HYDROLOGICAL PROCESSES
Hydrol. Process. **28**, 5081–5092 (2014)
Published online 28 August 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.9972

Interpretation of mean-field bias correction of radar rain rate using the concept of linear regression

Chulsang Yoo,* Cheolsoon Park, Jungsoo Yoon and Jungho Kim
School of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul 136-713, Korea

Abstract:

In this study, the correction problem of mean-field bias of radar rain rate was investigated using the concept of linear regression. Three different relationships were reviewed for their slopes to be used as the bias correction factor: Relationship 1 (R1) is based on the conventional linear regression, relationship 2 (R2) is forced to pass the origin and relationship 3 (R3) is the line whose slope is the *G*/*R* ratio. In other words, R1 is the regression line connecting the intercept and the mass centre of measurement pairs, R2 is the regression line forced to pass the origin, and R3 is the line connecting the origin and the mass centre. The slopes of all three relationships were reviewed analytically to compare them, and thereby, the effect of zero measurements could be evaluated. Additionally, the effect of using switched independent and dependent variables on the derived slopes was also evaluated. The theoretically derived results were then verified by analysing the rainfall event on 10–11 August 2010 in Korea. Finally, the difference between the bias-corrected radar rain rate and the rain gauge rain rate was quantified by root mean square error and mean error so that it could be used as a measure for the evaluation of bias correction factors. In conclusion, the slope of R2 was found to be the best for the bias correction factor. However, when deciding the slope of this R2, the radar rain rate should be used as the independent variable in the low rain rate region, and the rain gauge rain rate in the high rain rate region above a certain threshold.

KEY WORDS *G*/*R* ratio; radar rain rate; rain gauge rain rate; regression line

Received 2 August 2012; Accepted 8 July 2013

INTRODUCTION

Finding and correcting the bias of radar rain rate is one of the most important problems regarding the use of radar rain rate. Many researchers have been focusing on this problem for the past few decades. However, as mentioned by Krajewski and Smith (2002), this issue is not yet theoretically established, and thus, the bias correction of radar rain rate is still dependent on empirical methods. This study also concerns this systematic bias.

In fact, there are many different kinds of errors involved in the radar measurements. For example, one can find the error in that the relationship between the water drops within the air and the echo intensity is not linear and that they cannot be consistently related (Chiu *et al*., 1990; Kummerow, 1998; Lafont and Guillemet, 2004). Errors can also be introduced in relating the radar echo to the rainfall intensity (Creutin *et al*., 1997; Anagnostou *et al*., 1998; Steiner *et al*., 1999). Curvature of the Earth′s surface may be another reason why the radar information cannot be fully related to the rainfall on the ground, especially in a remote place from radar (Joss and Waldvogel, 1990). The mean-field bias of interest in this study can also be explained by different error sources (Borga *et al*., 2000, 2002). The mean-field bias is, in principle, dependent upon electrical calibration error and bias in the radar reflectivity measurement. Additionally, it may be combined with other kinds of bias (systematic errors in space), which most of the time are related to height sampling errors that result in a range-dependent bias (Fulton *et al*., 1998; Vignal and Krajewski, 2001).

The *G*/*R* ratio (the ratio between the ground rain gauge rain rate and the radar rain rate) may be the most general form currently used for removing the mean-field bias of the radar rain rate. This form of bias correction factor has been considered by many researchers such as Wilson (1970), Brandes (1975), Collier *et al*. (1983), Collier (1986), Lin (1989), Smith and Krajewski (1991), Dinku *et al*. (2002), and Chumchean *et al*. (2006). In fact, using the *G*/*R* ratio as the form of the bias correction factor seems very reasonable when considering the wide variation of rain rate. However, the bias is intrinsically defined as the difference between the true and sample means. The fact that the bias correction factor takes the form of the ratio between the rain gauge rate and the radar

*Correspondence to: Chulsang Yoo, School of Civil, Environmental and Architectural Engineering, College of Engineering, Korea University, Seoul 136-713, Korea.
E-mail: envchul@korea.ac.kr

rain rate instead of the difference between them indicates that both rain rates follow the log-normal distribution. This is because the logarithmic mean difference between the rain gauge and radar rain rates is converted into the ratio between means. Recently, Chumchean *et al*. (2003, 2006) suggested another form of bias correction factor. This correction factor is basically the same as the *G*/*R* ratio, but it is more faithful to the assumption that the data follow a log-normal distribution. That is, the *G*/*R* ratio was determined by considering both the logarithmic mean difference and the logarithmic mean dispersion, which is more theoretically accurate to convert the logarithmic mean difference into the arithmetic mean difference.

However, this log-normality assumption can cause a serious problem, which is how to handle the no-rain (i.e. zero) measurements. We believe that Seo *et al*. (1996) was the first research to contain comments on zero measurements. In fact, they recommended using only positive measurement pairs of radar and rain gauge rain rates. Their recommendation was based on the fact that the impact of zero measurements varies depending on their relative prevalence, causing inconsistency in the estimation of the bias correction factor. In Seo (1998), this issue was examined once again to emphasize the risk of simply considering all the data including the zero measurements without any theoretical review of the role of zero measurements. Their recommendation was to use only the positive measurement pairs when calculating the bias correction factor, which was the same conclusion reached by Fulton *et al*. (1998), Seo *et al*. (1999), Anagnostou and Krajewski (1999), and Seo and Breidenbach (2002). However, Yoo *et al*. (2010) showed that the exclusion of zero measurements can cause another bias when comparing the radar rain rate with the rain gauge rain rate. However, the role of zero measurements varies depending on their prevalence, which should also be explained by the complex bivariate mixed log-normal distribution. The effect of zero measurements on the bias correction factor needs to be more simply and clearly explained.

One possible way of explaining the bias correction factor can be found in the regression analysis. Simply, the *G*/*R* ratio can be interpreted as the slope of the line connecting the origin and the mass centre of measurement pairs. Seo *et al*. (1996) also tried to use the slope of a regression line forced to pass the origin. Accepting the possibility of a negative rain rate, the linear regression line itself can also be used as the bias correction factor. By interpreting the bias correction factor as the slope of one of these lines, the effect of zero measurement can be fully explained. In particular, it may be possible to check if the form of the *G*/*R* ratio is valid enough to be used as the bias correction factor.

In this study, the effect of zero measurements on the bias correction factor was investigated using the slopes of three different relationships: Relationship 1 (R1) is based on the conventional linear regression, relationship 2 (R2) is forced to pass the origin, and relationship 3 (R3) is the line whose slope is the *G*/*R* ratio. In other words, R1 is the regression line connecting the intercept and the mass centre of measurement pairs, R2 is the regression line forced to pass the origin, and R3 is the line connecting the origin and the mass centre. All of these three slopes can be reviewed analytically and compared, and thereby, the effect of zero measurements can be evaluated theoretically. Additionally, it is also interesting to compare the effect of using switched independent and dependent variables when deriving the slopes. Theoretically derived results in this study were then verified by analysing the rainfall event on 10–11 August 2010 in Korea. The radar data used in this study are those from the Gudeoksan Radar, and the rain gauge data are from 95 rain gauges within the radar umbrella.

