CHAPTER VI

EXPERIMENTAL RESULTS

6.1 Data Collection

Radar and rain gauge observations were collected during Applied Atmospheric Resource Research Program (AARRP) experiment conducted over Omkoi radar site Chiang Mai, Thailand in May 1996. The National Weather Service Weather Surveillance Radar-1988 Doppler (WSR-88D) with a recording resolution of 5 minute collected the radar data and 40 tipping bucket rain gauges with a recording resolution of 1 min collected the rain gauge data. The 3-D radar data are collected every five minutes sequence of radar maps in natural spherical coordinate, which are range, azimuth, and elevation angle. The typical dimension of such a spherical grid point is 1.0 km along the radial by 1° azimuth span by 0.7° - 1.0° vertical span. This spherical matrix is reduced to a polar matrix consisting of range and azimuth with the same resolution.

6.2 Preprocessing of Training and Testing Data

A representative training data set consisting of the radar data and corresponding ground rain gauge data are needed to develop a backpropagation neural network for rainfall prediction problem. The values of radar reflectivity corresponding to the centered and neighboring gauges are applied to the network as the input and the values of ground rain gauge as the output (target). The input vector is scaled so that the elements in the input vector are of similar magnitude. For instance, the value of reflectivity between 0 and 60 is normalized with the possible maximum reflectivity value. A *log* function transformation is then calculated to the 5-minute rainfall intensity values to provide a target output between 0 and 1, and to define the value of the output layer of the network.

6.3 Imputation Efficiency and Robustness

Table 6.1: Imputation Efficiency and Robustness among Expectation Maximization (EM), Similarity Measure (SM) and Neural Network (NN) for Gauge no. 071 and 081.

No. of Missing	% of Missing	Correlation Coefficient Comparison of					
		"No Rain" Considered			Without "No Rain" Considered		
		EM	**SM**	NN	EM	**SM**	NN
5	5	0.8517	**0.9547**	0.8903	0.8274	**0.9491**	0.8700
10	10	0.8073	**0.9113**	0.8682	0.7656	**0.9102**	0.8195
15	16	0.7778	**0.8588**	0.7729	0.7058	**0.8549**	0.8037
20	21	0.7554	**0.7690**	0.7331	0.6495	**0.7706**	0.6686
25	26	0.6324	**0.7475**	0.6077	0.5959	**0.7361**	0.6321
30	31	0.5945	**0.7471**	0.5724	0.4957	**0.5586**	0.4503

The algorithm was tested with two experiments. In first experiment, the comparisons of the proposed missing data estimation with the other techniques are summarized in Table 6.1. Two aspects, robustness and estimation techniques, are considered. The robustness of the proposed techniques is measured by gradually increasing the percentage of missing data and measuring the correlation coefficient between the complete data manifold and incomplete data manifold. The correlation coefficient, $r_{x,y}$, between manifolds $X = [x_1, x_2, \ldots, x_n]^T$ and $Y = [y_1, y_2, \ldots, y_n]^T$ is defined as follows.

$$r_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where $\bar{x}$ and $\bar{y}$ are the mean of X and Y, respectively. There are 48 patterns in the considered window. This magic number 48 is from the experiment. The window of this size gives the maximum similarity measure. The first column of Table 6.1 indicates the number of patterns being marked as missing data patterns. The second column is the percentage of missing data patterns with respect to the window size of 48 patterns. The rest of columns are the "No Rain" considered comparisons of the correlation coefficients among EM (Expectation Maximization), SM (Similarity Measure with GRG), NN (Neural Network), and the without "No Rain" considered comparison of EM, SM, and NN.

Again, in the second experiment, the comparisons of the proposed missing data estimation with the other techniques are summarized in Table 6.2. Two aspects, robustness and estimation techniques, are considered. The robustness of the proposed techniques is measured by gradually increasing the percentage of missing data and measuring the correlation coefficient between the complete data manifold and incomplete data manifold.

Table 6.2 Imputation Efficiency and Robustness among Expectation Maximization (EM), Similarity Measure (SM) and Neural Network (NN) (Gauge no. 062, 063 & 073).

