may be preferable to retain significant features resulting

from the radar imagery. In addition to Doppler signatures of convergence, divergence, and rotation, the columns of Z_{dr} extending above the freezing level are an example of a polarimetric signature that may be useful to mark the location of the updraft within a storm. The correct location of the updraft regions is indeed fundamental to correctly reproduce the main dynamics of the storm and forecast its very short-term evolution. Besides advection by mid-tropospheric winds, in fact, one of the primary mechanisms of propagation of convective systems is by the action of cold-pools [112]. A low-level analysis is needed to realistically represent the gust-front associated with the storm propagation.

Dual-Doppler analysis has been widely used to reconstruct the three-dimensional wind field in the storm, but mostly offline for specific case studies. In fact dual-Doppler is seldomly available or over only limited portions of the spatial domain, depending on the beam intersecting angles. The proposed feature-based approach exploiting the Z_{dr} column signatures is intended for application to individual radars, so it could be more easily implemented for real-time operation. In addition, inclusion of surface wind measurements is essential to reproduce the relevant flow around the storm, where Doppler observations are scarce due to the lack of scattering targets.

The purpose of the analysis procedure described in this Chapter is primarily to demonstrate the potential offered by an elaborated radar product for possible inclusion in data assimilation systems. This may represent a complementary option to assimilate the information from radar, in addition to the standard processing of the whole amount of raw observations on a regular grid.

5.1. Dual-pol features: columns of Z_{dr} and relation with updraft

The updraft within storm cells can reach velocities as high as 20-30 $\mathrm{m\,s^{-1}}$. This intense upward motion may then exceed the terminal fall velocity of the drops in the radar resolution volume. While the smaller drops are carried aloft and eventually evaporated in the intense air flow, the resulting DSD deprived of the smaller drops results in a distinctive Z_{dr} signature. In fact if the updraft is sufficiently strong, also the bigger drops in the distribution are carried aloft at heights even few kilometers above the environmental freezing level. This particular feature has been reported since the early nineties [93, 94] and named Z_{dr} column.

The connection of the Z_{dr} columns with the updraft offers interesting clues for nowcasting the storm evolution. Indeed the potential for short-term forecasting has been investigated over the years among others by Scharfenberg et al. [113] and Van Den Broeke et al. [114]. Picca et al. [115] have shown evidence of the lagged correlation between Z_{dr} columns and the increased production of precipitation at the surface, including hail. Snyder et al. [116] recently demonstrated the strong spatial association between updrafts and Z_{dr} columns using high-resolution numerical simulations with spectral bin microphysics and a polarimetric forward operator.

Fig. 5.1 shows the 2-hours track of a convective cell observed by the C-band radars of Bric della Croce (table 4.3) and Monte Settepani, another C-band system located 95 km South-South-East of the Bric radar. The storm produced large hail at the surface, as per local reports. The presence of large hail in the storm is also inferred from the three-body scattering occurring in the radar volume data between 13:50 and 14:20 UTC (not shown).

Fig. 5.2 shows the intensifying convective cell between 13:30 and 14:00 on 22 May 2005. At 13:30 UTC a Z_{dr} column is clearly visible on the Southern edge of the convective storm in the 4000 m CAPPI in fig. 5.3a. The Z_{dr} column extends up to $\sim$6 km (fig. 5.3b), about

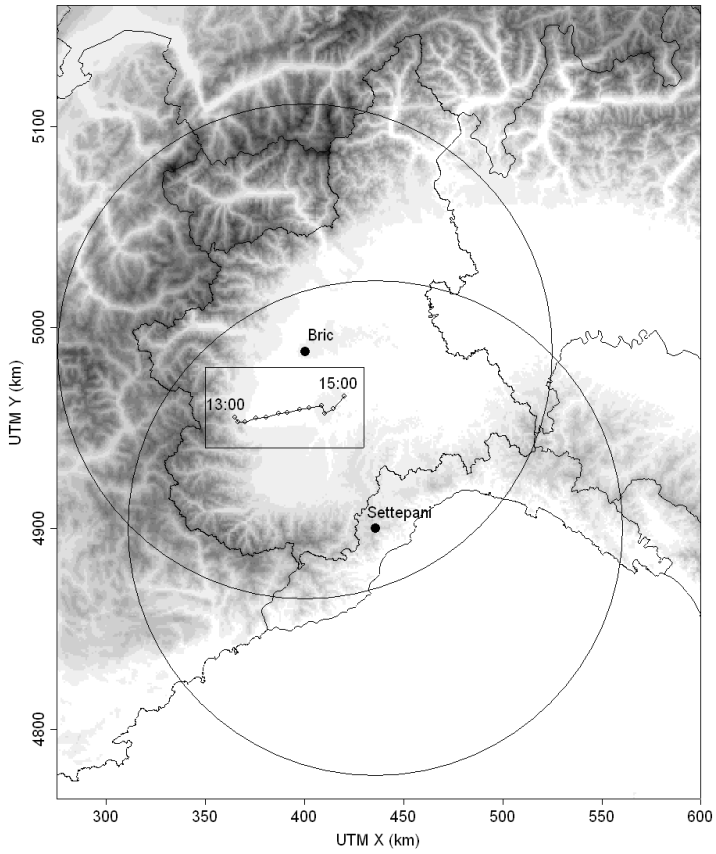
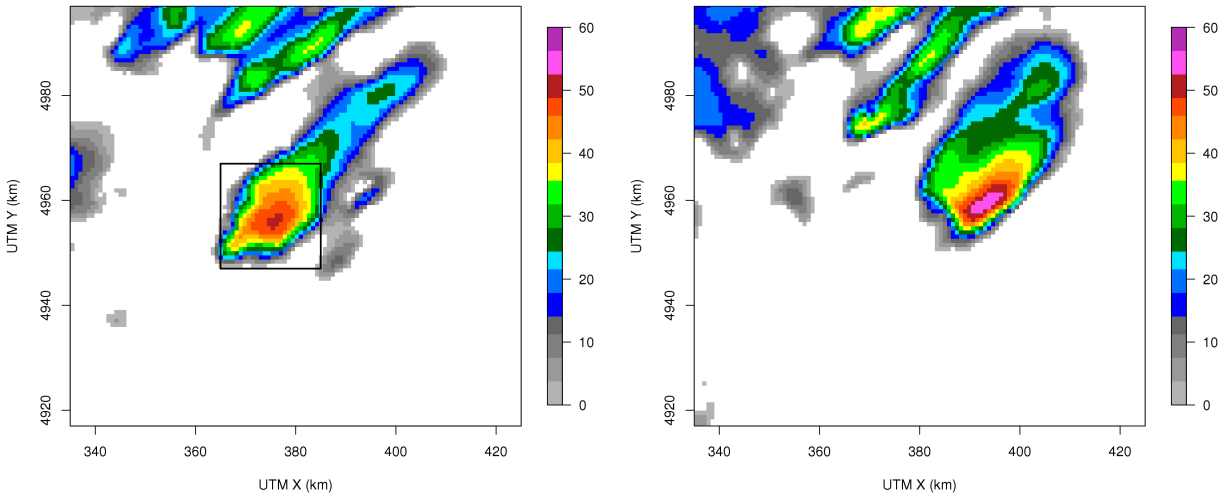


Figure 5.1: Domain of the two C-band radars and storm track between 13:00 and 15:00 UTC.

3 km above the freezing level (0°C was located at 3200 m from nearby sounding), where the ambient temperature was as low as -20°C . The super-cooled drops carried aloft by the updraft likely served as embryos for the growing hail [93].

The connection of the $Z_{\rm dr}$ column with the updraft is revealed by a dual-Doppler analysis. The additional Doppler observations provided by the Settepani radar allows a wind synthesis during the first part of the storm life, when the cross-beam angles were near 90° . The resulting wind analysis for the 2000 m height CAPPI at 13:30 UTC is shown in fig. 5.4, together with isolines representing the calculated divergence field, showing maximum values



(a) 13:30 UTC (b) 14:00 UTC

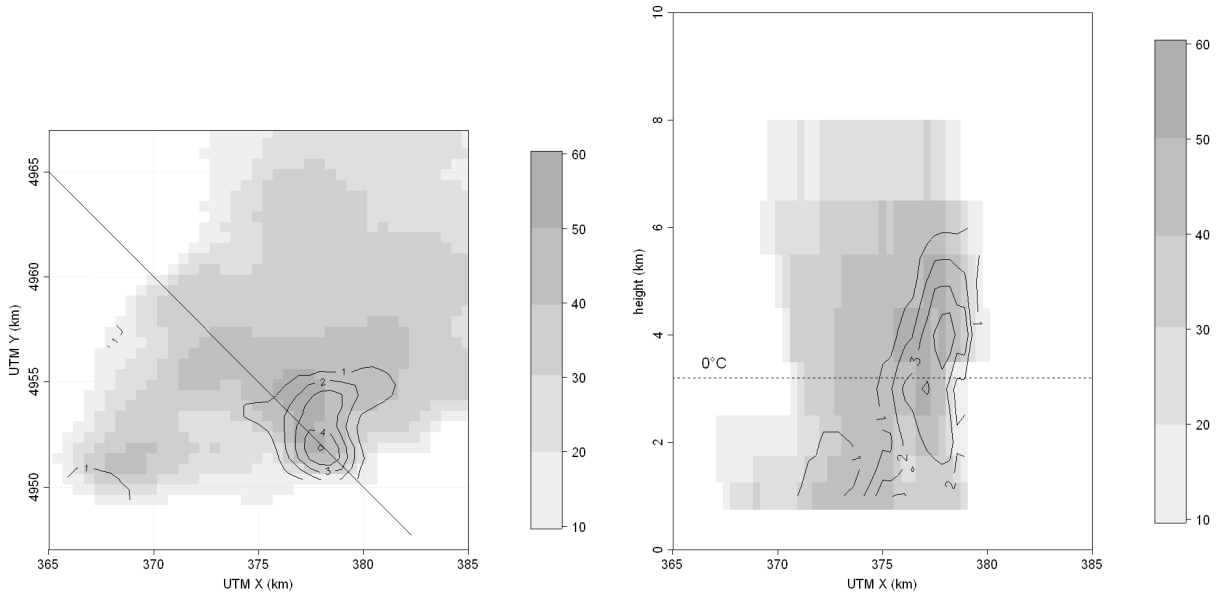
Figure 5.2: PPI of reflectivity at 0.5° elevation from Bric della Croce C-band radar, showing the marked precipitation intensity increase after 13:30 UTC. The rectangle in panel a) indicates the area of Fig. 5.3 (a).

of the order of $10^{-2}\mathrm{s}^{-1}$. It is possible to see the fairly good match with the location of the Z_{dr} column in fig. 5.3a.

5.2. Surface-based wind analysis incorporating radar features

The Z_{dr} column signature may be considered as a proxy for positive vertical velocity within the storm. Then, through a mass continuity constraint it can be related to horizontal convergence near the surface. The objective of this section is to describe a methodology to build a wind analysis from surface wind observations and divergence information from Z_{dr} columns.

Vector field interpolation may lead to different vector analyses, depending on the vector decomposition. For a wind vector $\mathbf{V}$ with zonal and meridional components (u, v), interpolation of the u- and v-components will lead to a wind analysis different from that obtained interpolating the direction and intensity. Interpolation of direction and intensity is in general



(a) CAPPI at 4000m. (b) Vertical cut through the line in panel a)

Figure 5.3: Reflectivity and Z_{dr} contours at 13:30 UTC.

avoided, since direction is a circular variable and could not be treated using standard linear methods. Most often the wind vector is analyzed from the u- and v-components separate interpolations. However for the purpose of this study and in general for many meteorological applications, a major aim is to preserve the convergence and rotational features in the wind field. According to Helmholtz's theorem, the wind vector field can be decomposed into an irrotational and a non-divergent component. For a horizontal wind vector $\mathbf{V}$ the vertical vorticity ξ and divergence δ are defined as:

$$\xi = \mathbf{k} \cdot \nabla \times \mathbf{V} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}, \qquad \delta = \nabla \cdot \mathbf{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{23}$$

The wind field can be expressed in terms of the stream function ψ and velocity potential χ:

$$\mathbf{V} = \mathbf{V}_\psi + \mathbf{V}_\chi = \nabla\chi + \mathbf{k} \times \nabla\psi \tag{24}$$

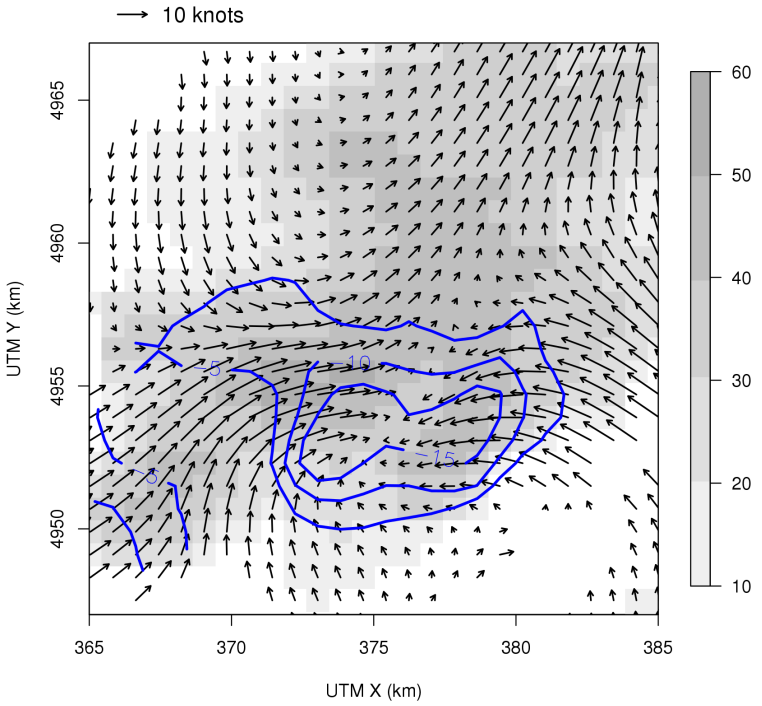


Figure 5.4: CAPPI (2000m) at 13:30 UTC: reflectivity (gray levels), divergence (blue contours, units $10^{-3}\,s^{-1}$) and wind retrieved from dual-Doppler analysis.

From eq. 23 and taking respectively the vertical component of the curl and the divergence of eq. 24 we get the Poisson equations for ξ and δ:

$$\nabla^2\psi = \xi, \qquad \nabla^2\chi = \delta \tag{25}$$

From equation 25, with proper boundary conditions over a limited domain, it is possible to reconstruct the wind field.

5.2.1. Estimation of divergence and vorticity. In addition to the standard definition in eq. 23, using Green's theorem the divergence and the vorticity can also be obtained from the line integral forms [117]:

$$\xi = \lim_{A \to 0} \frac{1}{A} \oint_{\Gamma} \mathbf{V} \cdot d\mathbf{r}, \qquad \delta = \lim_{A \to 0} \frac{1}{A} \oint_{\Gamma} \mathbf{k} \cdot \mathbf{V} \times d\mathbf{r} \tag{26}$$

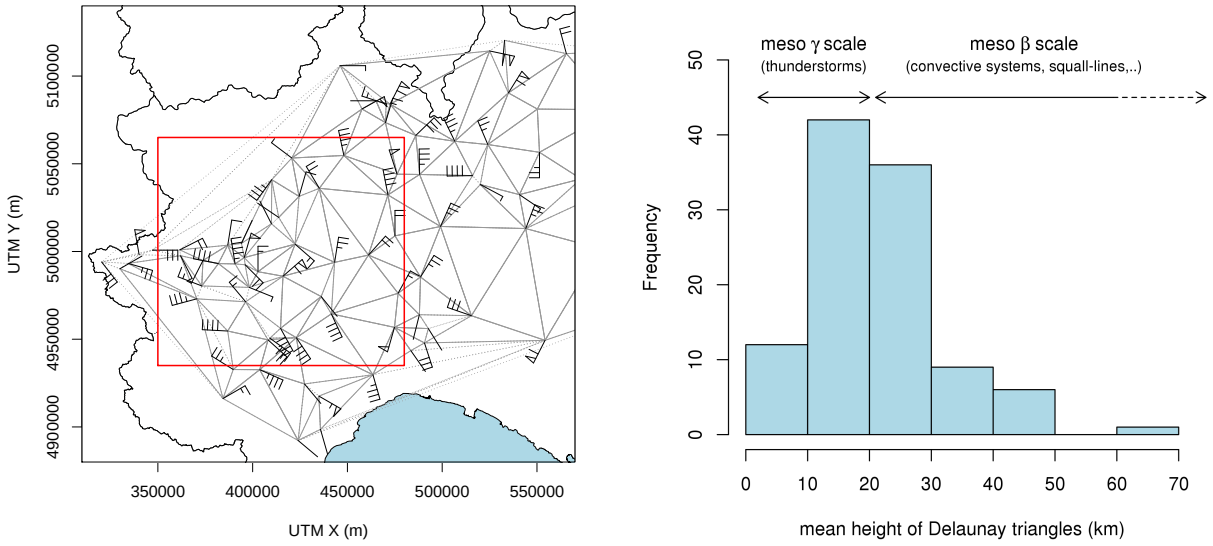
where A is the area of the surface S, $\mathbf{k}$ the unit vector perpendicular to S and $d\mathbf{r}$ the differential of the position vector along the curve Γ bounding the surface S in the horizontal plane.

Under a linearity assumption, the discretized forms of (26):

$$\hat{\xi} = \frac{\sum u\Delta x + v\Delta y}{A}, \qquad \hat{\delta} = \frac{\sum u\Delta y - v\Delta x}{A} \tag{27}$$

allow the calculation of the divergence and vorticity over a polygon region [117–119]. The error in the estimation of the divergence and vorticity using three observations goes to infinity when the three observations are aligned, while it is minimized for an equilateral triangle [120]. There are many possible ways to define the triangles in the network of surface observations. Delaunay triangulation is used in this context, because it allows a maximization of the minimum angle of all the angles of the triangles. In addition Delaunay triangle tessellation is unique (with the exception of the special case of four co-circular points), providing an objective and reproducible method for the specification of the triangle regions.

Fig. 5.5a shows an example of the Delaunay tessellation applied over the meteorological network of wind sensors in northwestern Italy (red dots in fig. 2.3). Mountain stations have been excluded in the current analysis, because the integral method is not intended for application over non flat areas, with triangles having vertices at different heights. The triangles obtained after tessellation are checked for their minimum angle and the ones having an angle lower that 20 deg are discarded, to avoid excessively noisy estimates. In fig. 5.5b,



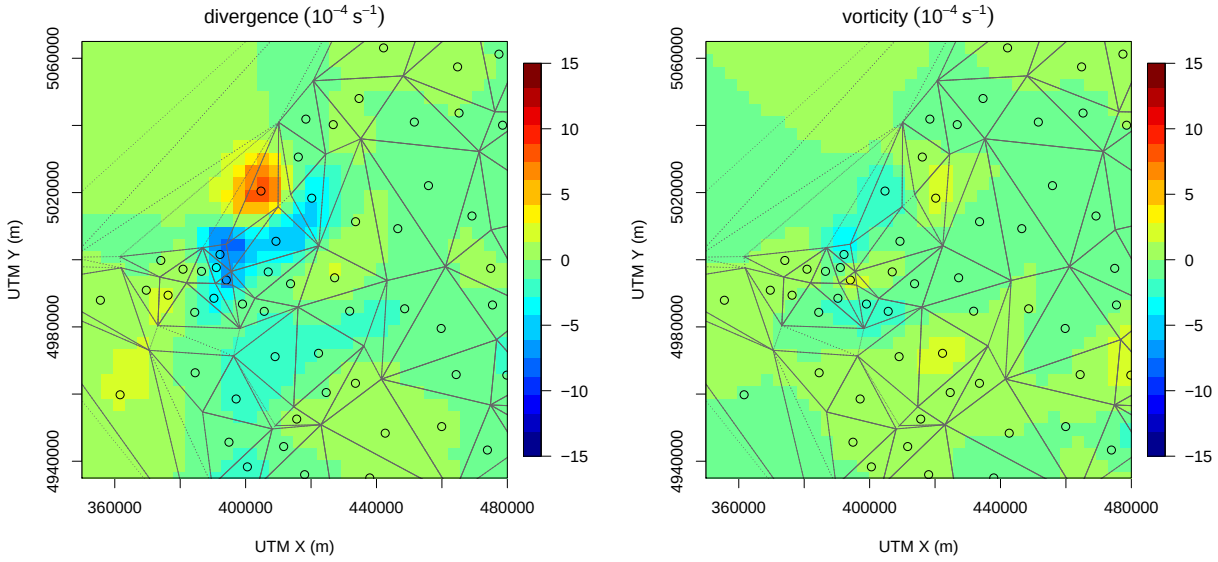
(a) Delaunay tessellation

(b) Size distribution

Figure 5.5: a) Example of Delaunay triangle tessellation over the northwestern Italy meteorological surface network (10 July 2011 at 17:30 UTC). The wind barbs intensity is magnified by a factor of 10 (one full barb = 1 knot) for display purposes. Solid gray lines indicate the triangles used in the divergence and vorticity calculation, while the dotted gray line mark the edges of the triangles discarded for having an angle lower than 20 deg. The red rectangle indicates the sub-area considered in subsequent figure 5.6. b) Distribution of the triangle dimensions.

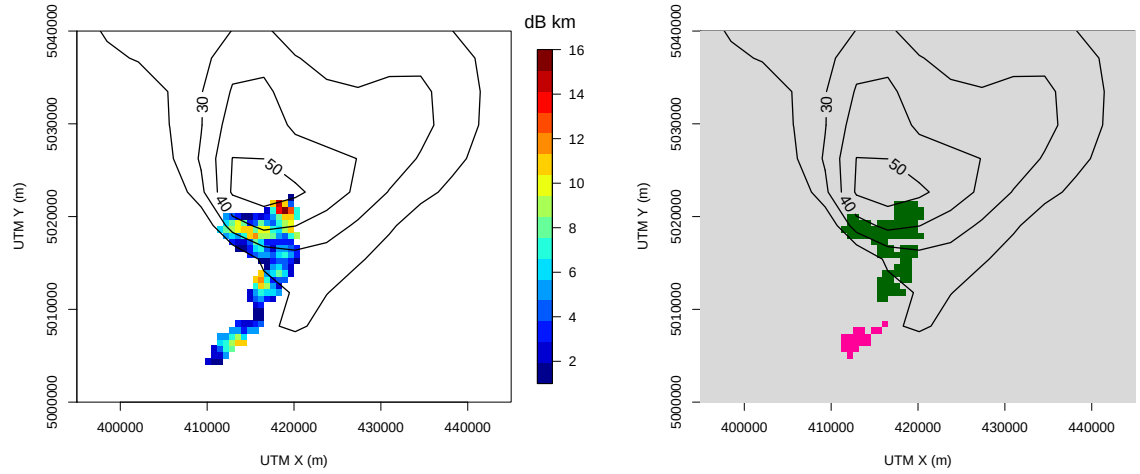
the distribution of the triangle dimensions is represented, together with the characterstic meso-γ and meso-β scales of atmospheric phenomena [121]. The network resolution appears sufficient to resolve the characterstic motions of mesoscale convective systems and squall-lines, but is on average too coarse to adequately represent the environment of individual thunderstorms. For this reason the input from radar observations is expected to be relevant in order to better resolve the motions near the surface in the environment of the storm.

The values of divergence and vorticity calculated with eq. (27) are referred to the centroid of the triangles and spatially interpolated using a IDW (Inverse Distance Weighting) method. The results obtained for the surface winds reported in fig. 5.5a are plotted in fig. 5.6



(a) divergence (b) vorticity

Figure 5.6: Divergence and vorticity fields interpolated from local estimates (open circles) using Inverse Distance Weighting, at 17:30 UTC.



(a) Z_{dr} integral (b) Z_{dr} connected regions

Figure 5.7: Z_{dr} integral (a) and column identification through connected-component labeling (b) at 17:30 UTC. The contours of Z_h (dBZ) are overplotted in both panels for reference.

5.2.2. Inclusion of Z_{dr} columns information. In order to identify the columns of Z_{dr} first a vertical integration ($Z_{\mathrm{dr}} \cdot dz$) is performed in the radar volume considering values

exceeding a given threshold (2 dB) above the freezing level and imposing a vertical contiguity constraint. The resulting $Z_{\rm dr}$ integral product, with units db km, is subsequently analyzed with a connected-component labeling [122] algorithm to identify the cores with maximum values exceeding 6 db km and minimum of at least 3 db km (fig. 5.7). Attributes of the identified connected regions ($Z_{\rm dr}$ columns) are calculated, including area, centroid coordinates, mean and max value.

Since a $Z_{\rm dr}$ column has in general a much smaller area than the Delaunay triangles over surface observations, a resampling over a larger area is necessary before the spatial interpolation is performed, to avoid mixing terms with very different spatial representativeness. The column is then assumed to correspond to a Gaussian spatial distribution of divergence with center corresponding to a fixed value of 10^{-2} s^{-1}, as per dual-Doppler observations (fig. 5.4), corresponding to about a 20 m s^{-1} increase in vertical velocity over a 2000 m deep vertical layer (under mass continuity hypothesis). The width of the Gaussian distribution (σ) is set equal to the radius of the equivalent circular area of the $Z_{\rm dr}$ column. Finally, the Gaussian distribution of divergence is integrated over an area equal to the average area of the five closest Delaunay triangles to yield a value of similar spatial representativeness.

Fig. 5.8 shows the same type of divergence IDW analysis as in fig. 5.6a, but in this case including the additional divergence information from $Z_{\rm dr}$ columns (marked with white crosses). The $Z_{\rm dr}$ column centroids (white crosses) lie in a region where the divergence was already estimated to have negative sign, using surface observations alone (fig. 5.6a). This is a confirmation of the connection of the $Z_{\rm dr}$ columns with the updraft. The inclusion of the $Z_{\rm dr}$ column information has the effect of enhancing the near surface convergence ahead of the storm.

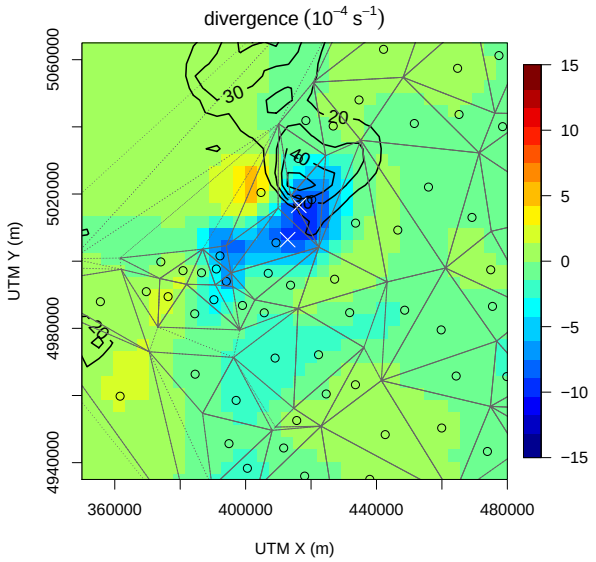


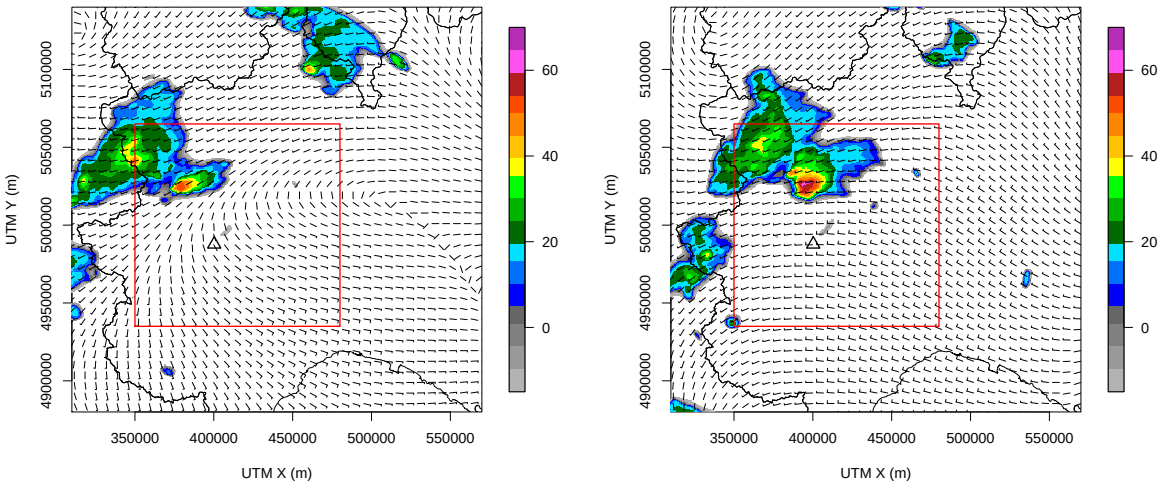
Figure 5.8: Divergence field interpolated from local estimates (open circles) and $Z_{\rm dr}$ column divergence information using Inverse Distance Weighting. The white crosses mark the centroids of the $Z_{\rm dr}$ column, while the black solid contours represent the Z_h field.