This paper is composed of a total of five sections including the Introduction and the Summary and Conclusion. The next section explains the data structure of radar and rain gauge rain rate, and the third section explains the theoretical background of three bias correction factors, or the slopes of three different lines considered in this study. The fourth section provides an application example with the rain rate data measured by the radar and rain gauges during 10–11 August 2010.

DATA PAIRS OF RADAR AND RAIN GAUGE RAIN RATES

Theoretically, depending on whether to consider the no-rain (or zero) measurements, there are four different methodologies available to construct the data pairs of radar and rain gauge rain rates. Among them, the methodology of using only positive measurement pairs (positive–positive pairs) and that of using all the measurement pairs including zero–zero pairs are more common. The other two methodologies are rather uncommon. For example, additional consideration of the data pairs of positive radar rain rates and zero rain gauge rain rates (positive–zero pairs) is theoretically feasible, but it is generally preferred to use all the data pairs including the zero–zero pairs. Even the extreme case of considering the data pairs of zero radar rain rates and positive rain gauge rain rates (zero–positive pairs) is, in fact, not feasible theoretically as the radar rain rate should be positive when the rain gauge rain rate is positive.

Among these four methodologies available to construct the data pairs, this study will focus on the two common methodologies that are assumed more practical. That is, the first methodology is to consider only the positive measurement pairs, and the second is to consider all the measurement pairs including no-rain measurements. Three bias correction factors considered in this study will also be applied separately to these two cases of data pairs to evaluate the effect of zero measurements as well as to compare the bias correction factors.

THEORETICAL ASPECTS OF THREE BIAS CORRECTION FACTORS

Bias correction factors for the positive measurement pairs

The analysis of the bias correction factors becomes relatively simple when only positive measurement pairs are used. As mentioned in the Introduction, the bias correction factor can be assumed to be the slope of a certain relationship between the radar rain rate and the rain gauge rain rate. First, we consider the regression line (R1), whose functional form is as follows:

$$y = \alpha + \beta x \tag{1}$$

where the radar rain rate is assumed to be the independent variable x and the rain gauge rain rate the dependent variable y. Also, α is called the intercept and β the slope. In fact, the regression line R1 passes the intercept and the mass centre $(\bar{x}, \bar{y})$ of the measurement pairs. The slope β is simply determined as $(\bar{y} - \alpha)/\bar{x}$; thus, both very low and very high rain rate measurements have a large effect on the slope of the regression line. For a given set of measurement pairs, the method of least squares is generally applied for the estimation of the two model parameters α and β:

$$\beta = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2}$$

$$\alpha = \bar{y} - \beta \bar{x} \tag{3}$$

Among the various characteristics of this regression line R1, the most important characteristic to be noted in this study is that the slope of this regression line R1 is determined by the weighted ratio between the dependent variable and the independent variable about the mass centre. The independent variable is used as the weighting factor. Thus, the roles of the independent variable and the dependent variable are different, and, obviously, the effect of the independent variable is more crucial to the slope. As the radar rain rate is used as the independent variable and also because the radar rain rate is generally far less than the rain gauge rain rate (Nam *et al.*, 2003; Jung *et al.*, 2005; Hong *et al.*, 2009), the derived slope is more gradual than that of the opposite case, where the rain gauge rate variable is used as the independent variable.

As a second case, when the intercept of the aforementioned linear regression line is forced to be zero (R2), the slope can also be determined by applying the method of least squares.

$$\beta = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2 \tag{4}$$

The line R2 determined in this way no longer passes the mass centre of the measurement pairs. This slope is calculated similarly to the slope of the linear regression line but about the origin rather than the mass centre. The slope of R2 is thus more influenced by the data far away from the origin than that of the linear regression line R1. The influence of very small measurements on the slope is minimal in this case. However, similar to the previous case, the role of the independent variable is still more crucial to the slope. As the radar rain rate is used as the independent variable and also because the radar rain rate is generally far less than the rain gauge rain rate, the slope derived is much more gradual than that of the opposite case.

Lastly, the straight line that passes the mass centre of the measurement pairs and the origin can be considered (R3). The slope of this straight line R3 can be determined as follows:

$$\beta = \sum_{i=1}^{n} y_i \Big/ \sum_{i=1}^{n} x_i \tag{5}$$

Compared with the other two slopes considered earlier, this slope is calculated simply as the ratio between the sum of the dependent variables and the sum of the independent variables. There is no weighting factor considered in this slope calculation. A specific effect of the independent variable on the slope calculation cannot be expected in this case, which makes it different from the previous two cases. Thus, in contrast to the previous two cases, it makes no difference in calculating the slope whether the radar rain rate or the rain gauge rain rate is considered as the independent variable.

Bias correction factors for all the measurement pairs

In the case of using all the measurement pairs, both the independent variable and the dependent variable include zero measurements to a certain degree, even though their relative proportions are different. This section examines how these zero measurements influence the slopes of the three lines considered.

First, Equation (2), the slope of the linear regression line R1, can be expressed a bit differently as follows:

$$\beta = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \tag{6}$$

As discussed before, this slope has the form of a weighted ratio between the dependent and independent variables. As it is represented by the means of the independent and dependent variables, along with the number of measurement pairs, it is not complicated to determine the impact of zero measurements on the slope. If the number of (+ radar rain rate, + rain gauge rain rate) measurement pairs is n, the number of (+, 0) pairs is m_1, the number of (0, +) pairs is m_2, and the number of (0, 0) pairs is m, then the slope of the regression line R1 is calculated as follows:

$$\beta' = \frac{\sum_{i=1}^{n}(x_i - \bar{x}')(y_i - \bar{y}') - \sum_{j=1}^{m_1}\bar{y}'(x_j - \bar{x}') - \sum_{k=1}^{m_2}\bar{x}'(y_k - \bar{y}') + \sum_{l=1}^{m}\bar{x}'\bar{y}'}{\sum_{i=1}^{n}(x_i - \bar{x}')^2 + \sum_{j=1}^{m_1}(x_j - \bar{x}')^2 + \sum_{k=1}^{m_2}(\bar{x}')^2 + \sum_{l=1}^{m}(\bar{x}')^2} \tag{7}$$

In the preceding equation, the means of the independent and dependent variables are different from those in the case considering only (+, +) pairs. The new slope is dependent on the relative proportions of zero measurements in the given data pairs. As a simple case, if zero measurements exist with equal prevalence in both the independent and dependent variables, then, as only m (0, 0) pairs are added to n (+, +) pairs, the slope of the regression line R1 becomes changed as follows:

$$\beta' = \frac{\sum_{i=1}^{n}(x_i - \bar{x}')(y_i - \bar{y}') + \sum_{l=1}^{m}\bar{x}'\bar{y}'}{\sum_{i=1}^{n}(x_i - \bar{x}')^2 + \sum_{l=1}^{m}(\bar{x}')^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{n^2}{n+m}\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - \frac{n^2}{n+m}\bar{x}^2} \tag{8}$$

In the preceding equation, $\beta' = \beta$ when $m = 0$. However, $\beta' < \beta$ when $m > 0$ and $\bar{x} > \bar{y}$, and $\beta' > \beta$ when $m > 0$ and $\bar{y} > \bar{x}$. As the mean radar rain rate is generally smaller than the mean rain gauge rain rate, the new slope calculated by considering the zero measurements is less than that calculated by considering only positive measurement pairs.

