Statistics of Precipitations on the Antarctic Plateau

Observations, Reanalyses, Climate Models

Ambroise Dufour
Mines Paristech
Cycle ingénieur civil
Option géostatistique

Supervisor : Dr Olga Zolina
Laboratoire de Glaciologie et de
Géophysique de l'Environnement

Abstract

Precipitation in the interior of the Antarctic continent plays a crucial role in the mass balance of the ice-sheet. Nevertheless, few observations have been made to quantify the snowfall and it is not clear how accurately the phenomenon is reproduced by climate models. For these reasons, a disdrometer (VPF-730, Biral) was set up on the Antarctic Plateau in 2009 at the French-Italian station Concordia. The observations were compared to a limited area model and a meteorological reanalysis. To account for the important differences between observations and models, a cluster analysis was performed on the disdrometer data. This method allowed to separate artefacts from actual snow events and to distinguish snow events among themselves. If the artefacts are excluded from the dataset, the agreement with the models is improved, at least in terms of the timing of events. A recalibration for specific snow types could potentially yield information on the magnitude of the events as well.

Acknowledgements

I would like to thank Olga Zolina for trusting me with this subject and for introducing me to the atmospheric sciences. I hope we will continue working together in the future.

Many thanks to Christophe Genthon, especially for his crash-course in the art of making observations in Antarctica. The word “crash” is appropriate : instruments breaking down is a recurring theme in polar science.

Specials thanks to Hans Wackernagel for agreeing to overlook this internship and for his support through the ups and downs of my journey in science.

Finally I am grateful to the other scientists and students at LGGE who made this internship an agreeable experience as well as an interesting one.

Contents

Introduction

Quantifying precipitation on the Antarctic Plateau is essential to predict the mass balance of the ice sheet. Upstream, snow accumulates on the continent and thickens the ice-sheet. Downstream, glaciers convey the ice to the coast where it is discharged as icebergs which then melt in the ocean. Estimates of either term have been put forward but important uncertainties undermine their significance. It is still not clear whether the snowfall compensates for the iceberg calving and how both mechanisms will be affected by climate change. The increasing temperatures could actually intensify precipitations thus withdrawing more water from the oceans.

Besides storing massive amounts of water, the ice-sheet also holds valuable information about past climates. Ice coring would benefit from an improved understanding of precipitation on the Antarctic Plateau. Indeed, to reconstruct past climates from the cores, one needs to know how the ice was formed and the nature of the precipitation regime. For instance, an important seasonal variability of the snowfall would introduce biases in the interpretation of the ice-core signals. Moreover, for a given ice depth, the glaciologists need to know where the snow accumulation is the smallest to find the oldest archives.

Together with the extreme conditions, the isolation and the complicated logistics, the measurement of precipitation itself is delicate. On the coast, the main challenge is to discriminate between falling snow and snow blown by the wind. In the interior, this problem is superseded by another one : the minute quantities are below the resolution of conventional instruments like snow gauges.

Optical instrument could provide a fitting alternative. A disdrometer (VPF-730, Biral) was set up at Concordia Station in January 2009. Two years of precipitation observations were compared to a limited area model (LGGE's *Modèle Atmosphérique Régional*) and the ERA-Interim reanalysis. The models are not fully reliable themselves which is the reason why in situ measurements are needed in the first place. However, the extent of the disagreement between observations and models suggests that the former may be even less reliable than the latter. We thus fell back on a different objective : to understand when the disdrometer gave such deviant results, and why. Hopefully, under certain conditions, some information can still be extracted from the dataset.

Chapter 1

Snowfall on the Antarctic Plateau

1.1 The Antarctic Plateau

Antarctica is usually divided into three distinct regions : the Antarctic Peninsula, West Antarctica and East Antarctica (Figure 1.1). East Antarctica is sometimes referred to as "Greater Antarctica" because it accounts for two thirds of the continent's surface. Apart for some parts of the coast, East Antarctica is covered by a massive ice sheet whose average elevation is above 2000 m. It reaches 4093 m at its highest point (Dome A) but its base is often well below the sea-level. There are several other prominences, including Dome C (3233 m). The term "dome" is misleading : the interior of the continent is extremely flat. Only the slopes surrounding the plateau and leading to the sea are actually steep.

As the Earth's southernmost continent, Antarctica is angled away from the sun. The same amount of sunlight spreads over a larger area than in lower latitudes. Even with the twenty-four hours of daylight in summer, the Sun remains low on the horizon. Most of that sunlight is reflected by the snow. Unlike in the Arctic, the tempering influence of the ocean is not felt inside the continent. On the contrary, the climate is harsher due to the altitude of the ice-sheet. As a result, the lowest temperature measured on Earth, -89.2°C, was recorded on the Antarctic Plateau, at the Russian station Vostok.

In spite of the amount of water in solid form, the plateau is technically a desert. Due to the cold, the air can hold very little moisture. The abrupt slopes on the coastal strip usually prevent depressions from penetrating inland. Each year, at Dome C approximately 30 mm of precipitation (equivalent water) accumulate as snow [Hou et al., 2007].

Dome C has been a site of ice core drilling since the seventies. Since 2002, it hosts the French-Italian station Concordia (Figure 1.2). The first winter-over took place in 2005. Research is led in a variety of disciplines besides glaciology : geophysics, astronomy, psychology and the atmospheric sciences.

1.2 Surface mass balance

The ice-sheet shrinks when icebergs are "calved" into the oceans. It grows when the surface mass balance (SMB) is positive, that is when snow accumulates on the continent. The SMB involves a number of other processes besides snowfall [Bromwich, 1988] :

$$A = P - S - \text{div } \mathbf{Q}$$

- A : accumulation rate
- P : precipitation rate
- S : net sublimation
- $\mathbf{Q}$: horizontal flux of blowing snow

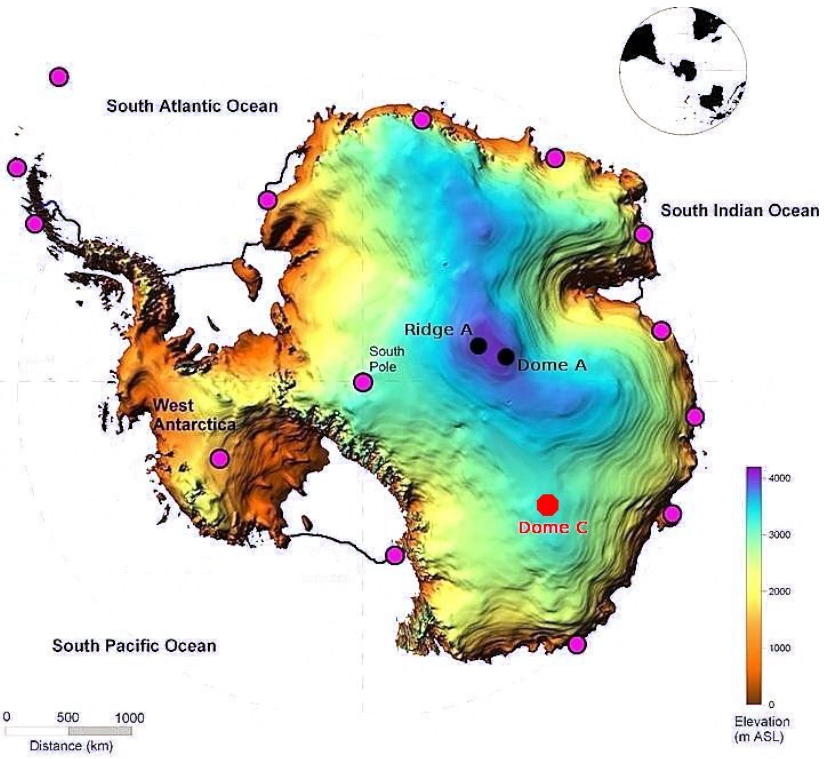

Figure 1.1: Topography of Antarctica (adapted from [Monaghan and Bromwich, 2008])

Figure 1.2: The French-Italian base of Concordia seen from the weather tower [S. Hudson]

S, the net sublimation, is equal to sublimation[1] minus deposition of hoar frost. Just like a puddle gradually dries by evaporation, so does snow by sublimation. The water molecules belonging to the liquid phase in the case of the puddle and to the solid phase in the case of snow progressively become part of the gas phase.

There is a maximum amount of water vapour that a given volume of air can hold. This amount increases with temperature therefore the moisture content of polar air is very low. Sublimation occurs until this maximum is reached. When it is, the air is said to be saturated. The lowest layers of the atmosphere are often saturated on the plateau. Indeed, these layers are very stable and stagnate above the surface for long periods of time.

For a variety of reasons, the air can also become super-saturated. In this case, "inverse sublimation" or "deposition" takes place : this is the origin of hoar frost. In the case of Dome A, [Hou et al., 2007] estimate that net sublimation is negative and that hoar frost represents 7% of the total accumulation. To perform a similar estimation at Dome C we would need to know when the humidity is above saturation using a hygrometer. The standard meteorological equipment is not adapted to the harsh conditions of the Antarctic. After only few weeks of measures, the hygrometer at Dome C had ceased to function [C. Genthon, personal communication].

Blowing snow is not as important on the plateau as it is on the coast, because it depends strongly on the wind speed. The air inside the continent is much colder than near the ocean. Being more dense, it flows down the steep slopes surrounding the plateau creating strong, persistent winds on the coast. These so-called *katabatic* winds can lift important quantities of snow some of which will end up in the sea. Measuring the amount lost is quite a challenge. The first disdrometer aquired by LGGE was used for this purpose [Bellot et al., 2011].

On the contrary, the winds are considerably milder at Dome C. Occasionally though, a frontal system penetrates inside the continent bringing winds strong enough to erode the snow surface. When the storm abates, the blowing snow settles down. The net impact of these events is difficult to evaluate, especially as the storm generates some fresh snow as well.

1.3 Precipitation mechanisms

The type of precipitation caused by the atmospheric fronts associated with cyclone systems is called stratiform precipitation. Such precipitation mostly occurs on the Antarctic Peninsula where midlatitudinal cyclones can pass through the continent. On the plateau, precipitation is often of a more local nature, caused by isolated clouds or even no cloud at all (Figure 1.3). It falls as a semi-continuous shower of ice crystals called "diamond dust."