No. of Missing	% of Missing	Correlation Coefficient Comparison of					
		“No Rain” Considered			Without “No Rain” Considered		
		EM	**SM**	NN	EM	**SM**	NN
7	5	0.7704	**0.9857**	0.7785	0.7050	**0.9323**	0.6547
13	10	0.5806	**0.9716**	0.4040	0.4588	**0.8296**	0.4720
18	18	0.5868	**0.8996**	0.3497	0.3974	**0.5245**	0.2269
24	25	0.2827	**0.8451**	0.3427	0.2412	**0.3543**	0.1500
30	33	0.2635	**0.5692**	0.2828	0.2051	**0.1603**	0.1463

There are 72 patterns in the considered window. This magic number 72 is from the experiment. The window of this size gives the maximum similarity measure. The first column of Table 6.2 indicates the number of patterns being marked as missing data patterns. The second column is the percentage of missing data patterns with respect to the window size of 48 patterns. The rest of columns are the "No Rain" considered comparisons of the correlation coefficients among EM (Expectation Maximization), SM (Similarity Measure with GRG), NN (Neural Network), and the without "No Rain" considered comparison of EM, SM, and NN. Then, I will give an example of how to apply the new obtained technique on actual data. The results are illustrated in Table 6.3.

Table 6.3: An example of a piece of manifold of imputing large missing rain data.

time	Z_1	Z_2	Z_3	G_max	time	Z_1	Z_2	Z_3	G_max
409	0.60833	0.45833	0.53333	0.31895	1200	0.275	0.60833	0.65	0
410	0.46667	0.5	0.56667	0.31895	1201	0.49167	0.51667	0.59167	0.42247
411	0.50833	0.575	0.45833	0.31895	1202	0.46667	0.58333	0.59167	0.49814
412	0.425	0.55833	0.55	0.31895	1203	0.475	0.51667	0.625	0.42247
413	0.58333	0.6	0.44167	0.44383	1204	0.50833	0.6	0.49167	0.31345
414	0.65	0.55	0.41667	0.31895	1205	0.53333	0.53333	0.48333	0.42247
415	0.65	0.575	0.475	0.44383	1206	0.43333	0.45833	0.43333	0.31345
416	0.5	0.55833	0.54167	0.53251	1207	0	0	0	0.31345
417	0.46667	0.59167	0.425	0.31895	1208	0.4	0.475	0.225	0
418	0.48333	0.575	0.525	0.44383	1209	0.5	0.525	0.26667	0.31345
419	0.60833	0.48333	0.55833	0.31895	1210	0.55	0.425	0	0.31345
420	0.51667	0.54167	0.425	0.44383	1211	0.5	0.325	0	0.44383
421	0.64167	0.46667	0.45	0.45162	1212	0.35	0.15833	0.23333	0.44383
422	0	0	0	0	1213	0	0	0	0.31345
423	0.50833	0.525	0.41667	0.29836	1214	0.21667	0.35	0.05833	0
424	0.50833	0.48333	0.325	0.31895	1215	0.15	0.275	0	0.31345
425	0.35	0.40833	0.31667	0	1216	0	0.11667	0	0
426	0	0.25833	0.23333	0	1217	0.04167	0.15	0	0
427	0	0	0.29167	0.29836	1218	0.05833	0	0	0
428	0	0.23333	0.28333	0	1219	0.375	0	0	0
429	0	0.175	0	0	1220	0.33333	0	0	0.44383
430	0	0.25	0.28333	0.31895	1221	0.18333	0.05	0	0.31345
431	0	0	0	0	1222	0.21667	0	0	0
432	0	0	0	0	1223	0	0.06667	0	0
433	0	0.31667	0	0	1224	0	0	0	0
434	0	0.275	0	0	1225	0	0	0	0
435	0	0	0	0	1226	0	0	0	0
436	0	0.19167	0	0	1227	0	0	0	0
437	0	0.15833	0	0	1228	0	0	0	0
438	0.16667	0	0.23333	0	1229	0	0	0	0
439	0	0	0	0	1230	0	0	0	0
440	0.175	0	0.35	0	1231	0	0	0	0
441	0	0	0.45833	0	1232	0	0	0	0
442	0.4	0.20833	0.275	0.31345	1233	0.375	0	0	0
443	0.44167	0	0.23333	0.44383	1234	0.16667	0	0	0