As discussed in the previous section, for the case in which the intercept of the linear regression line is forced to zero (that is, R2), the slope is derived as in Equation (4):

$$\beta = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n} x_i^2$$

This slope will not be changed when only (0, 0) pairs are added to the (+, +) pairs. However, when different proportions of zero measurements are added to the independent and dependent variables, this slope will change. More generally, if the number of (+, +) pairs is n, the number of (+, 0) pairs is m_1, the number of (0, +) pairs is m_2, and the number of (0, 0) pairs is m, this slope is calculated as follows:

$$\beta' = \sum_{i=1}^{n} x_i y_i \Big/ \sum_{i=1}^{n+m_1} x_i^2 \tag{9}$$

As can be seen in the preceding equation, only the (+, +) pairs are considered in the numerator. However, in the denominator, both (+, +) and (+, 0) pairs are considered. Thus, the new slope β' is always smaller than or equal to β regardless of whether the radar or the rain gauge rain rate is used as the independent variable.

Finally, the slope of the line passing the origin and the mass centre of the data (that is, R3) is calculated as follows:

$$\beta' = \left(\sum_{i=1}^{n} y_i + \sum_{j=1}^{m_2} y_j\right) \Big/ \left(\sum_{i=1}^{n} x_i + \sum_{j=1}^{m_1} x_j\right) \tag{10}$$

Obviously, the aforementioned slope depends on the relative proportions of zero (or positive) measurements in the independent and dependent variables. In an ideal case, as the rain gauge measures a point rain rate whereas the radar measures an areal average rain rate, the radar data will contain more positive measurements than the rain gauge data. Therefore, when taking the radar rain rate as the independent variable, the new slope β' is always smaller than β.

Effect of using switched independent and dependent variables

The bias correction factor can be determined as one of the slopes examined earlier. However, there is another problem requiring further examination. This problem is the effect of using switched independent and dependent variables. The switching problem described in this section refers to whether a different slope is derived when considering switched independent and dependent variables. In this case, the significance of the derived slope may or may not be the same as that of the original slope.

More fundamentally, it should also be reviewed what could be expected when switching the independent and dependent variable. In the regression analysis, the error involved in the independent variable is assumed negligible. However, this is not true for both the radar and rain gauge rain rate. The radar rain rate is the remotely sensed one given as the form of areal average value over the radar bin. The rain gauge rain rate seems more accurate but has the problem of representing the radar bin for the comparison. In general, the radar rain rate is used as the independent variable, as the purpose of the bias correction is to obtain the bias-corrected radar rain rate.

The *G*/*R* ratio is applied to simply balance the radar and rain gauge rain rate. There is not any concept of independent and dependent variables involved. Even though the cross-correlation between the radar and rain gauge rain rate is high, their relation is not a causal connection. As the goal of applying the *G*/*R* ratio is simply to obtain the bias-corrected radar rain rate, it is more important to obtain better bias-corrected radar rain rate. This is what is reviewed in this part of the study.

With the rain gauge data as the independent variable and the radar data as the dependent variable, one can consider the regression line as reviewed in previous sections.

$$x = \alpha' + \beta' y \tag{11}$$

Now, the inverse function of the regression line is expressed as

$$y = -\frac{\alpha'}{\beta'} + \frac{1}{\beta'}x \tag{12}$$

However, the preceding equation is very different from the regression line R1 derived by using the radar data as the independent variable and the rain gauge data as the dependent variable. Most of all, the slope was derived differently, as a different weighting factor was used. In the regression line in the previous section, the weighting factor used was the radar rain rate, but the rain gauge rain rate was used in this section. However, in both cases, the coefficient of determination remains the same (Park *et al*., 2002).

In practical cases of correcting the radar rain rate, as the radar rain rate is used as the independent variable, the derived regression line would be influenced largely by the radar rain rate. As the radar rain rate is generally estimated to be smaller than the rain gauge rain rate, the slope of the regression line would be rather gentle, so the bias-corrected radar rain rate could be overestimated in the low rain rate region and underestimated in the high rain rate region. Naturally, the opposite is also true; if the rain gauge rain rate is used as the independent variable, the regression line will be largely influenced by the rain gauge rain rate. The bias-corrected radar rain rate could be underestimated in the low rain rate region and overestimated in the high rain rate region. However, this characteristic of the regression line does not answer the question of which variable should be chosen as the independent variable in the derivation of the bias correction factor.

The quality of the observed data is another factor that should be considered under this circumstance. As is well known, the quality of the rain gauge rain rate data (hereafter the rain gauge data) is quite high, to the extent that they can be assumed to be almost true. However, there is some amount of uncertainty involved when the rain gauge data are assumed to represent the radar bin for the comparison. Similarly, the radar rain rate data (hereafter the radar data) are not of such high quality, as they are retrieved by a remote sensing tool. Between these two different rain rate data, the rain gauge data are assumed better with no systematic bias (MOCT, 2007; Yoo *et al*., 2012).

This issue of data quality is directly connected to the reliability of the slope of a derived regression line. If the rain gauge rain rate is considered as the independent variable, then because the rain gauge rain rate is also used as the weighting factor, the uncertainty of the radar rain rate may not be amplified higher than it already is. However, if the radar rain rate is used as the independent variable, then because the radar rain rate is considered in both the numerator and the denominator of the slope of a regression line, the reliability of the calculation could be worsened further. In that sense, the reliability of the slope is different from the coefficient of determination of the regression line itself.

This problem remains the same even when the regression line is forced to pass the origin (that is, R2). In fact, the role of extreme values (or very high rain rates) becomes greater in this case, as the slope is calculated by considering the deviation from the origin, not the mass centre as in the case of the regression line.

For the line passing both the mass centre of the measurement pairs and the origin (that is, R3), however, no difference can be found in the case of using switched independent and dependent variables. This is simply because this slope is calculated without considering any weight factors such as in the previous two cases.