For hydrometeors to grow, the humidity levels must be above saturation. On the Antarctic plateau, there are two main processes that lead to supersaturated conditions [King and Turner, 1997] :

- **Radiative cooling of the air** can make its temperature drop below its "dew point" (or its "frost point") in some circumstances e.g. at night, under a cloudless sky. Cold air cannot hold as much moisture : the air becomes supersaturated.
- **Incoming warm moist air** can already be supersaturated, or it can mix with cold air so that the overall temperature is below the frost point. The warm moist air could come from :
  - a distance : advection of maritime air
  - another altitude : vertical mixing

To explain the formation of diamond dust, two mechanisms have been put forward. Both have been observed thanks to a remote sensing tool called "lidar" (LIght Detection And Ranging). It was moved to Dome C after a successful campaign on the coast [Del Guasta et al., 1993]. Due to a laser crash, the instrument stopped working a week before the disdrometer was set up. This was quite a disappointment as

[1] Sometimes sublimation and evaporation are conflated

Figure 1.3: Clear-sky precipitation (Kohnen Station, source : H. Grobbe, AWI)

we had hoped to use the lidar data to distinguish between different modes of precipitation. The lidar has been partly repaired ever since and is currently operating at limited power.

- **Seeding by clouds** At usual temperatures, hydrometeors need a foreign particle to grow on. The air in Antarctica is very clean and devoid of such *nuclei* : supersaturated conditions can therefore persist. Thin water clouds have even been spotted [M. del Guasta, personnal communication] : the supercooled droplets were unable to freeze through lack of ice nuclei. However, a thin cirrus cloud can sprinkle minute crystals at high altitude. If they sublimate before reaching the ground they will nevertheless release their own ice nucleus. Either way, particles on which water vapour can depose will be falling through the super-saturated layers closer to the surface.

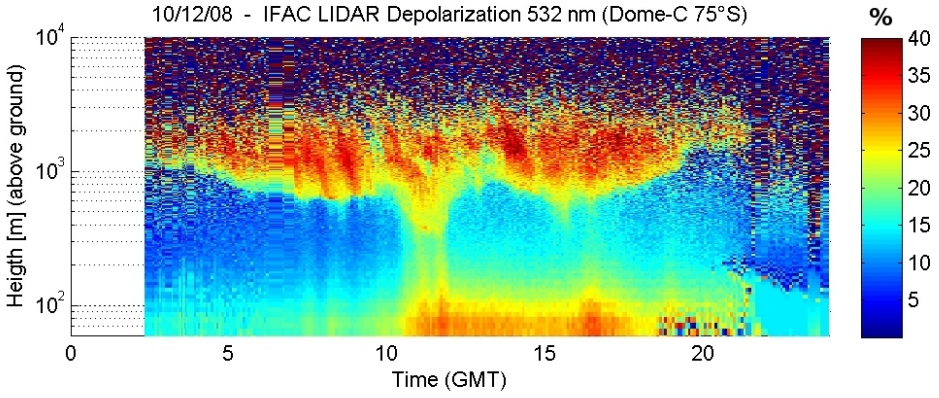

Figure 1.4: A high-level cloud over Dome C (in orange-red) seems to trigger precipitation at lower altitudes (M. del Guasta, personal communication)

- **"Freezing out"** This term has been coined by [Hogan, 1975][2]. Below -40°C, ice nuclei are no longer necessary for crystals to form. This can happen under a cloudless sky : the humidity in excess simply "freezes out." A particular case of freezing out occurs in spring and autumn. During the day, the air warms and becomes subsaturated : snow therefore sublimates. During the night, the air temperature falls below the frost point and diamond dust precipitates.

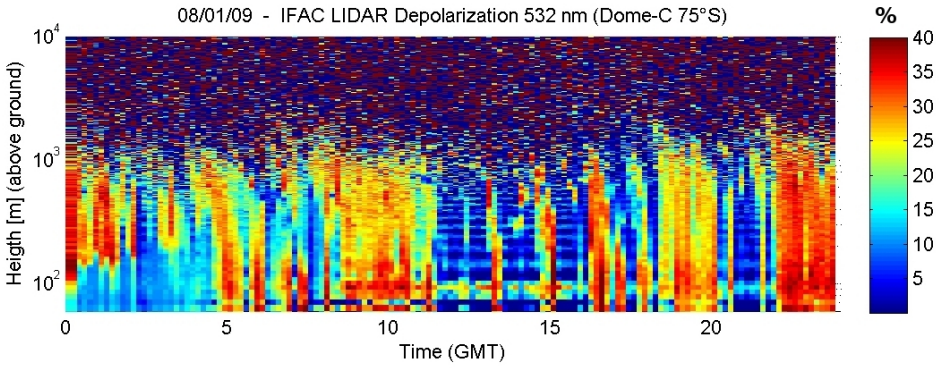

Figure 1.5: Ice crystals precipitating under no apparent cloud (M. del Guasta, personal communication)

Because of the low humidity, the crystals grow slowly and are unlike snowflakes in shape and size. They have well defined shapes : columns, plates, bullets. They glitter in the sunlight and can produce a variety of optical phenomena, the most spectacular being halos.

[2]Its use does not seem widespread

Chapter 2

Climate models

2.1 General circulation models

Since direct measurements are very difficult in Antarctica, numerical models remain our main source of knowledge. General Circulation Models (GCMs) simulate the atmosphere on a global scale. Some model the oceans, the land surfaces and the cryosphere as well. Ideally, the models should be solving directly the primitive equations of the underlying physical processes. In practice, a solution is obtained through a series of assumptions and approximations.

The numerical methods used often require a discretization of the atmosphere. The grid must be coarse for the computation to run in a realistic time. 100 km between gridpoints is considered "high resolution" [Genthon and Krinner, 2001]. This is already too wide to represent accurately the slopes that surround the Antarctic Plateau.

In addition, the scale of many phenomena is much smaller than the grid. Indeed, the processes inside clouds are adequately termed "microphysics." Even the cloud itself, a cumulonimbus for instance, is on average only a few kilometers wide. In other cases, the physics is not well understood and the primitve equations simply do not exist.

Consequently, the models rely on parametrizations. These simplified relations are adjusted to fit approximately the laws of physics or, more often, just observations. In extreme climates, they may no longer hold. The case of diamond dust is a forcible argument. Clear-sky precipitation has nothing to do with snow in the mid-latitudes. The precipitation predictions are usually lower than the observed snow accumulations.

Parametrizations may vary from one model to another. As a result, GCMs do not agree on the snowfall quantities in Antarctica [Gallée and Genthon, 2011]. The spread of the predictions is disturbing (Figure 2.1).

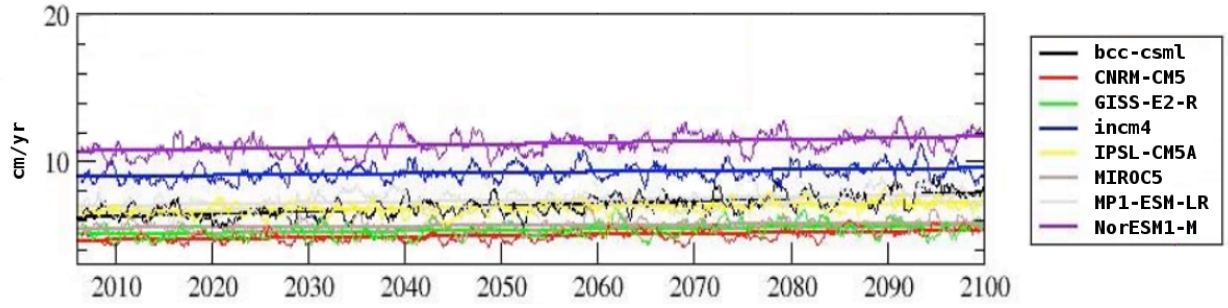

Figure 2.1: Integrated precipitation over the Antarctic ice sheet surface above 2250 m from 8 models running the RCP4.5 IPCC5 scenario, series of annual (12-month running) mean and linear regression [Gallée and Genthon, 2011]

2.2 Meteorological reanalyses

"Analyses" and "reanalyses" are the result of a data assimilation process that takes observations as input. Its output is the state of the atmosphere which is most likely to have generated the data. Fundamental parameters (pressure, temperature, humidity, etc.) can thus be estimated between the measurement points. Nevertheless analyses are more than just a formal interpolation : a physical model of the atmosphere is run during the process.

Operational analyses are computed for the practical purpose of initializing numerical weather prediction models. It is difficult to compare operational analyses produced in different periods or by different institutions. Indeed the data assimilation methods will probably be inconsistent. "Reanalyses" avoid this problem : they run through observation archives using the same protocol. We will be using the ERA-Interim reanalysis produced by ECMWF (European Center for Medium-range Weather Forcasting). ERA-Interim spans from 1979 to the present. It provides the state of the past atmosphere four times per day.

2.3 Limited Area Models

As we saw, one of the main drawbacks of GCMs is their use of inadapted parametrizations. It is also unnecessary to determine the climate globally when the inquiry concerns a specific region. Computer ressources could be allocated to refine the grid. Limited Area Models are usually more adapted to particular regions.

LGGE has its own model dedicated to Antarctica : the "Modèle Atmosphérique Régional" (MAR). It contains three sub-models : one for the atmosphere, one for the surface processes and one entirely devoted to blowing snow. The parametrizations for blowing snow and transport are developed in [Gallée et al., 2001] Other phenomenon specific to Antarctica are reproduced for example the very stable boundary layer or the thin clouds.

MAR needs to be provided with the initial state of the atmosphere and the boundary conditions (top and lateral limits). In [Gallée and Gorodetskaya, 2010], MAR was forced by the ECMWF operational analyses. We compared the precipitation predicted by MAR and by the ERA-Interim reanalyses (Figure 2.2). The timing and magnitude of the snow events are relatively similar. MAR's prediction is systematically higher. On several occasions, MAR predicts massive snowfalls. These extreme events are preceded by "negative precipitations" : this reproduces the erosion of the snow surface by the wind. The important positive snowfall that follows a few hours later represents the blown snow settling down.

