Database of Snowfall for Analyzing Vertical Structure of Cloud

A. Teras[1], K. Muramoto[1], H. Aoyama[1], M. Tamura[1], T. Koike[2], H. Fujii[2], T. Pfaff[2]

1 Department of Information and Systems Engineering, Kanazawa University, Kanazawa-shi, 920-8667 Japan

2 Department of Civil Engineering, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

muramoto@t.kanazawa-u.ac.jp

Abstract: This paper introduces a system based on a relational database containing data from a wide range of different instruments for analyzing precipitation and vertical cloud structure during snowfall. Temporal resolution of less than one minute allows distiguishing between different phases of a snowfall event. Using a database facilitates data retrieval and simultaneous visualization from multiple sources.

Keywords: relational database, radar, lidar, image processing, snowfall events, cloud structure

1. Introduction

International collaboration in the Coordinated Enhanced Observing Period (CEOP)[1)] has been designed to promote cooperation in global climate research with a particular focus on the water cycle. It started in July 2001 and will last for a total of 4 years.

On the Sea of Japan coast, rather heavy snowfalls from convective clouds are frequently brought by the prevailing north-west wind in winter. In Fukui, which is located near the middle coast of the Sea of Japan, the type of snow particles is usually graupel or snowflake aggregate, and it often changes from graupel to aggregate and sometimes from aggregate to graupel. Graupels are formed near the cloud top where updraft is prevailing, and they fall in the downdraft region below the cloud base. On the other hand, aggregates grow slowly in weak updraft between the cloud top and the ground[2)].

These fast changing conditions present a challenge to meteorologists. It is difficult to estimate the type and amount of precipitation from radar or satellite data alone. In order to examine the growth process of snowfall, it is essential to monitor the vertical structure of snow clouds from the ground and to analyze the relationship of these data.

This paper describes a system of several cloud observation devices installed at the same location with various instruments providing precipitation reference data on the ground level. The instruments operated using a temporal resolution of one minute or better, allowing to distinguish between different phases of an individual snowfall event. The data was stored in a relational database.

Using a database helps to manage the values from a large number of different instruments. Automatically generated visualizations give a quick look of the data during the whole time period. SQL language (Standard Query Language) provides a standard method to select and export data from the database for further analysis using external tools. The search capabilities of SQL can be used to find interesting precipitation events.

2. Instruments and Measured Elements

The Wakasa 2003 observation campaign was carried out at Fukui Airport, near Wakasa Bay, Japan, from January 13th to February 6th 2003, as a part of the AMSR/AMSR-E validation program[3)]. Measurements were done both at the ground level and at high altitudes (Fig. 1).

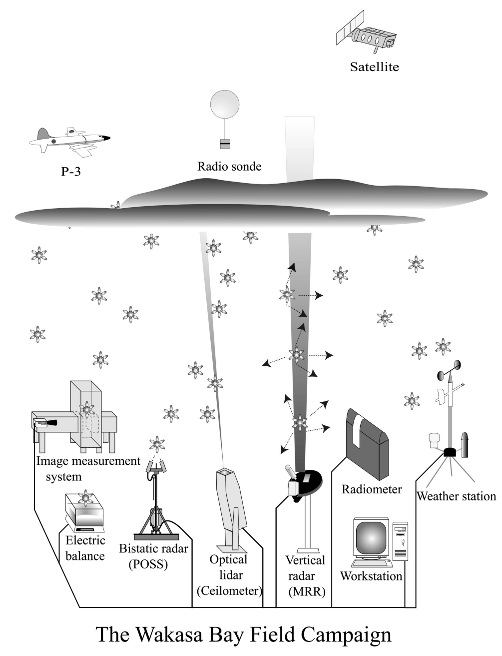

Fig. 1: Instruments used for observation

The following devices were installed at the ground level:

- Video camera and image processing based snowfall measurement system[4)5)]
    - number concentration of snow particles