Regardless of the aforementioned considerations, it is still not easy to choose the best way of determining the bias correction factor. Each method has its own characteristics, which may not be directly comparable with the others. However, it is clear that the relation between the radar rain rate and the rain gauge rain rate is not a causal connection. The radar rain rate is generally smaller than the rain gauge rain rate, and the quality of rain gauge data is believed to be higher than that of radar data. Especially, in Korea, as the rain gauges have been well maintained, the quality of rain gauge data is known to be much higher than the radar data (MOCT, 2007).

Also, it is important to remember that both the radar and rain gauge data include a significant portion of zero measurements, although the proportion is theoretically smaller in the radar data. The following section will review these issues by using the radar rain rates measured over the Korean Peninsula in 2010.

APPLICATION EXAMPLE

Rainfall event and rain rate data

This study analysed the rainfall event that occurred on 10–11 August 2010. This rainfall event arrived at the southern coast of the Korean Peninsula on 10 August at around 9 PM, moved north–northeast, and entered the East Sea around 3 PM of 11 August. The maximum rain rate recorded was 78.2 mm/h.

The rain rate data used in this study are the radar data from the Gudeoksan Radar and the hourly rain rate data from 95 rain gauges within the radar umbrella operated under the Korea Meteorological Administration (KMA). Only the rain gauges within the range of 120 km were selected to minimize the range-dependent error. The Gudeoksan Radar is an S-band radar. The transmitter type is klystron, the frequency is 2,718 MHz, and the peak power is 850 kW. The diameter of the antenna is 8.5 m, and the effective observation range is 240 km. The resolution of the radar data is 1 km × 1 km, which forms a 481 × 481 grid.

The radar data obtained were those quality controlled by the KMA. The radar rain rate data were derived by applying a certain *Z–R* relation to the 1.5-km constant altitude plan position indicator (CAPPI) data after completing the quality control (QC) procedure of KMA. The QC of KMA is very typical, including removal of ground clutter and sea clutter, velocity aliasing, and super refraction. The *Z–R* relationship ($Z = 300R^{1.4}$) was selected by considering the storm type. As the 1.5-km CAPPI data are those rather far from the ground (in fact, the use of 1.5-km CAPPI was an obligated decision because of the very irregular topography of the Korean Peninsula), the derived radar rain rate is rather underestimated in general (Hong *et al*., 2009). The location of the Gudeoksan Radar and the spatial distribution of the rain gauges are shown in Figure 1. Figure 2 shows one of the radar images taken during the rainfall event in this study.

The data structure of the radar and rain gauge data used in this study is summarized in Table I. To be brief, all the measurement pairs including zero measurements has a total of 14,232 (100%) measurement pairs, among which the proportion of the positive measurement pairs or (+, +) pairs is 49.7%. The proportion of (+, 0) pairs is 0.2%, and the proportion of (0, +) pairs is 6.0%.

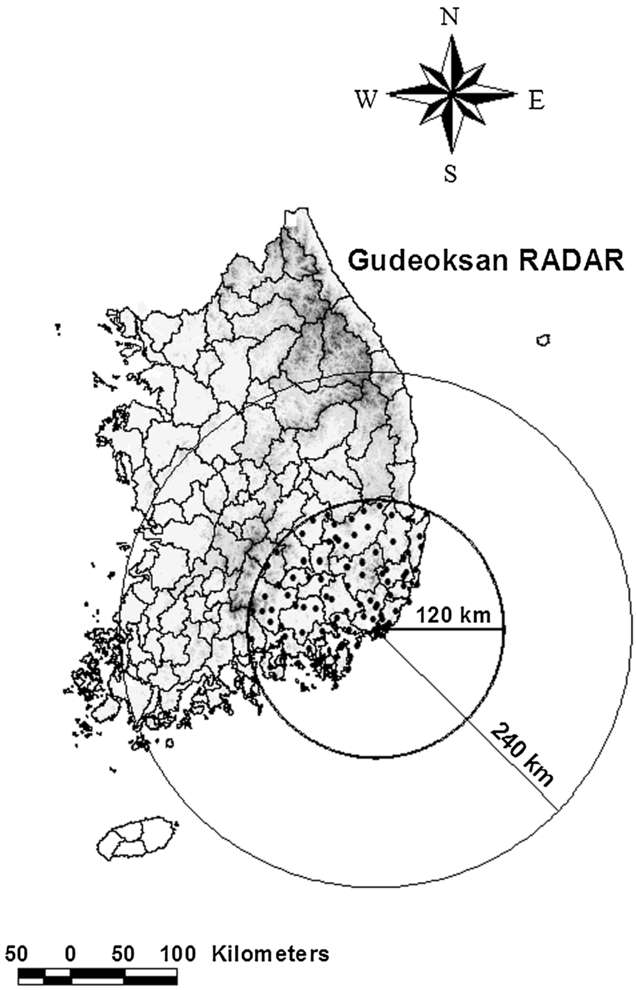


Figure 1. Location of rain gauges used in this study within the Gudeoksan Radar umbrella

Relationship between radar and rain gauge rain rates

This study evaluated the effect of switched independent and dependent variables and the impact of considering zero measurements as examined in the previous section. First, Figure 3(a) shows the case where the radar rain rate is taken as the independent variable, whereas Figure 3(b) shows the opposite case in which the rain gauge rain rate is taken as the independent variable. In each panel, R1 represents the linear regression line, and R2 represents the regression line that forces the intercept to be zero. R3 represents the line passing the origin and the mass centre. Figure 3(b) was prepared by taking the inverse function of each of the three derived lines to ease the comparison with Figure 3(a). The slopes of the relationships R1, R2, and R3 shown in both figures are summarized in Table II.

First, the effect of using switched independent and dependent variables was examined. According to Figure 3 (a), slopes of the three relationships derived by taking the radar rain rate as the independent variable are very gentle.