Here, the left manifold of time data 409-480 with some missing values similar to similar to the right manifold of these data of time data 1200-1271 with maximum similarity measure is equal to 0.76. The first column of Table 6.3 indicates time sequences. The second, third, and forth column is normalized radar reflectivity corresponding to centered and neighboring gauges. The fifth column is normalized gauge rain intensity at centered gauge which maximum value of neighboring gauge. The yellow color attributes being marked as simulated to be missing. The brown color attributes being marked as filling-in missing using

SM (Similarity Measure with GRG). The purple color attributes being marked as filling-in missing data with EM (Expectation Maximization).

6.4 Testing Accuracy

6.4.1 Testing Determination "Missing" or "No Rain" Condition Accuracy

The algorithm was tested with two experiments. In first experiment, a data set of 2-hour radar and rain gauge pairs during specific time between 16:00 hours and 18:00 hours daily for one month is considered. In our experiments, out of 744 patterns the training set of effective radar reflectivity factor (Z_e) and gauge rain intensity (G) pairs is equal to 446 patterns which is the percentage of 60 and the testing set of random test is equal to 298 patterns which is the percentage of 40. The efficiency of determination of "no rain" or "missing" conditions applied by neural network classification is summarized in Table 6.4. The percentage of the classification being predicted as False-True is considered. The first column indicates the number of times of random test for each gauge no. 071 and 081. The second column is the percentage of False-True being predicted as NF (Negative False)--the values of radar are missing which is approximately the percentage of 0.4-1.4. The third column is the percentage of False-True being predicted as PF (Positive False)--the values of rain gauge are missing which is approximately the percentage of 2.5-11.2. The forth column is the percentage of False-True being predicted as BR (Both Rain)--the values of radar and rain gauge are not missing which is approximately the percentage of 0.9-8.4. The fifth column is the percentage of False-True being predicted as NR (No Rain) which is approximately the percentage of 0-75.3. The rest of columns are the percentage of False-

True prediction of NF, PF, BR and NR which is approximately the percentage of 0.4-0.5, 3.0-10.1, 0.3-7.0 and 0-78.6, respectively.

Table 6.4: Efficiency of Classification of "Missing" or "No Rain" Condition using Neural Network for Gauge no. 071 and 081.

No. of Random Test	Gauge No. 071 % of False-True Prediction				Gauge No.081 % of False-True Prediction			
	NF	PF	BR	NR	NF	PF	BR	NR
1	0-2.0	1.6-9.6	1.2-8.8	0-76.8	1.2-0	2.4-9.6	0-6.0	0-80.0
2	1.7-1.7	1.7-10.8	1.3-9.6	0.73.3	0.8-0	3.8-8.3	1.3-7.9	0-77.9
3	0.4-1.6	2.0-10.0	1.2-9.2	0-75.5	0-0.4	4.0-11.2	0.4-4.8	0-79.1
4	0.4-1.6	2.9-11.8	1.2-7.3	0-74.7	0-0.4	2.4-9.0	1.2-6.5	0-80.4
5	0-1.2	2.4-9.8	1.2-8.5	0-76.8	0.8-0	2.8-9.3	0-8.9	0-78.0
6	0-1.7	3.3-12.5	0.4-7.9	0-74.2	0.8-0	2.5-9.2	0-9.6	0-77.9
7	0.4-0.4	4.2-10.5	0-7.1	0-77.4	0-1.3	4.2-10.9	0-5.9	0-77.8
8	0-1.2	1.7-12.9	0.4-8.3	0-75.5	0-1.7	2.5-10.0	0-7.1	0-78.8
9	0-1.2	3.3-14.2	0-8.1	0-73.2	0-0.8	2.4-11.4	0-8.9	0-76.4
10	0.8-1.2	2.0-10.0	1.6-8.8	0-75.5	0-0.4	3.2-12.0	0.4-4.8	0-79.1
Average	**0.4-1.4**	**2.5-11.2**	**0.9-8.4**	**0-75.3**	**0.4-0.5**	**3.0-10.1**	**0.3-7.0**	**0-78.6**