5.2.3. Wind field evaluation technique. When the global values of divergence and vorticity are available, the wind field can be retrieved within an additive constant. However, when the spatial domain is limited, boundary conditions need to be specified. Taking the partial derivatives of eq. 23, we obtain a new pair of Poisson equations:

$$\frac{\partial \delta}{\partial x} - \frac{\partial \xi}{\partial y} = \nabla^2 u, \qquad \frac{\partial \xi}{\partial x} + \frac{\partial \delta}{\partial y} = \nabla^2 v \tag{28}$$

These equations are solved using a SOR (Successive Over Relaxation) method, with boundary conditions specified by an auxiliary wind field estimate. In the following example we used the WRF analysis resulting from the 3D-Var assimilation of surface observations (Chapter 2).

5.2.4. Test on a real case. The outlined procedure is illustrated with an application to a real case. On 10 July 2011 a convective storm developed on the Western Alps and shortly after moved eastwards (fig. 5.9), gaining energy from the warm and humid air subsiding over the plains. The storm produced large hail and one tornado. The steps in the analysis procedure are summarized as follows:



(a) 16:20 UTC

(b) 16:50 UTC

(c) 17:20 UTC

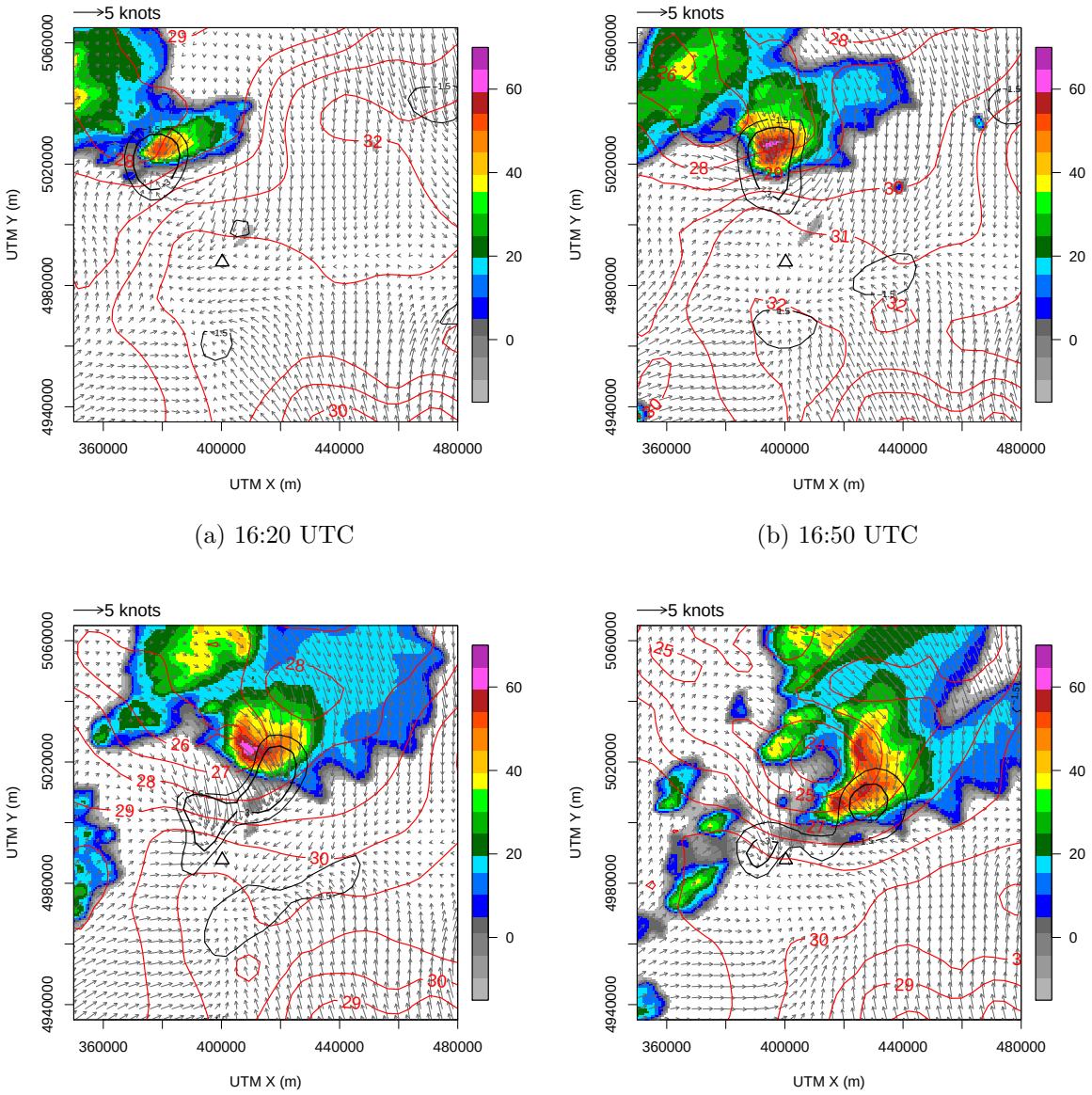
(d) 17:50 UTC

Figure 5.9: DARTS (Dynamic Adaptive Radar Tracking of Storms) motion vectors [5] and reflectivity between 16:20 and 17:50 UTC. The red rectangles define the sub-domain in the following fig. 5.10.

- A first-guess from WRF analysis is obtained upon 3D-Var assimilation of surface observations (winds and temperatures);
- Divergence and vorticity are estimated from surface observations, based on integral definitions (eq. 27) over Delaunay triangles;
- The location and extent of the Z_{dr} columns provide additional bogus observations of divergence;
- Divergence and vorticity observations are interpolated on a regular grid by means of IDW;
- Finally, a successive over-relaxation iterative method is applied to retrieve the u- and v-component of the wind field from divergence and vorticity analyses, using a first guess and boundary conditions provided by the WRF analysis.

Fig. 5.9 shows the reflectivity maps between 16:20 and 17:50 UTC, with overplotted motion vectors calculated using the frequency-domain estimation technique implemented in the CASA radars [5]. The initial eastward displacement is followed by a right-flank development of the storm determining a southeastward motion and south-west to north-east alignment of the radar echo pattern.

The corresponding surface-based wind analyses including Z_{dr} information are shown in fig. 5.10. In addition to the reflectivity and wind vectors, the divergence contours (black solid lines) and surface temperature (red isolines) are also represented. Temperature observations are interpolated with a kriging technique, using the *autoKrige* function (package *automap* [123]) within the R Project for Statistical Computing package [124].



(a) 16:20 UTC

(b) 16:50 UTC

(c) 17:20 UTC

(d) 17:50 UTC

Figure 5.10: Surface analysis of temperature (red contours) and winds with observed reflectivity between 16:20 and 17:50 UTC. The black contours represent divergence levels of -1.5 and -3.0 $10^{-4}\,s^{-1}$, the red contours the surface kriging-interpolated temperature. The radar location is marked with a triangle.

Auto-propagation (the process in which a storm can regenerate itself) by cold-pools is indeed a major mechanism for storm propagation [112], in addition to the advection by mid-tropospheric winds. The analyses in fig. 5.10 well represents this mechanism. The dominant

feature in the interpolated temperature field is represented by the minimum behind the cell core. The temperature minimum is already visible at 16.20 UTC (28°C) and subsequently deepens down to 24°C at 17:50 UTC. During the intense phase of the storm life-cycle, the retrieved regions of maximum negative divergence (black isolines) are always roughly co-located with the maximum gradient of the surface temperature. This is a clear indication of the vertical motion triggered by the gust-front. The movement of the advancing gust-front is close to the speed a the density current, which is typically of the order of 10 $\mathrm{m\,s^{-1}}$ [125]. A quantitative estimation of the gust-front North-West to South-East displacement is obtained from the analysis of the temperature gradient at successive times. The overall displacement of about 45 km corresponds to a fairly consistent speed of $\sim$8 $\mathrm{m\,s^{-1}}$. The apparent lagged displacement and modification of the reflectivity field is also apparently correlated with the estimated divergence field, providing further evidence of the overall reliability of the retrieved wind field.

It is finally worth noting how the cold-pool triggered auto-propagation has the important effect to modify the initial storm trajectory (West to East, fig. 5.9), making it difficult for reflectivity-based extrapolation methods to ascertain with reasonable accuracy the position of the storm in the next 20-30 minutes.

CHAPTER 6

Assimilation of dual-polarization information in VDRAS (Variational Doppler Radar Analysis System)

Four-dimensional data assimilation techniques have been introduced during the last couple of decades in global and regional models [126–128] to provide a temporal extension respect to previous three-dimensional variational systems, and as an alternative to other popular methods based on statistical estimation theory like the Ensemble Kalman filtering. In large-scale models the four-dimensional assimilation process is mainly intended as an optimal filter, but for the convective scale it becomes also relevant to retrieve the unobserved variables. In fact, while for the meso-synoptic scale the conventional meteorological observations (surface measurements and upper-air soundings) provide adequate resolution to represent the model variables, for the convective scale only weather radars are able to collect observations with enough spatial and temporal resolution, although only for a reduced set of variables. In addition, the variables observed by the weather radar (e.g. reflectivity and radial winds) are not among the model variables, but are generally related to the water content and winds through appropriate observation operators.

The Variational Doppler Radar Analysis System (VDRAS) is an advanced data assimilation system specifically designed for ingesting Doppler weather radar observations at the convective scale [7]. The system has been installed at many sites around the world and is typically running using long-range operational S-band or C-band radar networks. The core four-dimensional data assimilation scheme is based on a cloud-scale model and considers a 12-15 minutes time window for the radar assimilation, with 1-3 km spatial resolution.

The Dallas Fort Worth (DFW) Urban Demonstration Network project is centered on the deployment of a network of several dual-polarization, X-band radars to demonstrate improved hazardous weather forecasts, warnings and response in a densely populated urban environment (pop. 6.3 million in 2010). Specifically, low level wind analysis and forecast ranging from 10 minutes to 3 hours are among the main research areas of the project. The scanning strategy of the radars is inherited from the CASA project distributed collaborative adaptive sensing concept [106] and is intended to sample with high time resolution the lower atmosphere (1-3 km above ground level). During standard operation, 4 or more full PPI or sector scans at elevations ranging between 1 and 7 deg and range resolution of 50-100 m are performed within 1 minute.

In this chapter the feasibility of running VDRAS at high spatial resolution (2 km) with rapid update (5 minutes) is studied, exploiting the frequent low-level sampling of the atmosphere available within the DFW network of X-band radars.

6.1. VDRAS 4D-Var assimilation

The central process of VDRAS is the 4D-Var radar data assimilation, which includes a cloud-scale numerical model, the adjoint of the numerical model, a cost function and a minimization algorithm [7, 129]. The assimilation scheme, through the iterative minimization of a cost function, fits the model to the observations on the three-dimensional spatial domain and over a specified time window. Similarly to eq. (3) in Chapter 2 a cost function, measuring the distance between the model variables and the observations, is defined as:

$$J = \sum_{\sigma,\tau} [\eta_v (v_r - v_r^{obs})^2 + \eta_q (q_r - q_r^{obs})^2] + J_b + J_p \tag{29}$$

where σ and τ represent the spatial and temporal domains, the variable v_r is the radial velocity computed from the model velocity components and v_r^{obs} is the observed radar Doppler velocity; q_r is the rainwater mixing ratio from the model and q_r^{obs} is the rainwater mixing ratio estimated from radar observations. The quantities η_v and η_q are constant weighting coefficients. This formulation assumes that the observation errors of the radar fields are uncorrelated in space and time. The terms J_b and J_p represent respectively the background and the penalty term. The background term provides a measure of the distance between the analysis and a prior estimate, while the penalty term ensures a proper spatial and temporal smoothness. It is important to emphasize that, as opposed to eq. (3) in Chapter 2, the cost function in 4D-Var includes the summation over the temporal domain. The minimization of the cost function (eq. 29) is obtained using the limited-memory quasi-Newton conjugate method [130]. The search direction is calculated using the information of the cost function and its gradient. While the cost function is computed integrating forward in time the non-linear model, the gradient is derived from the integration backward in time of the adjoint model (the transpose of a tangent linear version of the non-linear cloud-scale model).

The specific implementation of the VDRAS system used in this work (appropriate also for real-time applications) is described in detail in the Sun and Crook paper [129].

6.2. Radar data pre-processing

The rainwater mixing ratio is conventionally derived in VDRAS from reflectivity observations using a power-law relation obtained assuming a MarshalPalmer raindrop size distribution:

$$Z = 43.1 + 17.5 \log_{10}(\rho q_r) \tag{30}$$

where Z is the reflectivity expressed in dBZ, ρ is the air density and q_r in in units of g kg^{-1}. However, for X-band systems path attenuation greatly affects the reliability of reflectivity-based estimates in heavy precipitation. Dual-polarization measurements allow to correct for path attenuation and to estimate the rain rate and the rainwater content with higher accuracy [25].

In this work a blended algorithm is adopted that combines the available dual-polarization observations (namely reflectivity, differential reflectivity and the specific differential phase shift) using different relations, providing a rainwater estimate less sensitive to DSD (Drop Size Distribution) variations and mostly un-affected by attenuation. The basis to apply different relations is provided by a preliminary hydrometeor classification [6] described in Chapter 4, which drives the choice of the most proper algorithm. The relations used are for the liquid water content (LWC), i.e. the rainwater content mixing ratio scaled by the air density: LWC $= \rho q_r$: the dual-polarization relations include functions of Z_h and $Z_{\rm dr}$, $K_{\rm dp}$ and $Z_{\rm dr}$, and $K_{\rm dp}$ only, with coefficients for the S-band as follows [25]:

$$
\begin{aligned}
\mathrm{LWC}(Z_h, Z_{dr}) &= 0.7\ 10^{-3} Z_h^{0.89} \xi_{dr}^{-4.16} \\
\mathrm{LWC}(K_{dp}, Z_{dr}) &= 6.05 K_{dp}^{0.88} \xi_{dr}^{-2.52} \\
\mathrm{LWC}(K_{dp}) &= 1.65 K_{dp}^{0.77}
\end{aligned}
\tag{31}
$$

where Z_h is expressed in mm^6 m^{-3}, $\xi_{dr} = 10^{0.1 Z_{dr}}$ is the differential reflectivity in linear units. For S-band, the LWC(K_{dp}) is always used when the hydrometeor classification indicates *heavy rain* or *rain + hail* mixture, while for the other liquid hydrometeors either LWC(Z_h, Z_{dr}) or LWC(Z_h) as in eq. (30) are used, according to predefined thresholds on the radar Z_h and $Z_{\rm dr}$ values [131]. AT X-band the attenuation and differential attenuation may severely affect the radar observations. In particular the correction for the differential

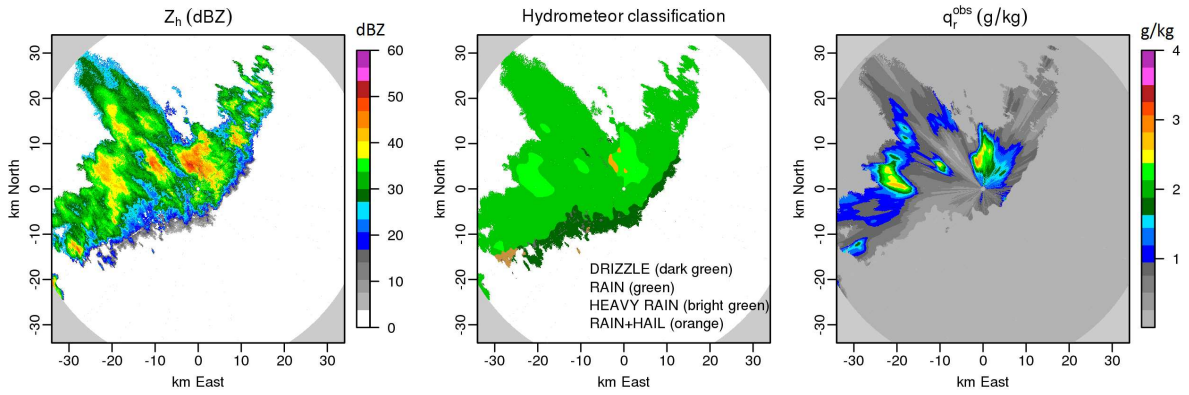


Figure 6.1: PPI from Midlothian radar at 20:39UTC. Left: reflectivity; center: hydrometeor classification; right: rainwater mixing ratio estimated from dual-polarization observations.

attenuation may introduce a relevant error on Z_{dr}. Considering the high sensitivity of the LWC(Z_h, Z_{dr}) and LWC(K_{dp}, Z_{dr}) estimators to small Z_{dr} biases, it is then preferred to avoid the use of such relations. So, for the small X-band systems, the LWC(K_{dp}) estimator:

$$\mathrm{LWC}(K_{dp}) = 0.64 K_{dp}^{0.77} \tag{32}$$

is used whenever the hydrometeor classification output reports *rain*, *heavy rain* or *rain + hail* mixture. In fact, the differential phase shift (K_{dp}) measurements are immune to both radar calibration and attenuation due to propagation, and are also mainly insensitive to partial beam blocking [25]. In the other cases the simpler LWC(Z_h) is adopted.

For assimilation in VDRAS, the radar observations of v_r^{obs} and the estimates of q_r^{obs} in the polar domain are first interpolated to Cartesian PPIs with 500 m-resolution. Figure 1 shows an example of a reflectivity PPI at 2.0deg elevation from the Midlothian X-band radar, with hydrometeor classification and the resulting q_r^{obs} estimation.

6.3. Model set up

Data assimilation experiments have been initially conducted for a hailstorm event occurred on 12 May 2014. For this case the data from the S-band NEXRAD KFWS radar have been considered, in addition to the X-band radars located in Arlington and Midlothian. In order to provide an initial condition (background) for the cost function minimization (eq. 29), a mesoscale analysis is preliminarily performed to start the model simulation (cold start), based on surface observations (METAR reports) and vertical sounding from a preliminary WRF model simulation on a larger domain. The experiment started at 19:25UTC, and then cycled every 5 minutes using the previous forecast as background. The cycling procedure is actually matched with the NEXRAD volume update frequency, which was slightly less than 5 minutes, ensuring the availability of a large scale 3-dimensional coverage for the analysis. Within the 5-min window, in addition to the NEXRAD volume scan, about 5 to 10 PPIs at the same low elevation angle (2deg) are available from each of the X-band radars. Adaptive sector scans at higher elevations were not considered in these first experiments.

In previous VDRAS applications assimilation windows of about 15-20 minutes have been used, including several radar volume scans. The reference time for all the observations during the whole volume scan was the beginning of the first elevation scan, disregarding the time differences between successive elevations. In the current set up, however, the short assimilation window required to consider the actual observation time of the individual PPIs within the NEXRAD volume scan, in order to deal more consistently with the frequent low-level scans of the X-band systems. Since the VDRAS version with warm-rain only microphysics is considered in this study, the assimilation of the radar observations is limited to data below 5 km height.

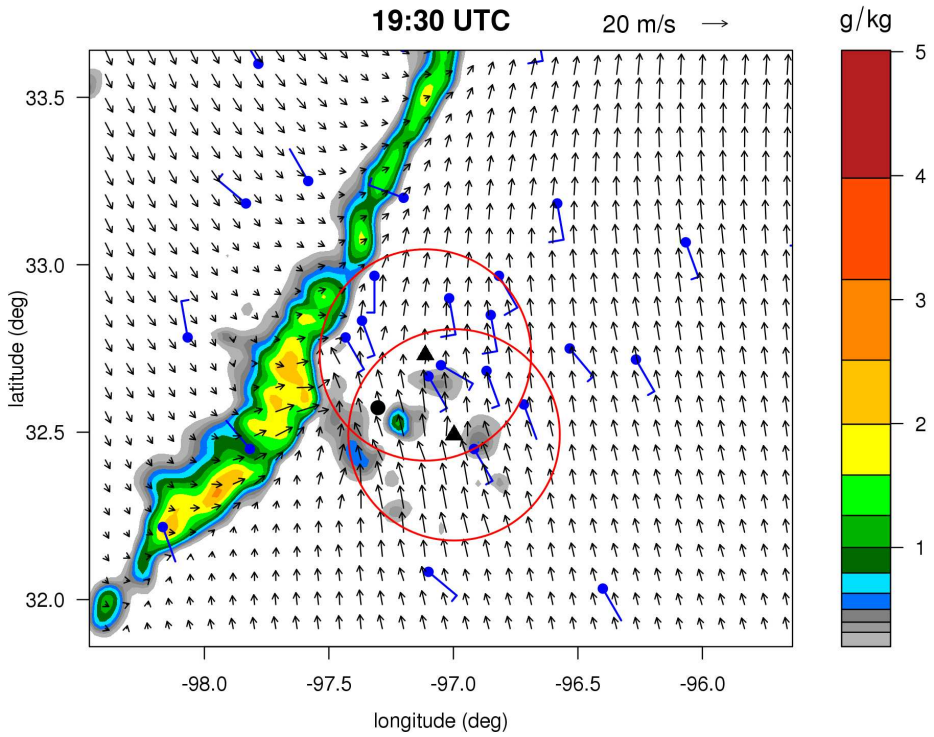
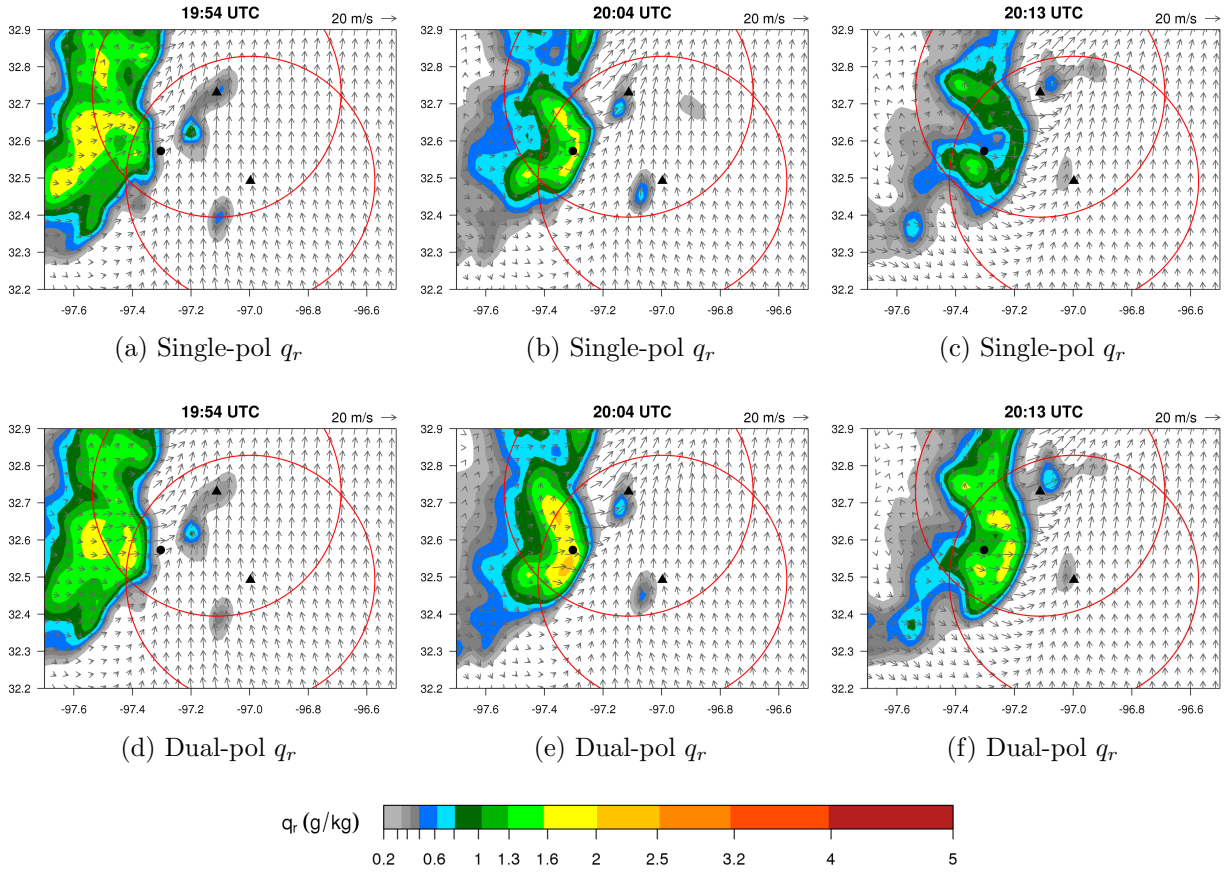


Figure 6.2: VDRAS analysis of q_r (colors) and winds at 19:30 UTC from the first cycle on the full model domain. The blue wind barbs represent the METAR surface observations, the red circles indicate the 40 km-range domain of the two X-band radar (black triangles). The small filled black circle indicates the position of the NEXRAD KFWS radar.

The model domain is 133x100x37 (nx,ny,nz) grid points, with horizontal resolution of 2 km and vertical resolution of 400 m, and the integration time step is 4 s. The runtime for the analysis is about 20 minutes on 8 processors. In fig. 6.2 the VDRAS analysis from the first (cold start) 4D-Var cycle is shown on the full model domain. At this stime, a squall line can be seen approaching from west-northwest the DFW region covered by the X-band radars (red circles). The wind analysis during the following period can then rely on multiple Doppler observations as well as on several surface METAR measurement.



(a) Single-pol q_r (b) Single-pol q_r (c) Single-pol q_r

(d) Dual-pol q_r (e) Dual-pol q_r (f) Dual-pol q_r

Figure 6.3: Analysis of rainwater mixing ratio (color) and winds for the lowest model level (200m). Panels a) to c) shows three successive analysis using the single-polarization q_r estimation algorithm (eq. 30), while panels d) to e) present the same analysis obtained with assimilation of the dual-polarization q_r estimates (eqs. 31 and 32).

6.4. Wind analysis

6.4.1. Dual-polarization vs. single-polarization rainwater estimation. One of the notable advantages of using dual-polarization relations for the estimation of the rainwater (eqs. 31 and 32) as opposed to the single-polarization algorithm (eq. 30) is illustrated in fig. 6.3. Panels a) to c) (single-polarization q_r estimation) show a suspect decrease of the rainwater intensity when the squall line passes over the KFWS radar and enter the domain of the smaller X-band systems. On the other hand, the dual-polarization assimilation experiments show an apparently more consistent evolution of the storm. The progressive intensity

weakening in panels a) to c) may not be suspect in other circumstances. What makes it suspicious in this case is the fact that the weakening coincides with the storm passage above one radar (KFWS) and when the X-band observations start to contribute significantly to the analysis in this region. Further investigation allowed to identify two distinct issues:

- the reflectivity was not completely recovered from path attenuation on the X-band measurements, using the rain profiling algorithm described in Chapter 4, section 4.2.1;
- up to ∼2 dB attenuation was estimated to affect the S-band observations, due to radome wetting.

While the former issue is not completely surprising, due to the known difficulties to correct large attenuation affecting radar operating at high frequencies, the second problem is somehow unexpected since the effect of radome wetting at S-band is generally neglected. However, Merceret and Ward [132] reported a similar attenuation of approximately 2 dB at S-band for high rain rates in the order of 50-100 $\mathrm{mm\,h^{-1}}$. This would produce an underestimation of the rainwater mixing ratio of about 25%. At X-band the effects of radome attenuation are more important [61], making the $R(K_{dp})$ estimator the first choice in the majority of situations. It is concluded that the use of dual-polarization algorithms, in particular $R(K_{dp})$, is relevant for its insensitivity to the wet radome attenuation (in addition to the previously mentioned advantages), also for S-band when heavy rain reaches the radar site.