  - size distribution
  - velocity distribution
- Electric balance
  - precipitation rate (calculated from weight data)
- Andrew POSS ground level bistatic radar
  - reflectivity, doppler velocity
  - estimated precipitation type and rate
- Metek MRR-2 vertically pointing doppler radar
  - reflectivity and doppler velocity up to 6000 m
  - rain rate, drop size distribution (calculated, for water only)
- Vaisala Ceilometer CT-25K optical lidar
  - optical backscatter profile up to 7500 m
  - cloud base height
- Radiometrics WVR1100 radiometer
  - brightness temperatures for 23.8 and 31.4 GHz
  - integrated water vapor and liquid water path
- Radiometer Physics GmbH RPG-LWP and RPG-TEMPRO 90 radiometers
  - brightness temperatures for 23.8, 36.5 and 90 GHz
  - oxygen line profile consisting of 10 frequencies between 50 and 60 GHz
- Snow heat capacity sensor
  - amount of heat to melt snow particles
- GPS sensor
  - integrated water vapor
- Weather stations (2 units)
  - rain rate by rain gauge
  - wind direction and strength
  - temperature, humidity, air pressure

In addition, radiosondes measuring temperature, humidity and wind profiles were released four times a day. NASA P3 Orion aircraft did periodic measurements using a dual frequency precipitation radar, a cloud radar and two radiometers. Data for the same region was also obtained from a Sankosha DWSR-2501C dual polarization doppler radar installed 10 kilometers away. The total amount of data produced by the instruments was about 2 GB / day.

3. Data Storage

3.1 Time Synchronization During the Observation Campaign

The timestamps of gathered data were synchronized with a server attached to a GPS clock. For most instruments, the timestamps were created by the computer storing the data. In those cases where it was not possible, the time synchronization of internals clocks of the devices was checked periodically.

After the observation campaign the data was in text files, with different naming schemes and mostly in formats specific to each instrument. This was a considerable inconvenience, especially considering that several groups of researchers were participating in the project, and each group was only familiar with the data formats of their own instruments. To achieve a more uniform access to data, the measurement values were transferred from the text files into a database.

3.2 Choosing the Database

In most database applications, it is known in the design phase how the data will be accessed. The main view to the system is normally an interface which hides all the internals of the data storage from users. However, in this case the users are scientists which will further analyze the data using external tools, such as spreadsheets, numeric computation packages and custom applications. Therefore, the role of the database system is to provide an overview of the data and then be an easy starting point to export data for use in other programs, or alternatively enable these programs to connect to the base for reading the data.

Relational databases[6)] provide a standard method for data retrieval using the SQL language (Standard Query Language). SQL allows to flexibly select the data which is needed in the current situation, limiting the query with conditions, for example specifying the desired time range. Relational databases are also supported by many generic graphical viewers and virtually all programming languages. Therefore, a relational system seemed appropriate for this project.

The relational model is based on two-dimensional tables, which have one value in each cell. Obviously, many types of data don't trivially fit into this model. The most popular databases today are all relational, but also include different object oriented features and other extensions to help users. Therefore, one important question is to choose whether to use strictly generic SQL or exploit database specific features.

In this project, extensions were avoided as much as possible to ensure interoperability. A popular open source database system PostgreSQL[7)] version 7.2.1 was chosen as the database engine in the initial implementation. However, it will be easy to change to another engine later if necessary.

3.3 Database Structure and Table Layout

The structure of the database is mainly organized by instrument. For each of them, there is at least one main table to store measurements, and another table for parameters which remained constant during the whole observation campaign. Depending on the type of device, there may be additional subtables for measurement values. Figure 2 shows an overview of the current structure of the base[1].