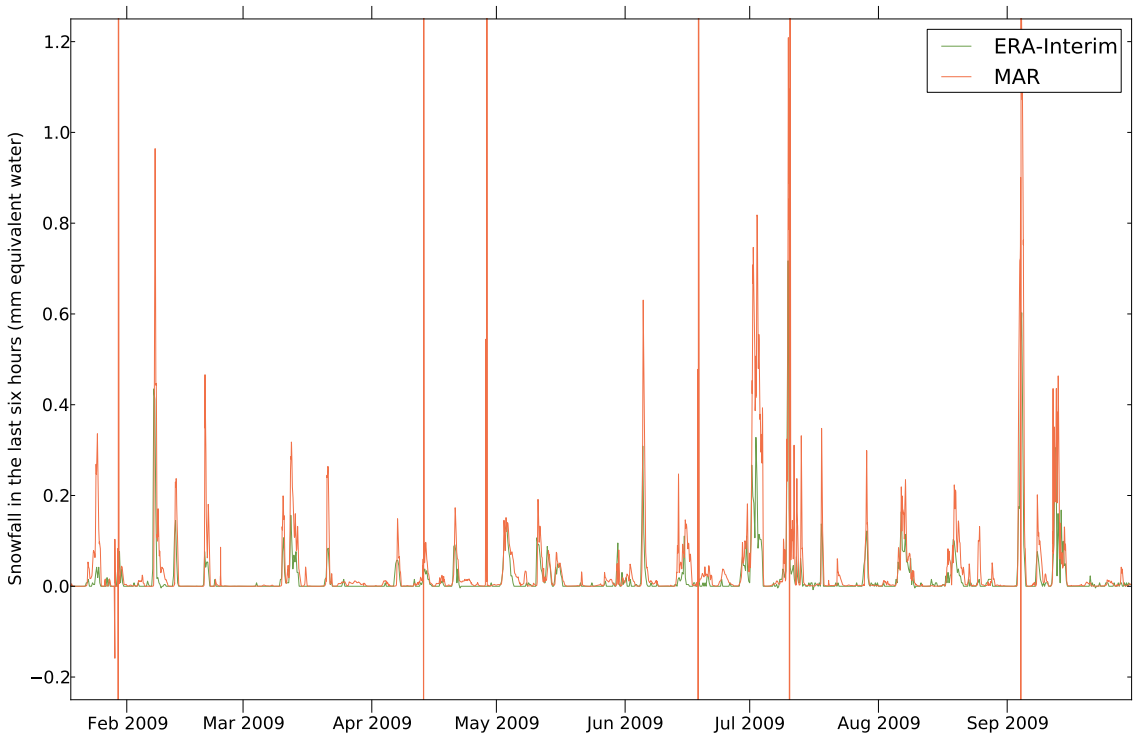

Figure 2.2: Snowfall according to MAR and ERA-Interim. The negative values correspond to wind erosion

Chapter 3

Precipitation measurements

3.1 Snow gauges

The obvious way to measure the precipitation is to use a snow gauge. It needs to be periodically emptied ; its content is then melted so the equivalent volume of liquid water can be measured. The measurement is flawed if the wind is too strong. It will stir the snow on the ground and in the container ; both will mingle in unknown proportions. In the interior of the continent, the snowfall is often below the minimum resolution of the snow gauges [Bromwich, 1988].

[Sato et al., 1981][1] describes a quantitative experiment to measure the precipitation rate during a diamond dust event. The protocol is labor intensive : it requires to count the crystals falling on Formvar-coated plates. The authors concluded that diamond dust fell at a rate of 0.1 mm/hour (equivalent water).

3.2 Satellite observations

Satellites have become a key tool in climate science, especially in Antarctica where measurements are rare. However, it is not yet within their means to measure snowfall. The remote sensing specialists still have difficulties identifying the clouds over the ice-sheet. A team has come up with a method to detect fresh snow by measuring variations in the surfaces' microwave emissions [Bindschadler et al., 2005]. We know so little about precipitation in Antarctica that detecting the snowfall events is already an achievement. Like all remote sensing methods, it requires a validation by field observations.

3.3 Optical precipitation sensors

"Recently developed optical precipitation sensors offer some potential for improving Antarctic precipitation measurements." [King and Turner, 1997] Biral's VPF-730 (Figure 3.2) is one of these sensors.

More precisely, the VPF-730 and VPF-710 are nephelometers : they detect suspended or precipitating particles. The VPF-710 (Figure 3.1) is specifically a visibility sensor : its output is the visual range, given in kilometers. The sensor head is a fork with a transmitter and a receiver on each end. The infrared beam of the transimitter and the receiver's field of view define the sample volume. The receiver measures the proportion of light scattered at small angles by the particles inside the sample volume. This quantity can then be related to the visual range [bir, 730].

The VPF-730 includes an additional receiver to determine present weather as well as visibility. The back scatter receiver helps in identifying the precipitation type. It can detect precipitation (liquid and frozen) and suspended particles (e.g. mist, haze, smoke).

[1] Cited by [Bromwich, 1988], original article not found, possibly in Japanese...

Figure 3.1: Biral VPF-730 combined visibility and present weather sensor

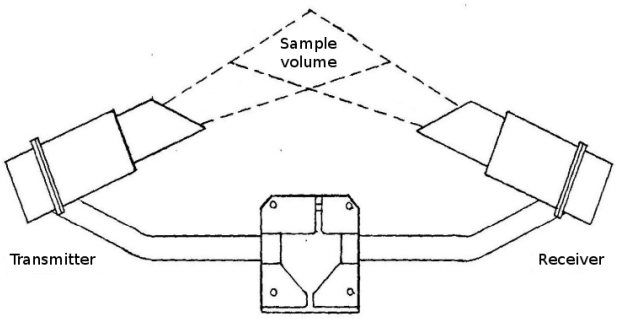

Figure 3.2: Top view of the sensor head (VPF-710 : visibility sensor only, no backscatter receiver)

According to the manual [bir, 730] :

> The VPF-730 measures the amplitude and duration of the light pulse created by each precipitating particle as it falls through the sample volume. From the amplitude and duration it then determines the particle size and velocity.

These abilities define disdrometers. The particles are dispatched into 336 bins according to their apparent speed and size : 16 speed bins and 21 size bins. Each snow event thus has its "Precipitation matrix" (Figure 3.3).

The recognition algorithm assumes that a particle's size/velocity "signature" is characteristic of one type of precipitation. For instance, the particles dispatched along the diagonal of the precipitation matrix are considered raindrops. Snow is assumed to be slower. The mass of each precipitation type can be computed knowing the density of each hydrometeor (1 g/cm^3 for raindrops, approximately 0.08 g/cm^3 for snow flakes, 0.9 g/cm^3 for hail, etc).

The sum of each contribution gives the "AWP" output variable, which stands for :

> Amount of water in precipitation in last measurement period (mm)

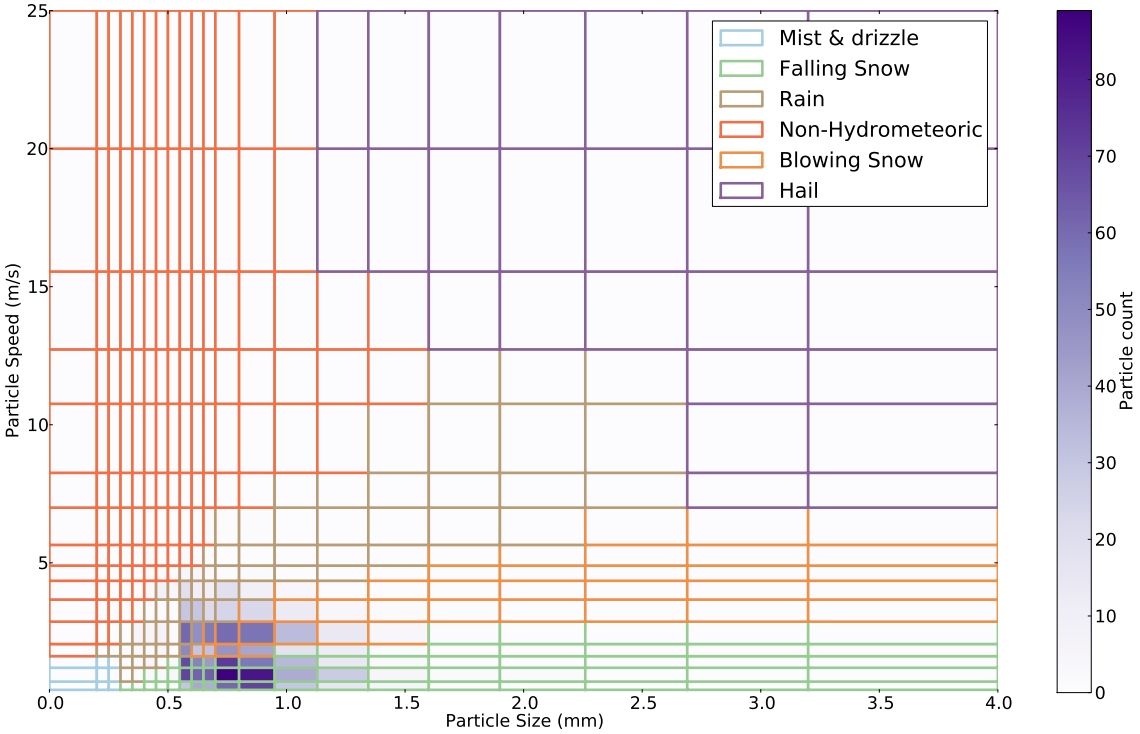

Figure 3.3: Example of a precipitation recognition matrix (2009-07-03 11:30 UTC)

Chapter 4

Exploration of the disdrometer data

4.1 Observations and models : first comparison

The disdrometer was initially deployed on the ground (Figure 4.1). It was then moved to the top of a shelter to limit the influence of blowing snow. The measurements started in January 2009. Our dataset ends in January 2011 but the instrument is still at Dome C accumulating data. The time step between observations is usually 30 minutes. The precipitation matrix only uses information from the last five minutes of the time step.