It is also intersting to look at the evolution of the cost function (eq. 29) during the minimization process. Fig. 6.4 shows the velocity (left panel) and rainwater (right panel) components of the cost function for the Midlothian radar analysis at 20:18 UTC. This is a representative example of the experiment running with the assimilation of dual-polarization

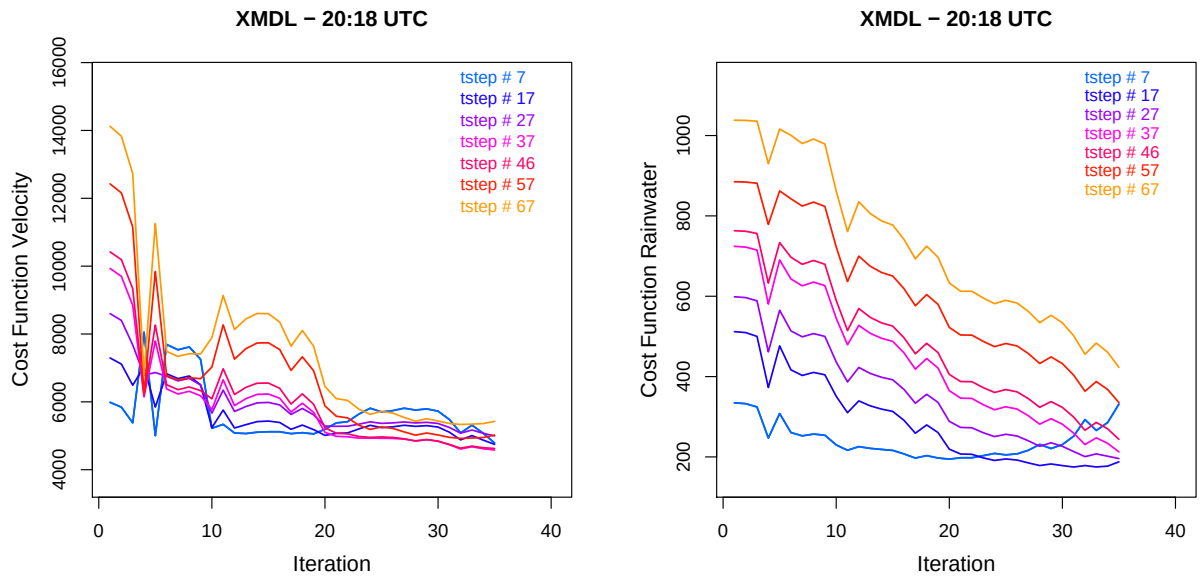
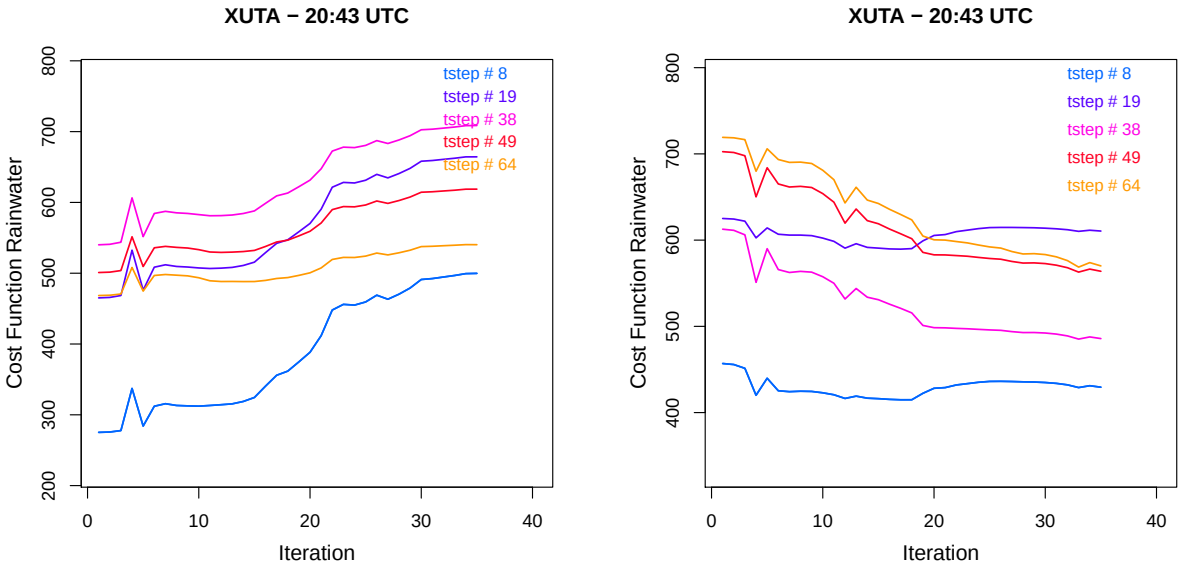


Figure 6.4: Cost function at 20:18 UTC for the Midlothian radar. The velocity and rainwater components as defined in eq. (29) are respectively plotted in the left and right panel.

q_r estimates, including seven low-level scans available within the five minutes assimilation window. The colored lines represent the cost function for the observations collected during each scan, whose time is indicated in units of time steps (4 seconds) relative to the beginning of the assimilation window. Both the velocity and rainwater components of the cost function show a decrease of at least a factor of two during 35 iterations. Similar performances in terms of convergence are obtained for most of the event, with an exception of about 10-15 minutes during tho most intense phase of the event, with the X- band observations suffering extreme attenuation and complicating the minimization process. The evolution of the cost function rainwater component for the Arlington radar during the strongest storm phase in the overlapping radar region (around 20:43 UTC, as depicted in fig. 6.7) is especially informative about the impact of the dual-polarization vs. the single-polarization rainwater estimation. The left panel in fig. 6.5 presents the cost function evolution for the five scans within the assimilation window, when the single-polarization algorithm (eq. 30) is used in



(a) Single-pol q_r

(b) Dual-pol q_r

Figure 6.5: Evolution of the cost function rainwater component for the Arlington radar at 20:43 UTC. The left and right panel respectively show the results obtained after application of the single-polarization and dual-polarization rainwater estimation algorithms.

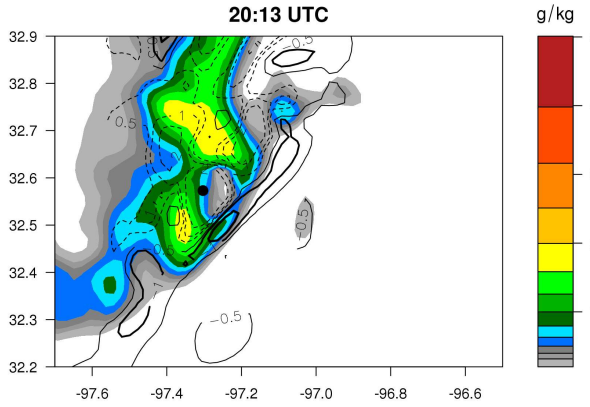
the assimilation. In this case it is evident that it was not possible to reach convergence, as the cost function shows an increase during the iterative process instead of the expected decrease. This is clearly a particularly difficult case for the X-band radars (the cost function for the NEXRAD radar on the other hand shows the anticipated decreasing trend), for which the differences deriving from the application of different rainwater estimators are more pronounced. The use of the blended dual-polarization rainwater algorithm is in fact associated with a better minimization of the cost function (right panel in fig. 6.5).

6.4.2. Single radar (S-band) vs. multiple radars (S-band + X-band) assimilation. The availability of multiple Doppler radial velocity observations is expected to allow a detailed retrieval of the 3-dimensional wind field. In the case considered three radars are available, the conventional S-band NEXRAD radar (KFWS) and two newly deployed

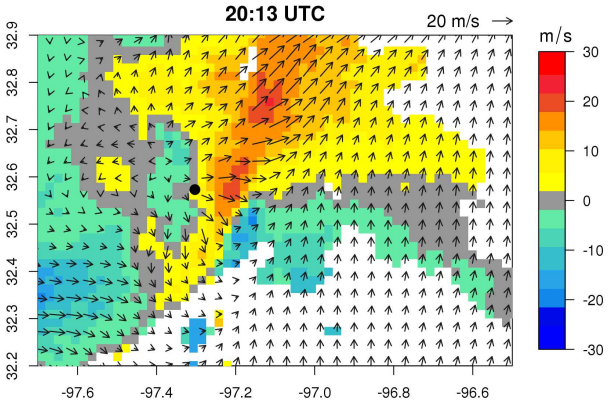
small X-band systems (XUTA and XMDL). The major benefit for the wind retrieval is expected during the period when the storm entered the central area of the model domain with relevant radar overlapping. It is therefore useful to compare the analysis obtained using only the long range S-band radar and the analysis based on the assimilation of all three radars.

As an example, fig. 6.6 shows the rainwater and wind retrievals obtained when only the NEXRAD observations are assimilated (panels a and b) and when the two X-band radars are additionally included in the assimilation process (panels c to f). In order to facilitate the interpetation of the resulting wind fields, the original radial Doppler wind observations are displayed (color) in panels b), and d) to f), after horizontal interpolation on the model grid. In the single radar analysis (panels a and b) the retrieved wind field presents a divergence just east of KFWS resulting form the model "interpretation" of the single Doppler velocities. In addition, as a consequence of the model physical balances a depression in the rainwater field is generated, which has no correspondence with the actual observations. When the Doppler observations from XUTA and XMDL are assimilated together with the NEXRAD data, the wind field appears noticeably different, in particular in the southern portion of the model domain. A band of intense convergence near the surface is identified from this updated analysis, crossing from south to north the ideal triangle formed by the three radars on the map. In this region the outflow generated by the evaporative cooling of the precipitation within the approaching squall line from the west meets the warmer southerly flow on the eastern portion of the domain. This convergence leads to relevant upward motions, which continued for about 30 minutes until a hailstorm eventually formed (fig. 6.7).

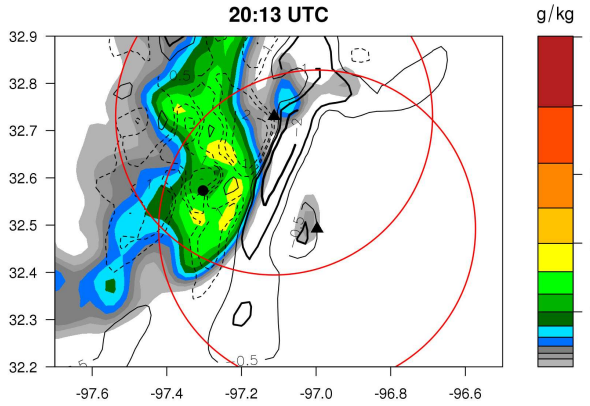
This analysis and the following ones are consistent with the subsequent development of the storm. In particular the wind analyses at 20:28 and 20:33, when the intensity of the



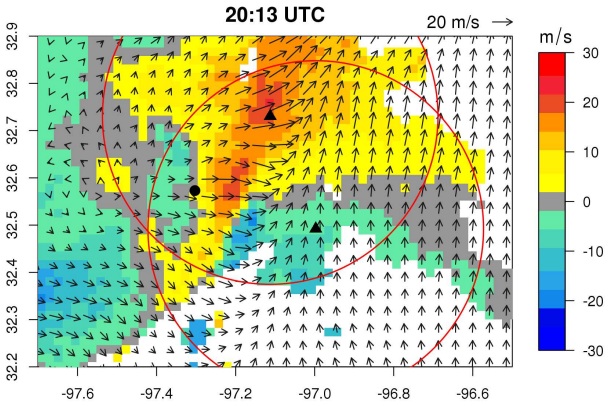
(a) q_r + δ contours (NEXRAD)



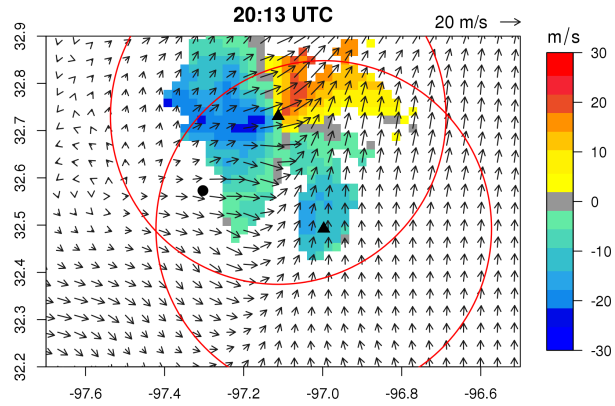
(b) Doppler NEXRAD+analysis (NEXRAD)



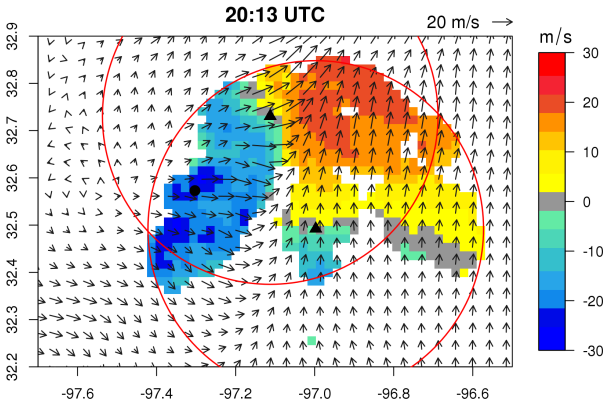
(c) q_r + δ contours (NEXRAD+XDFW)



(d) Doppler NEXRAD+analysis (NEXRAD+XDFW)



(e) Doppler XUTA+analysis (NEXRAD+XDFW)



(f) Doppler XMDL+analysis (NEXRAD+XDFW)

Figure 6.6: Analysis at 20:13UTC on a smaller portion of the model domain, for the lowest vertical level (200 m height). In brackets the observations used in the analysis are indicated. The Doppler observations are from the lowest PPI (0.5° for KFWS, 2.0° for XUTA, and 1.8° for XMDL), and interpolated on the model grid. The solid (dashed) contours in panels a) and c) represent negative (positive) divergence.

main convective core within the squall line was actually decreasing, show an updraft region ahead of the squall line with vertical velocities in excess of 16 $\mathrm{m\,s^{-1}}$.

At 20:43 UTC (fig. 6.7) the estimated rainwater mixing ratio is characterized by a well-defined convective core downwind of the region previously distinguished by the intense updraft. There were several hail reports around this time and location.

For this case the hydrometeor classification algorithm described in Chapter 4 is independently applied on the dual-polarization measurements collected by the three radars. The algorithm running on the one-minute scans of the X-band systems initially identified hail within a heavy rain region around 20:37 UTC (XUTA and XMDL) and during the subsequent 18 minutes (until 20:55 UTC). The achieved results are consistent with the lower sampling frequency S-band NEXRAD radar (hail identified in the scans at 20:38, 20:43 and 20:48 UTC), and with the hail reports obtained from the media. This analysis led to estimate the area interested by hail at the surface level. In fig. 6.7 the perimeter of the estimated hail swath is drawn as a solid contour, overplotted on the rainwater field estimated from the NEXRAD dual-polarization observations at 20:43 UTC.

6.5. VDRAS forecast

VDRAS was primarily developed as an analysis system and as such is being used for research purposes and also as an operational tool at several meteorological offices. However, the same cloud model used for the analysis retrieval in the four-dimensional assimilation scheme can be integrated further in time to provide a forecast. For the experiments described in the previous section 6.4.2 a forecast up to +58 minutes lead time was run, with a specific focus on evaluating the contribution to the forecast performance deriving from the use of the X-band radar observations in the assimilation.

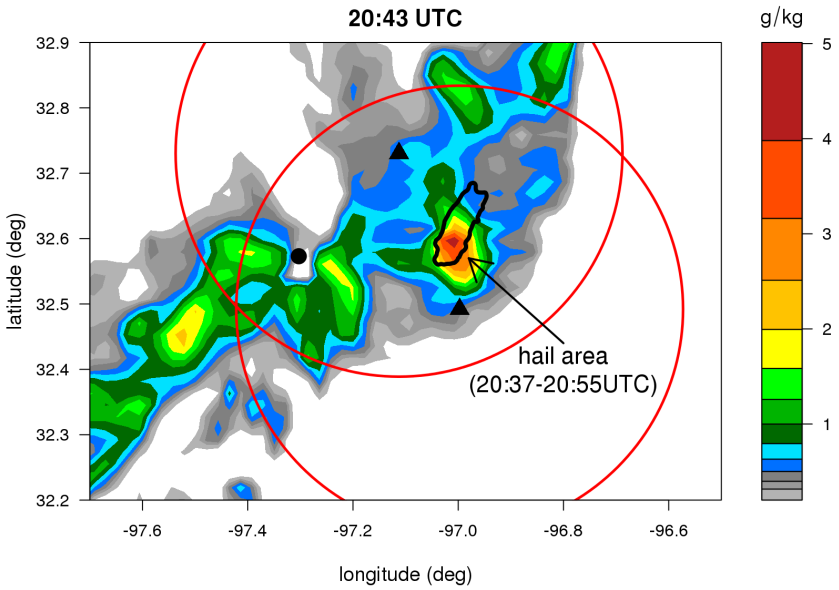


Figure 6.7: Estimated rainwater mixing ratio from KFWS dual-polarization measurements at 20:43 UTC, with overplotted the hail swath perimiter as inferred from hydrometeor classification [6].

As an example, fig. 6.8 shows the forecast comparison for the analysis at 20:13 UTC. The forecast starting from the analysis elaborated with the NEXRAD radar only is displayed in the left column, while the forecast obtained from the analysis including both the NEXRAD and the two X-band radars (NEXRAD+XDFW) is reported in the middle column. In the right column the analyses at the corresponding lead times are displayed for comparison. One notable difference between the two experiments can be seen early during the forecast in the rainwater fields. As already noted discussing fig. 6.6, the divergent wind field close to the KFWS radar, in the case of the single radar assimilation, produces an anomalous weakening of the rainwater in the analysis. Coherently with this context, the forecast shows a gradual splitting of the squall line, with the southern sector intensifying during the first ∼30 minutes and the central portion in the overlapping region progressively dissipating. On the other hand, the NEXRAD+XDFW experiment shows a remarkable ability to qualitatively simulate the evolution in the northern part of the domain during the first ∼30 minutes.

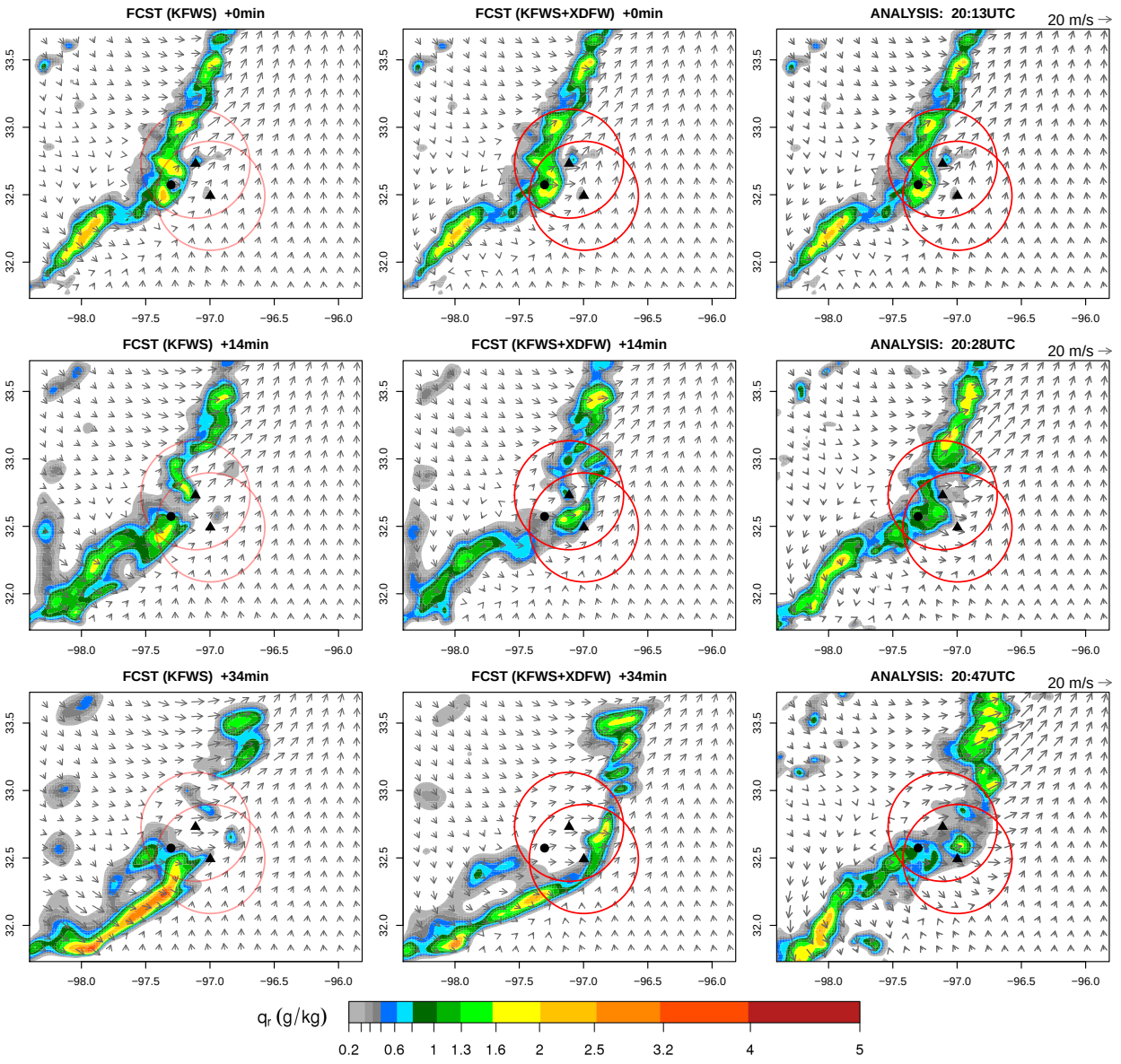
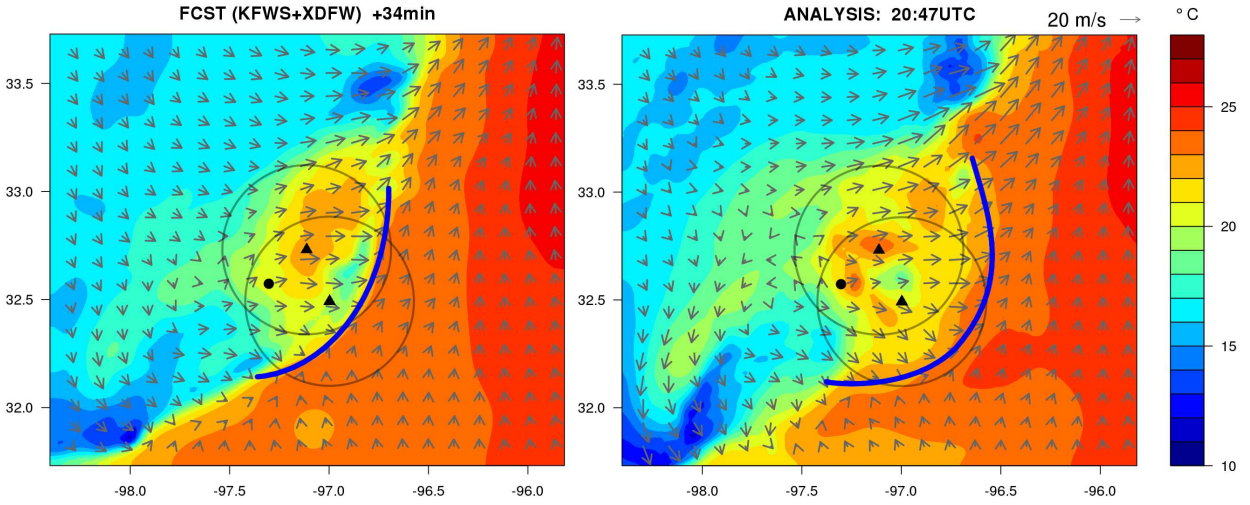


Figure 6.8: Forecast using VDRAS with assimilation of NEXRAD radar only (left column), multiple radars NEXRAD+XDFW (middle column), and corresponding analysis at the proper lead time (right column). In the first row the +0 min lead time forecast (analysis) is displayed, in the second row the +14 min forecast, and in the third row the +34 min forecast.

Since this portion of the squall line is located downwind respect to the region where multiple-Doppler observations are available, this can be interpreted as an indication of the X-band measurements significance for the improvement of the wind analysis.



(a) Forecast (b) Analysis

Figure 6.9: Liquid water potential temperature forecast at +34 minutes lead time (left) and corresponding analysis (right), with wind arrows, at 200m vertical level. The overplotted thick blue lines mark the approximate position of the gust front, based on the wind and temperature fields.

Regarding the wind prediciton, it is evident from the evolution of the wind field and the temperature (fig. 6.9), that the model forecast is not able to properly represent the advancing gust front associated with the cold pool propagation. The thick blue lines in fig. 6.9 mark the approximate position of the advancing gust front in the forecast (left) and in the analysis (right). In the forecast the gust front propagates very similarly as the rainwater field (fig. 6.8, bottom/middle panel), while the analysis depicts a faster propagation of the gust front relative to precipitation within the squall line (fig. 6.8, bottom/right panel). In fact in the analysis at 20:47 UTC (+34 min forecast in fig. 6.9) the near-surface convergence region and the strongest temperature gradient are located near the southern limit of the Midlothian radar domain (southern circle in the image), while the organized precipitation system is also moving eastward but has just barely passed the radar site. The forecast also

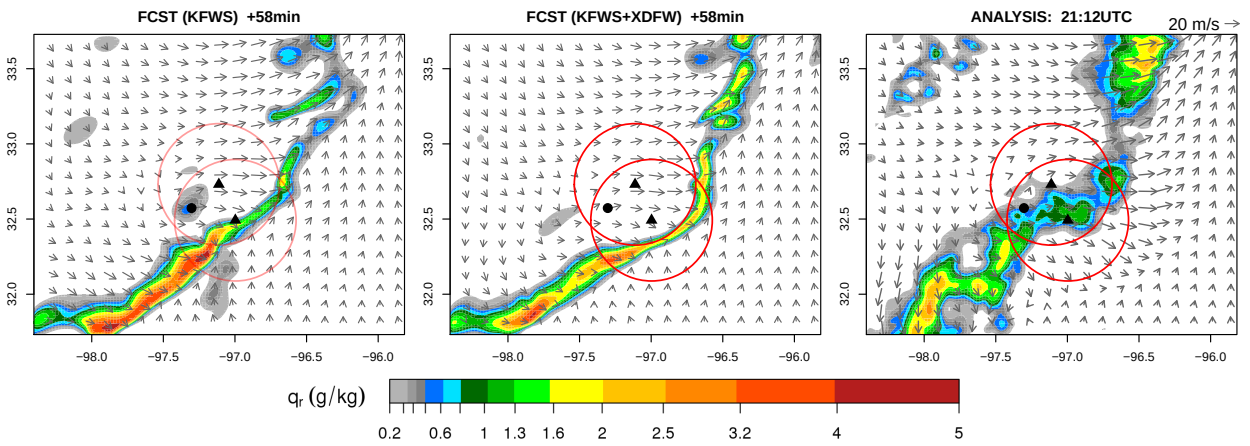


Figure 6.10: Same as fig. 6.8, but for the +58 min forecast.

presents a very pronounced intensification of the temperature gradient along the leading edge of the squall line, which is not found in the corresponding analysis.

As the forecast progresses, the squall line tend to narrow excessively, reducing the width along the propagation direction. The ∼1 hour forecast of rainwater presented in fig. 6.10 shows a considerably reduced width of the squall line, associated with an increased amount of rainwater. In particular values of q_r up to ∼5 g kg^{-1} are visible in the forecast starting from the single radar analysis (left panel in fig. 6.10), about a factor of two higher than the maximum found in the analysis. It seems plausible that the use of warm-rain only microphysics in the model may be responsible for the excessive narrowing in the forecast. The impact of the assimilation of the whole radar data set, including observations in the ice phase, will need to be evaluated using the recent VDRAS implementation of the ice-phase microphysical process [133].

In general the boundary conditions play a relevant role on the limited area model forecasts. In fact the uncertainties in the boundary conditions may have implications early in the forecast, sooner for smaller domains. The current version of the VDRAS system uses boundary conditions derived from the WRF model for the initialization and assumed fixed

during the assimilation time window and for the subsequent forecast. This is a viable solution if the assimilation system is only used for the analysis, as it was originally devised for. If the domain is sufficiently extended, in fact, the impact of using fixed boundary conditions over a typical time window of about 5-15 minutes is expected to be very limited. If the same model is run further ahead in time to provide forecast up to $\sim$1 hour or longer, the impact may start to become relevant. In addition to the uncertainties in the boundary at the time of the analysis, the steadiness of the model state on the borders starts introducing more errors as the forecast progresses.

For the specific case analyzed, some of the issues discussed and specifically the excessive intensification in the southwestern part of the domain, appear very likely ascribable to the steadiness of the boundary conditions.

6.6. Conclusions

VDRAS simulations using X-band radar data in addition to a single NEXRAD radar were performed in this preliminary study. In order to cope with path attenuation affecting short-wavelength radar a new estimation of the rainwater mixing ratio from radar has been implemented, exploiting dual-polarization observations and the hydrometeor classification described in Chapter 4.

The availability of multiple short-range radars is fundamental to provide radial velocity observations and rainwater estimates near the surface. In addition, the scan strategy implying very frequent low-level scans appears especially suitable for running an analysis system based on four-dimensional data assimilation. First results are encouraging and indicate the potential for low-level wind analysis over the DFW metropolitan area.

The same cloud model employed for the assimilation is also integrated forward in time in this study to evaluate its potential for short term forecasting. First results show lights and shadows. On the upside the model demonstrated an ability to provide reliable indications about the very short term storm development, beyond the simple information that is generally possible to infer from applying simple extrapolation on previous observations. This seems particularly true when multiple radar observations contribute to the analysis, as discussed with regard to the 0-30 min forecast in fig. 6.8 (middle column). The impact on the forecast deriving from the inclusion of the two X-band radars in the initial analysis can be quantitatively evaluated considering the classical summary measures (Appendix A). Fig. 7.14 reports the average Threat Score with increasing forecast lead times for two experiments (single and multiple radar assimilation) and two thresholds on the rainwater (0.1 and 0.4 g kg^{-1}). In this figure the light blue area area denotes an improvement of the forecast skill when the X-band observations are used in the VDRAS analysis. The forecasts are verified against the simple mosaicked rainwater field (the first guess from which the analysis is eventually obtained after the assimilation process). For this reason the initial (+0 min forecast) value of the Threat Score is less than one. The Threat Score plots indicate a small improvement associated with the NEXRAD+XDFW assimilation, for lead times approximately $\geq$30 minutes. However, the amount of the improvement does not appear significant, given the larger variability of the score during the event represented by the gray area in the plots. It is possible that the relative small coverage of only two X-band systems considerably limit the impact on the forecast when evaluated over the larger model domain during the whole event. For the case study analysed only two radars were operational over the Dalls-Fort Worth testbed, but at the time of writing the rest of the initially planned eight systems are being deployed. Future analysis of more case studies including the enlarged X-band radar

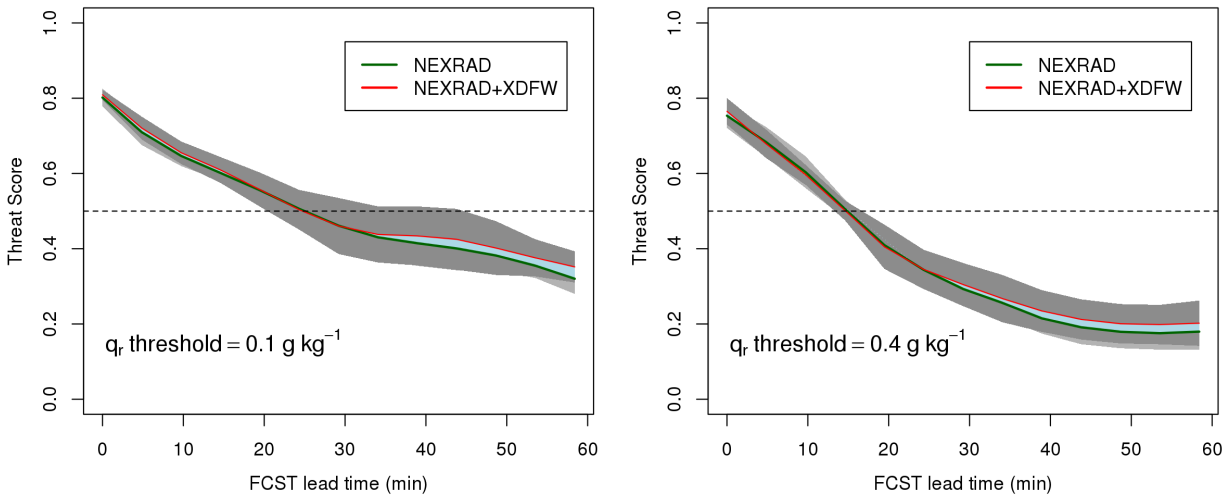


Figure 6.11: Threat score calculated for two different q_r thresholds (0.1 and 0.4 g kg^{-1} respectively in left and right panel), for the forecast based on assimilation of single NEXRAD radar (blue line) and both NEXRAD and the two X-band radars (red line). The gray area represents $\pm$ 1 σ (standard deviation), with light gray for NEXRAD and darker gray for NEXRAD+XDFW. The light blue area denotes an improvement of the forecasting skill when the X-band observations are ingested in the assimilation.

network will allow a more quantitative evaluation of the VDRAS potential for nowcasting in the DFW area.