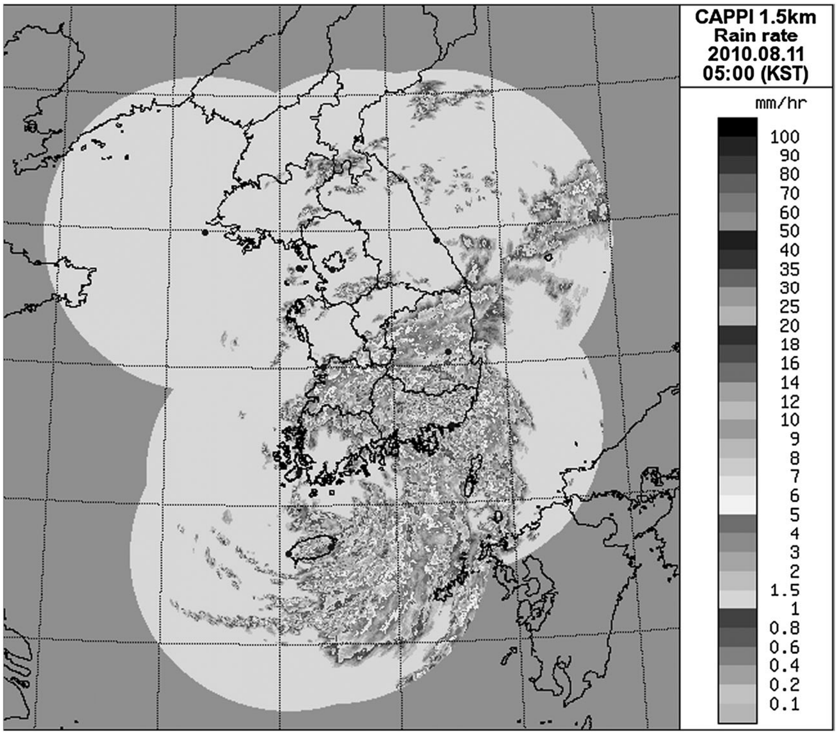


Figure 2. A radar image of the Gudeoksan Radar (5:00 AM, 11 August 2010)

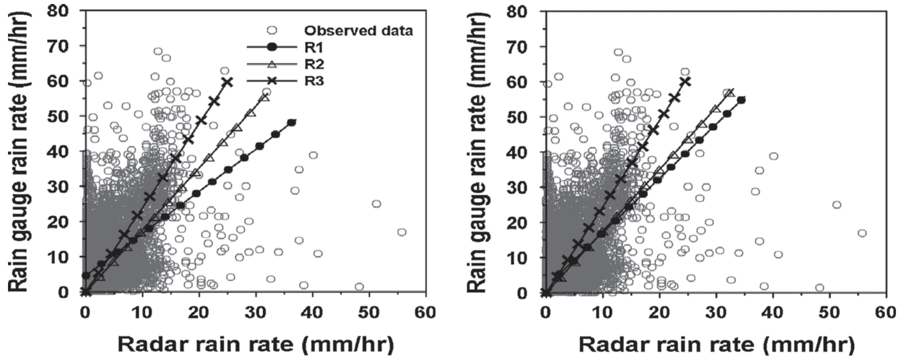
Table I. Data structures of the rainfall event considered in this study (R_i and G_i indicate the radar rain rate and the rain gauge rain rate, respectively)

Data structure	Number of data pairs (%)
$R_i > 0, G_i > 0$	7,068 (49.7)
$R_i > 0, G_i = 0$	23 (0.2)
$R_i = 0, G_i > 0$	859 (6.0)
$R_i = 0, G_i = 0$	6,282 (44.1)
$R_i \geq 0, G_i \geq 0$	14, 232 (100.0)

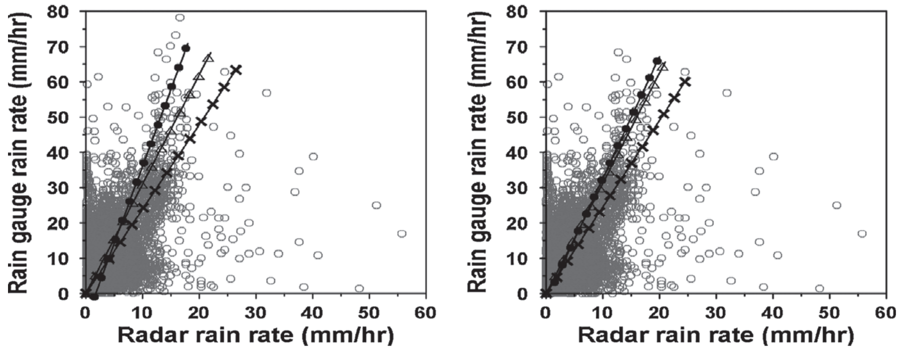
This is basically because the radar rain rate considered as the weighting factor for the determination of the slope is smaller than the rain gauge rain rate. The relatively large rain gauge rain rate did not have any significant impact on the determination of the slope of these relationships. As the slope of R3 is determined as the ratio between the rain gauge rain rate and the radar rain rate, it is steeper than the other two. On the contrary, three relationships derived by taking the rain gauge rain rate as the independent variable show relatively steep slopes (Figure 3(b)). It is because the rain gauge rain rate was considered as the weighting factor for the determination of the slope. Among the three relationships derived, the slope of R1 becomes the steepest and that of R3 the gentlest. In fact, the slope of R3 remains unchanged even when considering switched independent and dependent variables.

Among the slopes of these three relationships, because R1 provides a negative intercept that is far less than zero, it may not be used for the correction of the radar rain rate. The slope of R2 is dependent upon the independent variable taken. As the slope of R2 shown in Figure 3(b) is steeper than that of R2 in Figure 3(a), the bias correction calculated using the slope of R2 in Figure 3(a) provides a relatively underestimated value compared with the opposite case. For example, R2 in the right panel (all the measurement pairs) of Figure 3(a) suggests approximately 15 mm/h for the rain gauge rain rate (or the bias-corrected radar rain rate) when the original radar rain rate is 10 mm/h. In the opposite case (R2 in the right panel of Figure 3(b)), it is approximately 30 mm/h. This large difference also appears in the positive measurement pairs. However, for R3, the same result is suggested regardless of which data are taken as the independent variable, because it is the line that simply connects the origin and the mass centre.

Next, the impact of zero measurements was examined by comparing the three relationships derived for the positive measurement pairs and all the measurement pairs (Table II). The changes in the slopes can be predicted by considering the data structure of all the measurement pairs, including the relative proportions of zero measurements. First, the slope of the linear regression line (R1) derived using the positive measurement pairs becomes increased when considering all the measurement pairs.



(a) (radar, rain gauge) pairs: positive measurements (left) and all the measurements (right)



(b) (rain gauge, radar) pairs: positive measurements (left) and all the measurements (right)

Figure 3. Three relationships R1, R2, and R3 derived (a) by taking the radar data as the independent variable and (b) by taking the rain gauge data as the independent variable