With AWP, the disdrometer provides an apparently straightforward output variable. This variable can be readily compared to the "snow" variable in MAR or the "large-scale precipitation" in ERA-Interim. However visualizing both the models and the observations turned out to be impossible. The two models give values of comparable magnitude but the disdrometer values are off the chart. According to the disdrometer, the annual snowfall is 977 mm equivalent water whereas the models predict 19 mm (ERA-Interim) and 26 mm (MAR).

Perhaps the disdrometer detects the precipitation events accurately but systematically over-estimates the quantities. Unfortunately, even when the curves are scaled separately, they still have little to do (Figure 4.2).

4.2 Non-hydrometeoritic particles

As we saw previously, there is an intermediate stage between the detection of particles and the computing of AWP : the precipitation matrix. The attribution of precipitation types according to the size/velocity distribution is purely empirical. It holds approximately in the mid-latitudes where the instrument was calibrated. The recognition algorithms are not suited to polar regions. Indeed, when we compute the mean precipitation matrix (Figure 4.3), we find that almost all the particles fall in the "non-hydrometeoritic" section. In our regions, this could be anything from sand to pollen, insects perhaps : not water in any case. Particles falling in this section are not to be included in AWP. Following [Bellot et al., 2011], we compute manually the volume of particles in a matrix instead of using AWP :

$$V_{matrix} = \sum_{i=1}^{21} n_i \frac{4}{3} \pi R_i^3$$

n_i is the number of particles of a given size, and R_i is the average radius of that given size. Dividing V_{matrix} by the sampling time (5 minutes) yields the volumetric flux of particles through the sample volume. In spite of the correction, the fit between observations and models is still imaginary.

More preoccupying are the extreme speeds recorded in many matrices : over 20 m/s ! This is very suspicious : according to the weather tower, the wind at Dome C never exceeded 18 m/s during the two years of observations. When such speeds are measured by the disdrometer, the particle count is wild, often

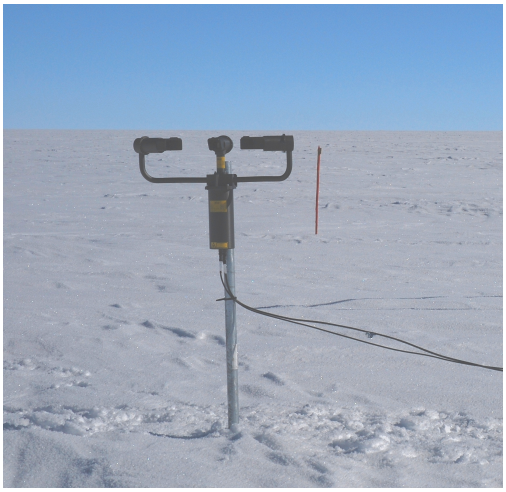

Figure 4.1: Initial location of the disdrometer : it was then moved to the top of a shelter [D. Six]

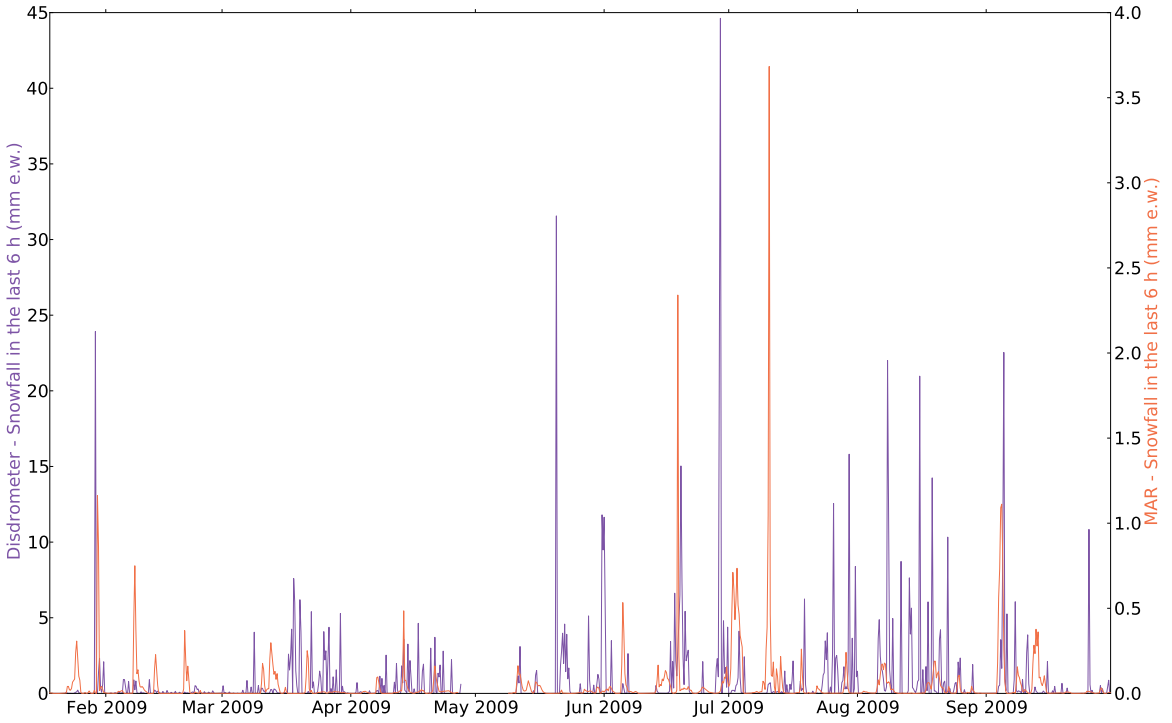

Figure 4.2: Precipitation predictions by MAR compared to the disdrometer observations (AWP)

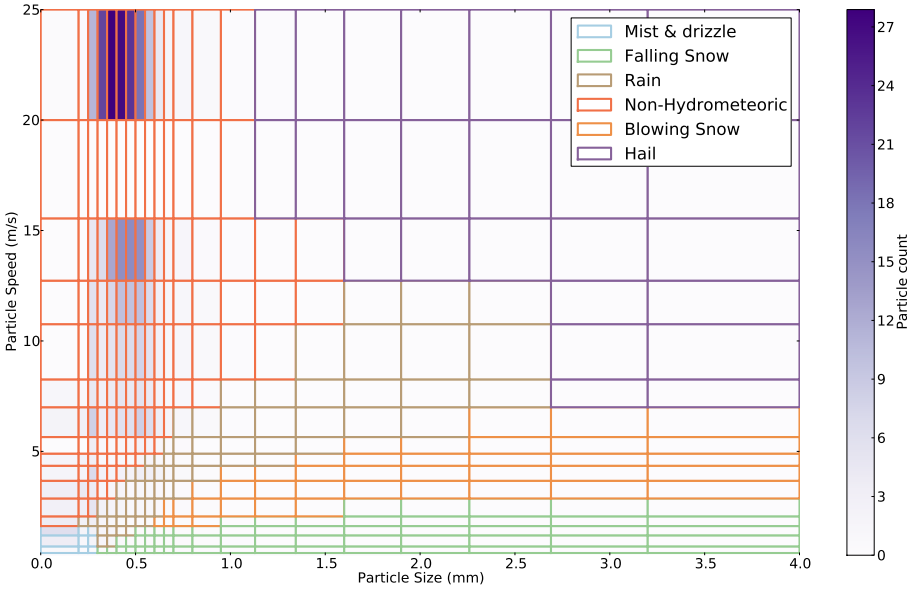

Figure 4.3: Average over all precipitation matrices (2009–2010)

saturating at 999 counts. These precipitation events are undoubtedly artefacts though they may be triggered by actual snow events. They tend to occur when the weather is especially cold so we will refer to them as the "temperature artefacts."

We came up with other explanations besides the temperature :

- frost blocking the receiver : this was indeed the case but for only for short periods of time
- the Sun hitting the receiver directly : the artefacts happen in winter... More generally, the difference between the models and the observations shows no obvious dependence on the time of the day.
- too many ice crystals fool the recognition process : the density of particles is much too low for this to happen [Walden et al., 2003].

According to the manufacturer, the "non-hydrometeoric particles" are ice-crystals [A. Bennet, senior scientist at Biral, personal communication]. The meteors are expected to *scatter* the light in all directions (Mie scattering). Instead, the ice crystals' well defined shapes *reflect* and *refract* the light in specific directions determined by geometrical optics. The sensor is probably fooled by the sparkle of the crystals. This explanation also accounts for the extreme speeds : if the ice crystals rotate as they fall through the sample volume, they send beams of light wildly across the sensor.

4.3 Self-test messages

Diamond dust is not the only one to blame : the artefacts and the extremely low temperatures tend to occur simultaneously. For the heated version of the disdrometer, the manufacturer guarantees a proper functioning for temperatures above -50°C. For colder weathers, the instrument will not necessarily break down but we cannot be sure that it will work properly. In our case, we we were well out of the temperature range : sometimes below -80°C according to the disdrometer (Figure 4.4). The thermometers on the weather tower are more conservative (and reliable) ; they register a minimum of -72 °C.