Another relevant aspect regards the VDRAS undergoing developments. Two major advancements potentially affecting the application in the DFW region are the implementation of the ice-phase microphysical process [133] and the inclusion of dynamic boundary conditions, currently being developed (Juanzhen Sun, NCAR, 2016, personal communication). The former development (implementation of the ice microphysics), in addition to represent an improvement for the model itself, will allow a more synergistic use in combination with the advanced hydrometeor classification technique based on dual-polarization radar observations (Chapter 4). Based on the radar classification it will be possible to provide better estimates of both liquid and ice water content to the four-dimensional data assimilation system. The inclusion of ice microphysics will also foster research on new approaches to assimilate the

information about the hydrometeor type. This is anticipated to be especially relevant for deep convection and the very short term prediction of hail. Finally, the inclusion of dynamic boundary conditions is also an awaited improvement that is expected to benefit in particular the forecasting phase.

CHAPTER 7

An attempt to bridge the gap between model and extrapolation nowcasting

The description of the atmospheric state evolution is inherently a three-dimensional problem involving several variables (e.g. for warm rain process: temperature, humidity, pressure, winds, water vapor, cloud water, rainwater). However the purpose of this work is specifically linked to the two-dimensional forecast of the rainwater (q_r) and winds (u, v) in the near-surface layer. One possible and increasingly pursued option for precipitation and wind nowcasting is to run a three-dimensional numerical model and then consider only a specific section of the forecasted atmospheric state, i.e. two variables on a single horizontal layer. This approach has been investigated in this research work in Chapter 2 using the WRF model, and in Chapter 6 with the VDRAS four-dimensional variational assimilation system. At the other extreme, extrapolation-based precipitation nowcasting exclusively relies on one variable (typically the radar reflectivity) observed on a given surface, e.g. the conical scan surface. Nowcasting by extrapolation often makes use of optical flow techniques to estimate the motion vectors based on two or more previous observations. The motion vectors are then used to advect in time the most recent observations. Given the huge difference in terms of complexity between the model and the extrapolation approaches, the performance attainable by adopting the latter technique for precipitation nowcasting is somehow surprising. The reason is that in general the advective component of the storm evolution prevails over the more complex dynamical and microphysical interactions within the storm and with the surrounding environment.

In the previous chapter it was shown how the use of a four-dimensional data assimilation system has the potential to provide an accurate low-level wind analysis. Unlike simpler multiple-Doppler techniques, the adoption of a cloud-scale numerical model in the assimilation process ensures the physical consistency between the wind and the rainwater fields over the assimilation time window. The same cloud model used in the assimilation phase can be let run further to provide a short range forecast. However, while the typical assimilation window only spans few to several minutes, the desired forecast range is quite larger for typical nowcasting applications ($\sim$60 minutes). Over this forecast range the model may actually show poorer performance respect to methods based on the simple extrapolation of the observed reflectivity or related quantity ([134, 135]). On the other hand, the extrapolation-based methods provide good results for very short term forecasting (up to about 20 minutes), but their performance tends to quickly decrease due to the absence of a representation of the atmospheric physics, and more specifically:

- the lack of source/sink terms: the most recent observations are simply advected, assuming Lagrangian persistence and irrespective of possible growth or decay affecting the storm evolution;
- the motion vectors are estimated from observations typically confined over a portion of the radar domain. When the storms are advected in regions originally not showing radar echoes, the motion vector estimates may loose their representativeness.

The availability of a reliable estimation of the low-level wind field may help to compensate the aforementioned limitations. In particular, the divergence of the low-level winds provides useful indications about the location of convergence (negative divergence) regions near the surface, where intensification of existing storm or new development is more likely. On the

other hand, the downdraft inferred from the positive divergence regions, in conjunction with the rainwater content evolution, may be used to attempt a simple representation of the outflow which is relevant to the evolution of the wind field in the lower level [125].

"Mean winds" methods have been used with some success to predict the storm motion. These methods rely on the empirical evidence that the motion of the storms correlates well with the average wind in the low- to mid-troposphere. For example Davies and Johns [136] used the average wind in the lowest 0-6 km AGL. The rainwater motion vectors estimated using the past storm location can therefore be regarded as representative of some weighted average of the lower tropospheric winds. It seems then consistent to attempt exploiting the low level wind prediction to update the initial estimate of the rainwater motion vectors during the forecasting phase, assuming that the upper level winds will evolve more slowly.

An attempt to model the interconnections between the rainwater and the low-level wind evolution is described in section 7.2. This model relies on the independent estimation of the motion vectors for the rainwater and for the components of the wind field (divergence and vorticity). In short, the approach described in this chapter can be regarded as an attempt to inject some basic elements of atmospheric physics in a model (pure extrapolation) otherwise completely lacking such information.

Preliminarily, the next section focuses on the specific technique employed to estimate the motion vectors.

7.1. Horn-Schunck optical flow with a multi-scale pyramidal approach

Many techniques exist to estimate the motion of objects or surfaces from a sequence of ordered images. These techniques are generally referred to as optical flow and can rely on different methods for the determination of the motion. In the atmospheric science the

most popular approaches are the block-based methods, minimizing the sum of squared differences or maximizing the normalized cross-correlation ([137, 138]) or the variational methods ([134]). Horn and Schunck [139] were the first to propose a variational method for optical flow estimation. In their seminal work, the basic optical flow constrain equation, frequently referred as the brightness constancy assumption, states that the apparent brightness of moving objects remains constant over time. This is expressed as:

$$\frac{dI}{dt} = \nabla I \cdot \mathbf{u}(\mathbf{x}) + \frac{\partial I}{\partial t} = 0 \tag{33}$$

This equation can not be solved pointwise. In fact, because of the two unknowns, only the magnitude of the motion in the gradient direction can be estimated. In order to solve this aperture problem, some additional constraint need to be introduced. Horn and Schunck proposed a variational method with global smoothing to ensure filling in the motion estimate from nearby gradient constraints. The variational problem is thus solved minimizing an energy functional:

$$J = \int\int \left[\left(\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} \right)^2 + \alpha^2 \left(|\nabla u|^2 + |\nabla v|^2 \right) \right] \mathrm{d}x \, \mathrm{d}y \tag{34}$$

where the parameter α is a regularization constant to control the smoothness of the motion estimates. Larger values of α lead to a smoother flow. The minimization of J can be achieved by solving the associated EulerLagrange equations:

$$\begin{aligned} &\frac{\partial I}{\partial x} \left(\frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} \right) - \alpha^2 \nabla^2 u \\ &\frac{\partial I}{\partial y} \left(\frac{\partial I}{\partial x} v + \frac{\partial I}{\partial y} v + \frac{\partial I}{\partial t} \right) - \alpha^2 \nabla^2 v \end{aligned} \tag{35}$$

where ∇^2 is the Laplace operator. With two equations and two unknowns, and approximating the Laplacian with $\nabla^2 u(x,y) = \bar{u}(x,y) - u(x,y)$ (with $\bar{u}(x,y)$ being the average of u in the neighborhood of (x,y)), it is possible to solve iteratively using the Jacobi method. For a given iteration number k:

$$
\begin{aligned}
u^{k+1} &= \bar{u}^k - \frac{\frac{\partial I}{\partial x}\left(\frac{\partial I}{\partial x}\bar{u}^k + \frac{\partial I}{\partial y}\bar{v}^k + \frac{\partial I}{\partial t}\right)}{\alpha^2 + \frac{\partial I}{\partial x}^2 + \frac{\partial I}{\partial y}^2} \\
v^{k+1} &= \bar{v}^k - \frac{\frac{\partial I}{\partial y}\left(\frac{\partial I}{\partial x}\bar{v}^k + \frac{\partial I}{\partial y}\bar{v}^k + \frac{\partial I}{\partial t}\right)}{\alpha^2 + \frac{\partial I}{\partial x}^2 + \frac{\partial I}{\partial y}^2}
\end{aligned}
\tag{36}
$$

The method relies on the proper estimation of the partial derivatives $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$, and $\frac{\partial I}{\partial t}$. If these can not be correctly estimated due to highly non linear gradients or excessively large displacements of the precipitation patterns between successive images, the motion vectors can not be correctly calculated. To overcome the possible issue related to non linear gradients, the radar reflectivity (or logarithmic rainwater) at a given vertical level is considered for the estimation of the motion vectors. In fact the rainfall rate (or liquid water) tend to show high peaks with exponential decay away from the precipitation core in convective precipitation. On the other hand logarithmic quantities like reflectivity present a more linear decay. The second issue may arise from either excessive physical displacements or too high grid resolution. In order to overcome this possible issue the multi-scale strategy approach of Meinhardt-Llopis et al. [140] is adopted. In their method a pyramidal structure provides a coarse-to-fine successive refinement of the flow field. The input reflectivity images are filtered and downsampled by a factor η using bi-cubic interpolation. Starting at the coarsest scale the optical flow equations (36) are solved and every intermediate solution is used as the initialization in the next scale. The factor η is chosen based on the expected

maximum storm motion and the grid resolution of the images, in order to keep the motion to be detected small at the coarsest scale.

Fig. 7.1 shows an example of motion vectors estimation for the case study discussed in section 7.3, for the rainwater (panels a and c) and the divergence (panels b and d).

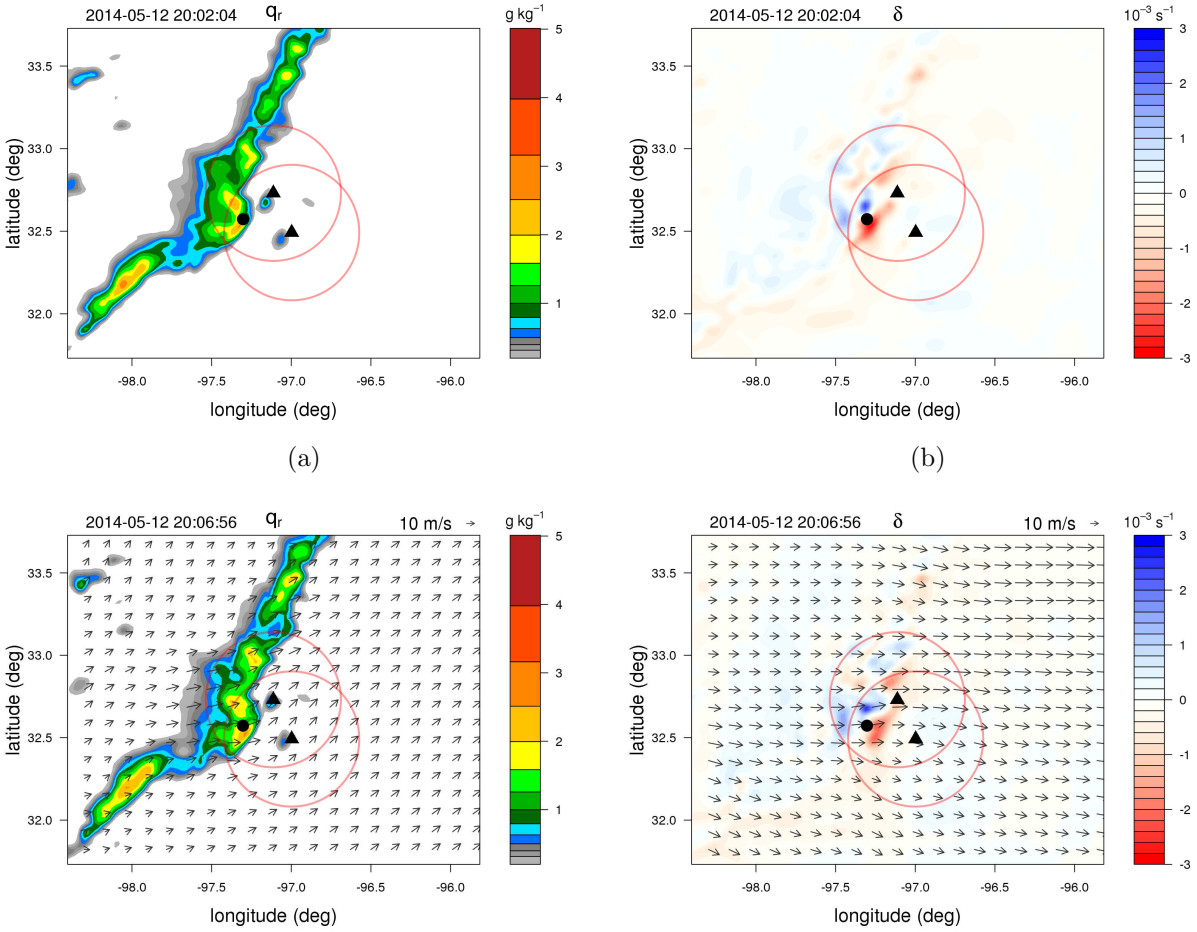


Figure 7.1: Example of motion vectors estimation. Panel a) and c) show the rainwater analysis at 600 m height for two successive time steps (20:02UTC and 20:06UTC). Based on the optical flow solution on this pair of images, the motion vectors in panel c) are estimated. Similarly, panels b) and d) show the corresponding motion vectors estimation for the divergence fields at the same vertical level. Only one vector every six grid points is plotted for clarity.

The motion vectors for the rainwater show a dominant westerly component in the western sector, behind the squall line (moving eastward), while a south-westerly flow is estimated ahead of the storm. This is in good agreement with the VDARS wind retrieval in the low levels (fig. 6.6 in Chapter 6). In this specific case the optical flow method is able to derive a reliable estimate ahead of the storm, exploiting the motion of the two small cells close to the two X-band radars (triangle symbol). The divergence (panels b and d in fig. 7.1) shows a westerly flow over most of the domain, also ahead of the storm, with a small modulation across the squall line. The motion fields of rainwater and the wind components (divergence and vorticity) may show relevant differences that are expected to bring useful complementary information for the forecast.

7.1.1. Semi-Lagrangian Advection. After the estimation of the motion vectors, the advection of the rainwater and wind components can be accomplished using either forward or backward schemes. Forward (in time) schemes foresee the distribution of the advected quantity among the neighboring grid points around the destination point (which in general does not coincide with a grid point). Forward schemes are therefore mass-conservative by definition. Another approach is to use backward advection, i.e. for a given grid point the origin at the previous time step is found by following the flow backward. This again will not coincide with a grid point, so in this case interpolation is necessary. Bi-linear interpolation is often used for this purpose [134]. However, in order to reduce the diffusion arising from the bi-linear scheme, a bi-cubic interpolation is adopted here.

The combination of the Lagrangian perspective and the use of a regular grid Eulerian framework is known as semi-Lagrangian scheme. This class of methods has the notable advantage of being particularly efficient, allowing the use of large time steps.

As opposed to the forward advection, the backward scheme is not mass-conservative, although it may be considered nearly mass conservative when the divergence of the flow field is negligible. So, in order to ensure mass conservation, the divergence component of the rainwater motion vectors field needs to be removed. This is accomplished relying basically on a technique widely used in fluid dynamics simulations [141]. In practice, referring to eqs. (24, 25, 28) in Chapter 5, the procedure can be summarized in three steps:

- calculate divergence from the motion vectors field
- solve for the velocity potential χ, given that $\nabla^2\chi = \delta$, using iterative relaxation
- subtract $\nabla\chi$ from original motion vectors field

This procedure to make the motion vectors of rainwater non-divergent is especially important when the motion vector field is mixed with the low level wind field during the forecast, as described later in section 7.2.3. In fact the low level wind field generally has a quite relevant divergence component, which would severely impact the stability of the rainwater advection in terms of mass conservation causing unphysical visual deformations (stretching/shrinking).

7.1.2. Image registration. As described in the next section, the parametric model relies on the analyses at two vertical levels for the estimation of the rainwater content gradient. The vertical gradient is used to estimate the contribution to the rainwater in the lowest level by vertical advection. However, possible tilting of the storm may affect this estimate introducing artifact gradients. In fact the two-dimensional model does not resolve the vertical wind shear, so the high-level rainwater need to be aligned with the low-level field to compensate the eventual tilting. In addition, depending on the scanning strategy and analysis method, an apparent additional tilting may be introduced by the delayed radar scanning of the higher elevations. Correction of the apparent misalignement can be performed through

image registration. Optical flow may also be used for this purpose, so the Horn-Schunck technique described in section 7.1 is also applied to determine the appropriate deformation (motion vectors) to align the upper level to the lower level rainwater field.

7.2. Parametric model for nowcasting

If the two basic assumptions for the optical flow, i.e. the stationarity of the motion vectors and the lack of a source term, are removed eq. (33) becomes:

$$\frac{dq_r}{dt} = \nabla q_r \cdot \mathbf{u}(t, \mathbf{x}) + \frac{\partial q_r}{\partial t} = S(t, \mathbf{x}) \tag{37}$$

If the brightness is not constant the motion estimate can be biased. Although attempts have been made to develop methods including brightness variation caused by time-dependent physical models [142], these were mainly limited to relatively simple applications such as changing illumination or thermal diffusion in infrared images.

The approach adopted here is instead based on a separate treatment of the optical flow derived from equation (33) and the local rate of change of the rainwater content and winds.

The proposed model relies on analyses from two vertical levels of the VDRAS assimilation system (Chapter 6):

- 600m MSL (rainwater and winds)
- 3400m MSL (rainwater)

The choice of the specific levels is dictated by the need to have a sufficient vertical spacing inside the liquid phase layer to calculate a reliable rainwater gradient (section 7.2.1). While the VDRAS analysis provides the full set of atmospheric variables, only the radar observable fields (rainwater and winds) are considered in this approach. The idea is to constrain the

observations (analyses) using a simplified physical model with adaptive parameters. It is argued that while the deficient physical description will inherently limit the validity of the forecast for large lead times, the adaptiveness of the model may help to improve the very short term forecast (0-60 minutes). The traditional use of extrapolation applied to reflectivity (rainwater) is here extended to the wind components and simple relations governing the rain growth and decay are defined and heuristically tuned through a set of adaptive parameters.

In an observational environment such as the DFW (Dallas-Fort Worth) testbed (Chapter 6), the architecture of the weather radar network provides the best coverage in the atmospheric layer closer to the surface. In this perspective, the aim of the proposed approach is to extract the most valuable information content from the available observations. In addition to the rainwater content, analysis and nowcast of the wind field near the surface has a special relevance on its own for the potential impact on human activities and infrastructures. A suitable representation of the dynamics taking place in the lowest layer is also important for the tight relation with the storm evolution.

The basic steps of the parametric model are described in detail in the following subsections. The divergence δ and the vertical component of the curl ξ (hereafter simply vorticity) are initially calculated from the low-level wind analysis. Considering a pair of observational time frames, typically lag-0 (time t_0) and lag-1 (time $t_0 - 1$), the optical flow technique is applied independently to the rainwater (q_r), divergence (δ) and vorticity (ξ), obtaining the respective motion vectors $\mathbf{U}_{qr}$, $\mathbf{U}_{\delta}$, $\mathbf{U}_{\xi}$. The divergence and vorticity are advected to the next time step using the respective motion vectors, while before applying advection to q_r, its local rate of change at the two vertical levels is estimated (section 7.2.1).

7.2.1. Growth and decay parameterization. The parameterization of the growth and decay local rate of change is realized according to the following equations:

$$\frac{\partial q_r}{\partial t} = \frac{\partial q_r}{\partial z}\,(w + V_t) + p_0\, w \tag{38}$$

$$V_t = 5.40\; q_r^{0.125} \tag{39}$$

where V_t is the terminal fall velocity of rain drops [7], while w is the vertical velocity of air obtained from the divergence in the hypothesis of mass conservation. Since the divergence at the near-surface level (600m) is being considered, assuming $w = 0$ below this level and zero divergence at the upper level (3400m), a positive (negative) divergence corresponds to negative (positive) vertical velocity in this atmospheric layer. Equation (38) is basically the continuity equation for precipitation originally derived by Kessler [143]. The first term on the right represents the sedimentation (vertical advection) of rainwater, and the second term the growth by condensation. Since cloud water (q_c) is not considered in this model, the microphysical contributions to precipitation due to auto-conversion of cloud to rain and accretion of cloud water by existing rain are not accounted for. Following [143] the depletion of rain by evaporation can be represented as:

$$\frac{\partial q_r}{\partial t} = p_1\; q_c\; q_r^{0.65} \tag{40}$$

where, with a strong approximation, q_c has to be assumed constant (= 1) for being not represented in this context. For the practical implementation, since only two vertical levels are considered, the above parametric equations are applied to the upper (superscript 1) and

lower (superscript 0) levels:

$$(41) \qquad (q_r)^1_{t+1} = (q_r)^1_t + p_0\ L^{p_2} w_t\ dt; \quad L^{p_2} w_t = w_{t-p_2}$$

$$(42) \qquad (q_r)^0_{t+1} = (q_r)^0_t + \frac{(q_r)^1_t - (q_r)^0_t}{\Delta z}\ (w + V_t)\ dt; \quad (q_r)^1_t > (q_r)^0_t \ \text{ and } \ (w + V_t) < 0$$

$$(43) \qquad (q_r)^{0,1}_{t+1} = (q_r)^{0,1}_t + p_1\ q_r^{0.65}\ dt;$$

where t and $t+1$ indicate the current and next time step, and dt is the forecast time interval. In eq. (41) L denotes the *lag operator*, i.e. L^{p_2} means lagging w by a period p_2. The lagged field is obtained by advecting backward or forward in time the divergence (vertical velocity) using the estimated motion vectors. This is a necessary and important feature of the model to compensate for the lack of three-dimensionality, in particular for squall lines with surface-based cold pools. In fact the regions of strong convergence near the surface, often associated with a gust front (fig. 7.2), may move significantly away (often downwind) from the main storm core. This may result in a tilted warm inflow current turning into the updraft. In this case the storm enhancement or new cell development will take place upwind respect to the observed low-level convergence. The role of the lagged vertical velocity is therefore to heuristically contemplate the possible spatial displacement. After all, although the storm evolution is greatly dependent on the environmental shear profile [112], it is believed that the relative position and propagation of the near-surface convergence respect to the precipitation core has a potential to provide valuable information to improve the very short-term prediction of the overall system motion and morphology.

The rainwater vertical advection is only applied when there is a positive q_r flux downward (eq. 42) in the current setting. Equation (42) may actually be applied irrespective of the

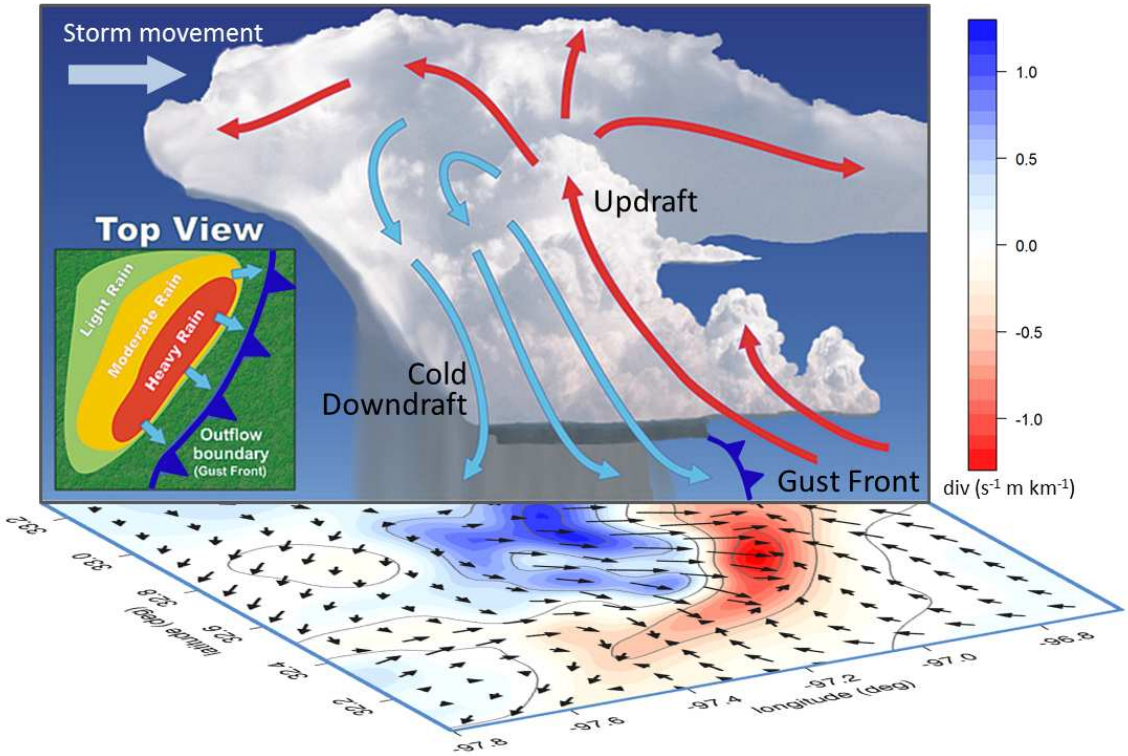


Figure 7.2: Idealized diagram of a squall line vertical structure showing updrafts, downdrafts, and a gust front. Precipitation forming in the tilted updraft falls into the downdraft. Beneath the cloud, the denser cool air of the downdraft spreads out along the ground. On the leading edge of the outflowing downdraft a gust fron may form, forcing the moist surface air to flow up into the cloud. In the lower horizontal plane oriented perpendicular to the diagram, a real VDRAS wind analysis from the case study discussed in section 7.3 (20:21UTC) is displayed with divergence in color. The diagram and the horizontal analysis are subjectively matched with the purpose of illustrating the relation between the low level wind and the storm vertical development. From the retrieved wind field, the average storm motion has been subtracted in order to show the storm-relative winds. The diagram is adapted from *www.srh.noaa.gov/jetstream/index.html*.

gradient and vertical velocity signs, although this may imply negative q_r values. Negative q_r can be dealt with and provide a way to inhibit new convection in regions where persisten downdraft have occurred. However some preliminary test showed that the adopted solution, exploiting the evaporation term to balance the overall rainwater budget, performed better.

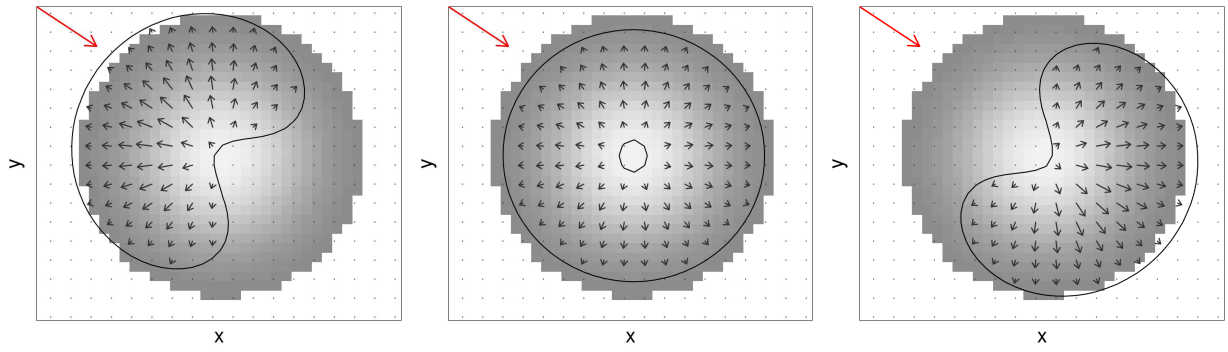
The coefficients p_0 (condensation), p_1 (evaporation) and p_2 (lagged vertical velocity) are part of the parameters set to be determined adaptively through the optimization described below in section 7.2.4.