A result of determination of "missing" or "no rain" condition as the percentage of "missing" condition being predicted as False-True is approximately the percentage of 4-19 and the percentage of "no rain" condition being predicted as False-True is approximately the percentage of 0-77. In the second experiment, a data set of radar-rain gauge pairs during 22 rainy day intervals for one month is considered. Out of 1,500 patterns, the training set of effective radar reflectivity and gauge rain intensity pairs is equal to 900 patterns which is the percentage of 60 and the testing set of random test is equal to 600 patterns which is the percentage of 60. The efficiency of determination of "no rain" or "missing" conditions applied by neural network classification is summarized in Table 6.5. The first column indicates the number of times of random test for each gauge no. 062 063 and 073. The second column is the percentage of "no rain" being marked as +1. The third column is the percentage of False-True being predicted as "No Rain" which is approximately the percentage of 6-94, 8-92 and 7-93 for gauge no. 062, 063 and 073, respectively. The forth column is the percentage of False-True being predicted as "Missing" which results in no false, is approximately the percentage of 28, 23 and 24 for gauge no. 062, 063 and 073, respectively.

Table 6.5: Efficiency of Classification of "Missing" or "No Rain" Condition using Neural Network for Gauge no. 062, 063 and 073.

No. of Random Test	Gauge No. 062 % of False-True Prediction			Gauge No. 063 % of False-True Prediction			Gauge No. 073 % of False-True Prediction		
	% NR	No Rain NF-TRUE	Missing TRUE	% NR	No Rain NF-TRUE	Missing TRUE	% NR	No Rain NF-TRUE	Missing TRUE
1	77	5 - 95	23	80	9 – 91	20	73	9 – 91	27
2	72	7 - 93	28	77	6 – 94	23	77	6 - 94	23
3	68	7 – 93	32	72	10 – 90	28	77	6 – 94	24
4	72	5 – 95	28	83	6 – 94	17	76	9 – 91	25
5	71	7 – 93	29	75	5 – 95	25	82	7 – 93	18
6	72	6 – 94	28	77	9 – 91	23	70	4 – 96	30
7	74	6 – 94	26	82	10 – 90	18	76	9 - 91	24
8	72	6 – 94	28	79	8 – 92	22	78	6 – 94	22
9	72	5 – 95	28	74	5 – 95	26	79	3 – 97	21
10	70	6 – 94	30	77	10 – 90	23	72	8 – 92	28
Average	**72**	**6 – 94**	**28**	**78**	**8 – 92**	**23**	**76**	**7 – 93**	**24**

6.5 Enhancing Reliability in Radar Rainfall Estimates

The relationship of $Z_e - R$ (measured reflectivity-estimated rain intensity) depends on several factors such as inaccuracy of radar observation process, incompatibility between radar and ground rainfall observation, uncontrollable physical mechanism environments, and uncontrollable human and equipment factors and large variation in time and space make the limitation of $Z_e - R$ relationship derived from theoretical considerations alone. Matching the radar-measured reflectivity (Z_e) to collocated and synchronized rain gauge measurement of G should provide a $Z_e - R$ relationship that solves implicitly unknown factors. Calheiros and Zawadzki [20] using Z_e- raingauge measured instantaneous data points, G in mm/hr applied the probability matching method (PMM) - bypassed the sampling volume, timing and collocation problems altogether by matching the unconditional (on G > 0) probabilities of nonsynchronous Z and G datasets. Krajewski and Smith [21] made simulation experiments that showed that regression methods are still significantly superior, providing much large rain estimation accuracy and smaller bias as compare to PMM for estimating $Z_e - R$ relationships of synchronous Z and G datasets because PMM does not use the information of joint probability. Rosenfeld and Amitai [11] matching Z and G pairs that have the same cumulative probability has more accuracy as compare to regression method when the scatter is caused by timing and collocation error. Lack of perfect synchronization, which is a major source of scatter between measured Z-G pairs. Solutions to this problem are needed.