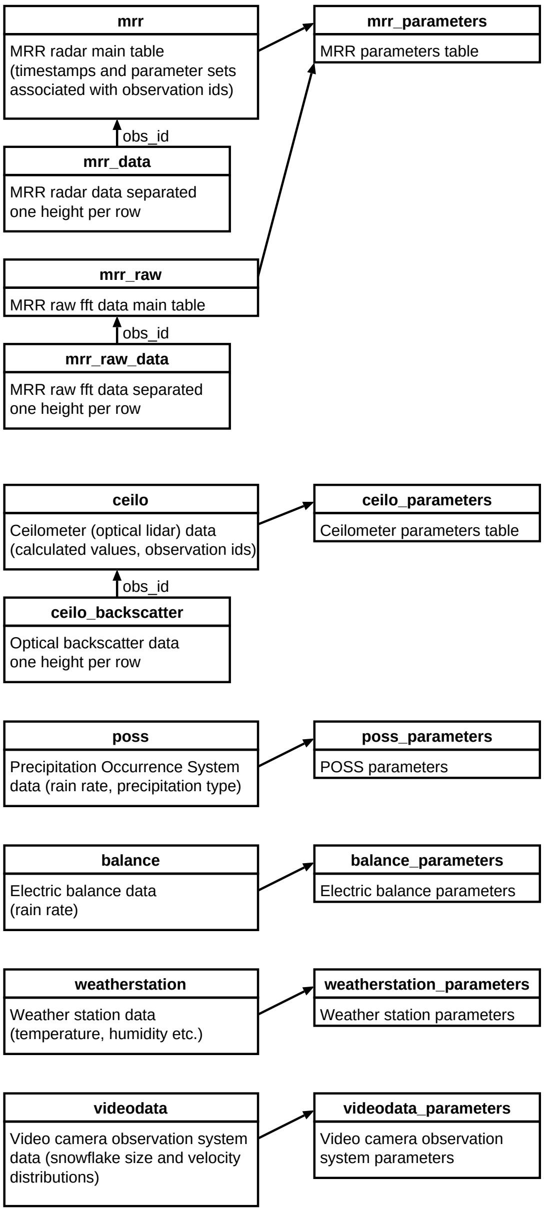

Fig. 2: Database structure overview.

In the main table of each device, one measurement is represented on one row. The unique identifier (key) of the row is the pair of a timestamp in Coordinated Universal Time (UTC) and a parameter set id number. Including the parameter set id as part of the key allows to run measurements with several similar devices simultaneously and still feed the results in the same table. Figure 3 shows the layout of tables for storing weather station data.

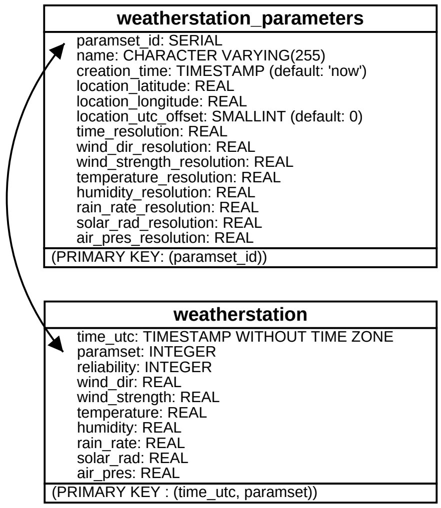

Fig. 3: Column names and data types in weather station data and parameter tables.

In most cases, the values of the parameters table are not required to interpret the measurements. For example, it is not necessary to know the time resolution of the device when reading data, but it may be useful for detecting possible gaps in measurement or unexpected observation intervals. Similarly, the location information is not yet used as all the devices were located in the same place, but may be useful in the future if data from several observations at different sites is inserted in the same database. Therefore, the majority of the values in parameter tables are optional, and can be left empty — marked as NULL in the database — if the correct values are not known when feeding the data in. NULL is also used in the main tables when a device produced incomplete data with some values missing from the output.

The reliability field is reserved for the future and has not been used yet. The idea is to provide a standard field to mark data which has been detected to be unreliable, or point out time ranges which have been manually checked and thus more reliable than others. Unreliable data can then be left out of analysis while still keeping it in the base. For example if heavy wind has caused a rain gauge to produce incorrect values they cannot be readily matched to other precipitation measurements, but one aspect of the research might be exactly to find out how the device failed and thus the data should be available when needed.

3.4 Modeling Matrix Values

In the Wakasa project, many devices do measurements of a vertical path in the atmosphere and produce an

[1]At this time, not all devices used during the observation have been included yet.

array of values, one for each height step. Additionally, several instruments give other types of array data such as FFT (Fast Fourier Transform) spectra, which combined with height information form a two-dimensional matrix for each time step.

