Gauge-radar adjustment by using multivariate kernel regression and spatiotemporal Kriging

Seppo Pulkkinen[1], Jarmo Koistinen[1], and Timo Kuitunen[1]

[1]*Finnish Meteorological Institute, Erik Palménin aukio 1, P.O. Box 503, FI-00101 Helsinki, Finland*

(Dated: 18 July 2014)

1. Introduction

We model the multiplicative error between gauge rainfall R_g and radar rainfall R_r above the gauge by using the logarithmic quantity

$$F = 10 \cdot \log_{10}\left(\frac{R_g}{R_r}\right). \tag{1.1}$$

Based on empirical evidence, this quantity is usually assumed to be normally distributed with zero mean and constant variance. Even in more advanced models, the mean and variance are assumed to be simple functions of one variable such as distance from the radar or the radar rainfall R_r (e.g. Villarini et al. (2009)). However, in practice both mean and variance of the error distribution have a complicated depencency on multiple factors (e.g. rainfall accumulation, distance from radar, attenuation effects and geographic factors). Though the uncertainty of radar-based rainfall measurements has been extensively studied, the existing models do not fully address this fact.

The aim of this research is to develop a multivariate model capable of explaining multiple factors that contribute to the gauge-radar error distribution. Another novel feature is the use of a spatiotemporal variogram model. This model is used for Kriging-interpolation of the residual error after systematic corrections. The combined regression and Kriging model gives a probability distribution for the expected ground rainfall for a given radar observation. Such distributions allow computation of bias-corrected radar rainfall fields with uncertainty estimates, and also ground rainfall probabilities.

2. The Statistical Model

The proposed gauge-radar error model is of the form

$$F(\boldsymbol{\theta}) = \mu(\boldsymbol{\theta}) + \sigma(\boldsymbol{\theta})\varepsilon. \tag{2.1}$$

The function μ represents systematic bias, and the remaining term consists of variance σ and a stochastic component ε that is assumed to follow the standard normal distribution. The factors affecting to the mean and variance of F are denoted by $\boldsymbol{\theta}$.

Generalizing the model proposed by Ciach et al. (2007) to the multivariate case, the idea is to use a *kernel regression* model for the conditional mean and variance of F for the given parameters $\boldsymbol{\theta}$. In this model, a Gaussian kernel is assigned for each observed value of F (denoted by f_i) and associated parameter values (denoted by $\boldsymbol{\theta}_i$). An estimate for the conditional mean is given by

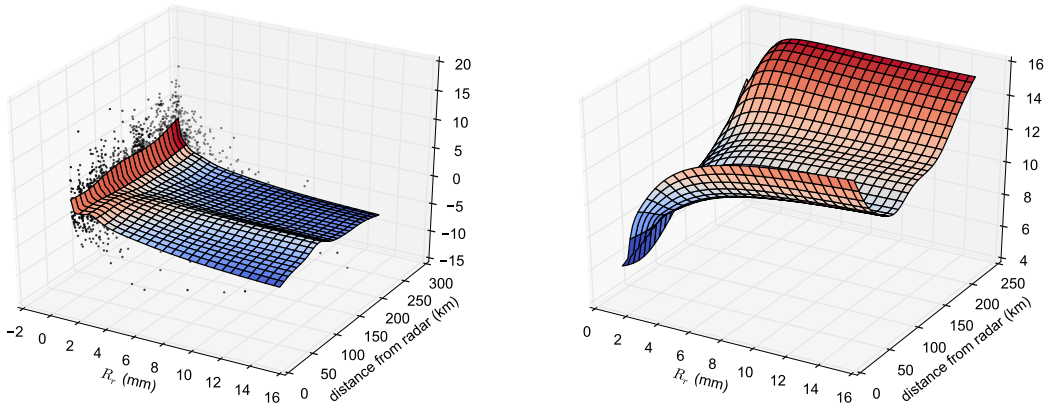
$$\hat{\mu}_{\boldsymbol{H}}(\boldsymbol{\theta}) = \frac{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{\theta} - \boldsymbol{\theta}_i) f_i}{\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{\theta} - \boldsymbol{\theta}_i)}, \tag{2.2}$$