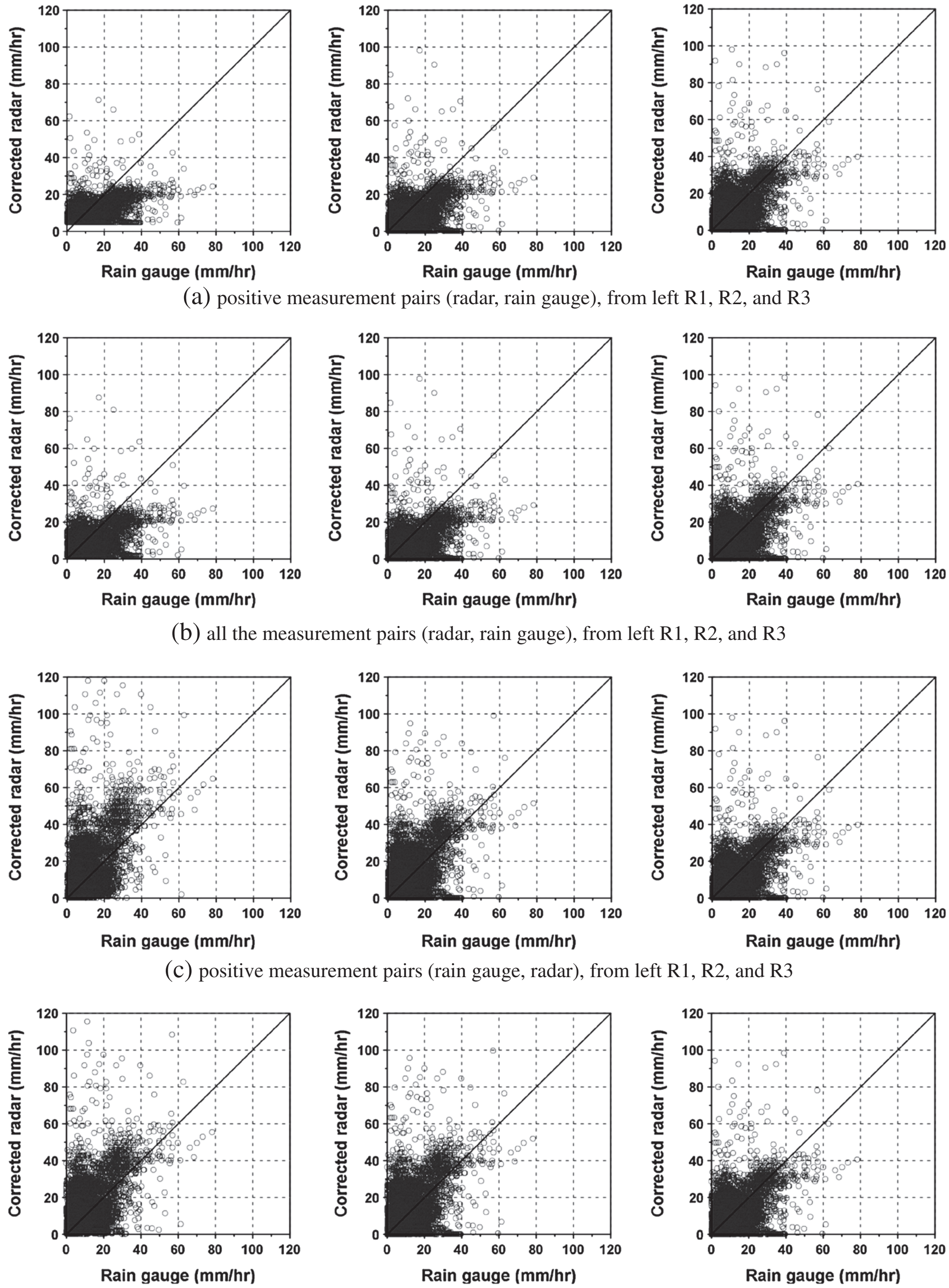
Table II. Slopes of three relationships (R1, R2, and R3) derived using the positive measurement pairs (positive) and all the measurement pairs (all)

		zPositive	All
(Radar, rain gauge)	R1	1.20	1.54
Figure 3(a)	R2	1.76	1.76
	R3	2.40	2.45
(Rain gauge, radar)	R1	4.36 (0.229)	3.46 (0.289)
Figure 3(b)	R2	3.10 (0.323)	3.13 (0.319)
	R3	2.40 (0.417)	2.45 (0.408)

This is because the portion of (0, +) measurement pairs is relatively large (Equation (8)). Second, the slope of R2 derived using the positive measurement was not changed when considering all the measurement pairs. This result can be expected, as the number of (+, 0) pairs is very small, as was already shown in Equation (9). Lastly, the slope of R3 derived using all the measurement pairs is fully dependent on the relative proportions of zero measurements in the independent and dependent variables. For the rainfall event considered in this study, as can be seen in Table I, the proportion of (+, 0) pairs is only 0.2% and that of the (0, +) pairs is 6.0%, which are, respectively, m_1 and m_2 in Equation (10). Thus, the slope becomes a bit steeper when considering all the measurement pairs compared with that derived using only the positive measurement pairs.

The preceding analysis can also be applied to cases in which the rain gauge rain rate is taken as the independent variable. As can be seen in Table II, the results are different from those for the previous cases, with the exception of R3. As the slope of R3 is calculated simply by calculating the ratio between the independent and dependent variables, switching the independent and dependent variables does not affect the slope of the line.

Lastly, we compared the rain gauge rain rate with the bias-corrected radar rain rate by the three relationships considered in this study. The results are shown in Figure 4, and the root mean square error (RMSE) for each case is provided in Table III. The RMSE indicates the mean difference between the bias-corrected radar rain rate and the rain gauge rain rate. In fact, it is very statistical that R1 gives the smallest RMSE and that R3 gives the largest RMSE. However, this result is not assured visually in Figure 4. As can be seen in Figure 4(a, b), the bias-corrected radar rain rate is smaller than the rain gauge rain rate for R1 and R2 cases when taking the radar data as the independent variable. This result is especially vivid when the rain rate is high. R3 appears much better, but its RMSE is much larger than that of the



(a) positive measurement pairs (radar, rain gauge), from left R1, R2, and R3

(b) all the measurement pairs (radar, rain gauge), from left R1, R2, and R3

(c) positive measurement pairs (rain gauge, radar), from left R1, R2, and R3

(d) all the measurement pairs (rain gauge, radar), from left R1, R2, and R3

Figure 4. Comparison of bias-corrected radar rain rate (Corrected radar) and rain gauge rain rate (Rain gauge) for the positive measurement pairs and all the measurement pairs with respect to the three relationships (R1, R2, and R3) and switched independent and independent variables

other two relationships. In fact, the bias-corrected radar rain rate is higher when the rain rate intensity is relatively low and lower in the opposite case. To a greater or lesser extent, this result is common in all three cases when taking the radar rain rate as the independent variable.

On the contrary, the results of applying the three relationships derived by taking the rain gauge rain rate as the independent variable appear very differently. First of all, the bias-corrected radar rain rate becomes much higher when taking the rain gauge data as the independent variable (Figure 4(c, d)). This is especially true for R1 and R2. Between R1 and R2, R1 provides a negative bias-corrected radar rain rate when the radar rain rate is very low but provides a much higher one when the radar rain rate is high (Figure 3(b)). That is, the results can be unrealistic in this case of R1. This problem could become

Table III. Quality of bias-corrected radar rain rate in RMSE estimated by applying three bias correction factors (R1, R2, and R3) to the positive measurement pairs (positive) and all the measurement pairs (all)

	Data structure			
	(Radar, rain gauge)		(Rain gauge, radar)	
	Positive	All	Positive	All
R1	7.9	5.9	15.1	8.9
R2	8.6	6.1	11.4	8.2
R3	9.3	6.7	9.3	6.7

more serious when only the positive measurements are considered. However, R2 provides much similar bias-corrected radar rain rate to the rain gauge rain rate both in the low and high rain rate regions. There is no possibility of negative bias-corrected radar rain rate. For R3, the result is the same as the case that takes the radar rain rate as the independent variable. A simple comparison of RMSEs for these three cases shows that the RMSE of R3 is the smallest and that of R1 is the largest (Table III). In fact, this result indicates that the simple comparison of RMSEs does not support the characteristics of each bias correction factor visually found in Figure 4(c, d).