These values cannot be directly inserted in one cell of a relational table while at the same time preserving the possibility to use standard SQL to select the desired range of the array.[2] The two main ways to handle such data are to encode it as one block inside the cell — losing the SQL search capabilities — and creating an additional table where one array is split on several rows.

In this case, it seemed important to be able to select values only from a certain height or from an arbitrary height range when retrieving data. Therefore the height arrays were split into separate tables, with one height on one row, and linked to the timestamps in the main table using observation ids (automatically generated by the database), as shown in Fig. 4. On the other hand, other arrays were encoded as strings containing comma separated values in table cells.[3]

Table "ceilo"

time_utc	paramset	obs_id	other columns...
14:01:00	1	1247	...
14:01:30	1	1248	...
14:02:00	1	1249	...
...	...	...	...

Table "ceilo_backscatter"

obs_id	height	bs
1247	30	0.0294
1247	60	0.0319
1247	90	0.3512
...	...	...
1248	30	0.0285
1248	60	0.0299
...	...	...

Fig. 4: Storing optical backscatter values from the optical lidar (Ceilometer).

This approach allows fetching individual values and specific height ranges quickly, but has shown performance problems when long time periods of the whole height spectrum is needed. The main issue is not the time consumed by the database for finding the right section of the table but rather the large number of rows to be processed when reading the data from the base. Therefore, depending on the type of use it may be preferable to store the height spectra as one block, either as text or binary string.

[2]Some databases provide this kind of functionality as SQL extensions.

[3]Depending on the database engine, this may cause the table size to grow quickly and cause performance degradation, but at least PostgreSQL stores internally such large data blocks separately and leaves only a pointer in the table.

4. Data Retrieval and Search Criteria

4.1 Basic Features

Flexible search functions in the SQL language are one key reason for using a database. Some types of selections and searches are naturally provided by the database:

- Selecting only the desired columns from a table.
- Imposing different conditions over one column, using logical and comparison operators

For example, to select timestamp and lowest cloud base height columns from the ceilometer table, limiting the retrieval to time between January 24 and 27, 2003, one can issue the following SQL command:[4]

```
SELECT time_utc, cb_height_1 FROM ceilo WHERE
       time_utc >= "2003-01-24 00:00:00" AND
       time_utc < "2003-01-28 00:00:00";
```

Similar conditions (equal to, greater than etc.) can be imposed on any column, thus limiting the rows included in the result. These operations are normally quite efficient. In the case of using mathematical operations on columns which are not part of the key — for example, the `time_utc` column is part of the key in all tables, but `cb_height_1` column in ceilo table is not — it may be necessary to manually instruct the database to create an index for the column. Without the index, the search works but less efficiently.

4.2 Comparing Data Between Tables

In an SQL query, values from two tables can be joined based on a column which is common to both. For example, it is trivial to join timestamps and cloud base height from table `ceilo` with height and optical backscatter values from table `ceilo_backscatter`, based on the observation id.

In principle, timestamps can be used to join tables between different devices, but this becomes a bit trickier. If the time stamps are different by only one second, the join operation does not place values from the source tables on the same row. SQL outer join can be used to get a result where the rows from two (or several) tables are interleaved and ordered by time, but naturally this is not as convenient as directly matching values.

To resolve this problem, one possibility would be to round timestamps to specific resolution when reading data in the base, for example one minute. The disadvantage of this is that the time resolution of fast devices

[4]The time stamp values are stored in UTC, this must be taken into account when giving the time range. The parameter tables contain a field location_utc_offset which can be used by scripts to alternatively use the local time at the place where the instrument was located when doing the measurements (which is not necessarily the same time zone than at the site where the analysis is done).

is reduced, while devices recording data slower (not providing a result every minute) would still have either gaps or duplicated data. Also, it would be difficult to choose the best resolution. Therefore, the data is initially stored with time stamps as exact, without rounding. For cross-analysis, scripts can be written which go through the whole range of values once doing the averaging at desired time resolution and write the results in other tables or files.

5. Precipitation Event Analysis

5.1 Overview

Monitoring the conditions in the clouds using radio sondes or aircraft relies on direct measurements of physical parameters. An obvious disadvantage of this method, although accurate, is that because of the quick change of cloud condition, continuous data of vertical structure of snow clouds is difficult to obtain. Satellites have the convenience of being able to observe a large continuous area, but for the interpretation of this data, field experiments at reference sites are essential.