where $K_{\boldsymbol{H}}$ denotes a Gaussian kernel

$$K_{\boldsymbol{H}}(\boldsymbol{r}) = \frac{1}{\sqrt{(2\pi)^d |\boldsymbol{H}|}} \exp\left(-\frac{\boldsymbol{r}^T \boldsymbol{H}^{-1} \boldsymbol{r}}{2}\right)$$

with bandwidth matrix $\boldsymbol{H}$. Determining the model from the given observations amounts to finding a matrix $\boldsymbol{H}$ that is optimal with respect to some regression error measure (e.g. sum of squared residuals). Likewise, the variance σ^2 is estimated from the squared residuals by using an estimate of the form (2.2).

Examples of kernel regression surfaces fitted to observed F-values are shown in Figure 1. The values represent hourly rainfall accumulations between 0.25-15 mm obtained from the lowest elevation angle of the Ikaalinen radar and rain gauges within the radar range. The observation period is June-September 2013. Attenuation and VPR corrections are applied to the radar data. A median filter is also used to suppress residual ground echoes. For the bandwidth estimation we use the np package for the R software. Clearly, the regression surfaces shown in Figure 1 provide additional information that cannot be

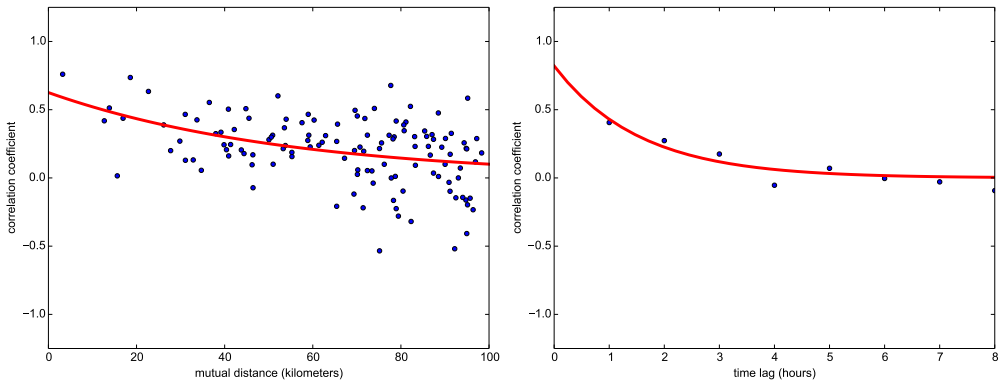


(a) mean (b) variance

Figure 1: Regression surfaces for the conditional mean and variance. The regression parameters $\boldsymbol{\theta}$ are hourly radar rainfall accumulation (mm) and distance from radar (km).

obtained from a univariate model. For instance, the range-dependent behaviour of the systematic bias depends on the rainfall accumulation.

The regression model described above is a static model, as it does not include any time-dependent behaviour of gauge-radar errors. Therefore it is included into the stochastic component ε in the model (2.1). As illustrated in Figure 2, the residual values ε computed from (2.1) by using the regression estimates for μ and σ exhibit significant spatial and temporal correlation structure. Utilizing this fact, we compute spatiotemporal variograms and use Kriging to obtain gridded estimates of this error component. Although the temporal correlation is restricted to short accumulation periods (e.g. Villarini et al. (2009)), including it in the model allows using the information provided by previous observations. This gives an advantage because of sparsity of rainfall observations and the gauge network.



(a) spatial correlation (b) temporal correlation (averaged for each time lag)

Figure 2: Spatial and temporal correlations of the stochastic component ε computed using the regression surfaces shown in Figure 1.