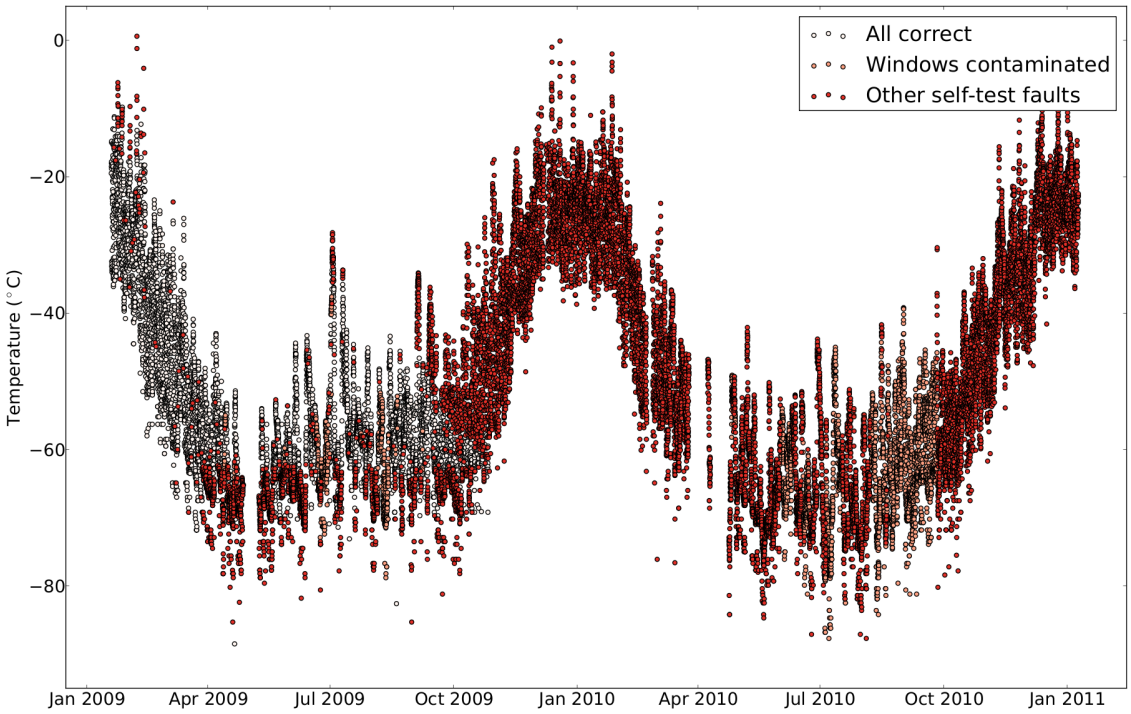

Figure 4.4: Temperature measurements and self-test messages

For every measurement, the disdrometer performs a number of self-tests. We realised to our dismay that after four months on the field, the observations where plagued with error messages. The occurrence of faults is closely related to temperature (Figure 4.4). In the first year, the messages appeared when the temperature dropped below -60°C but the apparatus recovered when the weather warmed up again. As of October 2009, the messages persisted even in the summer months.

We do not know which functions of the disdrometer have been damaged and how that affected the data. Yet up to this day and in spite of the warnings, the device still continues to produce measurements. Out of caution, we only studied the first 2000 precipitation events thus discarding the observations made after October 2009.

Chapter 5

Different clustering methods

The disdrometer dataset aggregates information of different quality :

- **artefacts** : a very unlikely precipitation event
- **unidentified snow event** : the timing is accurate but the magnitude is dubious (no adequate calibration)
- **ordinary snow event** : both the timing and the magnitude of the event are assumed to be accurate

The latter is close to mid-latitude snow, for which the disdrometer was initially calibrated. The unidentified snow events correspond to an exotic type of snow that falls in the "non-hydrometeoritic" bins of the precipitation matrix. Quantitative information (particle count, AWP) could be computed manually by determining the right calibration coefficients. Artefacts may correspond to actual precipitation events, however the data is so distorted that no information seems retrievable.

The natural tool to group the snow events is "cluster analysis" which is a form of unsupervised statistical learning. The use of clustering methods is widespread in the atmospheric sciences, for example to define weather regimes.

We have a number of datasets at our disposal :

- **precipitation matrices** a 336-dimensional data vector for each snow event
- **other VPF-730 outputs** including air temperature and visibility
- **weather tower observations** especially wind speed[1]
- **radiation measurements**

A cluster analysis over all the data available is technically feasible. However, it means manipulating cumbersome data vectors containing variables of different units and magnitudes. It is unlikely that the result will be meaningful. Instead we preferred to use only the information in the precipitation matrix. The matrices are scaled and expressed in percentages of the maximum particle count. This way, only the "signature" of the event is taken into account, not its magnitude. Once clusters of signatures are formed, the extra variables are used to validate the relevance of the partitioning.

We tried a number of clustering methods before finding the appropriate one. For educational purposes, they were all implemented from scratch in FORTRAN. The subroutines were called from Python using the F2PY interface and the results were displayed with Matplotlib.

We did not immediately use all the 336 variables of the matrix. We first formed a 2D summary of every snow event by calculating the average particle size and the average particle velocity. This provided a simpler training dataset.

[1]The grease lubricating the mobile parts of the anemometers thickened and became viscous with the cold. They now run without lubrication and an empirical formula is applied to compensate for the increase in friction [A. Trouvilliez, personal communication].

5.1 Hierarchical agglomerative clustering

In this method, each data point begins in its own cluster [Wilks, 2011]. At each step, the two nearest clusters are merged. The iteration is stopped when a relevant number of clusters have emerged.

Two different metrics need to be specified : one for the points and one for the clusters. Within points, the Euclidian distance is an obvious choice but it is by no means the only one. Within clusters, "Average linkage" is also one of many possible metrics. It defines the distance between clusters G_1 and G_2 as :

$$d(G_1, G_2) = \sum_{x_1 \in G_1} \sum_{x_2 \in G_2} d(x_1, x_2)$$

The method gave some decent results in two dimensions. However, when the whole matrix is used, the output displays no apparent structure. Hierarchical clustering does not have a notion of noise : the results are cluttered with singletons. Furthermore, once a point has been assigned to a cluster, there is no way to reallocate it if the choice seems sub-optimal in retrospect. Finally, the algorithm is too slow for large datasets.

5.2 k-means

k-means divides the points into k clusters that minimize the sum of all intra-cluster variance [Wilks, 2011]. Clusters are defined by their center of mass, or "centroid." A point is allocated to the cluster whose centroid is the closest. The clusters are therefore Voronoi cells (Figure 5.1). k-means is a difficult problem computationally speaking.

Luckily, there are heuristics. Llyod's algorithm is one of them. It converges rather rapidly (compared to agglomerative clustering) but only to a local maximum. The rigid shape of the Voronoi cells is the main drawback. They have no regard for the density troughs though this is where we would place the cluster boundaries intuitively.

5.3 DBSCAN

DBSCAN stands for Density Based Clustering of Applications with Noise [Ester et al., 1996]. This method is closer to the aforementioned "intuitive clustering" but paradoxically it gave the poorest results on our dataset. Roughly speaking, it divides the dataset in dense points and noise. Density is defined by a minimum number of neighbours in a given radius. The clusters are connected dense regions.

The main weakness of this method is that the density cutoff is uniform. In our dataset, the density is heterogeneous. On the one hand, if we pick a stringent density criterion, we label as noise entire patches of data (Figure 5.3). On the other hand, if the criterion is more lenient, the output is made up of few giant clusters (a phenomenon known as "chaining").

5.4 DENCLUE

DENCLUE is short for "DENsity-based CLUstEring" [Hinneburg and Gabriel, 2007]. DENCLUE addresses DBSCAN's main drawback : the rigid density threshold. DENCLUE relies on "Kernel density smoothing," a method to estimate the probability density function of a random variable given a finite number of its realizations. It uses characteristic shapes called kernels. We will be using Gaussian kernels :

$$K(\mathbf{x}) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\mathbf{x}^2/2}$$

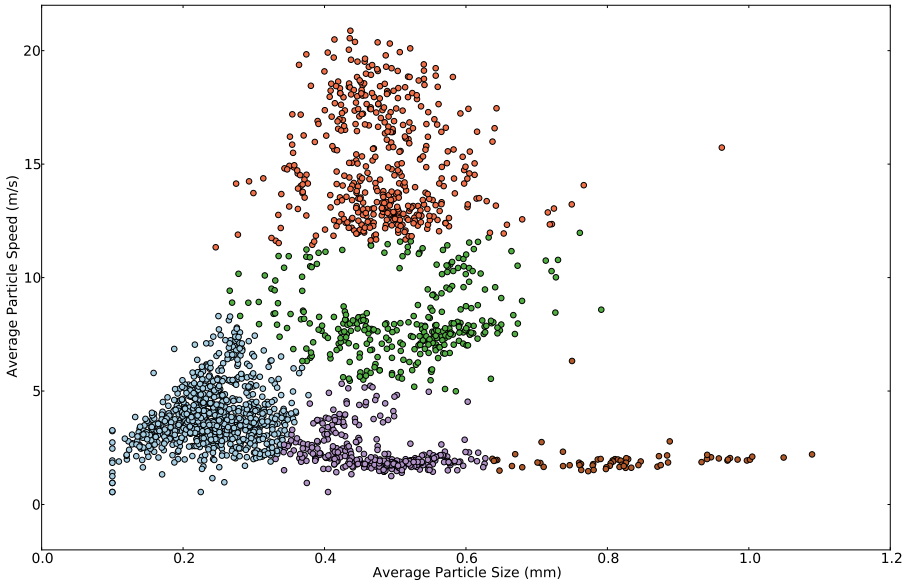

Figure 5.1: k-means clustering : the clusters are Voronoi cells

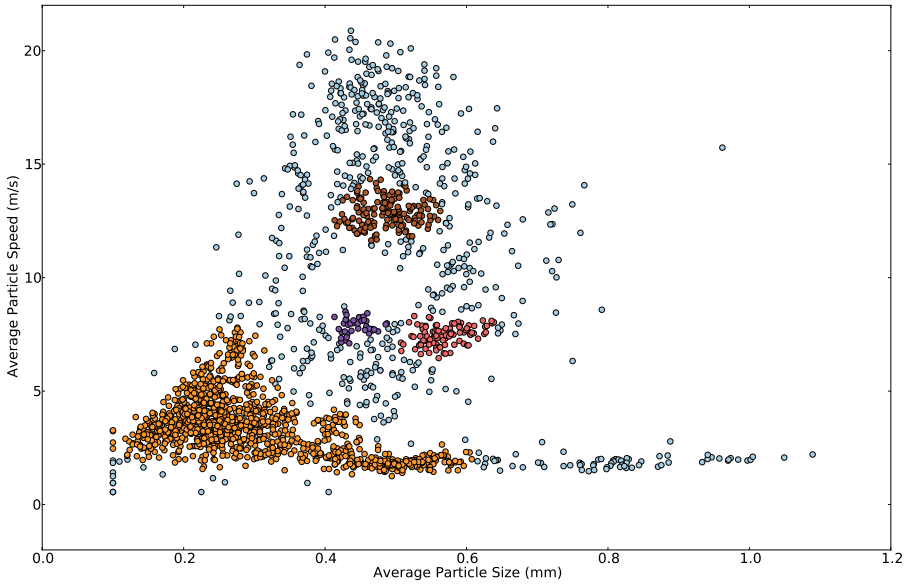

Figure 5.2: DBSCAN counts most of the data-points as noise (light blue) and is prone to “chaining”