6.5.1 Algorithm for Similarity Alignment

A data set of 2-hourly radar-raingauge pairs during specific time between 16:00 hours and 18:00 hours daily for one month is considered. Two aspects, maximum similarity measure and correlation coefficient, are considered. Consider an example of similarity manifold matching which having maximum similarity measure (SM) ***before*** and ***after*** applied similarity alignment algorithm shown in Figure 6.1 and 6.2. The detail of similarity alignment algorithm is given as follows.

1. Similarity manifold matching with the same time domain behavior is used along 2-dimensional space. This is an important feature why applied neighboring gauge correlation approach.

2. To determine window size we consider convective rainfall type corresponding to maximum gauge rain intensity which is the same sequence. In our experiment, the window size is set at 2-hourly--24 patterns. The similarity measure is equal to 0.7815 and correlation coefficient is equal to 0.73 and 0.62 for gauge no. 071 and 081, respectively.

3. To solve time synchronization error caused by delayed time, we shift or slide 5-minute time of the similarity manifold matching to place maximum gauge rain intensity at the same position.

4. To solve collocation error caused by navigation mismatch, we only shift 5-minute radar reflectivity which having gauge rainfall measurement in sequence of time for both manifolds. The objective function is maximum similarity measure(SM). Here, the similarity measure is equal to 0.8864 and correlation coefficient is equal to 0.91 and 0.92 for gauge no. 071 and 081, respectively.

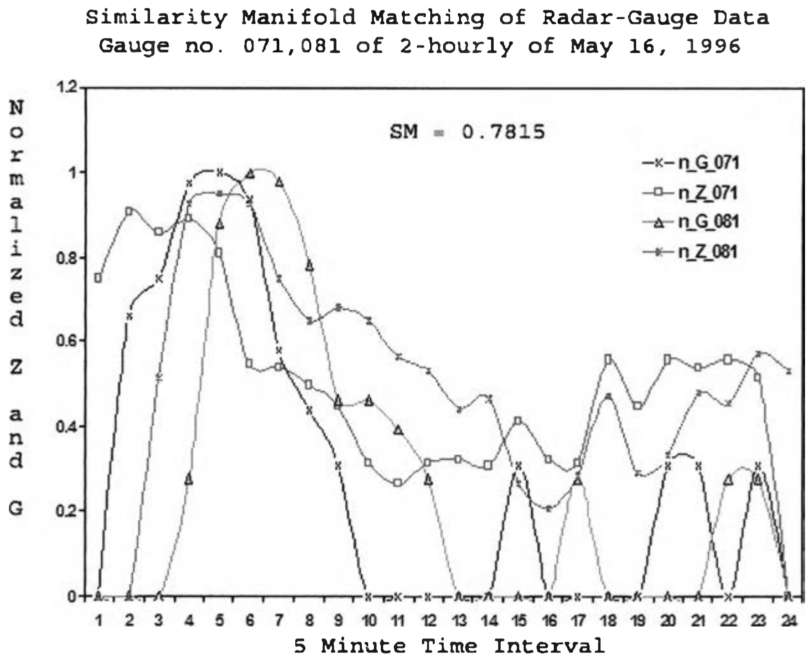

Figure 6.1: An example of similarity manifold matching which is having maximum Similarity Measure (SM) *before* applied similarity alignment algorithm.