Thus, in this study, we separately estimated the RMSE of the region where the rain rate is relatively high and where it is low. Additionally, the mean errors (MEs) of the two regions were also estimated for the comparison. Because the accuracy of the relatively high rain rate region is very important when the radar rain rate is used to calculate flood runoff, the RMSE for all the data may lose its significance. This study divided the rain rate region into two by using a threshold rain rate of 10 mm/h. Here, the threshold of 10 mm/h is a subjectively selected criterion, which may be the minimum rain rate that initiates the surface runoff. In Figure 4, the threshold of 10 mm/h seems to be located in the middle of all the data, but, in fact, it lies near the top 30% of all the data. The result for each case is summarized in Table IV.

First of all, as can be expected from Figure 4, the RMSEs in the low rain rate region are small, and those in the high rain rate region are very large. Also, the MEs in the low rain rate region are all positive, and those in the high rain rate region are all negative. The smallest RMSEs are found in the low rain rate region when taking the radar rain rate as the independent variable, and the highest RMSEs are found in the high rain rate region when taking the rain gauge rain rate as the independent variable. It is also noticeable that the RMSEs in the high rain rate region are found not so sensitive to the independent variable taken as in the low rain rate region. However, the MEs in the high rain rate region are found very different depending on the independent variable taken. The MEs estimated when taking the rain gauge rain rate as the independent variable are much smaller than those in the opposite case. Among the three relationships considered, the smallest ME can be found for R1 and R2 derived by taking the rain gauge rain rate as the independent variable. However, as the RMSE of R1 is much higher, R2 can be selected as the better one. In fact, this result corresponds well to that in Figure 4. The effect of zero measurements seems to be not as serious for R2 and R3 as it is for R1.

The aforementioned results show that it is advantageous to consider the slope of R2 as the bias correction factor. When calculating the slope of R2, the radar rain rate should be taken as the independent variable when the target rain rate is rather low, such as a rain rate of less than 10 mm/h, and the rain gauge rain rate should be taken when the target rain rate is high, such as a rain rate higher than 10 mm/h. Thus, if the purpose of using the radar rain rate is to calculate the flood runoff, it is good to use the slope of R2 as the bias correction factor. R2 should be derived by taking the rain gauge rain rate as the

Table IV. Quality of bias-corrected radar rain rate in RMSE and ME (inside the bracket) in the low rain rate region (less than or equal to 10 mm/h) and in the high rain rate region (higher than 10 mm/h) depending on the three relationships (R1, R2, and R3) applied to the positive measurement pairs (positive) and all the measurement pairs (all)

		Data structure			
		(Radar, rain gauge)		(Rain gauge, radar)	
		Positive	All	Positive	All
<10 mm/h	R1	5.1 (3.8)	3.2 (1.7)	12.0 (0.4)	6.7 (0.2)
	R2	4.8 (0.7)	3.2 (0.2)	9.5 (4.5)	6.2 (1.7)
	R3	6.9 (2.5)	4.6 (1.0)	6.9 (2.5)	4.6 (1.0)
>10 mm/h	R1	12.0 (–8.3)	13.0 (–9.0)	20.4 (–0.9)	16.3 (–1.0)
	R2	13.6 (–9.4)	13.7 (–9.5)	14.8 (–1.1)	14.9 (–1.1)
	R3	13.1 (–5.5)	13.2 (–5.2)	13.1 (–5.5)	13.2 (–5.2)

independent variable and by considering only the high rain rate region above a certain threshold. The threshold rain rate could be decided by considering the initial infiltration rate, which is generally higher than 10 mm/h in Korea (Yoon, 1998).

SUMMARY AND CONCLUSION

In this study, the problem of correcting the mean-field bias of the radar rain rate was investigated using the concept of linear regression. We analysed three different relationships and considered which of their slopes should be used as the bias correction factor: R1 is based on the conventional linear regression, R2 is forced to pass the origin, and R3 is the line whose slope is the *G*/*R* ratio. In other words, R1 is the regression line connecting the intercept and the mass centre of measurement pairs, R2 is the regression line forced to pass the origin, and R3 is the line connecting the origin and the mass centre. The slopes of all three relationships were reviewed analytically to compare them, and thereby, the effect of zero measurements could be evaluated. Additionally, the effect of using switched independent and dependent variables on the derived slopes was also evaluated. The theoretically derived results were then verified by analysing the rainfall event on 10–11 August 2010 in Korea. Finally, the difference between the bias-corrected radar rain rate and the rain gauge rain rate was quantified by RMSE so that it could be used as a measure for the evaluation of bias correction factors. The results are summarized as follows.

First, the impact of which data are taken as the independent variable was found to be very significant in the determination of the slope. When taking the radar rain rate as the independent variable, all three slopes considered were estimated to be very gentle. This is basically because the radar rain rate considered as the weighting factor for the determination of the slope is smaller than the rain gauge rain rate. The relatively large rain gauge rain rate did not have any significant impact on the determination of the slope of these relationships. Very typical results could also be found, such as overestimation in the low rain rate region and underestimation in the high rain rate region. On the contrary, three relationships derived by taking the rain gauge rain rate as the independent variable show relatively steep slopes as the rain gauge rain rate was considered as the weighting factor for the determination of the slope.

Second, the impact of zero measurements was examined by comparing the slopes derived using only positive measurement pairs and those derived using all the measurement pairs. Changes in the slopes could be predicted by considering the relative proportions of zero measurements. For the rainfall event analysed in this study, the slope of R1 when taking the radar rain rate as the independent variable was increased when considering zero measurements, the slope of R2 was unchanged, and the slope of R3 was increased.

Third, this study compared the bias-corrected radar rain rate with the rain gauge rain rate for all three relationships considered in this study. A simple comparison of RMSEs showed that there is no significant difference among the three slopes when taking the radar rain rate as the independent variable. However, the RMSE for the case of using switched independent and dependent variables was found to be significantly increased, except for R3. When considering only the low rain rate regions separately, both the small RMSEs and MEs are found for R2 when taking the radar rain rate as the independent variable. However, in the high rain rate region, both the small RMSEs and MEs are found for R2 when taking the rain gauge rain rate as the independent variable.