In our setup, size and velocity distributions of snowflakes can be obtained from the video camera based measurement system. Snowfall rate on the ground level is given by the rain gauge, the electric balance and the POSS bistatic radar. The optical lidar measures cloud base height while the MRR-2 radar gives reflectivity profile and particle falling velocity up to 6000 m of altitude. All these operate with a time resolution of one minute or better, allowing us to distinguish between different phases of each snowfall event during analysis.

The radiometers retrieve total integrated water wapor and liquid water contents from ground level through the atmosphere. This can be compared with data from satellites, which are equipped with radiometer components.

5.2 Sharing Data Between Groups

The Wakasa 2003 observation was a joint project of several groups of researchers. Each group brought their own instruments and expertise to the observation site. The measurement data is shared between these groups.

The database can be easily accessed over a network. However, due to the large amount of data involved, it is not possible to have just one centralized server. It would be sufficient for quick looks of parts of the data, but exporting long time ranges of values over the Internet would not be feasible, as one export can grow to even hundreds of megabytes. The total amount of data in the base is currently about 3 GB.

Until now, the database has been in development and therefore not yet used as the data source for analysis. It is planned that each group would have one server installation in their local network, and each installation would contain all the data. When new data is added to the bases, the contents can be periodically mirrored between groups.

5.3 Automatically Generated Visualizations

The ability to generate automatic first phase quicklooks of the data from each instrument is an important advantage of the measurement database. This functionality has been implemented by some relatively simple scripts. Compared to text files, the database makes the task easier by avoiding the need of parsing through a set of files and finding the right time range and columns of data, as these can be specified using a short SQL query. The visualizations can be used to quickly see what was going on and which events to pick for more detailed analysis.

Figures 5 and 6 present an example of data gathered during January 28, 2003 from 17:00 to 19:00 JST. This period includes two clearly separated precipitation events, one from about 18:00 to 18:20 and another from 18:30 to 18:40.

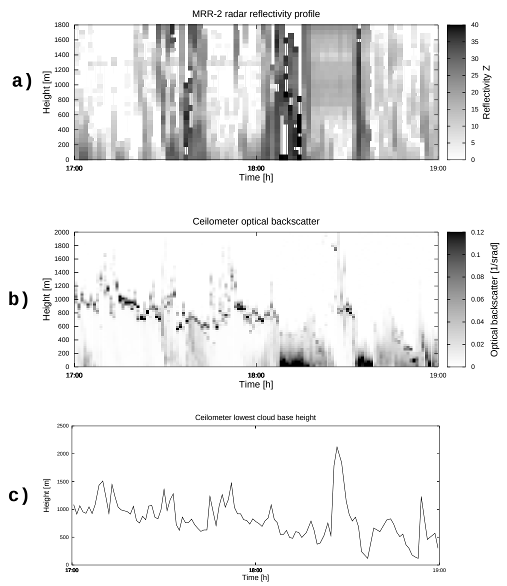

Fig. 5: Measurements during January 28, 2003 from 17:00 to 19:00 JST. a) MRR-2 reflectivity profile, b) Ceilometer (optical lidar) backscatter profile, c) Lowest cloud base height.

From the optical lidar cloud base data it can be seen that the area was covered by clouds during most of the period. However, just after the first precipitation event there is a short gap in the clouds. During this period also the wind calmed down before another cloud came over the site and precipitation started again. This can be clearly seen from the lidar backscatter profile. The velocity distributions give additional insight to the events. While the first event contains mainly small velocities, the second event shows a large number of relatively fast particles. This indicates graupel, which also matches our handwritten notes during that period.