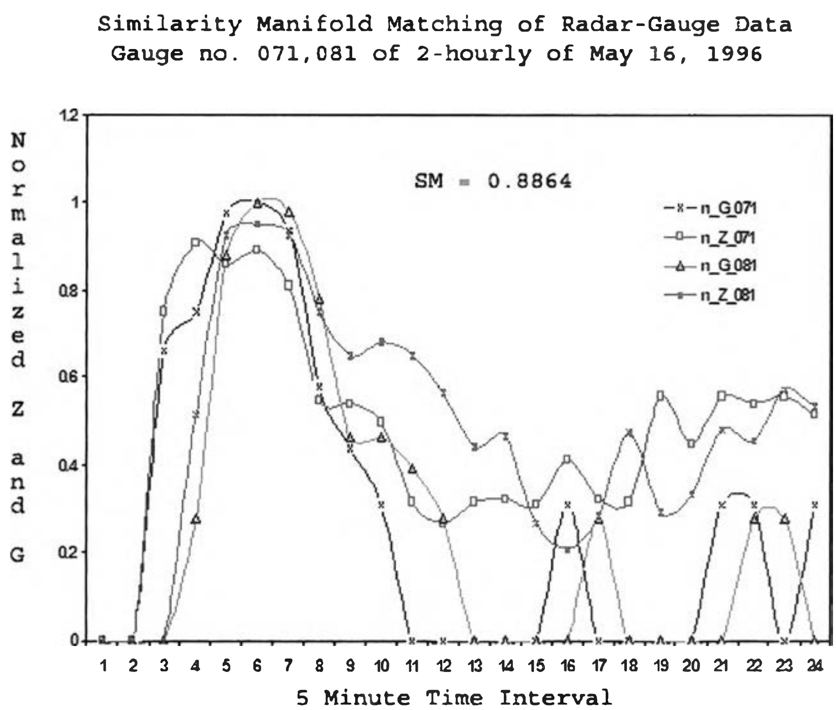

Figure 6.2: An example of similarity manifold matching which is having maximum Similarity Measure (SM) *after* applied similarity alignment algorithm.

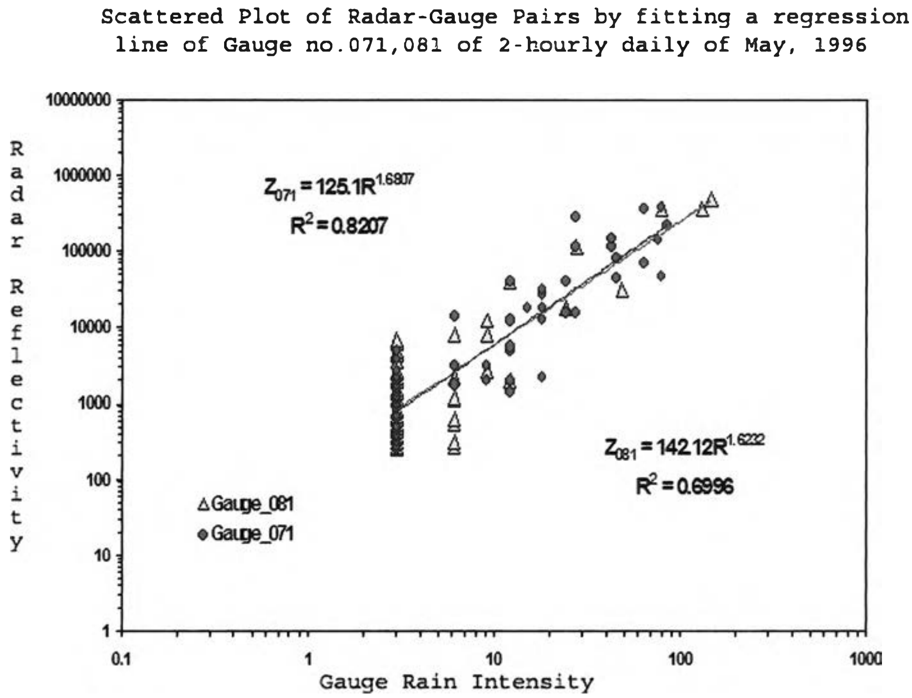

Figure 6.3: A result of 16:00-18:00 hours daily monthly of a difference Z-R parameter at cloud base with radar reflectivity threshold value of 24 dB.