7.2.2. Outflow parameterization. A downdraft generally develops within a thunderstorm when precipitation falls through an unsaturated layer and evaporation starts cooling the air. The combined effect of precipitation loading (drag of liquid water) and evaporative cooling can lead to the formation of a cold pool associated with damaging winds at the surface. In fact the downdraft approaching the surface causes a divergent flow and a gust front (outflow boundary) propagates, separating the cooler air below the storm core from the environmental warmer air [125]. The speed of the advancing gust front relative to the ambient flow is found to be close to the speed of a density current and can be expressed in terms of the density difference between the surrounding air and the more dense air within the cold pool [144]. However, the lack in this context of any information about density (or pressure) prompts for an attempt to represent the flow associated with the cold pool in terms of the vertical velocity and the evaporative cooling (eq. 40). Due to the non-uniform surface winds and the three-dimensional flow organization, in particular the presence of a rear inflow jet in squall lines causing cold air to be drawn in on the rear side of the storm [125], the divergent flow on the two-dimensional plane near the surface is in general not symmetric. In order to mimic this near-surface two-dimensional structure of the flow originating in downdraft regions, a term defining the outflow strength is first introduced:

$$DIV_{out} = p_3\,(w + V_t)\,dE\,dt; \quad dE = ((q_r)^0_{t+1})^{0.65} - ((q_r)^0_t)^{0.65} \tag{44}$$

$$dE > 0 \text{ and } (w + V_t) < 0$$

where dE represents the incremental rainwater mass loss owing to evaporation between time steps t and $t+1$. Eq. (44) is based on the knowledge that the outflow produced at the surface is the consequence of both the negative vertical velocity (producing divergence) and cooling due to evaporation (producing a spreading density current). The initial wind analysis already implicitly includes the outflow term, while for the next predicted time steps the evolution of both the downdraft velocity and the rainwater mass needs to be considered.



(a) p_4 = -1.0 (b) p_4 = 0.0 (c) p_4 = +1.0

Figure 7.3: Qualitative illustration of the flow representation within the cold-pool on an arbitrary xy-plane. The gray shading represents the intensity of the divergence, the black arrow the storm-relative wind in the cold-pool. The red arrows indicates the direction of the divergence motion and the solid contour represents the outflow boundary.

From the divergence term associated with the downdraft (eq. 44) the corresponding irrotational flow ($\mathbf{U}_{out}$) is estimated using a Successive Over Relaxation (SOR) technique and recalling eq. (24) and (25) in Chapter 5:

$$\mathbf{U}_{out} = \nabla\chi; \qquad \nabla^2\chi = \delta \tag{45}$$

where χ is the velocity potential and assuming Dirichlet boundary condition (null velocity on the boundary).

The combination of the mean wind and the rear inflow with the outflow causes the circulation within the cold pool to divert from the symmetric pattern arising from the SOR retrieval. In practice over time the cold pool tends to elongate in the direction of the mean wind [145], with segments of the gust front oriented parallel to the mean wind remaining quasi stationary, while segments perpendicular to the mean wind move faster downwind. This behavior can be portrayed considering the angle θ between the unit vector representing the motion of the wind field divergence (indicative of the gust front propagation) and the irrotational flow vector associated with the outflow (eq. 45). A damp factor is defined based on the dot product between these two vectors as:

$$
\begin{aligned}
d &= d_0 - (d_0 - 0.5)\,(1 - p_4); \quad d \geq 0.5 \\
d &= d_0 + (0.5 - d_0)\,(1 - p_4); \quad d < 0.5 \\
d_0 &= (cos(\theta) + 1)/2
\end{aligned}
\tag{46}
$$

and applied to the outflow vector:

$$
\mathbf{V}^{'}_{out} = d \cdot \mathbf{U}_{out} \tag{47}
$$

For p_4=0, the flow is unaffacted except for a 0.5 scaling factor (middle panel in fig. 7.3), while values of $-1 \leq p_4 < 0$ and $0 < p_4 \leq 1$ are associated respectively with backward and forward propagation (left and right panel in fig. 7.3). The resulting flow $\mathbf{V}^{'}_{out}$ is finally added to the low-level wind field:

$$
\mathbf{U}^{'} = \mathbf{U} + \mathbf{V}^{'}_{out} \tag{48}
$$

Table 7.1: Synthetic description of the parameters in the model.

Parameter	**Description**
p_0	Condensation
p_1	Evaporation
p_2	Vertical velocity Lag
p_3	Outflow divergence
p_4	Forward-Backward outflow factor
p_5	Mix-winds weight factor

7.2.3. Wind advection and mixing with rainwater motion vectors. Vorticity and divergence are calculated from eq. (48) and advected using the respective motion vectors estimates. The wind field at time step $t+1$ ($\mathbf{U}'_{t+1}$) is then calculated applying the SOR technique from the divergence and vorticity components (eq. 28 in Chapter 5), relying on the current time wind field ($\mathbf{U}'_t$) as first guess. The updated low-level wind field is also used to adjust the rainwater motion vectors, based on the previously discussed assumption that the storm motion is influenced by the mean wind in the low to medium troposhere:

$$\mathbf{U}_{qr} = (p_5 \cdot \mathbf{U}' + \mathbf{U}_{qr})/(p_5 + 1) \tag{49}$$

The flow-related coefficients p_3, p_4, p_5 complete the set of six parameters (table 7.1) that need to be determined.

The general procedure is represented by the flow diagram in fig. 7.4 and summarized hereafter:

(1) the analysis (rainwater and wind) at the current (t_0) and previous (t_0-1) time steps are considered as input for the model;

a) image registration is applied in order to align the rainwater fields at the two vertical levels (600 and 3400m), compensating for possible real (wind shear) or apparent (inter-scan delays) tilting of the storm;

b) the intensity of the outflow, and relative flow vectors, is estimated from the rainwater and divergence at time t_0 according to eq. (44) (blue blocks in the flow diagram);

(2) optical flow estimation of the motion vectors is performed separately for the three fields: rainwater (q_r), and the two derived components of the wind field, i.e. divergence (δ) and vorticity (ξ);

(3) growth/decay terms are applied to the q_r fields (lower and upper levels) according to eqs. (41, 42, 43).

(4) the fields q_r, δ and ξ at t_0 are advected to time step $t_0 + 1$ using the respective motion vectors (point 2), relying on a backward advection scheme with bicubic interpolation.

(5) from the updated q_r and δ at time $t_0 + 1$, the new intensity of the outflow and relative flow vectors is also estimated.

(6) the advected δ and ξ are combined with the outflow from point 1b (also advected using the motion vectors $\mathbf{U}_\delta$) to retrieve the update low-level wind field at time step $t_0 + 1$

(7) the updated low-level wind field is mixed with the q_r motion vectors at time t_0 to provide new displacement vectors to advect q_r forward from $t_0 + 1$ to $t_0 + 2$.

For the successive time steps, points 3 to 7 are cyclically repeated incrementing the time indexes. The time step used in the forecast is the same as the time interval between the pair

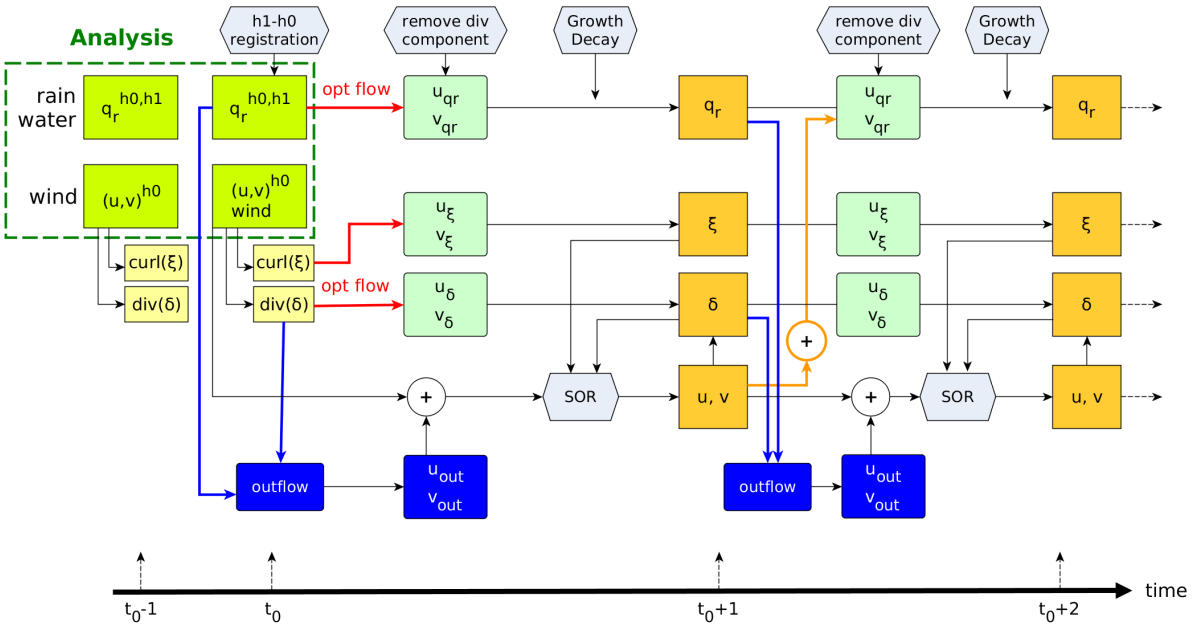


Figure 7.4: Schematic flow diagram of the nowcasting model. The Growth/Decay process icludes sedimentation, condensation and evaporation (section 7.2.1). SOR means Successive Over-Relaxation. $h0$ and $h1$ respectively refer to the lower (600m) and upper (3400m) level.

of initial observations t_0 and $t_0 - 1$ (292 seconds). Experiments using shorter time steps for the forecast resulted in small differences.

7.2.4. Optimization of the model parameters. In order to find the optimal set of parameters, the Nelder-Mead (NM) downhill simplex method [146] is adopted. The NM method belongs to the class of direct search methods and is suitable for multidimensional unconstrained optimization. The simple grounding idea and ease of implementation makes it a very popular method, used in a wide range of scientific applications. The method relies on the concept of a simplex, which is a special polytope with N+1 vertices in N dimensions (e.g. a triangle on a plane for N=2). A simplex-based direct search method begins with a set of N+1 points x_0,...,$x_n \in R^n$ that are considered as the vertices of a working simplex S, and the corresponding set of function values at the vertices f_j:=$f(x_j)$, for $j = 0, ..., n$. A starting point is defined corresponding to the first vertex (x_0), and the initial S is constructed to be a

regular simplex using an arbitrary scaling factor. Then, a series of transformation (reflection, expansion, contraction, reduction) on S is performed, aimed at decreasing the funcion values at the vertices of the simplex. The type of transformation required is determined at each step by computing the function values at one or more test points, and by comparison of these function values with those at the vertices. This process is eventually terminated when the the function values f_j are close enough (the variance of f_j is used, with a given tolerance) or a predefined maximum number of iterations is reeached.

The NM method is not a true global optimization algorithm, although in practice it tends to work reasonably well for problems that do not have many local minima. The objective function to minimize in the current application is defined by the following sum of root mean square errors:

$$f = \frac{\left\langle (q_r^{obs} - q_r^{fcst})^2 \right\rangle^{0.5}}{\sigma_{qr}} + \frac{\left\langle (u^{obs} - u^{fcst})^2 \right\rangle^{0.5}}{\sigma_u} + \frac{\left\langle (v^{obs} - v^{fcst})^2 \right\rangle^{0.5}}{\sigma_v} \tag{50}$$

$$\sigma_{qr} = 0.2 \text{ g kg}^{-1}; \quad \sigma_u = 3.0 \text{ m s}^{-1}; \quad \sigma_v = 3.0 \text{ m s}^{-1}$$

where the average is calculated over the space-time validation domain. The normalization factors in the denominator are assumed constant. The spatial domain is a sub-domain of the whole model domain, to avoid boundary effects, while the temporal domain extends from the analysis time until a given forecast lead time (e.g. 60 minutes).

7.3. Results

Considering the case study of Chapter 6 the period between 19:47 and 21:00UTC is selected, including a total of 16 analyses (one every 292 seconds). The analyses have been processed running VDRAS on a slightly modified domain respect to the experiment in Chapter 6. In this case the model domain is 122x112x30 (nx,ny,nz) grid points, with horizontal

resolution of 2 km, vertical resolution of 400 m, and an integration time step of 4 s. In addition, the assimilation window is reduced from five to three minutes, allowing inclusion of one S-band volume scan and three to five low level PPI scans from each X-band system. These modified parameters allowed to reduce the wall-clock time for the generation of a single analysis to about 5 minutes, using 16 processors.

For each analysis a forecast up to +58 minutes (12 time steps with dt=292s) is started, using the current and the lag-1 fields for the optical flow estimation. The forecasted q_r and low-level winds $\mathbf{U}$ are compared with the corresponding analyses for the estimation of the function f in the iterative optimization procedure (eq. 50). In the current implementation all the parameters in the model (table 7.1) are scaled during the forecast by a factor of 1 (forecast +0 min) linearly decreasing with lead time until zero (forecast +120 min). In fact, since the model greatly relies on extrapolation (q_r and winds) and the domain is partial (both the spatial domain and the variable space), the simple physical relations adopted (eqs. 41-49) will inherently loose their adequacy during the forecast. The decreasing scaling factor is then adopted to give more confidence to the physical relations during the first stages of the forecast, while trying to keep the performance robust for longer lead times.

The validation domain is represented by the rectangle in fig. 7.6 (89x94 grid points). The results of the current method are compared with both the simple Horn-Schunck Optical Flow based on q_r motion vectors only (hereafter simply OF) and with the Dynamic Adaptive Radar Tracking of Storms (DARTS) method [5]. For DARTS, the previous 20 minutes (5 reflectivity images) are considered for the estimation of the motion verctors, while the forecast is performed using a backward bi-linear advection scheme.

For each of the 16 analyses in the study period, the Nelder-Mead optimization is performed to determine the optimal set of model parameters. The optimization is performed

without constraints, but the choice of the initial value of the parameters has an impact on the result, since NelderMead is not a true global optimization algorithm and it may converge to different local minima depending on the initial setting. One way to overcome this limitation would be to perform an outer loop utilizing many initial simplices in order to find the most suitable part of the parameter space to start with. In this preliminary evaluation, the initial parameters are set to "reasonable" values based on the physical meaning of the processes involved (p_0 and p_5 are expected to be positive, while p_1, p_2 and p_3 should assume negative values) and trial and error forecast runs with varying configurations. The adopted initial set of values is: $p_0 = +2.0 \cdot 10^{-5}$, $p_1 = -0.5 \cdot 10^{-5}$, $p_2 = -600$ s, $p_3 = -1.5 \cdot 10^{-7}$, $p_4 = +0.8$, $p_5 = +2.0$. During the optimization a 58 minutes forecast is run iteratively until convergence is reached. On average a single optimization loop took about 18 iterations and 39 function evaluations. The resulting parameter are plotted in fig. 7.5 for every independent forecast. It can be seen that, although the 16 optimizations at successive times are performed independently [1], the resulting parameters are substantially stationary or smoothly changing. The parameter showing the most significant relative variation is p_1 (evaporation), passing from about $-1.0 \cdot 10^{-5}$ at 19:47UTC to ~ 0 around 20:50UTC. As discussed further later in this section, the evaporation term contributes little or nothing to the skill of the forecast when evaluated using the Threat Score (eq. 55), but is only useful to maintain the average rainwater level close to the observations.

As an example, fig. 7.6 shows the parametric model forecast starting at 20:26UTC, denoted Enhanced Optical Flow (EOF) hereafter. In this and the following analysis/forecast

[1]However, the initial analyses are not completely independent. In fact for a given analysis time the VDRAS assimilation relies on the the background forecast from the previous cycle. This contributes to guarantee physical consistency among successive analyses, and can also reflect in the smooth evolution of the parameters resulting from the optimization.

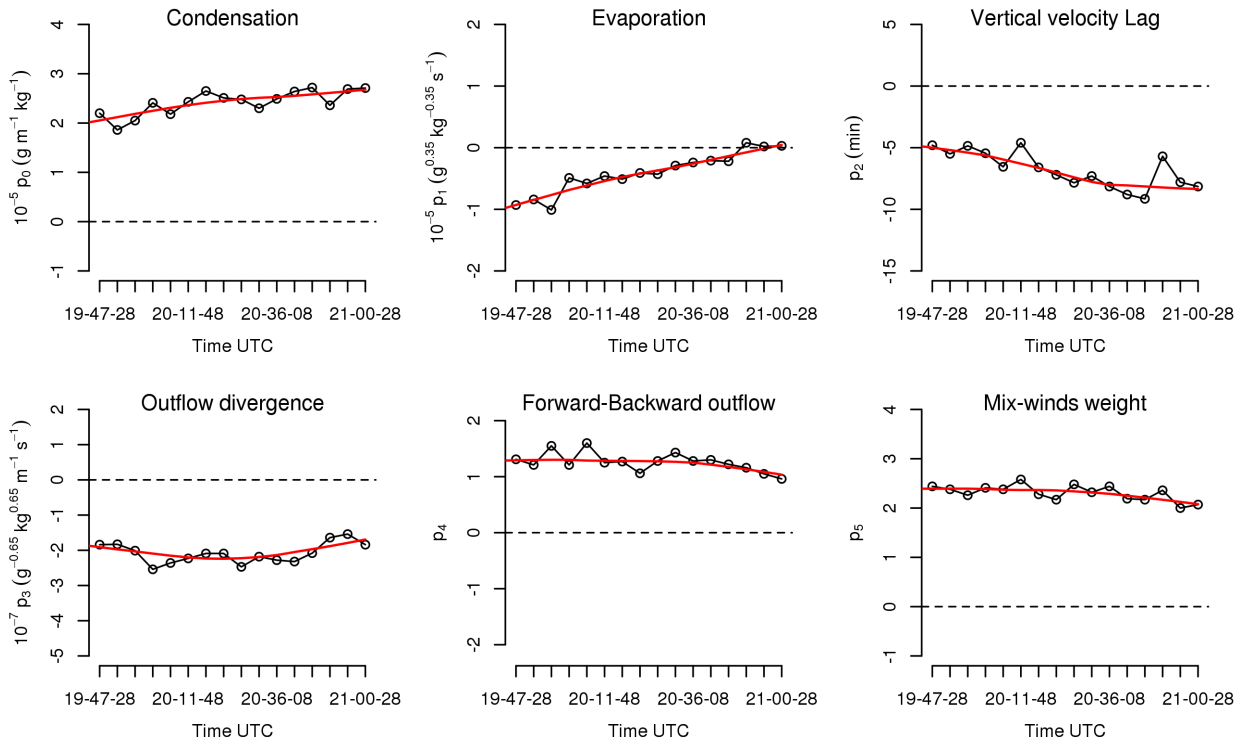
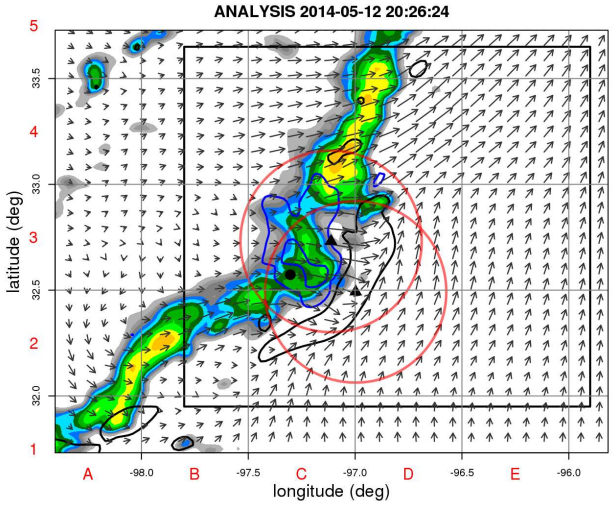
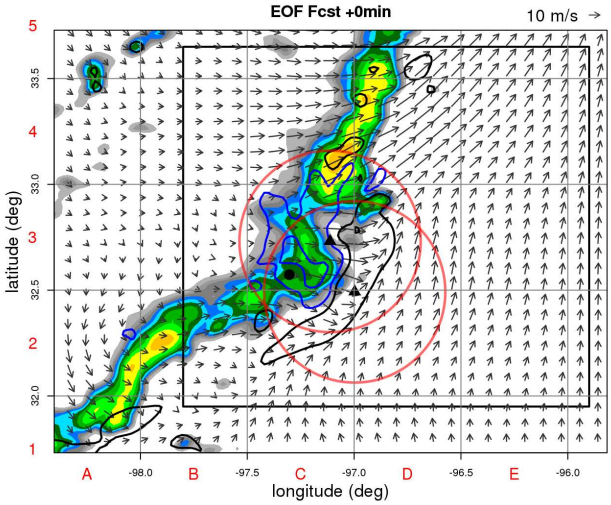
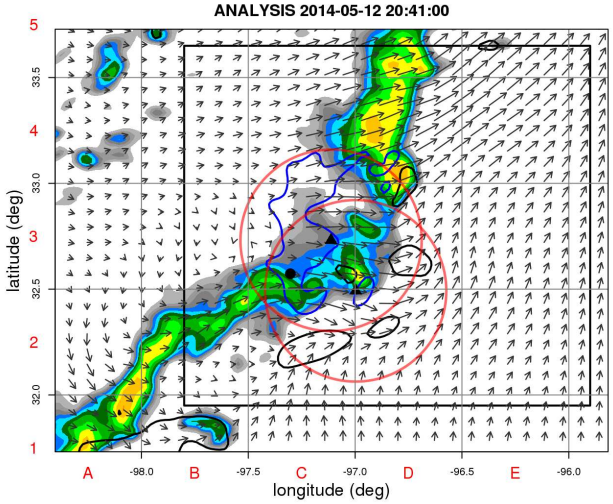


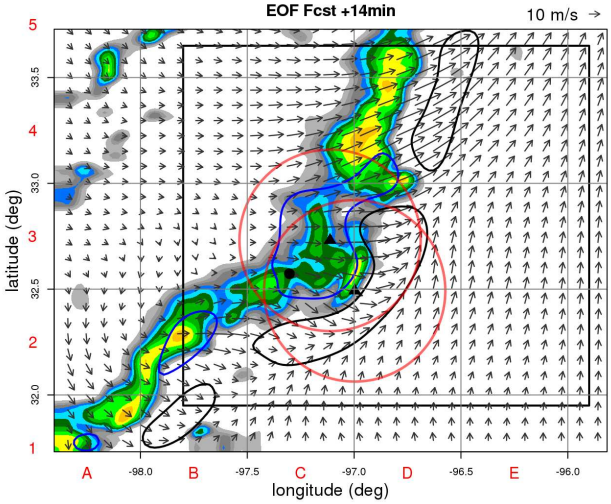
Figure 7.5: Parameters obtained after Nelder-Mead optimization over 58-minutes forecasts between 19:47 and 21:00UTC. The red line represents a spline fit overplotted to highlight the trend of the parameters with time.

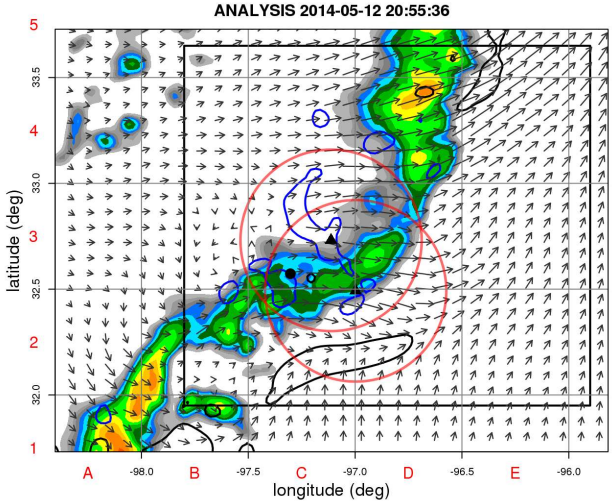
images a coarse grid with 0.5° lat/lon spacing is superimposed with letters/numbers coordinates in red to facilitate the comparison. The forecast for successive lead times are displayed in the righ column, while the left column reports the corresponding analysis. In addition to the rainwater (color palette), the analyzed and forecasted low-level (600m) winds are displayed (arrows). In the +14 min forecast the cell development associated with the hailstorm described in Chapter 6 is located fairly well just north of the Midlothian radar (lower-right triangle). This local enhancement associated with the gust front propagation is rather continuous in time in the forecast, while the analyses show a more pulsed behavior. In general the larger scale morphology is depicted reasonably well until ∼1-hour lead time. In particular the model seems able to reproduce the increasingly faster movement of the northeastern

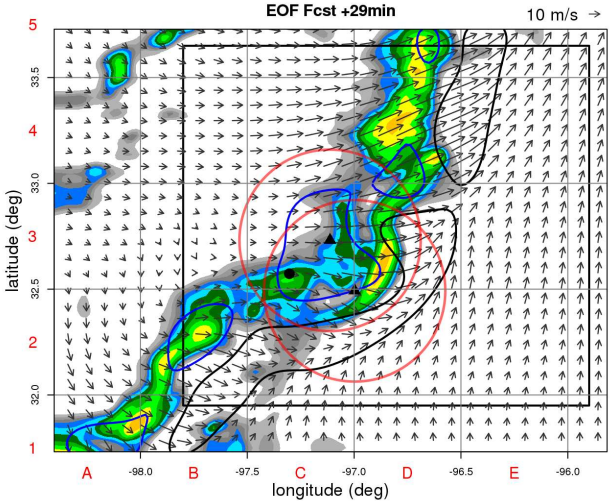

ANALYSIS 2014-05-12 20:26:24
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0


EOF Fcst +0min
10 m/s
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0


ANALYSIS 2014-05-12 20:41:00
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0


EOF Fcst +14min
10 m/s
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0


ANALYSIS 2014-05-12 20:55:36
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0


EOF Fcst +29min
10 m/s
latitude (deg)
longitude (deg)
A
B
C
D
E
-98.0
-97.5
-97.0
-96.5
-96.0

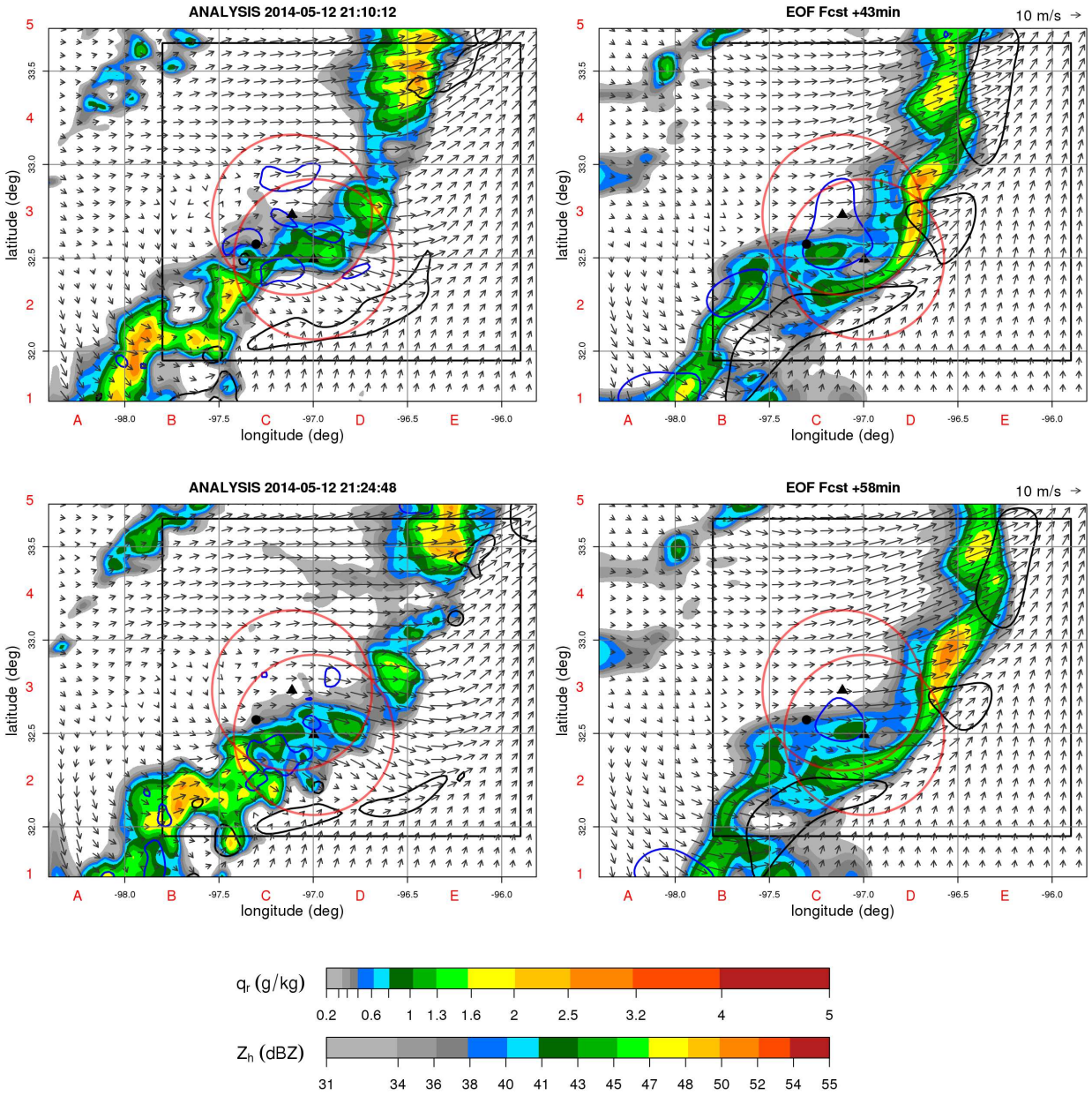


Figure 7.6: Analysis (left column) and corresponding forecasts every $\sim$ 15 min (right column) of rainwater and low-level winds starting at 20:26UTC. The black (blue) solid contour represents the 0.5 $(-0.5)\mathrm{s}^{-1}$ m km^{-1} divergence level. The domains of the X-band radars are marked with red circles, while the rectangle encloses the validation area. The color palette for reflectivity (dBZ) is defined assuming a $Z(q_r)$ relation as in [7].