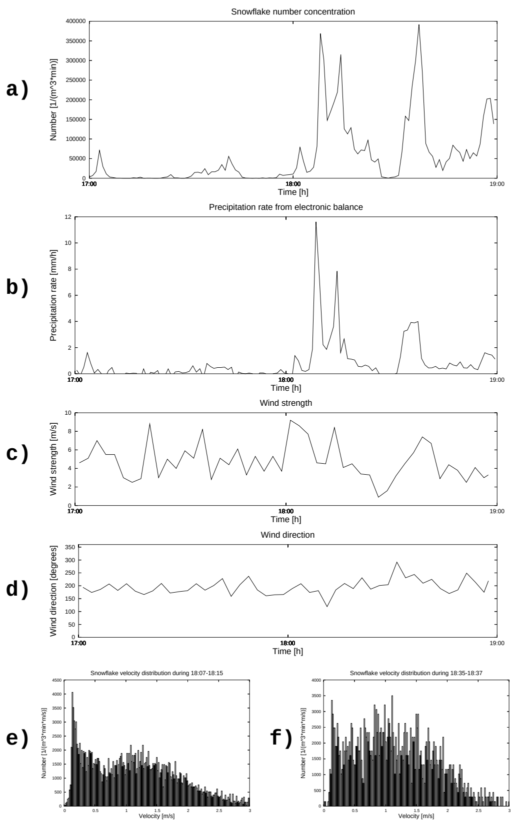

Fig. 6: Measurements during January 28, 2003 from 17:00 to 19:00 JST. a) Snowflake number concentration, b) Precipitation rate, c) Wind strength, d) Wind direction, e) Velocity distribution during 18:07-18:15, f) Velocity distribution during 18:35-18:37.

5.4 Observation Campaigns in the Future

The database layout has been designed so that data from future experiments can be added to the same tables. Parameter sets (stored in parameters tables) can be used to distinguish data from different experiments and different locations. Existing scripts can be used to generate visualizations already during the field work, which will also help detecting failures in equipment.

When new types of instruments are introduced, it is of course necessary to add new tables in the database for them and write a parser which will read the data in from files produced by the instrument. This will take some time, but only once. Indeed, the main goal of the database project is to make basic tasks easier, allowing the researchers to concentrate on the actual analysis and developing new algorithms.

Over time, as measurement data from future experiments is added, the system will develop into a large pool of data about rain- and snowfall. This will give a good opportunity to see how the precipitation events differ between various locations. It will also provide a good basis to develop more robust statistical algorithms for automatic classification of precipitation events.

6. Conclusion

We have proposed a database based system containing data from a wide range of different instruments for analyzing the relation between cloud structure and ground precipitation data during snowfall. The good temporal resolution of the setup allows us to distinguish between different phases of a single snowfall event.

We have designed a database for snowfall measurement and populated it with data from one large observation campaign. The system allows us to easily retrieve and visualize data simultaneously from all relevant instruments. This helps to understand the precipitation events which often occurred in complicated weather conditions. Various search criteria can also be used to find interesting events. Data from future experiments can be added to the same base, which will save time and give further opportunities to compare precipitation events between different locations.

7. Acknowledgement

The authors want to thank Fukui Airport Construction Planning Section for having the possibility to use Fukui airport facilities during the observation campaign.

References

[1] Coordinated Enhanced Observing Period home page. `http://monsoon.t.u-tokyo.ac.jp/ceop/`.

[2] T. Harimaya, Y. Nakai, "Riming growth process contributing to the formation of snowfall in orographic areas of Japan facing the Japan Sea", J. Meteorol. Soc. Japan, Vol.77, No.1, pp.101-115, 1999.

[3] AMSR/AMSR-E validation program home page. `http://www.ghcc.msfc.nasa.gov/AMSR/validation.html`

[4] T. Shiina, K. Muramoto, "Image Processing of Snow Particles and Classification into Snowflakes and Graupel", Systems and Computers in Japan, Vol.33, No.2, pp.43-53, 2002.

[5] H. Servomaa, K. Muramoto, T. Shiina, "Snowfall Characteristics Observed by Weather Radars, an Optical Lidar and a Video Camera", IEICE Trans. Inf. & Syst. Vol.E85-D, No.8, 2002.

[6] T. Connolly, C. Begg, "Database Systems: A Practical Approach to Design, Implementation and Management", Third Edition, Addison-Wesley, 2002.

[7] PostgreSQL database system home page. `http://www.postgresql.org`