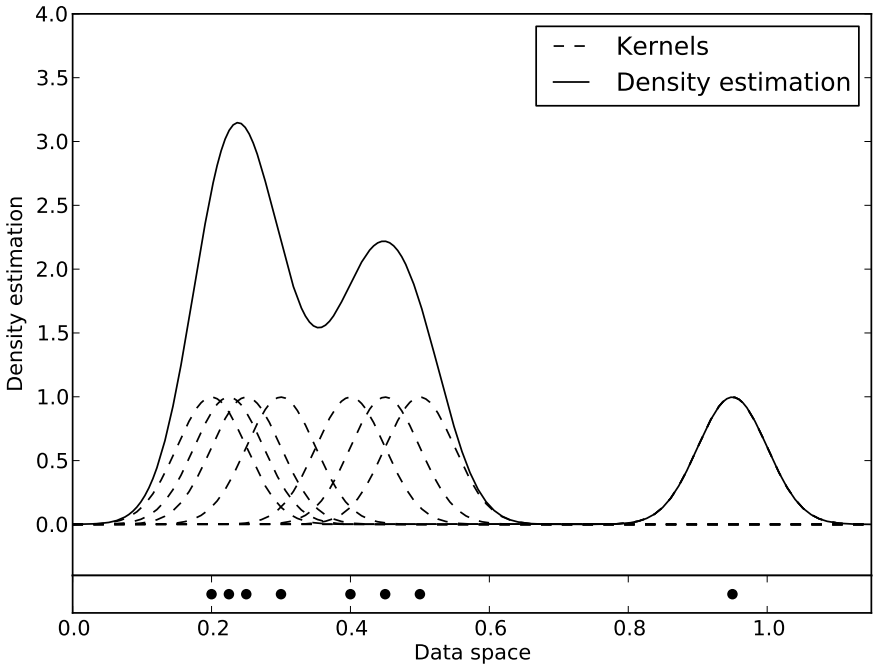

Figure 5.3: Kernel density estimation : a 1D example

To obtain the density estimation, we stack kernels centered on the data values (Figure 5.3) :

$$\hat{f}(\mathbf{x_0}) = \frac{1}{n \cdot h^d} \sum_{k=1}^{n} K\left(\frac{\mathbf{x}_0 - \mathbf{x}_k}{h}\right)$$

n is the number of points, d is the dimension of the data space. The spread of the kernels is governed by a smoothing parameter h. Clusters are defined by local maximums of the density function. The cluster boundaries are density troughs. Points are assigned to a local maximum by hill-climbing.

The metaphorical "hill-climber" starts from a data point and "climbs" until he reaches a summit. DENCLUE 1.0 approaches the local maximums by gradient ascent. At every iteration, the hill-climber moves by a distance δ in the direction of the greatest slope (hence the gradient). Let $(\mathbf{u}_n)_{n\in N}$ be the hill-climber's positions. The new position of the hill-climber, $\mathbf{u}_{l+1}$ is given by :

$$\mathbf{u}_{l+1} = \mathbf{u}_l + \delta \cdot \frac{\nabla \hat{f}(\mathbf{u}_l)}{||\nabla \hat{f}(\mathbf{u}_l)||}$$

The method we use is actually DENCLUE 2.0, which offers an improved hill-climbing algorithm. The step size is no longer fixed but is adjusted automatically. When the hill-climber no longer moves we assume that he has reached the summit. The update formula is now :

$$\mathbf{u}_{l+1} = \frac{\sum_{k=1}^{n} K\left(\frac{\mathbf{u}_l - \mathbf{x_k}}{h}\right) \cdot \mathbf{x}_k}{\sum_{k=1}^{n} K\left(\frac{\mathbf{u}_l - \mathbf{x}_k}{h}\right)}$$

In 2D, for an appropriate smoothing parameter, DENCLUE grouped the data the way we would have done intuitively. However, we prefer to illustrate the report with a case where the smoothing parameter is excessive. In Figure 5.5 and 5.6 a potentially valuable cluster is completely absorbed.

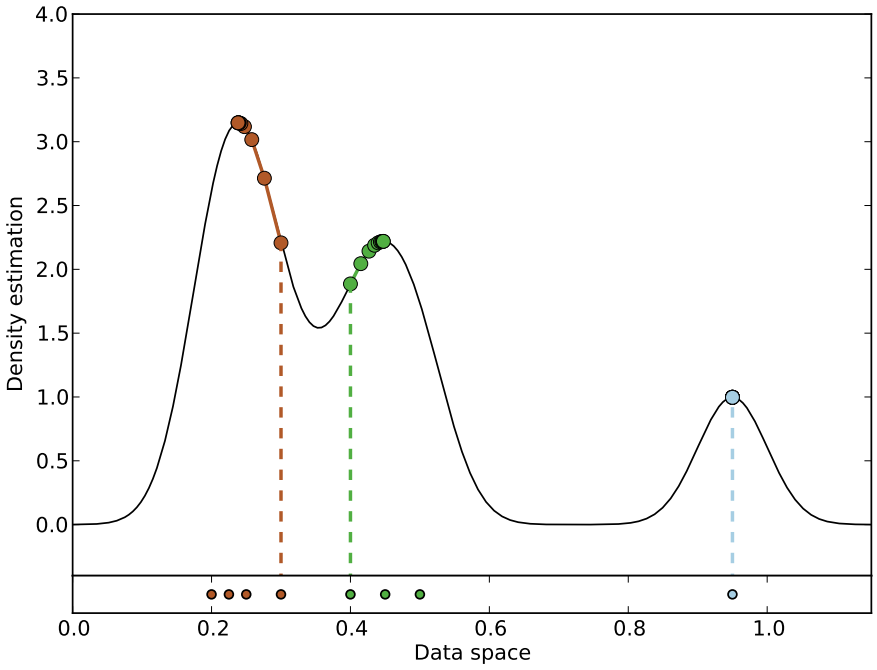

Figure 5.4: Hill-climbing in DENCLUE 2.0 : a 1D example

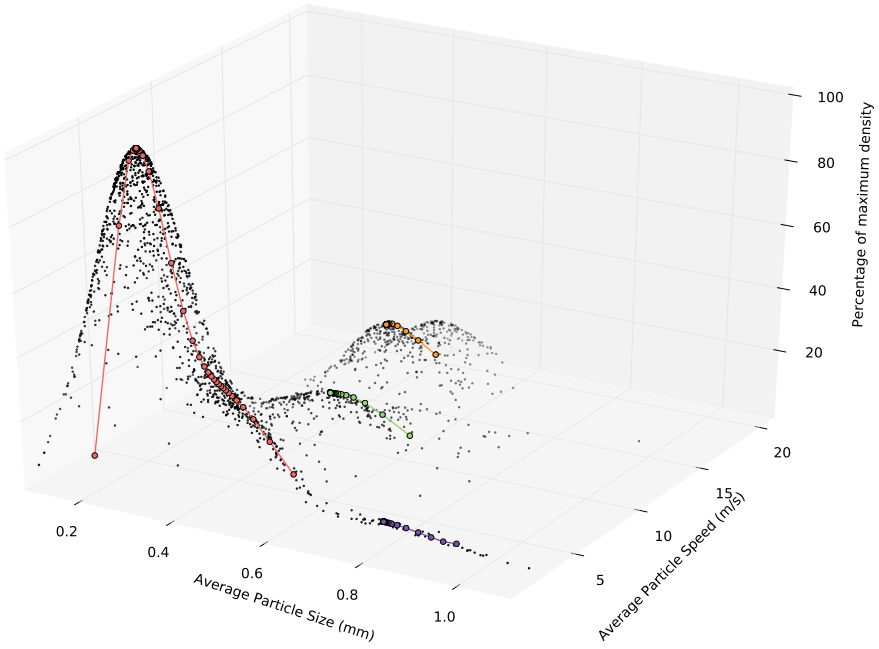

Figure 5.5: Density estimation and hill-climbing : 2D training dataset (excessive smoothing parameter)

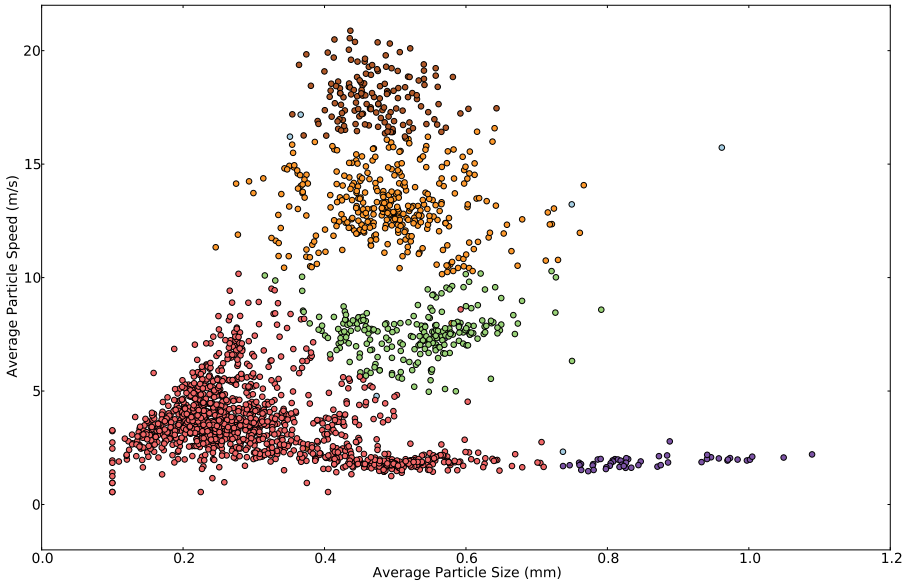

Figure 5.6: DENCLUE 2.0 : 2D training dataset (excessive smoothing parameter)