portion of the storm and the broadening taking place south-southwest of the three radar, where a cyclonic circulation developed. This is more evident when the parametric model

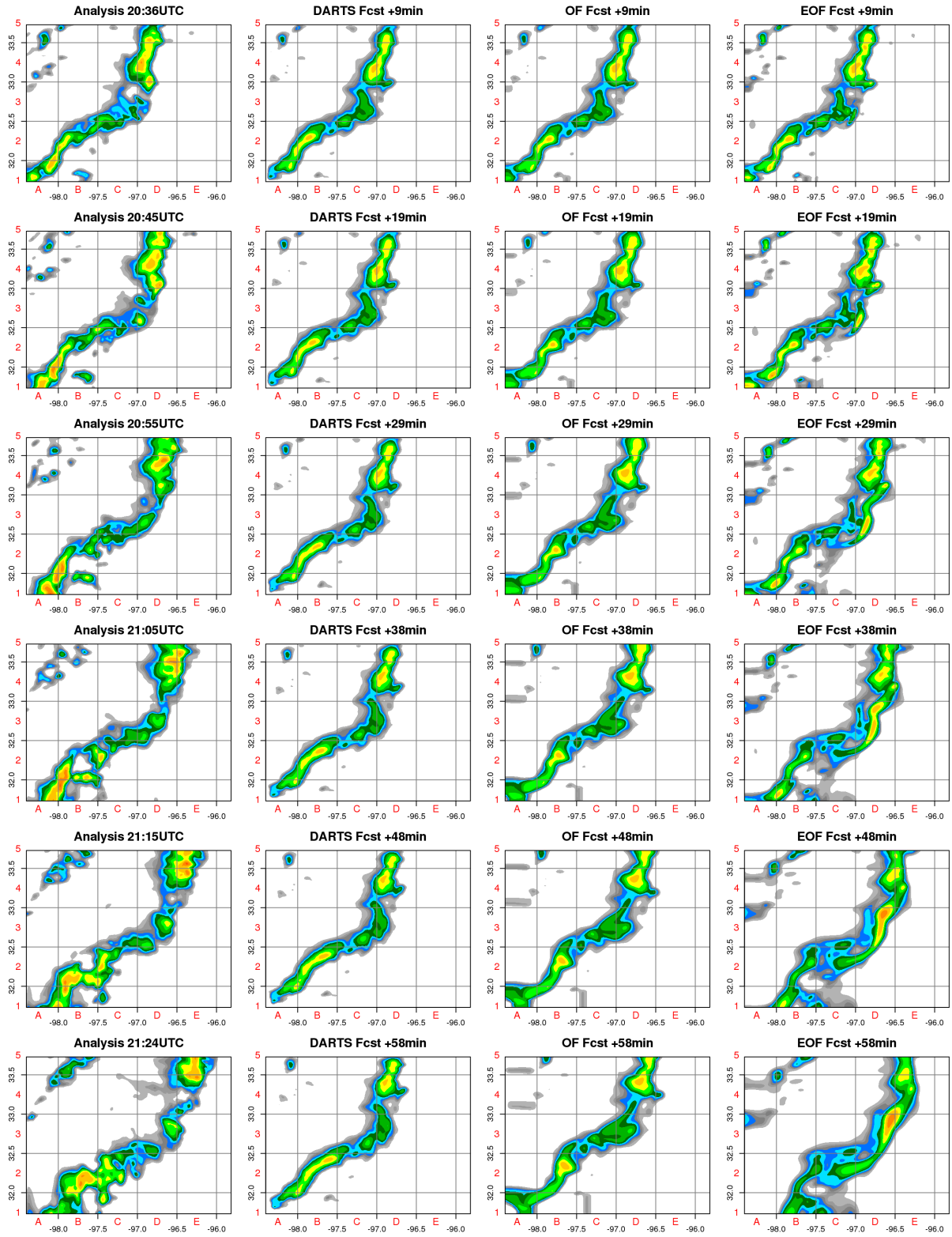


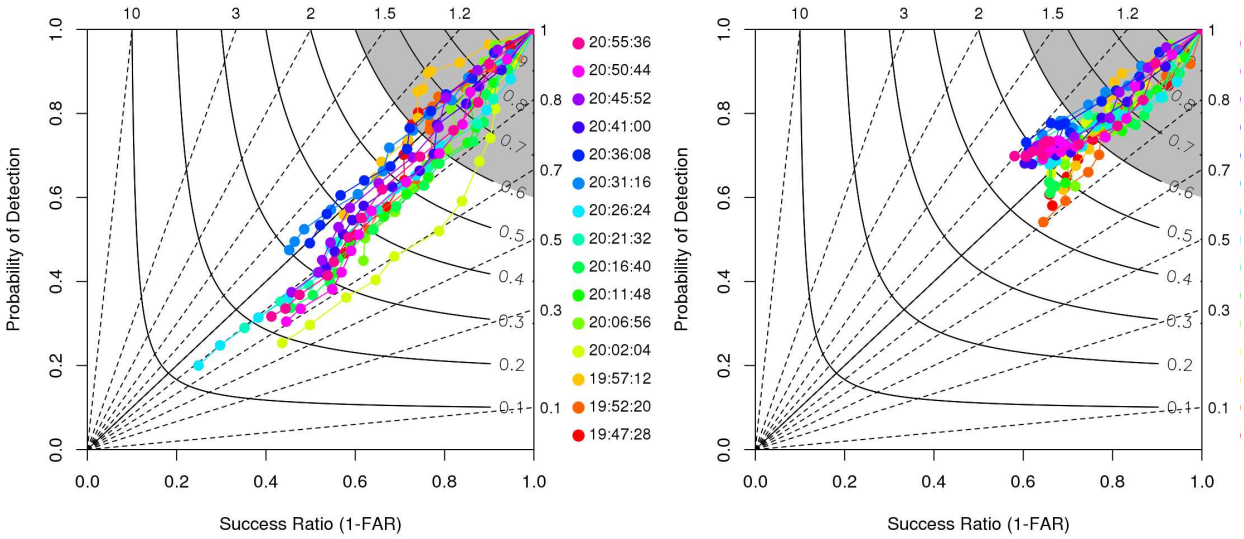
Figure 7.7: Analysis (leftmost column) and corresponding forecasts every $\sim$ 10 min using different techniques. From left to right: DARTS, standard optical flow (OF), parametric model forecast (PMF).

forecast (EOF), is compared with the image processing techniques in fig. 7.7. Both DARTS and OF tend to greatly underestimate the westward motion of the northeastern part of the storm, which appears to be caused by a combination of auto-propagation [125] and the stronger winds flowing in the region. In fact at lead time +58 min the grid points E4-E5 are still empty for DARTS and OF, while the PM forecast is in better agreement with the truth. Another relevant difference is in grid point C2. The southwestern part of the storm has a marked elongated shape at the analysis time, and the simple image processing methods tend to simply displace this pattern forward. The slower velocity respect to the rest of the storm is well captured by these techniques, as demonstrated by the corrrect position of the rear boundary of the storm. However the interaction of the southerly flow with the advancing storm is determining a broadening and instensification in this region which can not be reproduced by a simple motion vectors advection. On the other hand the parametric model is triggering new convection in this region. Although there are significant differences in magnitude and small scale organization respect to the actual analysis, the general agreement of the large scale pattern appears valuable for nowcasting applications.

From a qualitative perspective the forecasted wind field in fig. 7.6 appears reliable over much of the domain until $\sim$ 30 min lead time, with a reasonable location of the main convergence regions. For larger lead times however the actual gust front located over the overlapping region between the two X-band radars at 20:26UTC propagates faster than depicted in the forecast. At 21:24UTC the gust front in the analysis extends from C2 to E2, while in the forecast is located significantly behind (B2 to D2).

For a quantitative evaluation of the proposed method the well-known summary measures POD, TS (or CSI), FAR and bias (Appendix A) are considered. Fig. 7.8 shows the performance diagram, which allows to visualize multiple measures of forecast quality on the same

diagram [147] (Appendix A). Left panel in fig. 7.8 shows the performance of the simple optical flow (OF). The Threat Score decreases from 1 (+0 min lead time) to 0.2-0.3 (+58 min lead time) for most of the forecasts. The forecast starting at 20:21UTC has the worst performance with a TS reaching almost 0.1, while the 20:02UTC forecast shows a significant bias (0.6) after about 30 minutes into the forecast. In the right panel the same diagram for the parametric model (EOF) depicts a significantly better peformance, with a TS never reaching below 0.4 for the longest forecast. The bias is also very close to 1 for the first 30 minutes forecast and within the 0.8-1.2 range afterwards.



(a) Optic Flow (OF)

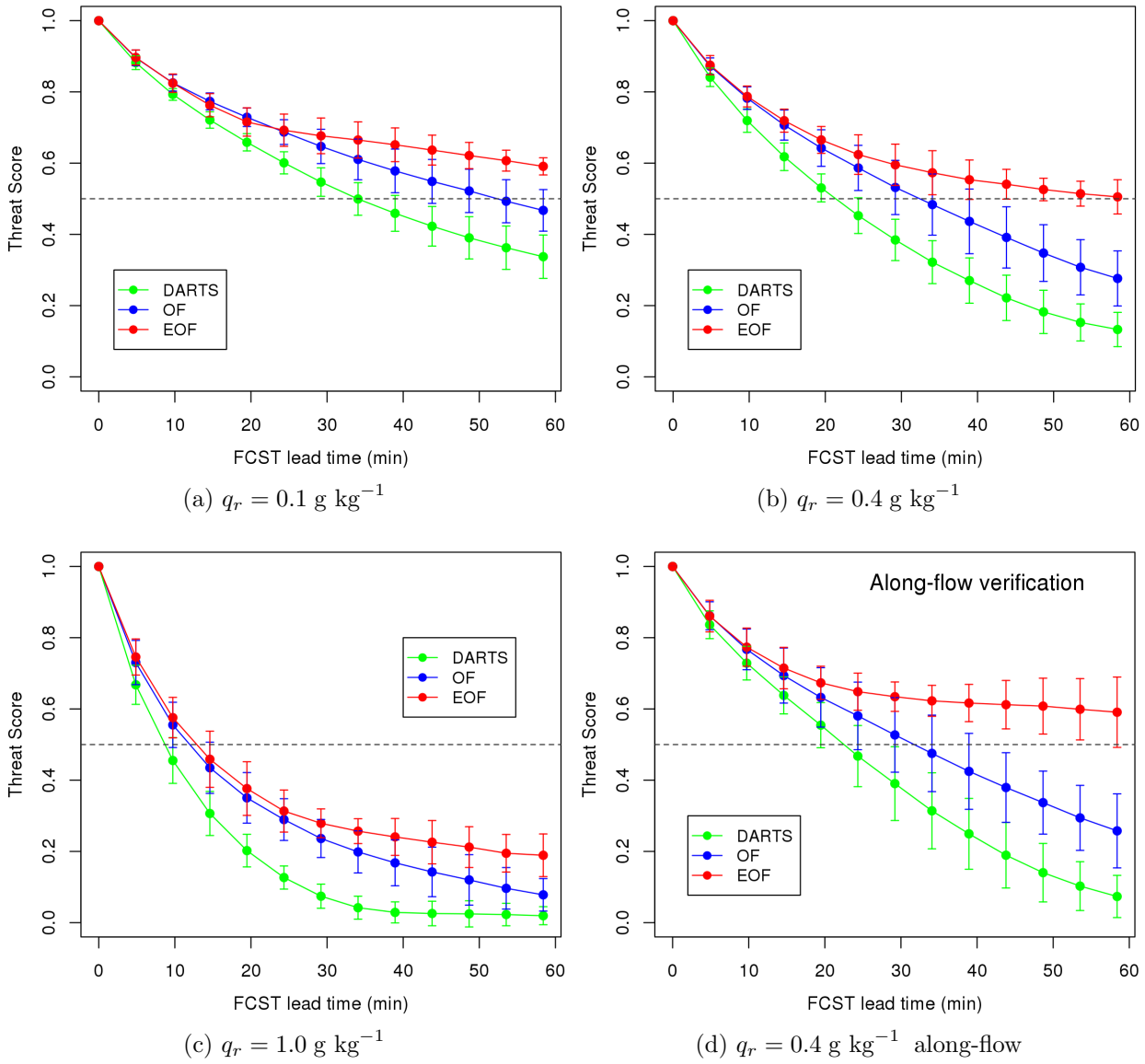
(b) Enhanced Optic Flow (EOF)

Figure 7.8: Performance diagram for q_r threshold of 0.4 g kg^{-1}, corresponding to a reflectivity of $\sim$ 36 dBZ. The colors represent the forecasts starting at the analysis time in the legend and the circles along a line indicate the successive forecast steps (dt=292s). The final circle along each line corresponds to the +58 minutes forecast.

The average TS over the different forecast runs is also summarized in fig. 7.9, which reports the results for the EOF together with the corresponding performance of DARTS and OF. Panels a), b) and c) are for three different thresholds on q_r, namely 0.1, 0.4 and 1.0 g kg^{-1}, corresponding respectively to a reflectivity of approximately 26, 36 and 43 dBZ.

DARTS was originally developed to perform extremely short term forecasts (few minutes) in order to accordingly adapt in real time the radar scan. It has not been used for longer forecast and it actually shows a poorer performance compared to both OF and EOF for this kind of application. The OF and EOF forecast show similar skill until about 20-25 minutes, after which the EOF performs better. In particular for the 0.4 g kg^{-1} threshold, the TS score for the EOF forecast lowers to 0.5 about 20 minutes later than for the OF forecast. For the other thresholds a similar improvement is also observed for the larger lead times. This kind of performance would reflect in a sizeable impact in terms of advanced warning for real time applications. The reults displayed in panel d) are for the same threshold as in panel b), i.e. 0.4 g kg^{-1}, but considering a validation area moving along the q_r flow. Specifically the verification is performed on a portion of the whole validation domain, determined by the grid points that track back (using the model motion vectors) to the region in the analysis where co-located Doppler observations from at least two radars were available. In this way it is possible to attempt sizing the impact of multiple Doppler observations on the forecast. It is in fact beneficial for the VDRAS assimilation to ingest Doppler observations from different radars, allowing an accurate retrieval of the 2-dimensional wind field. The density of the Doppler observations is advected similarly to q_r and is used to define the along-flow validation domain for every forecast lead-time. The result in fig. 7.9d shows an increased improvement upon the standard OF forecast (which did not changed substantially), corroborating the idea that a good wind analysis from multiple radars is crucial for the quality of the nowcasting.

The results illustrated so far represent the maximum achievable performance of the model for the given case study, since the results are optimized for each individual forecast using the future analyses. In a real-time application this is of course not possible, so the variability of the model parameters will have to be further analyzed using a comprehensive dataset to assess



(a) $q_r = 0.1$ g kg^{-1} (b) $q_r = 0.4$ g kg^{-1}

(c) $q_r = 1.0$ g kg^{-1} (d) $q_r = 0.4$ g kg^{-1} along-flow

Figure 7.9: Average Threat Score (19:47-21:00UTC) plotted for increasing forecast lead time and for three diferent q_r thresholds (panels a, b, c), corresponding respectively to $\sim$ 26dBZ, $\sim$ 36dBZ and $\sim$ 43dBZ. In panel d) the Threat Score is calculated for the same threshold as in b), but "along the flow", i.e. over grid points originating from regions in the analysis where Doppler observations from at least two radars were available. This moving sub-region extends over $\sim 20 - 25\%$ of the whole validation domain.

their validity for a wider range of meteorological situations. However, for this specific event it is evident from fig. 7.5 that most of the parameters do not show important variations during the event considered. As previously noted, this is likely ascribable to the specific

VDRAS assimilation technique, ensuring the proper physical consistency in space, time and among the model variables. The observed slow time change of the parameters is encouraging for a hypothetical real-time application, when one can not perform the optimization using future observations. In real-time it may be possible to evaluate the set of parameters on the previous 40-60 minutes and use it for the current forecast.

To show the impact of neglecting the parameters variation during the event, the performance using fixed parameters is evaluated. The parameters are simply set from the arithmetic average of the values resulting from the optimization (fig. 7.5), i.e.: $p_0 = +2.4 \cdot 10^{-5}$, $p_1 = -0.5 \cdot 10^{-5}$, $p_2 = -394$ s, $p_3 = -2.0 \cdot 10^{-7}$, $p_4 = +1.3$, $p_5 = +2.3$.

The performance diagram corresponding to the forecasts with fixed parameters is presented in fig. 7.10. Not surprisingly given the limited variations of the optimized parameters, these summary measures indicate a performance actually very similar to the reference in fig. 7.8, except for a more pronounced bias earlier in the forecast for some specific runs. For example, the 19:57UTC (light orange color) forecast presents a positive bias, which is attributable to the lower evaporation coefficient p_1 ($+0.5 \cdot 10^{-5}$) respect to the optimized value ($+1.0 \cdot 10^{-5}$).

For the wind verification no other nowcasting reference is available, so the parametric model results are compared with simple persistence, i.e. assuming the t_0 analysis wind does not change during the forecast. Fig. 7.11 shows the average Root Mean Square Error (RMSE) for the zonal (u) and meridional (v) wind components, for increasing forecast lead times. These are the same errors used in eq. (50) (2^{nd} and 3^{rd} term) for the optimization of the model parameters, so the clear improvement upon persistence is anticipated. The performance using fixed parameters (orange lines) is very similar, as the parameters directly affecting the wind forecast (p_3 and p_4) are relatively constant during the event (fig. 7.5).

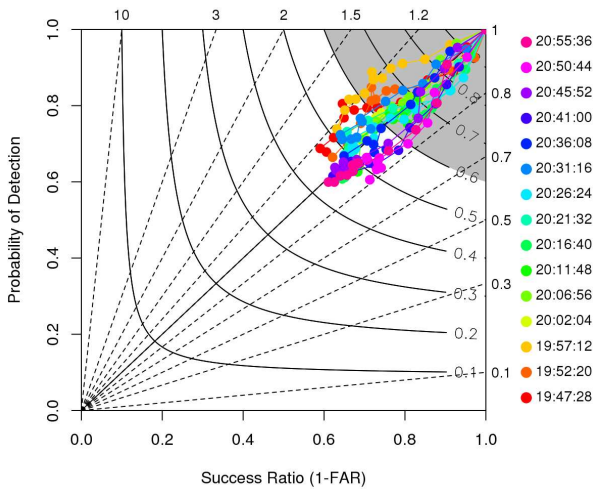


Figure 7.10: As in fig. 7.8, but using fixed parameters in the model.

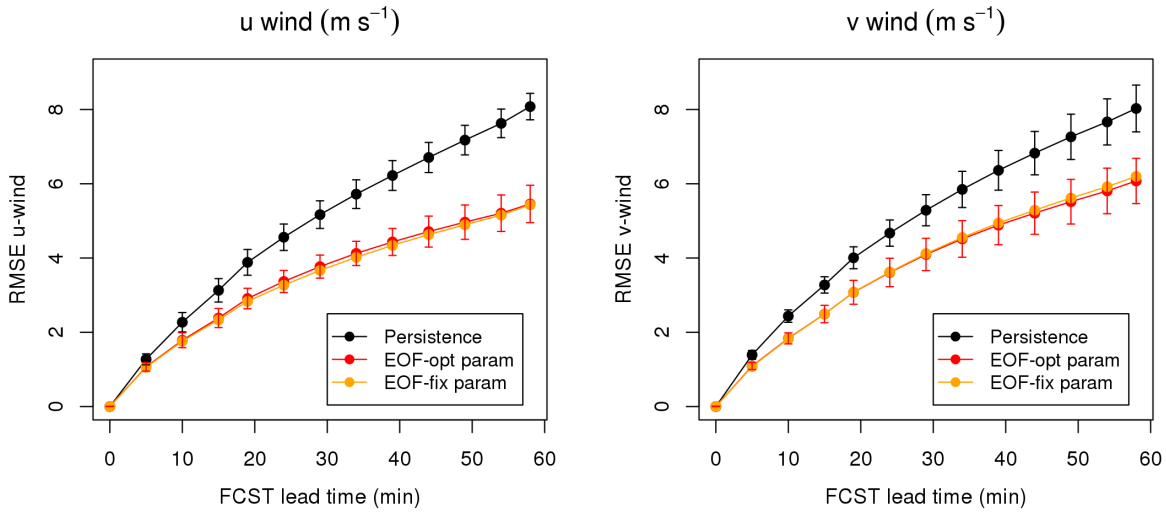


Figure 7.11: Average RMSE for zonal (u) and meridional (v) wind components for increasing forecast lead time. The black line represents the error associated with the wind prediction assuming persistence, while the red (orange) line shows the error for the extrapolation model forecast with optimized (fixed) parameters. The error bars, corresponding to $\pm\sigma$, are omitted for clarity for the orange line.

The improvement of the wind components (u,v) forecast accuracy also reflects on the wind direction, and to a lesser degree on the wind velocity (fig. 7.12). In precentage the

relative improvement for the wind direction is over 30% at ∼60 min lead time, while only 10-15% for the wind velocity.

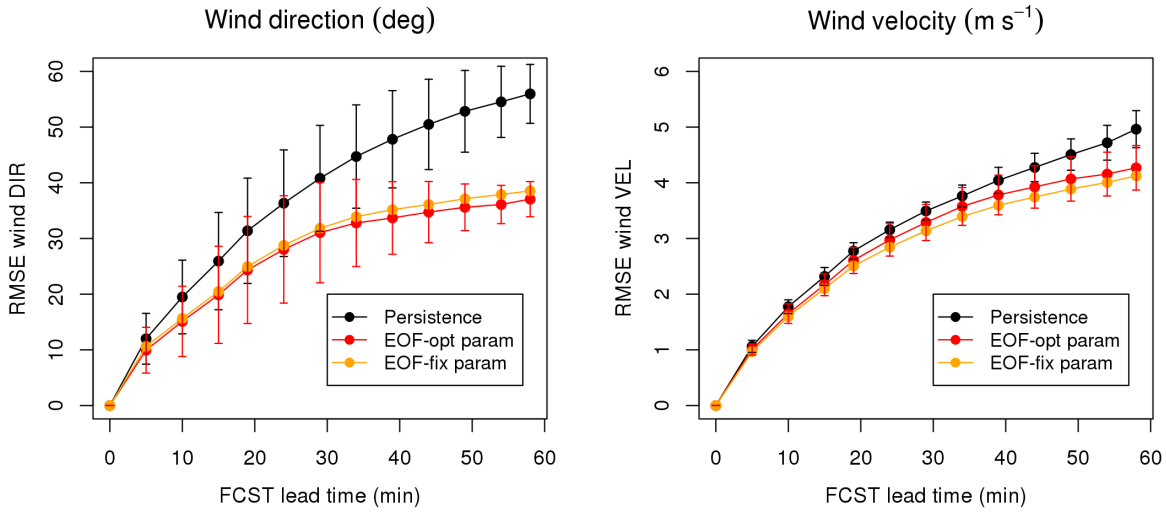


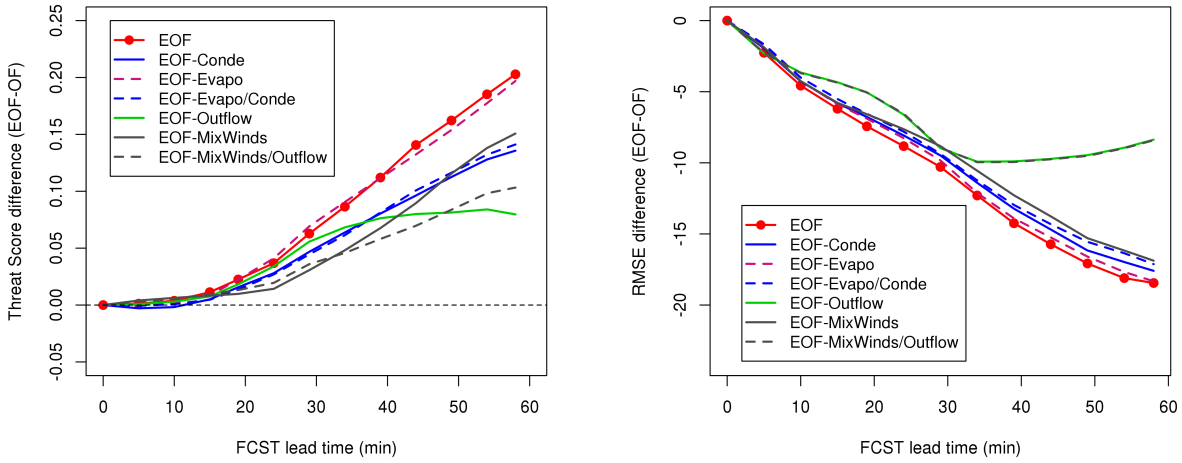
Figure 7.12: Same as fig. 7.11, but for wind direction (left) and intensity (right).

In order to assess the relative impact of the individual terms in the parametric model, a number of experiments is performed by selectively suppressing some of the processes. This is realized by setting to zero the parameter(s) controlling a given process and re-running the optimization procedure on a reduced number of parameters. In this way, although the subset of the remaining parameters may converge to different values respect to the full model configuration, the summary statistical measures allows to evaluate the relevance to the nowcasting of the single components of the model. Fig. 7.13 reports the results for the performance of rainwater nowcasting (left panel) and wind direction (right panel). The impact on the wind velocity is not considered because the improvement obtained with the full model (fig. 7.12) is quite small (the difference between the RMSE of Persistence and EOF is within $\pm\sigma$), and the even smaller differences achievable with the partial model can not be reliably evaluated.

Instead of reporting the TS as in fig. 7.9 for q_r, in the left panel of fig. 7.13 the difference between the EOF and OF Threat Score for the 0.4 g kg^{-1} threshold (i.e. the improvement upon simple optical flow) is displayed to better appreciate the smaller differences. The solid red line with circles is the difference between the Threat Score of EOF and OF as in fig. 7.9b and represents the reference performance with the full model. The lower the TS difference for the other experiments, the more relevant the process removed from the forecast.

It is clear that the removal of the evaporation process alone (dashed pink line) has a negligible impact on the Threat Score, as previously anticipated. The bias (not shown) remains very close to 1 because the reduced condensation parameter p_0, close to zero during the first part of the event, compensates the lack of evaporation. On the other hand, removing the condensation process (alone or together with the evaporation, solid and dashed blue line respectively) has a sizeable effect on the overall performance of the rainwater forecast. For the wind forecast (right panel in fig. 7.13), evaporation and condensation have little impact. This is somehow expected, since these processes directly affect the q_r distribution and only secondarily the wind field through the ouflow parametrization. This last process (outflow) is indeed the most relevant for the wind field, since it explicitly influences the flow around the cold pool region below the storm core. Less obviously the outflow modeling is also very important for the rainwater nowcasting. This highlights that a proper forecast of the gust front propagation is essential to ascertain the location where the storm enhancement or new convection development will occur. In the proposed approach, this aim is pursued through 1) the forecast of the divergence field using the optical flow technique and 2) the parameterization (parameters p_3 and p_4) of the outflow below the precipitation cores. The last parameter considered in the model, the mix-winds weighting factor p_5, has also a relevant impact on the rainwater forecasting skill. Compared to the outflow modeling, the update of

the rainwater motion vectors using the low-level winds seems to have a larger impact earlier in the forecast. In fact the solid gray line in fig. 7.13 starts deviating from the red line around 20 minutes lead time and then remains almost parallel to the reference performance. On the other hand the forecast without the outflow parameterization becomes increasingly worst than the reference after 30 minutes lead time. The solid gray (mix winds) and green (outflow) lines intersect around 40 minutes lead time. This indicates that the outflow representation has indeed the longer lasting effect on the quality of the forecast.



(a) q_r: TS difference (q_r threshold = 0.4 g kg^{-1}) (b) Wind Direction: RMSE difference (°)

Figure 7.13: Performance of the forecast runs suppressing selected processes. The red line with circles indicate the reference performance of the full model as in fig. 7.9b, relative to the Optical Flow forecast. The other lines represent the forecasts without given processes (the name in the legend indicates the process suppressed).

7.4. Conclusions

The basic idea behind the nowcasting approach discussed in this section is to extend the Lagrangian persistence concept, traditionally applied to reflectivity only (or related quantities such as the rainfall rate, rainwater,..), also to the components of the wind field, namely divergence and vorticity. This clearly presumes the availability of a reliable wind analysis

consistent with the radar observations in terms of coverage, spatial and temporal resolution. In addition, some basic interactions between the wind and the rainwater advection have to be represented in the nowcasting model to allow the forecasted variables to evolve coherently, at least for short lead times.

The motivation for this work originates in part from the frustration of not being able in many instances to fully exploit the large amount of information provided by a Doppler weather radar. In the large majority of radar nowcasting applications nowadays reflectivity is still the only observation being contemplated. While methods including dual-polarization variables have been proposed to improve the diagnosis and forecast of rainfall [148], the use of the Doppler information is in general largely unexploited.