In conclusion, first, it seems advantageous to divide the rain rate data into two regions, low and high, for the bias correction. Also, the slope of R2 was found to be the best for the bias correction factor. However, when deciding the slope of this R2, the radar rain rate should be used as the independent variable in the low rain rate region, and the rain gauge rain rate in the high rain rate region. Thus, if the purpose of using the radar rain rate is to calculate the flood runoff, it is good to use the slope of R2 as the bias correction factor derived by taking the rain gauge rain rate as the independent variable.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (no. 2010-0014566).

REFERENCES

Anagnostou EN, Krajewski WF. 1999. Real-time radar rainfall estimation. Part I: algorithm formulation. *Journal of Atmospheric and Oceanic Technology* **16**: 189–197.

Anagnostou EN, Krajewski WF, Seo DJ, Johnson ER. 1998. Mean field rainfall bias studies for WSR88D. *Journal of Hydrologic Engineering* **3**(3): 149–159.

Borga M, Anagnostou EN, Frank E. 2000. On the use of radar rainfall estimates for flood prediction in mountainous basins. *Journal of Geophysical Research* **105**(D2): 2269–2280.

Borga M, Tonelli F, Moore RJ, Andrieu H. 2002. Long-term assessment of bias adjustment in radar rainfall estimation. *Water Resources Research* **38**(11): 1226.

Brandes EA. 1975. Optimizing rainfall estimates with the aid of radar. *Journal of Applied Meteorology* **14**: 1339–1345.

Chiu LS, North GR, McConnel A, Short DA. 1990. Rain estimation from satellite: effects of finite field of view. *Journal of Geophysical Research* **95**: 2177–2186.

Chumchean S, Sharma A, Seed A. 2003. Radar rainfall error variance and its impact on radar rainfall calibration. *Physics and Chemistry of the Earth* **28**: 27–39.

Chumchean S, Seed A, Sharma A. 2006. Correcting real-time radar rainfall bias using a Kalman filtering approach. *Journal of Hydrology* **317**: 123–137.

Collier CG. 1986. Accuracy of rainfall estimates by radar, part I: calibration by telemetering raingauges. *Journal of Hydrology* **83**: 207–233.

Collier CG, Larke PR, May BR. 1983. A weather radar correction procedure for real-time estimation of surface rainfall. *Quarterly Journal of the Royal Meteorological Society* **109**: 589–608.

Creutin JD, Andrieu H, Faure D. 1997. Use of a weather radar for the hydrology of a mountainous area. Part II: radar measurement validation. *Journal of Hydrology* **193**: 26–44.

Dinku T, Anagnostou EN, Borga M. 2002. Improving radar-based estimation of rainfall over complex terrain. *Journal of Applied Meteorology and Climatology* **41**: 1163–1178.

Fulton RA, Breidenbach JP, Seo DJ, Miller DA, O'Bannon T. 1998. The WSR-88D rainfall algorithm. *Weather and Forecasting* **13**: 377–395.

Hong JH, Kim BS, Bae YH. 2009. Comparison of rainfall field for flood simulation using GDK meteorological radar and ground gauge. Proceedings of 2009 KSCE Annual Conference; 552–556.

Joss J, Waldvogel A. 1990. Precipitation measurements and hydrology. Radar in Meteorology. Battan Memorial and 40th Anniversary Radar Meteorology Conference, Atlas D (ed), American Meteorological Society; 577–606.

Jung SH, Kim KE, Ha KJ. 2005. Real-time estimation of improved radar rainfall intensity using rainfall intensity measured by rain gauges. *Journal of Korea Meteorology Association* **41**(5): 751–762.

Krajewski WF, Smith JA. 2002. Radar hydrology: rainfall estimation. *Advances in Water Resources* **25**: 1387–1394.

Kummerow C. 1998. Beam filling errors in passive microwave rainfall retrievals. *Journal of Applied Meteorology* **37**: 356–370.

Lafont D, Guillemet B. 2004. Subpixel fractional cloud cover and inhomogeneity effects on microwave beam-filling error. *Atmospheric Research* **72**: 149–168.

Lin DS. 1989. An application of recursive methods to radar-rainfall estimation. MS thesis, University of Iowa, Iowa Institute of Hydraulic Research: Iowa.

Ministry of Construction and Transportation (MOCT). 2007. Development of Quality Control System for National Hydrologic Data: Seoul; 515.

Nam KY, Song BH, Seo EK, Choi JH, Lee JS, Nam JC. 2003. A study on *Z–R* relationship of the summertime convective system occurred around Cheonju. *Journal of Korea Meteorology Association* **39**(4): 431–440.

Park S, Kim S, Kang M. 2002. *Regression Analysis*. KNOU Press: Seoul; 50–52.

Seo DJ. 1998. Real-time estimation of rainfall fields using rain gauge data under fractional coverage conditions. *Journal of Hydrology* **208**: 25–36.

Seo DJ, Breidenbach JP. 2002. Real-time correction of spatially nonuniform bias in radar rainfall data using rain gauge measurements. *Journal of Hydrometeorology* **3**: 93–111.

Seo DJ, Fulton RA, Breidenbach JP, Taylor M, Miller DA. 1996. *Interagency Memorandum of Understanding among the NEXRAD Program*. WSR-88D Operational Support Facility, the National Weather Service Office of Hydrology, Hydrologic Research Laboratory, National Weather Service: Silver Spring, Maryland.

Seo DJ, Breidenbach JP, Johnson ER. 1999. Real-time estimation of mean-field bias in radar rainfall data. *Journal of Hydrology* **223**: 131–147.

Smith JA, Krajewski WF. 1991. Estimation of mean-field bias of radar rainfall estimates. *Journal of Applied Meteorology* **30**: 397–412.

Steiner M, Smith JA, Burges SJ, Alonso CV, Darden RW. 1999. Effect of bias adjustment and rain gauge data quality control on radar rainfall estimation. *Water Resources Research* **35**(8): 2487–2503.

Vignal B, Krajewski WF. 2001. Large-sample evaluation of two methods to correct range-dependent error for WSR-88D rainfall estimates. *Journal of Hydrometeorology* **2**: 490–504.

Wilson JW. 1970. Integration of radar and rain gauge data for improved rainfall measurement. *Journal of Applied Meteorology* **9**: 489–497.

Yoo C, Kim K, Choi J, Ha E. 2010. Importance of no-rain measurements on the comparison of radar and rain gauge rain rate. *Hydrological Processes* **24**: 924–933.

Yoo C, Yoon J, Kim J, Park C, Jun C. 2012. A quality evaluation criteria for radar rain-rate data. Weather radar and hydrology. *IHAS Publication* **351**: 382–387.

Yoon Y-N. 1998. *Engineering Hydrology*. Chungmungak: Seoul; 139–155.