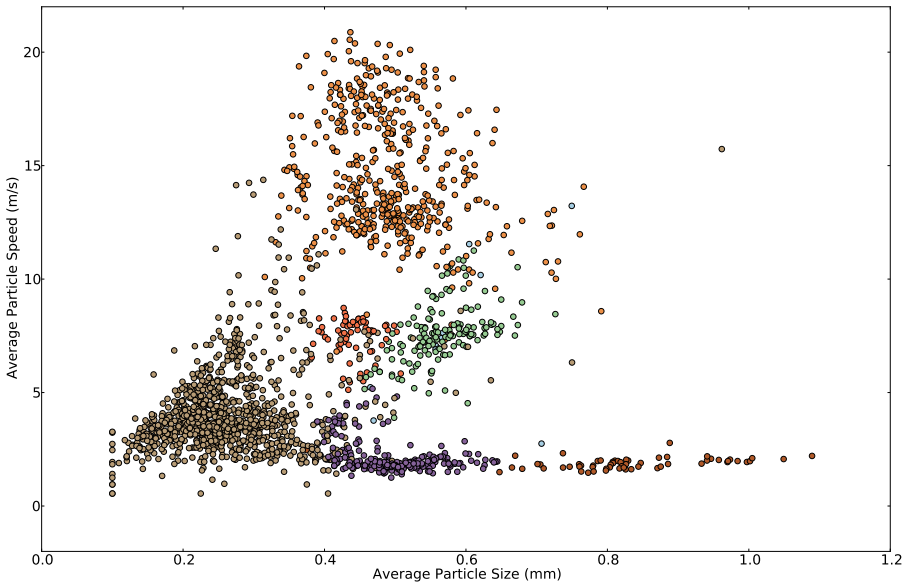

Figure 5.7: DENCLUE 2.0 : final result in 336D

Chapter 6

Clustering of precipitation events

We applied DENCLUE 2.0 on the 2000 first precipitation events using all the information from the matrices (Figure 5.7). This allowed us to identify three categories :

- blowing snow
- temperature artefacts
- the remaining events, which we assume are actual precipitation

The third category can potentially be split into more relevant clusters.

6.1 Blowing snow

We surmise that the brown and purple points in Figure 5.7 are cases of blowing snow. The strip of points with relatively large sizes but small velocites are reported only when the wind velocity is important (Figure 6.1). It would seem that this is drifting snow : the disdrometer measures a horizontal flux of particles. We cannot explain why the particle speed is so low compared to the wind's. It also seems impossible to determine the net snowfall during these events. At least we know the timing.

6.2 Temperature artefacts

The events showing extreme speeds (yellow in Figure 5.7) were good candidates for the artefacts. This intuition was confirmed when we plotted the temperature against time (Figure 6.2).

Our personal opinion is that this cluster can be removed from the dataset without a major loss of information. Although the artefacts may correspond to real snowfall, the amount is bound to be very small because the cold limits the air's water-holding capacity. Once the artefacts are removed, the agreement with the models improves, especially in winter (Figure 6.3). Nevertheless, accurate measurements are bound to be lost during this procedure.

The remaining disparities could be events that the model does not reproduce properly. MAR is constrained by observations only on the boundaries. Large scale phenomena display enough structure to be modelled correctly far inside the continent. However, local phenomena, like the more or less random vertical mixing of atmospheric layers are harder to predict from a distance.

6.3 Perspectives

The smoothing parameter we used ($h = 70$) is quite important and conceals smaller clusters. As a matter of fact, we did not quite benefit from using the whole precipitation matrix except at the boundaries between clusters. Information from the extra-dimensions would allow the clusters to overlap in 2D.

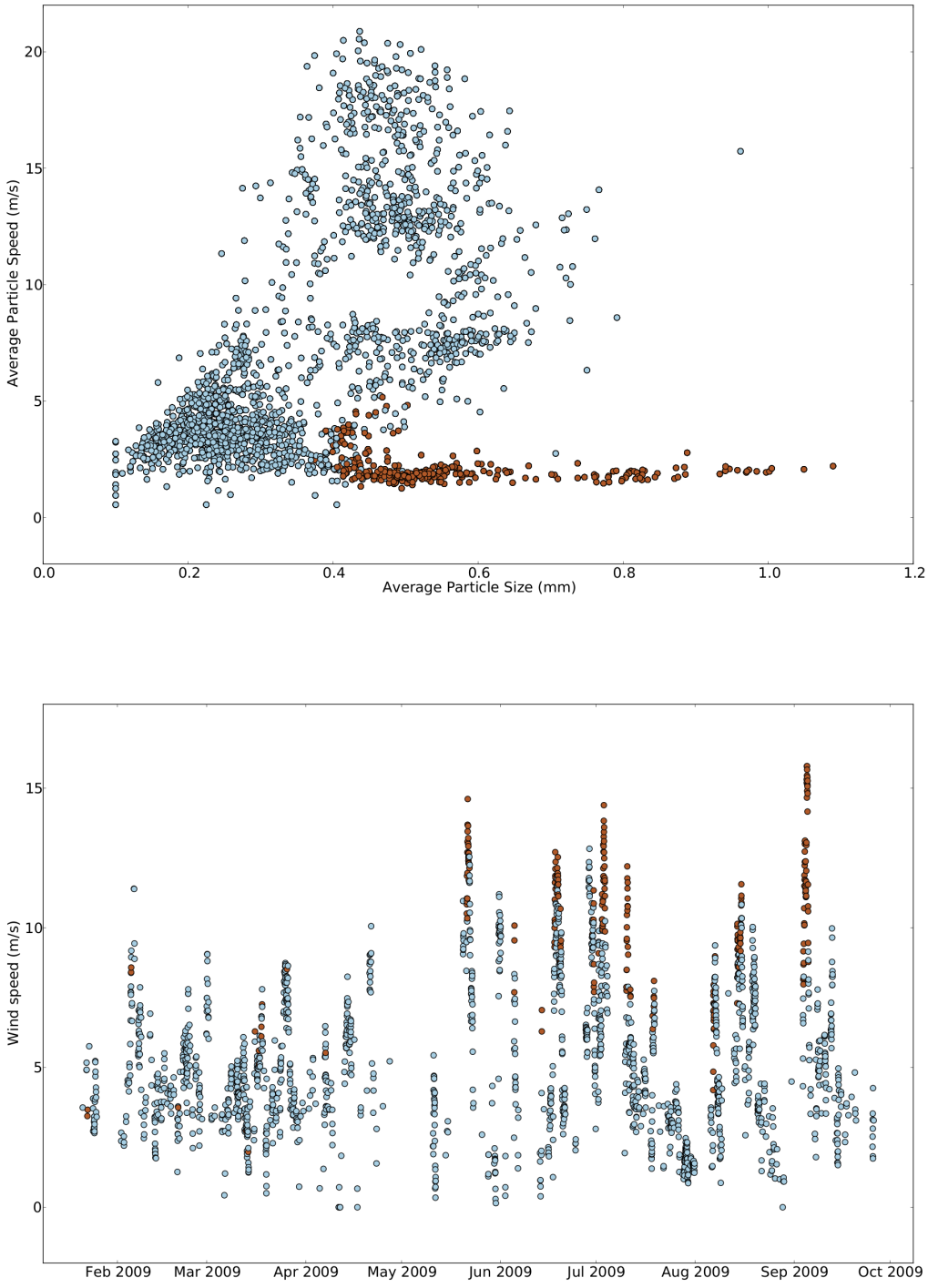

Figure 6.1: Identification of blowing snow events (in brown)

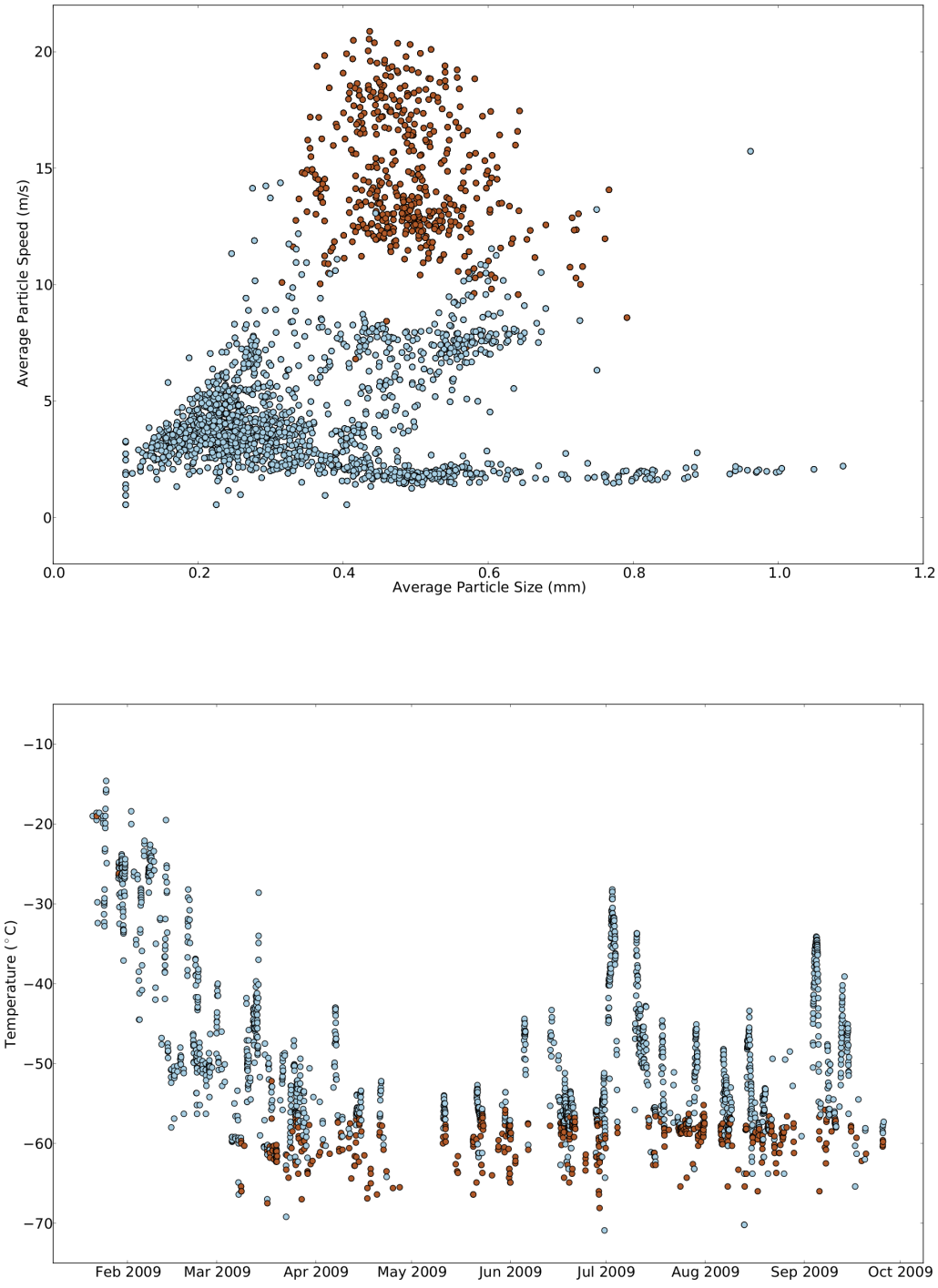

Figure 6.2: Identification of temperature artefacts (in brown)