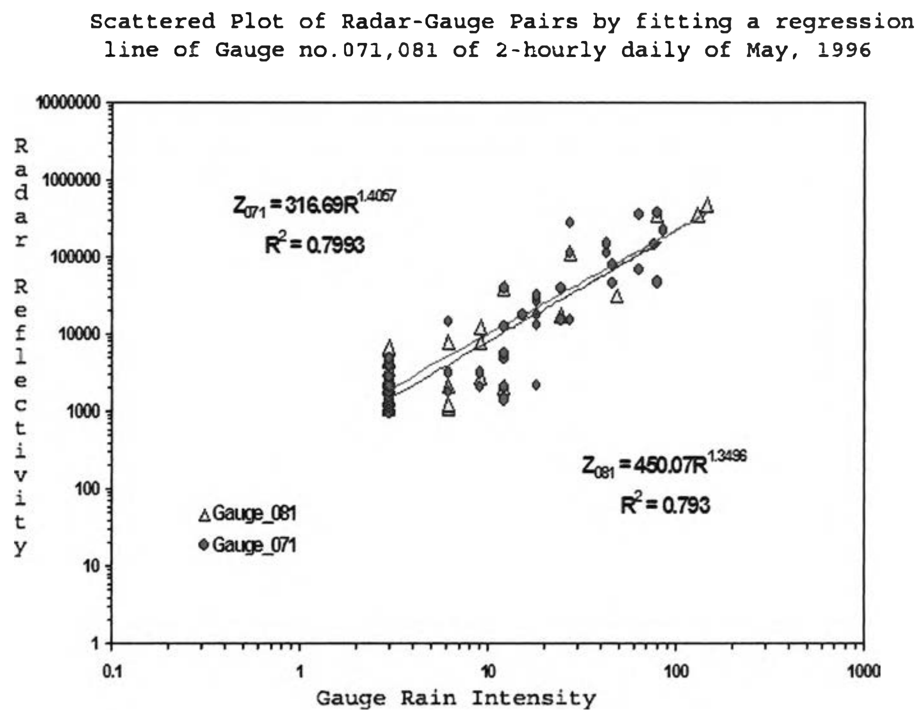

Figure 6.4: A result of 16:00-18:00 hours daily monthly of a difference Z-R parameter at cloud base with radar reflectivity threshold value of 30 dB.

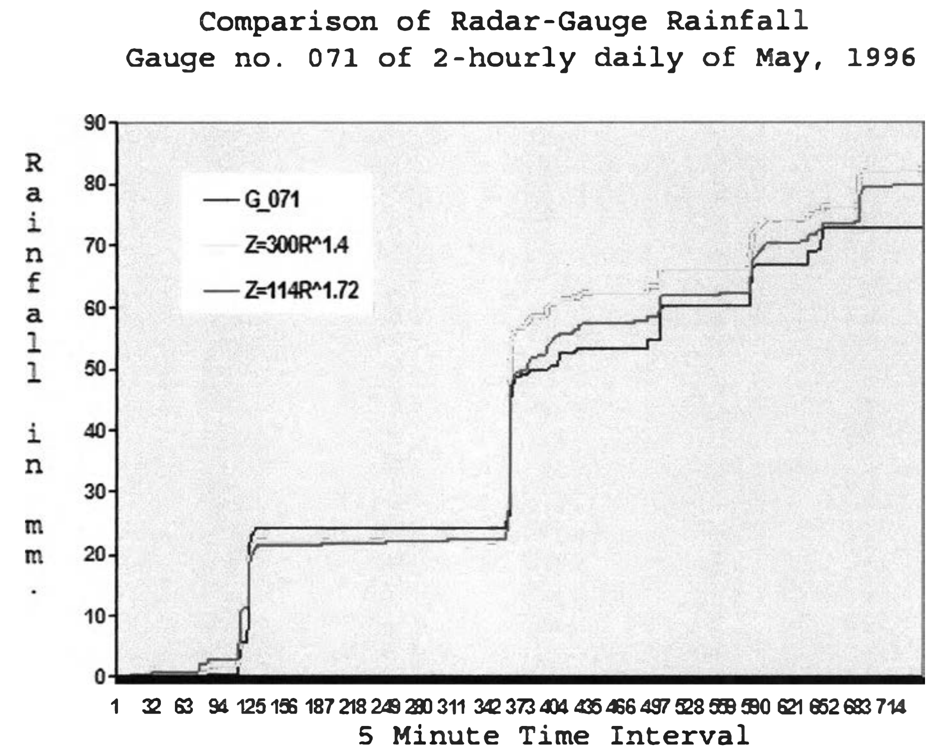

Figure 6.5: A result of 16:00-18:00 hours daily monthly comparison of radar-gauge rainfall accumulation for gauge no. 071.

