CHAPTER I

INTRODUCTION

A set of data is called a sparse data set if most of the data have zero values and these zero value data occur in a long sequence of time. To ensure the accuracy, the zero value data must be imputed prior to the prediction. However, this problem of sparse data is more difficult than the problem of imputing missing or incomplete data reported in [1, 5], only some small portions of data are incomplete or missing and they do not occur in a long sequence of time. Although, several efficient imputing algorithms have been reported to cope with incomplete data successfully, they are not efficient enough to impute the sparse data if determination of "missing" data or "no rain" condition has not been revealed prior to the data imputation. It is an assumption that to enhance the greater degree of accuracy and reliability of the data imputation, "missing" data or "no rain" condition must be determined.

1.1 Problem Identification

Problems of related algorithms of missing data are categorized into two approaches-- statistical and neural approaches-- as follows. This topic-related research is undertaken in reference to statistical computations [2], namely, the most popular expectation maximization (EM) algorithm designed by Demspster, Laird, and Rubin [3] and the fuzzy c-means clustering employed to evaluate the missing data by Hathway and Bezdek [4]. Supervised neural networks are also applied to solve the missing data problem, or a classification problem, in nearly all neural approaches. Training any supervised neural network involves not only input but also target data. The chance for missing data seems possible in these three

main parts: the missing input data [6], the missing target data [7], and the missing both input and target data [8].

In this type of topic-related research, two aspects of data prediction - time-series and time-independent - are generally emphasized. Due to the relative complexity and frequent occurrence in nearly all prediction problems, time-series aspect is the only focus on this study [1, 9]. The time-series prediction using a neural network approach is probable, considering its potential for statistic inference in the time-sequence. Nevertheless, we should be aware that using this method to achieve time-series prediction with missing data could be a problem. This is due to the incompatibility between two types of equipment for data collection and the inaccuracy of the data collecting process.

1.2 Objectives

The main goals of this study are:

1. To impute rain data from radar reflectivity corresponding to spatial and temporal variability of gauge rain intensity.

2. To develop an algorithm to improve the accuracy of the estimation of $Z_e - R$ relation.

1.3 Scope of Work and Constraints

The following conditions are considered in this study:

1. Every precipitating cumulus cloud produces a rain shaft that occupies a certain area of rain gauge location.

2. Reflectivity at radar's lowest elevation step that is close to the cloud base level is set at 2.5 km.

This dissertation is organized as follows. Chapter II reviews the literature. Chapter III describes the theoretical background. Chapter IV describes the problem definition and determination of "Missing" or "No Rain" condition. Chapter V gives the detail of our proposed algorithms. The experimental results of our evaluation are given in Chapter VI. Chapter VII concludes the dissertation.