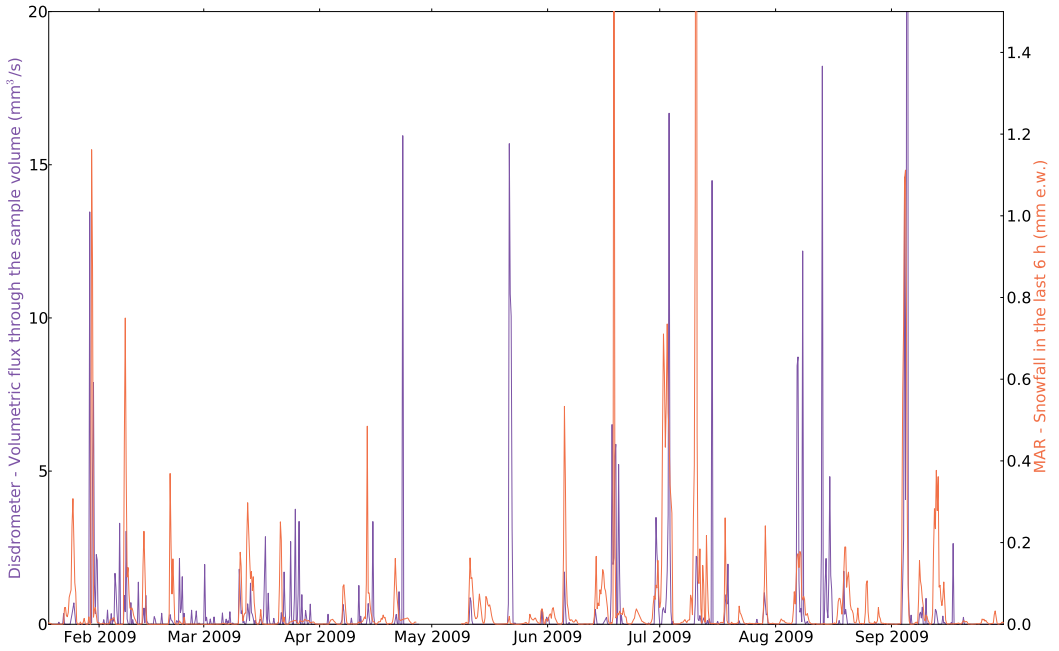

Figure 6.3: Comparison between MAR and the corrected disdrometer observations

However, in this exercise, the challenge is to explain the relevance of the new clusters. For example, a smoothing parameter of 65 yields a spectacular but useless result. The grey cluster in Figure 5.7 (small and slow particles) is split into two overlapping clusters. One group is dates from the summer and the other from the winter but nothing more can be said.

Moreover, even with $h = 70$, the three central clusters are not unaccounted for. We can reasonably guess that the small and slow particles are "diamond dust," although we were unable to prove this claim. Diamond dust is expected to fall from clear-skies so we hoped to spot it using radiation measurements. The raw data from the Baseline Surface Radiation Network (BSRN, source : V. Vitale) did not yield any result.

Clouds block the sun : they create an anomaly in the direct solar radiation measurements. This only holds during the day obviously. All year round, the most useful variable would be the Long-wave Downward radiation (LWD). When a cloud is directly above the pyrgeometer it radiates heat towards the ground : this is the so-called "blanket effect." LWD also depends on the air temperature so it follows a seasonal cycle which needs to be substracted. Large patches of data are missing in the BSRN LWD data yet this could be a promising lead.

This cloud detection problem would be solved should the lidar measurements resume. The different precipitation modes could be related to their matrices. Even if we do not manage to automate the procedure, a manual case study would provide a reference set. The remaining precipitation events could then be classified on the basis of this training set.

If we manage to distinguish the different types of snow, we could find inspiration in [Sato et al., 1981]'s experiment. We could follow the same protocol to calibrate the disdrometer for every class of exotic snow. This way we would dispose of a corrected "AWP" variable adapted to the Antarctic Plateau.

Conclusion

"It is extremely difficult to make measurements of precipitation in the Antarctic" [King and Turner, 1997]. Halfway through the internship, our work looked very much like an autopsy. The cold was our main suspect but there were others. The disdrometer was to be shipped back for repair. We ventured to retrieve at least some information from our dataset : not all the data was corrupted, so we thought. The cold did affect the instrument, but not to the point of disrupting the observations. However, flawed and accurate measurements were mingled.

By definition, the disdrometer can measure the size and velocity of precipitating particles. Thanks to clustering techniques, this information allowed us to identify precipitating particles from aberrations and drifting snow. Once we had removed the artefacts, we reached the modest goal of detecting the snow events. This is by no means an end in itself : our method will have to be used in conjunction with others. For instance it can increase the confidence in remote sensing methods like the one described in [Bindschadler et al., 2005]. However, due to the error messages, the instrument cannot be considered completely reliable – if such a notion exists.

During our study, we missed two valuable instruments : the lidar and the hygrometer. They could offer insights in the unusual precipitations mechanisms peculiar to Antarctica. Hopefully, these devices will be operational in the future and their measurements will be combined with the disdrometer's. Although this seems a little out of reach, we could replicate [Sato et al., 1981]'s experiment for each snow type and deduce quantitative information from the disdrometer as well.

Bibliography

[bir, 730] (VPF-730). *Operation and Maintenance Manual for the VPF-730*. Biral.

[Bellot et al., 2011] Bellot, H., Trouvilliez, A., Naaim-Bouvet, F., Genthon, C., and Gallée, H. (2011). Present weather-sensor tests for measuring drifting snow. *Annals of Glaciology*, 52(58):176–184.

[Bindschadler et al., 2005] Bindschadler, R., Choi, H., Shuman, C., and Markus, T. (2005). Detecting and measuring new snow accumulation on ice sheets by satellite remote sensing. *Remote Sensing of Environment*, 98(4):388–402.

[Bromwich, 1988] Bromwich, D. H. (1988). Snowfall in the high southern latitudes. *Reviews of Geophysics*, 26(1):149–168.

[Del Guasta et al., 1993] Del Guasta, M., Morandi, M., Stefanutti, L., Brechet, J., and Piquad, J. (1993). One year of cloud lidar data from Dumont d'Urville (Antarctica). *Journal of Geophysical Research*, 98(18):575–587.

[Ester et al., 1996] Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press.

[Gallée and Genthon, 2011] Gallée, H. and Genthon, C. (2011). Evaluation of uncertainties on the surface mass balance (SMB) of Antarctica and related contribution to sea-level. *ICE2SEA Report*.

[Gallée and Gorodetskaya, 2010] Gallée, H. and Gorodetskaya, I. V. (2010). Validation of a limited area model over Dome C, Antarctic Plateau, during winter. *Climate Dynamics*, 34(1):61–72.

[Gallée et al., 2001] Gallée, H., Guyomarc'h, G., and Brun, E. (2001). Impact of snow drift on the Antarctic ice sheet surface mass balance: Possible sensitivity to snow-surface properties. *Boundary-Layer Meteorology*, 99(1):1–19.

[Genthon and Krinner, 2001] Genthon, C. and Krinner, G. (2001). Antarctic surface mass balance and systematic biases in general circulation models. *Journal of Geophysical Research*, 106(D18):653–664.

[Hinneburg and Gabriel, 2007] Hinneburg, A. and Gabriel, H. H. (2007). Denclue 2.0: Fast clustering based on kernel density estimation. In *Proceedings of the 7th International Symposium on Intelligent Data Analysis*, pages 70–80.

[Hogan, 1975] Hogan, A. W. (1975). Summer ice crystal precipitation at the South Pole. *Journal of Applied Meteorology*, 14:246–249.

[Hou et al., 2007] Hou, S. G., Li, Y. S., Xiao, C. D., and Ren, J. W. (2007). Recent accumulation rate at Dome A, antarctica. *Chinese Science Bulletin*, 52(3):428–431.

[King and Turner, 1997] King, J. and Turner, J. (1997). *Antarctic Meteorology and Climatology*. Cambridge University Press.

[Monaghan and Bromwich, 2008] Monaghan, A. J. and Bromwich, D. H. (2008). Advances in describing recent Antarctic climate variability. *Bulletin of the American Meteorological Society*, 89:1295–1306.

[Sato et al., 1981] Sato, N., Kikuchi, K., Barnard, S., and Hogan, A. W. (1981). Some characteristic properties of ice crystal precipitation in the summer season at South Pole Station, Antarctica. *Bulletin of the Japanese Meteorological Society*, 59:772–780.

[Walden et al., 2003] Walden, V. P., Warren, S. G., and Tuttle, E. (2003). Atmospheric ice crystals over the Antarctic Plateau in winter. *American Meteorological Society*.

[Wilks, 2011] Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

List of Figures