The reasons are partly residing in the objective technical and conceptual difficulties, in particular when only single radar Doppler measurements are available. In addition to the advantages for the detection and quantitative estimation of precipitation, a dense networked radar system like the Dallas-Fort Worth test bed represents a unique tool to accurately sample the winds near the ground. Overlapping coverage allows each point in the three-dimensional network domain to be simultaneously viewed by two or more radars, allowing for multiple-Doppler wind vector retrievals. Several geometric techniques for multiple-Doppler retrieval exist and can provide wind retrievals for various applications. However, only using a numerical model for the assimilation and analysis allows to properly retain the physical consistency between the winds and the precipitation fields.

As a matter of fact the radar Doppler observations have been traditionally exploited in numerical models, being "naturally" prone to be ingested into data assimilation systems. The Doppler wind observations have in fact notable advantages. Radar wind measurements are immune to biases, while reflectivity are prone to calibration errors. In addition the wind

is defined continuously all over the model domain (unlike rainwater, which is subject to the "on-off" switches in the warm rain process), facilitating the assimilation of the radial wind component.

The four-dimensional assimilation approach realized in VDRAS and described in Chapter 6 is specifically devised for the assimilation of weather radar data and is able to largely retain the observation component in the final analysis. However, while NWP data assimilation systems can provide fairly accurate and complete analyses, the model forecasts are still beaten by extrapolation techniques when it comes to very short-term prediction. It appears then reasonable to attempt taking the best of the two approaches, i.e. the robustness and consistency of data assimilation for the analysis and the simplicity of extrapolation techniques for the subsequent nowcast. The method illustrated in this chapter represents a preliminary attempt to realize this idea.

The VDRAS data assimilation system is used to prepare the initial analyses (time t_0) of wind and rainwater. This ensures the proper environment for the subsequent forecast. A simple two-level, three-variables model (rainwater and two wind components) contemplating vertical advection of rainwater, condensation, evaporation and a simplified representation of the outflow originating from downdraft regions is constructed and expressed in terms of six adaptive parameters. While the local rate of change is determined according to these basic physical processes, the variables are advected using motion vectors established after application of an optical flow method on the previous observations. Despite a certain naivety of the basic model construction, the preliminary results seem to indicate an ability to generate a positive impact on the nowcast accuracy, in comparison with standard extrapolation methods. One peculiar aspect of this technique is the use of a parametric model, with parameters

adaptable to the specific meteorological situation. For this preliminary application the parameters have been determined a posteriori using an optimization method (Nelder-Mead), based on the deviation of the forecast from the corresponding analysis. However it seems possible to devise a real-time application where the parameters are evolving based on the past observations.

The achieved superior performance of the EOF parametric model compared to standard optical flow methods is ultimately ascribable to the inclusion of the radar Doppler wind observations in the VDARS analysis, which allows to reliably represent the wind field around and ahead of the storm. This favors the representation of the condensation process leading to local storm enhancement, and the evolution of the wind field in the outflow region. On the other hand, the prediction of the local dissipation appears more difficult, and the simple evaporation process currently included did not contributed to the forecast skill. A possible improvement, to be investigated in the future, may come from the inclusion of the cloud water q_c from the VDRAS analysis, allowing application of the Kessler parameterization of evaporation (eq. 40).

Apart from the actual chances of this substantially heuristic model to find its place as a nowcasting operationsl tool, the overall good performance documented in this study may also be interpreted as a demonstartion of the accuracy and potential of the VDRAS wind analysis system. With this respect it is also interesting to compare the forecasting skill of the EOF model depending on the initial VDRAS analysis setting (Chapter 6). Fig. 7.14 shows the average Threat Score obtained after running the model with optimization, starting from a first analysis with assimilation of single NEXRAD radar (dark green line) and a second analysis using multiple radars (NEXRAD+XDFW, red line). In order to perform a fair comparison, given two different initial analyses, the forecasts are verified against the simple

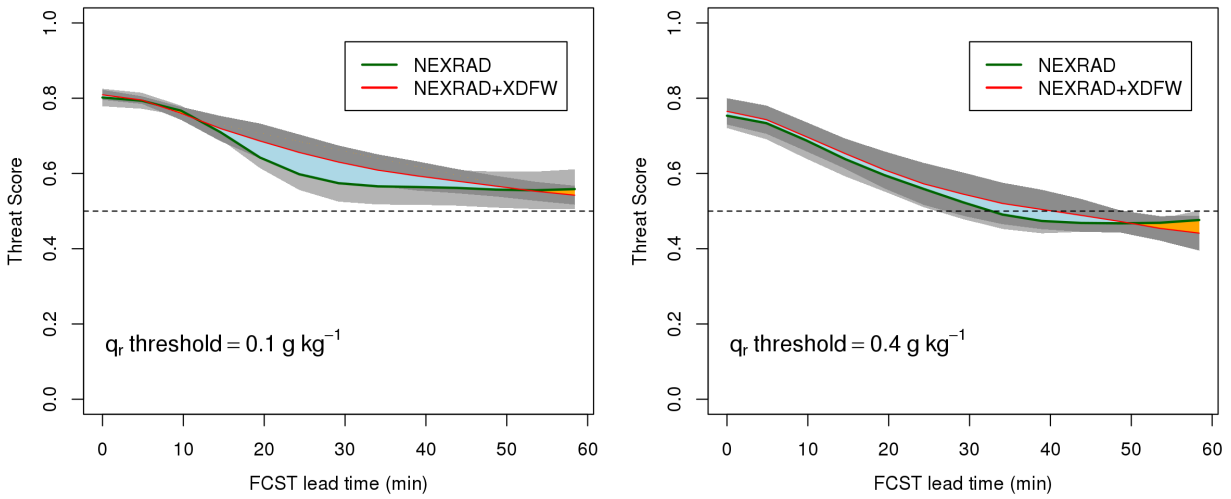


Figure 7.14: Threat score calculated for two different q_r thresholds (0.1 and 0.4 g kg^{-1} respectively in left and right panel), for the forecast based on assimilation of single NEXRAD radar (dark green line) and both NEXRAD and the two X-band radars (red line). The gray area represents $\pm$ 1 σ (standard deviation), with light gray for NEXRAD and darker gray for NEXRAD+XDFW. The light blue (orange) area denotes an improvement (worsening) of the forecasting skill when the X-band observations are used in the VDRAS analysis.

mosaicked rainwater field (the first guess from which the analysis is eventually obtained after the assimilation process). For this reason the initial (+0 min forecast) value of the Threat Score is less than one. The differences highlighted by the light blue area seem to indicate a small improvement of the forecast skill until ∼50 min lead time when the X-band observations are included in the analysis. This appears coherent with the overall qualitative evaluation of the analysis in Chapter 6. However, the improvement is objectively limited and relatively small when compared to the variability of the score (gray area) during the event.

As already noted, the portion of the model domain with overlapping Doppler observatons is quite limited, given that only two X-band systems were available for the event considered. For an overall quantitative evaluation and possible confirmation of the added value arising

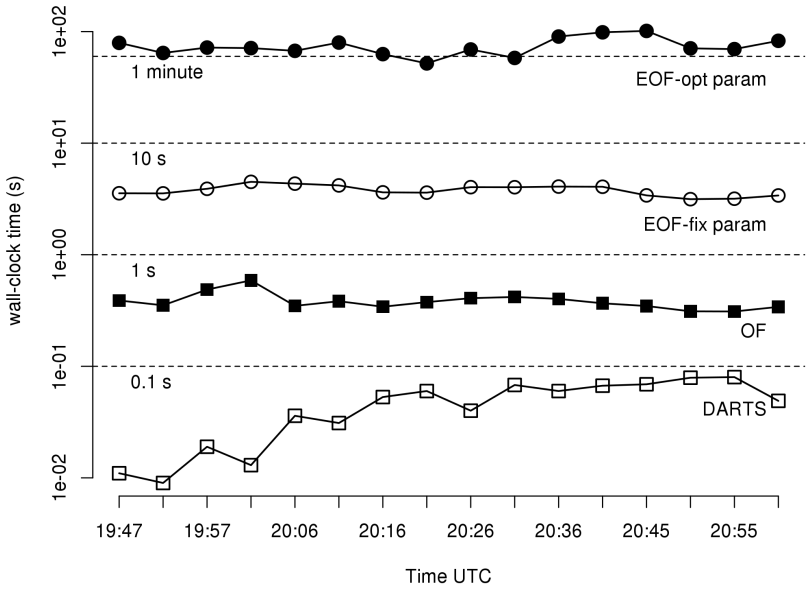


Figure 7.15: Wall-clock execution times in seconds for different nowcasting methods. Parametric model with optimized coefficients (filled circles), parametric model with fixed coefficients (empty circles), Optical flow (filled squares), DARTS (empty squares). The y-axis in in logarithmic units.

from the assimilation of X-band observations, new cases will need to be considered when the complete 8 radar DFW network will be completed.

In terms of computational requirements, this method is more demanding compared to extrapolation techniques based on optical flow techniques. Fig. 7.15 shows a comparison between the wall-clock run times for the different techniques during the case study considered. From bottom to top the time required to complete a one-hour forecast goes from the order of 0.1 seconds or less for DARTS to about 1-2 minutes for the parametric model including optimization of the parameters (EOF-opt param). When using fixed parameters (EOF-fix param) the run time is about 5 seconds. These performances refer to the domain with $N_{tot} = 13664$ grid points ($NX = 122$, $NY = 112$) and 2 km spatial resolution, running on a laptop hosting a Intel $i7$ CPU M 620 @2.67GHz. The computational complexity is of the

order of $O(N_{tot})$ so doubling the spatial resolution would result in about four times longer run times. For the parametric model method, however, the time required to elaborate the wind analysis (VDRAS) has to be added. The MPI (Message Passing Interface) version of VDRAS was used in this study, running on a parallel computer using 16 processors. For the setting described in Chapter 6 the time required to complete an analysis with a three minutes assimilation window is about five minutes. Given the focus on the low-level wind analysis it seems plausible to further decrease the needed run time, adopting a modified configuration with fewer vertical levels. These overall computational time requirements appear compatible with possible real-time implementation.

CHAPTER 8

Summary and Conclusions

The research activities pursued in this thesis work had the common goal of investigating the potential of Doppler and dual-polarization radar observations for microphysics retrievals and application to nowcasting, i.e. 0-2 hours short-term forecast. Although nowcasting is most widely associated with convective storms, stratiform precipitation systems may also have relevant short-term impacts. In particular during the winter season, stratiform systems associated with mid-tropospheric mesoscale forcing can lead to large snowfall accumulations at the ground affecting infrastructures, ground transportation and aviation. The first part of this study then focused on this type of events, considering a widely used numerical atmospheric model for the assimilation of radar observations (Chapter 2), and analyzing the dual-polarization radar signatures in the ice region of precipitation systems (Chapter 3). The three-dimensional assimilation scheme developed within the Weather Research and Forecasting (WRF) model (WRF 3D-Var) is used to evaluate the impact of the assimilation of Doppler radar observations for a case study in northern Italy. As already evidenced in previous studies, the assimilation of Doppler velocity has in general the larger and consistent impact on the quality of the analysis and the subsequent forecast, as compared with reflectivity. The reasons for this may be attributed to the fact that the core of the atmospheric model is the dynamics, i.e. the compressible, non-hydrostatic Euler equations, including conservation of mass (continuity), momentum, and energy. So the benefit introduced by an adjustment of the wind field using radar velocity observations may easily propagate to the other model state variables. Secondarily, reflectivity is assimilated through a non linear observation operator that relates its intensity to the amount of rainwater in the model. This

is currently realized in WRF 3D-Var using a partitioning of the total water mixing ratio implemented by a warm-rain process [22], including condensation, evaporation, accretion (collection of cloud droplets by precipitation) and auto-conversion (conversion of cloud into precipitation by coalescence of cloud droplets).

The lack of the ice phase inclusion in the current WRF 3D-Var assimilation system may certainly limit to some extent the potential benefit of the radar data assimilation. This explains the efforts undertaken over the past few years to investigate the information content inherent in dual-polarization observations in the ice phase and to develop suitable forward operators for ice processes [149]. In fact, while the strengths of dual-polarization radar observations have been widely demonstrated for rainfall, it is less clear how much information can be inferred from observations in ice regions. In this respect, the observational study presented in Chapter 3 focused on the characterization of the radar dual-polarization signatures in the ice region of stratiform precipitating clouds. This study, supported by a large statistical analysis, revealed a significant connection between dual-polarization signatures aloft and the precipitation intensity near the surface. In particular, the high specific differential phase shift (K_{dp}) observations in the ice portion of stratiform clouds are explained by the vapor depositional growth of dendritic crystals in weak ascending flow, associated with mid-tropospheric mesoscale forcing. It is argued that the improved understanding of the information content of dual-polarization radar observations and their relationship to model physics will contribute to future microphysically consistent assimilation approaches for both liquid and ice water content.

The purpose of establishing a more general framework for the microphysical interpretation of radar dual-polarization measurements led to the development of an improved method for the classification of hydrometeors from radar echoes (Capter 4). The proposed technique

presents important novelties with respect to the state-of-the-art radar classification methods. The traditional bin-based classification is replaced with a semi-supervised approach which combines cluster analysis, spatial contiguity, and statistical inference to assign the most likely class to a set of identified connected regions. The main salient points include a degree of self-adaptiveness (through the unsupervised intermediate cluster analysis), a noiseless classification (by means of spatial contiguity), and the inclusion of microphysical constraints to improve the consistency of the hydrometeors spatial distribution. This method, in addition to a smoother and easier to interpret classification, ensures a lower sensitivity to noise in the observations and also to small biases in the differential reflectivity measurements.

For convective precipitation, the connection between dual-polarization measurements and the storm dynamics is initially considered analyzing a well known polarimetric signature, the columns of differential reflectivity. The observation of vertically contiguous positive Z_{dr} values extending between the lower layers and above the environmental 0° level is extensively documented and associated with the presence of a well developed updraft capable of lifting oblate liquid water and partially frozen hydrometeors [116]. The estimation of the vertical velocity from the Z_{dr} columns is applied on a case study to demonstrate the potential for improving the analysis of the low-level wind field, in conjunction with surface observations. The preliminary results indicate a general consistency of the retrieved wind analysis with the temperature spatial distribution and the evolution of the precipitation.

In parallel with this heuristic approach for the wind field estimation in the storm environment, data assimilation of dual-polarization observations is investigated. The assimilation of radar observations for convective storms, as opposed to stratiform systems, presents specific challenges. In particular, some serious issue may arise if the background (a model state used as first guess) has no convection whereas the radar indicates precipitation, or vice versa.

Sugimoto et al. (2009) suggested to use a cloud analysis procedure to modify the background and mitigate huge departures from the observations. In practice, some model variables are modified based on the observed reflectivity before performing the 3D-Var assimilation. However, a difficulty with this approach is that the radar only observes quantities related to the rainwater content and winds, while it gives no information about temperature and relative humidity.

Another approach is to use a four-dimensional data assimilation system with the purpose of fitting the observations to a model trajectory as close as possible over a given time window, while at the same time retrieving the unobserved variables. The Variational Doppler Radar Assimilation System (VDRAS) [7] is an advanced data assimilation system specifically designed for the assimilation of Doppler radar observations. Its potential for the assimilation of small short wavelength radars (X-band) with dual-polarization capabilities is investigated in this work, over the Dallas-Fort Worth urban demonstration testbed. Recent attempts to assimilate the radar dual-polarization observations have been shown to improve the performance of the short-term precipitation forecast [150]. In their work Li and Mecikalski [150] have defined Z_{dr} and K_{dp} forward model operators according to the well assessed dual-polarization rainfall estimation algorithms described in [25]. In this work, the reverse approach is taken, partially shifting the assimilation problem from the model domain (observation operators) to the pre-processing of the radar observations (2.1). In fact, prior to the assimilation, the dual-polarization measurements are processed in order to correct for attenuation and differential attenuation, which may have a relevant impact for high frequency radars (f $\gg$ 3 GHz), and the hydrometeor classification is used as a basis for the application of a blended dual-polarization algorithm for the estimation of the liquid water content. The results of the VDRAS assimilation have shown the potential for a detailed and consistent

wind analysis when overlapping small radars are included in the assimilation process, in addition to the conventional long-range systems. Although for the event considered in this work only the first two radars were available, a significant improvement for the precipitation and wind analysis can be anticipated upon completion of the X-band network.

The final part of the thesis work focused on the development and assessment of a simple parametric model for nowcasting, starting from the VDRAS rainwater and wind analysis. While the standard approach of variational data assimilation (implying fitting the observations to a physical model) is maintained for the diagnostic phase (analysis), the approach is substantially reversed for the prognostic phase. A very simple, partially heuristic parametric model is built considering only the relevant processes for convection, with parameters dynamically tuned through an optimization process over a defined temporal period. The whole procedure is then split in two parts:

(1) the analysis phase, which is based on variational assimilation to ensure proper initial conditions and physical consistency among the diagnosed variables;
(2) the forecasting phase, which relies on extrapolation techniques to propagate into the future the past filtered observations, assuming that the optimized relevant interconnections among the model variables keep their validity as the storm evolves (persistency of the model setting).

The preliminary results are encouraging, having shown a clear improvement upon standard nowcasting by extrapolation techniques for the prediction of a squall line evolution in Texas. This approach also appears naturally prone to incorporate additional information from radar observations, like the dual-polarization signature related to the updraft location and intensity discussed in Chapter 5.

In conclusion, the research activities illustrated in this thesis have examined the potential of dual-polarization measurements for the microphysical characterization of precipitation, winds and their short-term prediction. The study focused in particular on some specific observational features and on the use of state-of-the-art radar data assimilation techniques. The main achievements, expected impact, and future developments are summarized as follows:

- Characterization of dual-polarization observations in the ice phase
  - This study contributes to the ongoing efforts towards a better understanding of the polarimetric information content for solid hydrometeors and their relation with model physics. The improved description of these relationships will benefit the development of enhanced radar data assimilation methodologies for ice-phase microphysical processes.
- Semi-supervised hydrometeor classification method
  - This technique aims at extracting and synthesize as much microphysical information as possible from the set of polarimetric observations. It is considered a necessary preliminary step for future applications ranging from data assimilation to hail nowcasting. Future developments will include extension to four dimensions (3D volume scans and time), and a solution to explicitly deal with the interconnection between attenuation correction and the classification.
- Assimilation of X-band dual-polarization observations in a four-dimensional variational system (VDRAS)
  - The first evaluation in the context of the developing X-band radar network in the Dallas-Fort Worth metroplex demonstrated the effectiveness for improving

the low-level analysis, specifically for the wind field. Currently undergoing improvements in the VDRAS assimilation process (ice phase microphysics, evolving boundary conditions) will be included in future experiments. At the same time, the contribution of the complete eight radar network sampling the lower atmospheric layer with unprecedented detail will need to be carefully evaluated, in particular for the near-surface retrievals and short-term forecast.

- Parametric adaptive model for precipitation and wind nowcasting
  - The proposed hybrid approach extends the conventional extrapolation-based nowcasting by including the wind analysis and basic modeling of relevant physical processes. The improved performance initially demonstrated by this methodology in comparison with conventional nowcasting techniques will need to be confirmed with an extensive analysis including a range of case studies representative of different meteorological situations. This will allow to asses the effective potential of this method as an operational nowcasting tool.

Bibliography

[1] B. W. Golding, "Nimrod: a system for generating automated very short range forecasts," *Meteorological Applications*, vol. 5, pp. 1–16, 1998.

[2] T. W. Schlatter, "Variational assimilation of meteorological observations in the lower atmosphere: A tutorial on how it works," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 62, no. 12, pp. 2147–2167, 2000.

[3] V. N. Bringi, T. D. Keenan, and V. Chandrasekar, "Correcting C-band radar reflectivity and differential reflectivity data for rain attenuation: A self-consistent method with constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 1906–1915, 2001.

[4] H. Liu and V. Chandrasekar, "Classification of hydrometeors based on polarimetric radar measurements: development of fuzzy logic and neuro-fuzzy systems, and in situ verification," *Journal of Atmospheric and Oceanic Technologies*, vol. 17, pp. 140–164, 2000.

[5] E. Ruzanski, V. Chandrasekar, and Y. Wang, "The CASA nowcasting system," *J. Atmos. Oceanic Technol.*, vol. 28, pp. 640–655, 2011.

[6] R. Bechini and V. Chandrasekar, "A semi-supervised robust hydrometeor classification method for dual-polarization radar applications," *J. Atmos. Oceanic Technol.*, vol. 32, pp. 22–47, 2015.

[7] J. Sun and N. A. Crook, "Dynamical and microphysical retrieval from Doppler radar observations using a cloud model and its adjoint. Part I: Model development and simulated data experiments," *Journal of the Atmospheric Sciences*, vol. 54, pp. 1642–1661, 1997.

[8] N. A. Crook and J. Sun, "Analysis and forecasting of the low-level wind during the Sydney 2000 forecast demonstration project," *Wea. Forecasting*, vol. 19, pp. 151–167, 2004.

[9] E. Lorenz, "Three approaches to atmospheric predictability," *Bulletin of the American Meteorological Society*, vol. 50, pp. 345–349, 1969.

[10] E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 2003.

[11] J. Sun, "Convective-scale assimilation of radar data: progress and challenges," *Quarterly Journal of the Royal Meteorological Society*, vol. 131, no. 613, pp. 3439–3463, 2005.

[12] S. Sugimoto, N. A. Crook, J. Sun, Q. Xiao, and D. M. Barker, "An examination of WRF 3DVAR radar data assimilation on its capability in retrieving unobserved variables and forecasting precipitation through observing system simulation experiments," *Monthly Weather Review*, vol. 137, pp. 4011–4029, 1992.

[13] H. Wang, J. Sun, X. Zhang, X.-Y. Huang, and T. Aulign, "Radar data assimilation with WRF 4D-Var. Part I: System development and preliminary testing," *Monthly Weather Review*, vol. 141, pp. 2224–2244, 2013.

[14] X.-Y. Huang, X. Yang, N. Gustafsson, K. S. Mogensen, and M. Lindskog, "Four-dimensional variational data assimilation for a limited area model," *HIRLAM Technical Report 57*, 2002.

[15] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, "A description of the advanced research WRF version 2," *NCAR Tech. Note NCAR/TN-4681STR*, p. 88, 2005.

[16] D. M. Barker, W. Huang, Y.-R. Guo, A. Bourgeois, and X. N. Xiao, "A three-dimensional variational data assimilation system for MM5: Implementation and initial results," *Monthly Weather Review*, vol. 132, pp. 897–914, 2004.

[17] X. Huang, Q. Xiao, D. Barker, X. Zhang, J. Michalakes, W. Huang, T. Henderson, J. Bray, Y. Chen, Z. Ma, J. Dudhia, Y. Guo, X. Zhang, D. Won, H. Lin, , and Y. Kuo, "Four-dimensional variational data assimilation for WRF: Formulation and preliminary results," *Monthly Weather Review*, vol. 137, pp. 299–314, 2004.

[18] D. Parrish and J. Derber, "The National Meteorological Center's spectral statistical-interpolation analysis system," *Monthly Weather Review*, vol. 120, pp. 1747–1763, 1992.

[19] R. Bechini and R. Cremonini, "The weather radar system of north-western Italy: an advanced tool for meteorological surveillance," in *Proc. 2nd European Conference on Radar in Meteorology and Hydrology, Delft, The Netherlands, 1822 November 2002*, pp. 400–404, 2002.

[20] B. E. Mapes and J. Lin, "Doppler radar observations of mesoscale wind divergence in regions of tropical convection," *Monthly Weather Review*, vol. 133, pp. 1808–1824, 2005.

[21] P. Davini, R. Bechini, R. Cremonini, and C. Cassardo, "Radar-based analysis of convective storms over northwestern Italy," *Atmosphere*, vol. 3, pp. 33–58, 2012.

[22] J. Dudhia, "Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model," *Journal of the Atmospheric Sciences*, vol. 46, pp. 3077–3107, 1989.

[23] P. Tabary, G. Scialom, and U. Germann, "Real-time retrieval of the wind from aliased velocities measured by doppler radars," *Journal of Atmospheric and Oceanic Technologies*, vol. 18, pp. 875–882, 2001.

[24] H. Wang, J. Sun, S. Fan, and X.-Y. Huang, "Indirect assimilation of radar reflectivity with WRF 3D-Var and its impact on prediction of four summertime convective events," *J. Appl. Meteor. Climatol.*, vol. 52, pp. 889–902, 2013.

[25] V. N. Bringi and V. Chandrasekar, *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge University Press, 2001.

[26] M. J. Bader, S. A. Clough, and G. P. Cox, "Aircraft and dual polarization radar observations of hydrometeors in light stratiform precipitation," *Q. J. Roy. Meteor. Soc.*, vol. 113, pp. 491–515, 1987.

[27] M. Wolde and G. Vali, "Polarimetric signatures from ice crystals observed at 95 GHz in winter clouds. Part I: Dependence on crystal form," *J. Atmos. Sci.*, vol. 58, pp. 828–841, 2001.

[28] M. Wolde and G. Vali, "Polarimetric signatures from ice crystals observed at 95 GHz in winter clouds. Part II: Frequencies of occurrence," *J. Atmos. Sci.*, vol. 58, pp. 842–849, 2001.

[29] R. J. Hogan, P. R. Field, A. J. Illingworth, R. J. Cotton, and T. W. Choularton, "Properties of embedded convection in warm-frontal mixed-phase cloud from aircraft and polarimetric radar," *Quart. J. Roy. Meteor. Soc.*, vol. 128, pp. 451–476, 2002.

[30] S. Y. Matrosov, R. F. Reinking, R. A. Kropfli, and B. W. Bartram, "Estimation of ice hydrometeor types and shapes from radar polarization measurements," *J. Atmos. Oceanic Technol.*, vol. 13, pp. 85–96, 1996.

[31] S. Y. Matrosov, R. F. Reinking, R. A. Kropfli, E. Martner, and B. W. Bartram, "On the use of radar depolarization ratios for estimating shapes of ice hydrometeors in winter clouds," *J. Appl. Meteor.*, vol. 40, pp. 479–490, 2001.

[32] T. A. Seliga and V. N. Bringi, "Differential reflectivity and differential phase shift: applications in radar meteorology," *Radio Science*, vol. 13, no. 2, pp. 271–275, 1978.

[33] A. Hendry and G. C. McCormick, "Radar observations of the alignment of precipitation particles by electrostatic fields in thunderstorms," *J. Geophys. Res.*, vol. 81, pp. 5353–5357, 1976.

[34] I. J. Caylor and V. Chandrasekar, "Time-varying ice crystal orientation in thunderstorms observed with multiparameter radar," *IEEE Trans. Geosci. Remote Sensing*, vol. 34, pp. 847–858, 1996.

[35] L. D. Carey and S. A. Rutledge, "A multiparameter radar case study of the microphysical and kinematic evolution of a lightning producing storm," *Meteorol. Atmos. Phys.*, vol. 59, pp. 33–64, 1996.

[36] A. V. Ryzhkov and D. S. Zrnic, "Depolarization in ice crystals and its effect on radar polarimetric measurements," *J. Atmos. Oceanic Technol.*, vol. 24, pp. 1256–1267, 2007.

[37] A. Hendry, G. C. McCormick, and B. L. Barge, "Ku-band and S-band observations of the differential propagation constant in snow," *IEEE Transactions on Antennas and Propagation*, vol. AP-24, pp. 521–525, 1976.

[38] J. Vivekanandan, V. N. Bringi, M. Hagen, and P. Meischner, "Polarimetric radar studies of atmospheric ice particles," *IEEE Trans. Geosci. Rem. Sens.*, vol. 32, pp. 1–10, 1994.

[39] A. V. Ryzhkov and D. S. Zrnic, "Discrimination between rain and snow with a polarimetric radar," *J. Appl. Meteor.*, vol. 37, pp. 1228–1240, 1998.

[40] J. R. Trapp, D. M. Schultz, A. V. Ryzhkov, and R. L. Holle, "Multiscale structure and evolution of an Oklahoma winter precipitation event," *Mon. Wea. Rev.*, vol. 129, pp. 486–501, 2001.

[41] P. C. Kennedy and S. A. Rutledge, "S-band dual polarization radar observations of winter storms," *J. Appl. Meteor. Climatol.*, vol. 50, pp. 844–858, 2011.

[42] D. S. Zrnic and A. V. Ryzhkov, "Polarimetry for weather surveillance radars," *Bull. Amer. Meteor. Soc.*, vol. 80, pp. 389–406, 1999.

[43] S. E. Giangrande and A. V. Ryzhkov, "Calibration of dual-polarization radar in the presence of partial beam blockage," *J. Atmos. Oceanic Technol.*, vol. 22, pp. 1156–1166, 2005.

[44] A. V. Ryzhkov, S. E. Giangrande, V. M. Melnikov, and T. J. Schuur, "Calibration issues of dual-polarization radar measurements," *J. Atmos. Oceanic Technol.*, vol. 22, pp. 1138–1155, 2005.

[45] K. Friedrich, U. Germann, J. J. Gourley, and P. Tabary, "Effects of radar beam shielding on rainfall estimation for the polarimetric C-band radar," *J. Atmos. Oceanic Technol.*, vol. 24, pp. 1839–1859, 2007.

[46] A. R. Jameson and E. A. Mueller, "Estimation of propagation-differential phase shift from sequential orthogonal linear polarization radar measurements," *J. Atmos. Oceanic Technol.*, vol. 2, pp. 133–137, 1985.