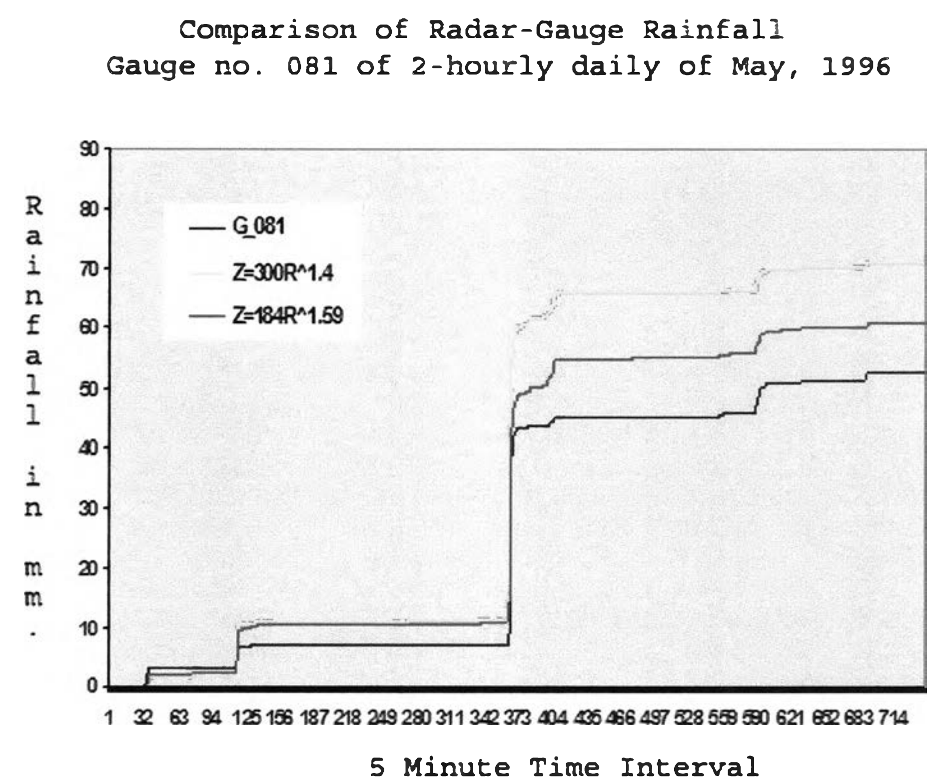

Figure 6.6: A result of 16:00-18:00 hours daily monthly comparison of radar-gauge rainfall accumulation for gauge no. 081.

One application of determination of "missing" or "no rain" condition prior to data imputation is to classify type of convective and stratiform rainfall which results in both reducing $Z_e - R$ conversion error in radar rainfall estimates as illustrated in Figure 6.3 and 6.4, and radar-gauge rainfall accumulation as shown in Figure 6.5 and 6.6. Consider a result of a difference $Z_e - R$ parameter of Figure 6.3 and 6.4. As known a $Z_e - R$ relationship, the rainfall rate R is measured in units of millimeters per hour (mm/hr). $Z_e - R$ Relationships are somewhat variable; Batten [17] states that fairly typical relationships are as follows: for stratiform rain, $Z = 200R^{1.6}$; for orographic rain, $Z = 31R^{1.7}$; for thunderstorm rain, $Z = 486R^{1.37}$. In our studies, a result of 16:00-18:00 hours daily monthly of a difference $Z_e - R$ parameter at cloud base with 24 dB of radar reflectivity threshold value is illustrated in Figure 6.4. As a result of 16:00-18:00 hours daily monthly of a difference $Z_e - R$ parameter at cloud base with 30 dB of radar reflectivity threshold value is illustrated in Figure 6.5. The $Z_e - R$ relationships are as follows: for orographic rain, $Z = 125R^{1.68}$; for stratiform rain, $Z = 142R^{1.62}$; for convective rain, $Z = 317R^{1.41}$; for thunderstorm rain, $Z = 450R^{1.35}$.