[47] J. C. Hubbert and V. N. Bringi, "An iterative filtering technique for the analysis of copolar differential phase and dual-frequency radar measurements," *J. Atmos. Oceanic Technol.*, vol. 12, pp. 643–648, 1995.

[48] J. Testud, E. L. Bouar, E. Obligis, and M. Ali-Mehenni, "The rain profiling algorithm applied to polarimetric weather radar," *J. Atmos. Oceanic Technol.*, vol. 17, pp. 332–356, 2000.

[49] Y. Wang and V. Chandrasekar, "Algorithm for estimation of the specific differential phase," *J. Atmos. Oceanic Technol.*, vol. 26, pp. 2565–2578, 2009.

[50] E. R. Williams, D. J. Smalley, M. F. Donovan, R. G. Hallowell, K. T. Hood, B. J. Bennett, R. Evaristo, A. Stepanek, T. Bals-Elsholz, J. Cobb, and J. M. Ritzman, "Dual polarization radar winter storm studies supporting development of NEXRAD-based aviation hazard products," in *Proc. 35th Conference on Radar Meteorology, Amer. Meteor. Soc., Pittsburgh, PA.*, 2011.

[51] I. Zawadzki, F. Fabry, and W. Szyrmer, "Observations of supercooled water and of secondary ice generation by a vertically pointing X-band Doppler radar," *Atmos. Res.*, vol. 59-60, pp. 343–359, 2001.

[52] K. A. and I. P. Mazin, "Supersaturation of water vapor in clouds," *J. Atmos. Sci.*, vol. 60, pp. 2957–2974, 2002.

[53] K. A., "Limitations of the Wegener-Bergeron-Findeisen mechanism in the evolution of mixed-phase clouds," *J. Atmos. Sci.*, vol. 64, pp. 3372–3375, 2007.

[54] H. Sauvageot, K. Kouadio, and C.-A. Etty, "The influence of temperature and supersaturation on the polarization of radar signals," in *Proc. 23rd Conference on Radar Meteorology, Amer. Meteor. Soc., Snowmass, CO., Sept. 22-26*, pp. 173–176, 1986.

[55] J. Andric, D. Zrnic, and V. Melnikov, "Two-layer patterns of enhanced ZDR in clouds," in *Proc. 34th Conf. on Radar Meteorology, Amer. Meteor. Soc., Williamsburg, VA.*, 2009.

[56] M. Heymsfield, G, "Doppler radar study of a warm frontal region," *J. Atmos. Sci.*, vol. 36, pp. 2093–2107, 1979.

[57] C. Magono and C. W. Lee, "Meteorological classification of natural snow crystals," *Journal of the Faculty of Science, Hokkaido University. Series 7, Geophysics*, vol. 2, pp. 321–335, 1966.

[58] M. Bailey and J. Hallett, "A comprehensive habit diagram for atmospheric ice crystals: Confirmation from the laboratory, AIRS II and other field studies," *J. Atmos. Sci.*, vol. 66, pp. 2888–2888, 2009.

[59] J. R. and I. Chlamtac, "The p2 algorithm for dynamic calculation of quantiles and histograms without storing observations," *Communications of the ACM*, vol. 28, no. 10, pp. 1076–1085, 1985.

[60] K. H. Lo and R. E. P. Jr., "The growth of snow in winter storms: An airborne observational study," *J. Atmos. Sci.*, vol. 39, pp. 697–706, 1982.

[61] R. Bechini, V. Chandrasekar, R. Cremonini, and S. Lim, "Radome attenuation at X-band radar operations," in *Proc. 6th European Conference on Radar in Meteorology and Hydrology (ERAD 2010), Sibiu, Romania*, 2010.

[62] M. Steiner, R. A. Houze, Jr., and S. E. Yuter, "Climatological characterization of three-dimensional storm structure from operational radar and rain gauge data," *J. Appl. Meteor.*, vol. 34, pp. 1978–207, 1995.

[63] S. E. Yuter and J. R. A. Houze, "Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part I: Spatial distribution of updrafts, downdrafts, and precipitation," *Mon. Wea. Rev.*, vol. 123, pp. 1921–1940, 1995.

[64] L. Lei, G. Zhang, R. J. Doviak, R. Palmer, B. L. Cheong, M. Xue, Q. Cao, and Y. Li, "Multilag correlation estimators for polarimetric radar measurements in the presence of noise," *J. Atmos. Oceanic Technol.*, vol. 29, pp. 772–795, 2012.

[65] L. Baldini and E. Gorgucci, "Identification of the melting layer through dual-polarization radar measurements at vertical incidence," *J. Atmos. Oceanic Technol.*, vol. 23, pp. 829–839, 2006.

[66] A. Korolev, G. A. Isaac, and J. Hallett, "Ice particle habits in stratiform clouds," *Quart. J. Roy. Meteor. Soc.*, vol. 126, pp. 2873–2902, 2000.

[67] J. Houze, R. A., *Cloud Dynamics*. Academic Press, 1993.

[68] J. Koistinen, "Operational correction of radar rainfall errors due to the vertical reflectivity profile," in *Preprints, 25th Int. Conf. on Radar Meteorology, Paris, France, Amer. Meteor. Soc.*, pp. 91–94, 1991.

[69] B. Vignal, H. Andrieu, and J. D. Creutin, "Identification of vertical profiles of reflectivity from volume scan radar data," *J. Appl. Meteor.*, vol. 38, pp. 1214–1228, 1999.

[70] Y. L., "Statistical theory of the Marshall-Palmer distribution of raindrops," *Atmos. Environ. A-Gen.*, vol. 27, pp. 15–19, 1993.

[71] R. R. Rogers and M. K. Yau, *A Short Course in Cloud Physics*. Elsevier Science, Oxford, UK, 1988.

[72] A. V. Ryzhkov and D. S. Zrnic, "Assessment of rainfall measurement that uses specific differential phase," *J. Appl. Meteor.*, vol. 35, pp. 2080–2090, 1996.

[73] E. Gorgucci, G. Scarchilli, and V. Chandrasekar, "Practical aspects of radar rainfall estimation using specific differential propagation phase," *J. Appl. Meteor.*, vol. 39, pp. 945–955, 2000.

[74] Y. Wang and V. Chandrasekar, "Quantitative precipitation estimation in the CASA X-band dual-polarization radar network," *J. Atmos. Oceanic Technol.*, vol. 27, pp. 1665–1676, 2010.

[75] J. Vivekanandan, S. M. Ellis, R. Oye, D. S. Zrnic, A. V. Ryzhkov, and J. Straka, "Cloud microphysics retrieval using S-band dual-polarization radar measurements," *Bull. Amer. Meteor. Soc.*, vol. 80, pp. 381–388, 1999.

[76] J. L. Myers and A. D. Well, *Research Design and Statistical Analysis*. Lawrence Erlbaum, 2nd ed., 2003.

[77] J. M. Straka, D. S. Zrnic, and A. V. Ryzhkov, "Bulk hydrometeor classification and quantification using polarimetric radar data: Synthesis of relations," *J. Appl. Meteor.*, vol. 39, pp. 1341–1372, 2000.

[78] H. Liu and V. Chandrasekar, "Classification of hydrometeor type based on multiparameter radar measurements," in *Proceedings, International Conference on Cloud Physics*, pp. 253–256, 1998.

[79] D. S. Zrnic, A. Ryzhkov, J. Straka, Y. Liu, and J. Vivekanandan, "Testing a procedure for automatic classification of hydrometeor types," *J. Atmos. Oceanic Technol.*, vol. 18, pp. 892–913, 2001.

[80] B. Dolan and S. A. Rutledge, "A theory-based hydrometeor identification algorithm for X-band polarimetric radars," *J. Atmos. Oceanic Technol.*, vol. 26, pp. 2071–2088, 2009.

[81] H. Al-Sakka, A.-A. Boumahmoud, B. Fradon, S. J. Frasier, and P. Tabary, "A new fuzzy logic hydrometeor classification scheme applied to the french X-, C-, and S-band polarimetric radars," *J. Appl. Meteor. Climatol.*, vol. 52, pp. 2328–2344, 2013.

[82] C. C. Aggarwal and C. K. Reddy, eds., *Data Clustering: Algorithms and Applications*. Chapman and Hall/CRC, 2013.

[83] R. Gonzalez and R. Woods, eds., *Digital Image Processing*. Prentice Hall, 2nd edition ed., 2002.

[84] R. Fisher, S. Perkins, A. Walker, and E. Wolfart, "Connected Components Labeling," 2003.

[85] L. Baldini, E. Gorgucci, V. Chandrasekar, and W. Peterson, "Implementations of CSU hydrometeor classification scheme for C-band polarimetric radars," in *Proc. 32nd AMS Conf. on Radar Meteorology, Amer. Meteor. Soc., Albuquerque, N.M.*, 2005.

[86] F. Marzano, D. Scaranari, and G. Vulpiani, "Supervised fuzzy-logic classification of hydrometeors using C-band weather radars," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, pp. 3784–3799, Nov 2007.

[87] B. Dolan, S. A. Rutledge, S. Lim, V. Chandrasekar, and M. Thurai, "A robust C-band hydrometeor identification algorithm and application to a long-term polarimetric radar dataset," *J. Appl. Meteor. Climatol.*, vol. 52, pp. 2162–2186, 2013.

[88] V. Chandrasekar, R. Keranen, S. Lim, and D. Moisseev, "Recent advances in classification of observations from dual polarization weather radars," *Atmospheric Research*, vol. 119, pp. 97–111, 2013.

[89] S. Lim, R. Cifelli, V. Chandrasekar, and S. Y. Matrosov, "Precipitation classification and quantification using X-band dual-polarization weather radar: Application in the hydrometeorology testbed," *J. Atmos. Oceanic Technol.*, vol. 30, pp. 2108–2120, 2013.

[90] R. Bechini, L. Baldini, and V. Chandrasekar, "Polarimetric radar observations in the ice region of precipitating clouds at C-band and X-band radar frequencies," *J. Appl. Meteor. Climatol.*, vol. 52, pp. 1147–1169, 2013.

[91] H. S. Park, A. V. Ryzhkov, D. S. Zrnic, and K.-E. Kim, "The hydrometeor classification algorithm for the polarimetric WSR-88D: Description and application to an MCS," *Wea. Forecasting*, vol. 24, pp. 730–748, 2009.

[92] J. J. Gourley, P. Tabary, and J. P. du Chatelet, "A fuzzy logic algorithm for the separation of precipitating from nonprecipitating echoes using polarimetric radar observations," *J. Atmos. Oceanic Technol.*, vol. 24, pp. 1439–1451, 2007.

[93] J. W. Conway and D. S. Zrnic, "A study of embryo production and hail growth using dual-Doppler and multiparameter radars," *Monthly Weather Review*, vol. 121, pp. 2511–2528, 1993.

[94] E. A. Brandes, J. Vivekanandan, J. D. Tuttle, and C. J. Kessinger, "A study of thunderstorm microphysics with multiparameter radar and aircraft observations," *Monthly Weather Review*, vol. 123, pp. 3129–3143, 1995.

[95] S. Lim, V. Chandrasekar, and V. N. Bringi, "Hydrometeor classification system using dual-polarization radar measurements: model improvements and in situ verification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, pp. 792–801, 2005.

[96] T. Otto and H. W. J. Russchenberg, "Estimation of specific differential phase and differential backscatter phase from polarimetric weather radar measurements of rain," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, pp. 988–992, 2011.

[97] M. Schneebeli, J. Grazioli, and A. Berne, "Improved estimation of the specific differential phase shift using a compilation of Kalman filter ensembles," *IEEE T. Geosci. Remote Sens.*, vol. 52, pp. 5137–5149, 2014.

[98] V. N. Bringi, V. Chandrasekar, N. Balakrishnan, and D. S. Zrnic, "An examination of propagation effects in rainfall on radar measurements at microwave frequencies," *J. Atmos. Oceanic Technol.*, vol. 7, pp. 829–840, 1990.

[99] S. E. Giangrande, J. M. Krause, and A. V. Ryzhkov, "Automatic designation of the melting layer with a polarimetric prototype of the WSR-88D radar," *J. Appl. Meteor.*, vol. 47, pp. 1354–1364, 2008.

[100] R. Hooke and T. Jeeves, ""Direct search" solution of numerical and statistical problems," *Journal of the Association for Computing Machinery*, vol. 8, pp. 212–229, 1961.

[101] G. Wolberg, "Cubic spline interpolation: A review," tech. rep., 1998.

[102] V. Lakshmanan, R. Rabin, and V. DeBrunner, "Segmenting radar reflectivity data using texture," in *Proc. 30th International Conference on Radar Meteorology, 19-24 July 2001, Munich, Germany*, 2001.

[103] P. Mahalanobis, "On the generalised distance in statistics," *Proceedings of the National Institute of Sciences of India*, vol. 2 (1), pp. 49–55, 1936.

[104] E. Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, vol. 33, pp. 1065–1076, 1962.

[105] CSU-CHILL, "National Weather Radar Facility," June 2014.

[106] F. Junyent, V. Chandrasekar, D. McLaughlin, E. Insanic, and N. Bharadwaj, "The casa integrated Project 1 networked radar system," *J. Atmos. Oceanic Technol.*, vol. 27, pp. 61–78, 2010.

[107] W. D. King, "Air flow and particle trajectories around aircraft fuselages. I: Theory," *J. Atmos. Oceanic Technol.*, vol. 1, pp. 5–13, 1984.

[108] W. D. King, "Air flow and particle trajectories around aircraft fuselages. IV: Orientation of ice crystals," *J. Atmos. Oceanic Technol.*, vol. 3, pp. 433–439, 1986.

[109] A. Korolev, E. Emery, and K. Creelman, "Modification and tests of particle probe tips to mitigate effects of ice shattering," *J. Atmos. Oceanic Technol.*, vol. 30, pp. 690–708, 2013.

[110] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1 (2), pp. 224–227, 1979.

[111] S. Openshaw, "The modifiable areal unit problem," *Concepts and Techniques in Modern Geography*, vol. 38, 1983.

[112] R. Rotunno, J. B. Klemp, and M. L. Weisman, "A theory for strong, long-lived squall lines," *J. Atmos. Sci.*, vol. 45, pp. 463–485, 1988.

[113] K. A. Scharfenberg, P. T. Schlatter, D. J. Miller, and C. A. Whittier, "The use of the Zdr column signature in short-term thunderstorm forecasts," in *Proc. 11 Conf. on Av. Range and Aerospace Meteor., Hyannis, MA, Amer. Meteor. Soc.*, 2004.

[114] C. A. Van Den Broeke, C. D. Payne, and L. R. Lemon, "Zdr column characteristics and trends during the 10 may 2010 severe weather outbreak," in *Proc. 25th Conference on Severe Local Storms, Denver, CO, Amer. Meteor. Soc.*, 2010.

[115] J. C. Picca, M. Kumjian, and A. V. Ryzhkov, "Zdr columns as a predictive tool for hail growth and storm evolution," in *Proc. 25th Conference on Severe Local Storms, Denver, CO, Amer. Meteor. Soc.*, 2010.

[116] J. C. Snyder, A. V. Ryzhkov, M. R. Kumjian, A. P. Khain, and J. Picca, "A ZDR column detection algorithm to examine convective storm updrafts," *Weather and Forecasting*, vol. 30, no. 6, pp. 1819–1844, 2015.

[117] J. Schaefer and C. D. III, "On the interpolation of a vector field," *Monthly Weather Review*, vol. 107, pp. 458–476, 1979.

[118] J. A. Dubois and P. L. Spencer, "Computing divergence from a surface network: Comparison of the triangle and pentagon methods," *Wea. Forecasting*, vol. 20, pp. 596–608, 2005.

[119] C. N. Helms and R. E. Hart, "A polygon-based line-integral method for calculating vorticity, divergence, and deformation from nonuniform observations," *J. Appl. Meteor. Climatol.*, vol. 52, pp. 1511–1521, 2013.

[120] R. Davies-Jones, "Useful, formulas for computing divergence, vorticity, and their errors from three or more stations," *Mon. Wea. Rev*, vol. 121, pp. 713–725, 1993.

[121] I. Orlanski, "A rational subdivision of scales for atmospheric processes," *Bulletin of the American Meteorological Society*, vol. 56, pp. 527–530, 1975.

[122] F. Chang, C. J. Chen, and C. J. Lu., "A linear-time component-labeling algorithm using contour tracing technique," *Comput. Vis. Image Underst.*, vol. 93, pp. 206–220, 2004.

[123] P. Hiemstra, E. Pebesma, C. Twenh"ofel, and G. Heuvelink, "Real-time automatic interpolation of ambient gamma dose rates from the dutch radioactivity monitoring network," *Computers and Geosciences*, 2008. DOI: http://dx.doi.org/10.1016/j.cageo.2008.10.011.

[124] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

[125] W. R. Cotton, G. Bryan, and S. C. van den Heever, *Storm and Cloud Dynamics, 2nd Edition*. Academic Press, 2010.

[126] F. Rabier, H. Jrvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, "The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics," *Quarterly Journal of the Royal Meteorological Society*, vol. 126, no. 564, pp. 1143–1170, 2000.

[127] F. Rawlins, S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, "The Met Office global four-dimensional variational data assimilation scheme," *Quarterly Journal of the Royal Meteorological Society*, vol. 133, no. 623, pp. 347–362, 2007.

[128] M. Tanguay, L. Fillion, E. Lapalme, and M. Lajoie, "Four-dimensional variational data assimilation for the Canadian regional deterministic prediction system," *Monthly Weather Review*, vol. 140, no. 5, pp. 1517–1538, 2012.

[129] J. Sun and N. Crook, "Real-time low-level wind and temperature analysis using single WSR-88D data," *Wea. Forecasting*, vol. 16, pp. 117–132, 2001.

[130] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528.

[131] R. Cifelli, V. Chandrasekar, S. Lim, P. C. Kennedy, Y. Wang, and S. A. Rutledge, "A new dual-polarization radar rainfall algorithm: Application in colorado precipitation events," *Journal of Atmospheric and Oceanic Technology*, vol. 28, no. 3, pp. 352–364, 2011.

[132] F. J. Merceret and J. G. Ward, "Attenuation of weather radar signals due to wetting of the radome by rainwater or incomplete filling of the beam volume," Tech. Rep. NASA/TM 2002-211171, NASA/Kennedy Space Center, FL, 2002.

[133] S.-F. Chang, Y.-C. Liou, J. Sun, and S.-L. Tai, "The implementation of the ice-phase microphysical process into a four-dimensional Variational Doppler Radar Analysis System (VDRAS) and its impact on parameter retrieval and quantitative precipitation nowcasting," *Journal of the Atmospheric Sciences*, vol. 73, no. 3, pp. 1015–1038, 2016.

[134] U. Germann and I. Zawadzki, "Scale-dependence of the predictability of precipitation from continental radar images. Part I: Description of the methodology," *Monthly Weather Review*, vol. 130, no. 12, pp. 2859–2873, 2002.

[135] Y. Hwang, A. J. Clark, V. Lakshmanan, and S. E. Koch, "Improved nowcasts by blending extrapolation and model forecasts," *Weather and Forecasting*, vol. 30, no. 5, pp. 1201–1217, 2015.

[136] J. M. Davies and J. R. H., "Some wind and instability parameters associated with strong and violent tornadoes, 1. Wind shear and helicity. the tornado: Its structure, dynamics, prediction, and hazards," *Geophys. Monogr.*, vol. 79, pp. 573–582, 1993.

[137] R. E. Rinehart and E. T. Garvey, "Three-dimensional storm motion detection by conventional weather radar," *Nature*, vol. 273, pp. 287–289, 1978.

[138] M. A. Chornoboy, E.S. and J. Morgan, "Automated storm tracking for terminal air traffic control," *MIT Lincoln Laboratory Journal*, vol. 7, pp. 427–448, 1994.

[139] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[140] E. Meinhardt-Llopis, J. Snchez Prez, and D. Kondermann, "Horn-Schunck Optical Flow with a Multi-Scale Strategy," *Image Processing On Line*, vol. 3, pp. 151–172, 2013.

[141] J. Stam, "Stable fluids," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, (New York, NY, USA), pp. 121–128, ACM Press/Addison-Wesley Publishing Co., 1999.

[142] H. W. Haussecker and D. J. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 661–673, Jun 2001.

[143] E. Kessler, "On the continuity and distribution of water substance in atmospheric circulations," *Atmospheric Research*, vol. 38, pp. 109–145, 1995.

[144] R. M. Wakimoto, "The life cycle of thunderstorm gust fronts as viewed with Doppler radar and rawinsonde data," *Monthly Weather Review*, vol. 110, no. 8, pp. 1060–1082, 1982.

[145] S. Corfidi, "Cold pools and MCS propagation: Forecasting the motion of downwind-developing MSCs," *Wea. Forecasting*, vol. 18, pp. 997–1017, 2003.

[146] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, pp. 308–313, 1965.

[147] P. Roebber, "Visualizing multiple measures of forecast quality," *Wea. Forecasting*, vol. 24, pp. 601–608, 2009.

[148] E. Ruzanski and C. V., "Nowcasting rainfall fields derived from specific differential phase," *J. Appl. Meteor. Climatol.*, vol. 51, pp. 1950–1959, 2012.

[149] D. J. Posselt, X. Li, S. A. Tushaus, and J. R. Mecikalski, "Assimilation of dual-polarization radar observations in mixed- and ice-phase regions of convective storms: Information content and forward model errors," *Monthly Weather Review*, vol. 143, no. 7, pp. 2611–2636, 2015.

[150] X. Li and J. R. Mecikalski, "Impact of the dual-polarization doppler radar data on two convective storms with a warm-rain radar forward operator," *Monthly Weather Review*, vol. 140, pp. 2147–2167, 2002.

[151] E. R. Williams, S. G. Geotis, N. Renno, S. A. Rutledge, E. Rasmussen, and T. Rickenbach, "A radar and electrical study of tropical hot towers," *J. Atmos. Sci.*, vol. 49, pp. 1386–1395, 1992.

APPENDIX A

Evaluation of forecast accuracy

For a quantitative evaluation of a forecast the well-known summary measures obtainable from the entries in the contingency table A.1 are considered:

$$\text{POD (Probability Of Detection)} = \frac{A}{A+C} \tag{51}$$

$$\text{FAR (False Alarm Ratio)} = \frac{B}{A+B} \tag{52}$$

$$\text{SR (Success Ratio)} = 1 - \text{FAR} \tag{53}$$

$$\text{bias} = \frac{A+B}{A+C} \tag{54}$$

$$\text{TS (Threat Score)} = \text{CSI (Critical Success Index)} = \frac{A}{(A+B+C)} \tag{55}$$

Mathematical relations can be easily derived among several of the above summary measures, and help construct a geometrical representation of the forecast accuracy in the so

Table A.1: 2x2 contingency table.

		Event Observed Yes	No
Event	Yes	A	B
Forecast	No	C	D

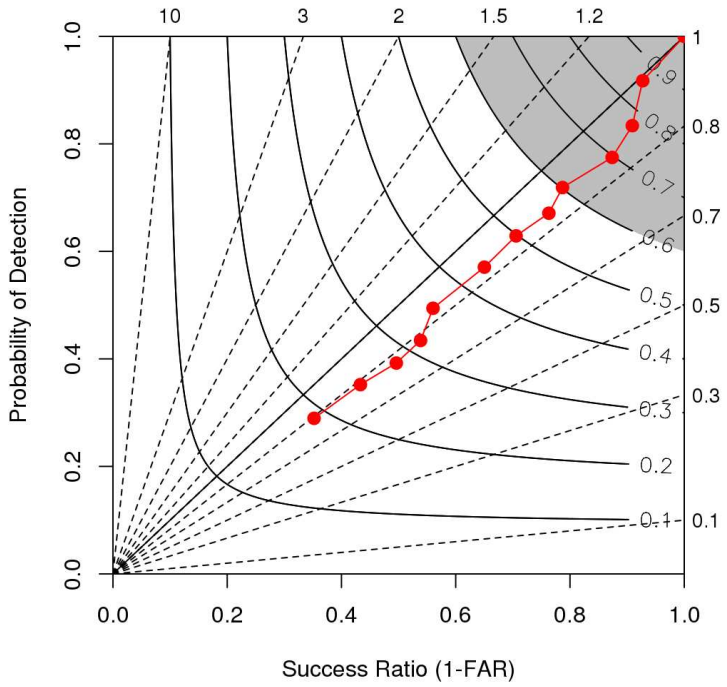


Figure A.1: Example of performance diagram for an extrapolation forecast from 0 to ∼60 minutes, involving ∼5 min time step (from chapter 7). The 13 dots along the red line shows the evolution of the performance during the forecast, starting from the initial upper right corner (POD=SR=TS=1) which corresponds to the analysis (perfect forecast).

called performance diagram (SR, POD plane) [147]. The following relations are used to overplot isolines of TS and bias on the performance diagram:

$$\text{TS} = \frac{1}{\frac{1}{\text{SR}} + \frac{1}{\text{POD}} - 1} \tag{56}$$

$$\text{bias} = \frac{\text{POD}}{\text{SR}} = tan\ \theta \tag{57}$$

where θ is the angle from the horizontal axis counterclockwise in the performance diagram.

The performance diagram then allows to easily visualize multiple measures of forecast quality on the same diagram. Fig. A.1 shows an example of the diagram for a precipitation forecast from chapter 7. The diagram summarizes the SR (x-axis), POD (y-axis), bias

(dashed lines, labeled on the outward extension), and TS (solid lines). For good forecasts, POD, SR, bias, and TS approach unity, such that a perfect forecast lies in the upper right of the diagram. The gray area is added to visually represent the region with Threat Score > 0.6 in the upper right portion of the diagram, which normally represents a good forecast.

APPENDIX B

Radar Convective Parameter (RCP)

Following Houze [67] stratiform precipitation is defined in terms of the vertical air motion leading to the formation of precipitation particles, i.e.:

$$|w| \gg |V_{ice}| \tag{58}$$

where w is the vertical air velocity and V_{ice} is the terminal fall speed of ice precipitation particles ($\sim$1-3 $\mathrm{m\,s^{-1}}$). Conversely, the precipitation type is defined to be convective when (58) does not hold. Unfortunately observations of the vertical air velocity are scarcely available, so a definition in terms of the radar reflectivity is needed for many applications. The condition (58) implies that all hydrometeors are constantly falling, therefore avoiding larger and denser particles growing by aggregation and riming to be found in the upper atmospheric levels. On the other hand in convective precipitation, the relationship between upward vertical air velocity and the development of ice condensate aloft results in much larger particles at sub-freezing temperatures [151]. The main consequence of the limited vertical air velocity in stratiform precipitation is the highly stratified reflectivity structure above the melting layer. The subsequent melting of the ice crystals and aggregates leads to the characteristic bright band signature, whose presence (absence) is often used to classify a reflectivity profile as stratiform (convective).

The bright band-based classification is dichotomic, i.e. it does not allow a continuous parameterization of the degree of convectivity. In addition the detection of the bright band is typically based on several empirical thresholds on the reflectivity values and gradients, which are also sensitive to the radar operation frequency due to complex electromagnetic

scattering properties of mixed-phase particles. For the scope of this study we wanted instead to perform an objective continuous classification of the average reflectivity profiles, either daily or hourly. We therefore defined the Radar Convective Parameter (RCP), a simple parameter to describe the degree of convectivity of a given reflectivity vertical profile as

$$\mathrm{RCP(dB)} = 10\log_{10}\left[\frac{\int_{h(0\,^\circ\mathrm{C})}^{h(-15\,^\circ\mathrm{C})} \left(\frac{\langle Z_{lin}\rangle}{median(Z_{lin})}\right)\mathrm{d}h}{h(-15\,^\circ\mathrm{C}) - h(0\,^\circ\mathrm{C})}\right] \tag{59}$$

where $\langle Z_{\mathrm{lin}}\rangle$, the average of the reflectivity in linear units ($\mathrm{mm^6\,m^{-3}}$), and median(Z_{lin}) are calculated over all the observations available at a given vertical level h and the integral is over the atmospheric layer bounded by the 0°C and -15°C temperature levels. In practice the integral in eq. 59 is replaced by a summation over the available discrete height levels.

The summation is performed on the ice portion of the vertical profile only, despite the fact that higher reflectivity differences between convective and stratiform situations are expected below the 0°C level [62]. Likewise the reflectivity enhancement in the bright band is not involved in the definition of the RCP to eliminate the variable depth of the liquid and mixed-phase layer, especially when dealing with a large dataset encompassing a wide range of freezing level heights (0.4 to 4.3 km MSL for the profiles analyzed in this study). It should be noticed that eq. 59 is actually very similar to the definition of the Pearson's second skewness coefficient:

$$\mathrm{skewness} = 3\frac{(\mathrm{mean} - \mathrm{median})}{\mathrm{standard\ deviation}} \tag{60}$$

The main difference is that in eq. 59 there is no normalization by the standard deviation. The higher vertical air motions in convective precipitation allow mixed phase particles (graupel)

and supercooled liquid water (with approximately five-fold higher refractive index compared to ice, for the 5-10 GHz frequencies considered in this work) to reach heights with temperatures below freezing. The presence of scatterers with higher equivalent reflectivity values in convective situations contributes to broaden the distribution, whose variance increases. The difference between the mean and the median also notably increases, due to the highly positively skewed distribution of the reflectivity in linear units. The use of eq. 59, compared to eq. 60, is therefore preferred to enhance the dynamic range of the convective parameter.