The interpolated values at location (x, y) at time t are obtained by using the observed values $\varepsilon(x_i, y_i, t_i)$ according to

$$\varepsilon(x, y, t) = \sum_{i=1}^{n} \lambda_i \varepsilon(x_i, y_i, t_i),$$

where the weights λ_i are computed from the Kriging equations by using a spatiotemporal variogram model. For hourly rainfall accumulations, the observed values $\varepsilon(x_i, y_i, t_i)$ are taken from a six hour time window preceding the current time t. They are computed by normalizing the observed values F_i according to

$$\varepsilon(x_i, y_i, t_i) = \frac{F_i - \hat{\mu}_{\boldsymbol{H}}(\boldsymbol{\theta}_i)}{\hat{\sigma}_{\boldsymbol{H}}(\boldsymbol{\theta}_i)}$$

by using the observed parameter values $\boldsymbol{\theta}_i$ and the estimates $\hat{\mu}_{\boldsymbol{H}}(\boldsymbol{\theta}_i)$ and $\hat{\sigma}_{\boldsymbol{H}}(\boldsymbol{\theta}_i)$ obtained from the regression model.

Accounting for multiple factors that affect to the mean and variance of F, the model is better able to separate the systematic bias and residual errors than univariate models or those that assume constant variance. In addition, the Kriging approach pro-

vides a fully statistical method with uncertainty estimates. Such estimates cannot be obtained when using interpolation methods that are not based on a statistical model, such as the commonly used Barnes interpolation (e.g. Michelson and Koistinen (2000) and Gregow et al. (2013)).

By combining the mean and variance estimates from the regression and Kriging models, we obtain the distribution for F. More specifically, at location (x, y) at time t and with parameters $\boldsymbol{\theta}$, F has estimated mean and variance

$$\hat{\mu}(x, y, t, \boldsymbol{\theta}) = \hat{\mu}_{\boldsymbol{H}}(\boldsymbol{\theta}) + \hat{\sigma}_{\boldsymbol{H}}(\boldsymbol{\theta})\varepsilon(x, y, t) \quad \text{and} \quad \hat{\sigma}^2(x, y, t, \boldsymbol{\theta}) = \hat{\sigma}_{\boldsymbol{H}}(\boldsymbol{\theta})^2 \hat{\sigma}^2_{\text{kr}}(x, y, t),$$

respectively. The quantity $\hat{\sigma}^2_{\text{kr}}$ denotes the uncertainty estimate obtained from Kriging interpolation. The estimate for the expected ground rainfall $\hat{R}_g(x, y, t, \boldsymbol{\theta})$ can then be obtained from

$$\hat{R}_g(x, y, t, \boldsymbol{\theta}) = 10^{\frac{\hat{\mu}(x,y,t,\boldsymbol{\theta})}{10}} R_r(x, y, t), \tag{2.3}$$

which is the inverse transformation of (1.1).

An important application of the model is computation of probabilistic ground rainfall estimates. That is, the probability that a given radar rainfall R_r above location (x, y) at time t with additional information incorporated in the parameters $\boldsymbol{\theta}$ corresponds to a ground rainfall in a given interval between $R_{\min}$ and $R_{\max}$. Assuming that F is normally distributed, such a probability can be obtained from the formula

$$P(R_{\min} \leq R_g(x, y, t, \boldsymbol{\theta}) \leq R_{\max}) = \hat{\Phi}\left(10 \log_{10}\left(\frac{R_{\max}}{R_r(x, y, t)}\right), x, y, t, \boldsymbol{\theta}\right) - \hat{\Phi}\left(10 \log_{10}\left(\frac{R_{\min}}{R_r(x, y, t)}\right), x, y, t, \boldsymbol{\theta}\right),$$

where

$$\hat{\Phi}(z, x, y, t, \boldsymbol{\theta}) = \Phi\left(\frac{z - \hat{\mu}(x, y, t, \boldsymbol{\theta})}{\hat{\sigma}(x, y, t, \boldsymbol{\theta})}\right)$$

and Φ denotes the CDF of the standard normal distribution.

The model can also be used for testing the credibility of individual gauge-radar ratios. When the deviation of an observed F-value from the expected value $\hat{\mu}_{\boldsymbol{H}}$ falls outside $c\hat{\sigma}_{\boldsymbol{H}}$ for some $c > 0$, either an incorrect radar or gauge measurement can be suspected. By using the proposed model, the accuracy of such tests is greatly improved compared to simpler models that cannot explain multiple factors contributing to the mean and variance of the error distribution.

3. Illustrations and Validation Results

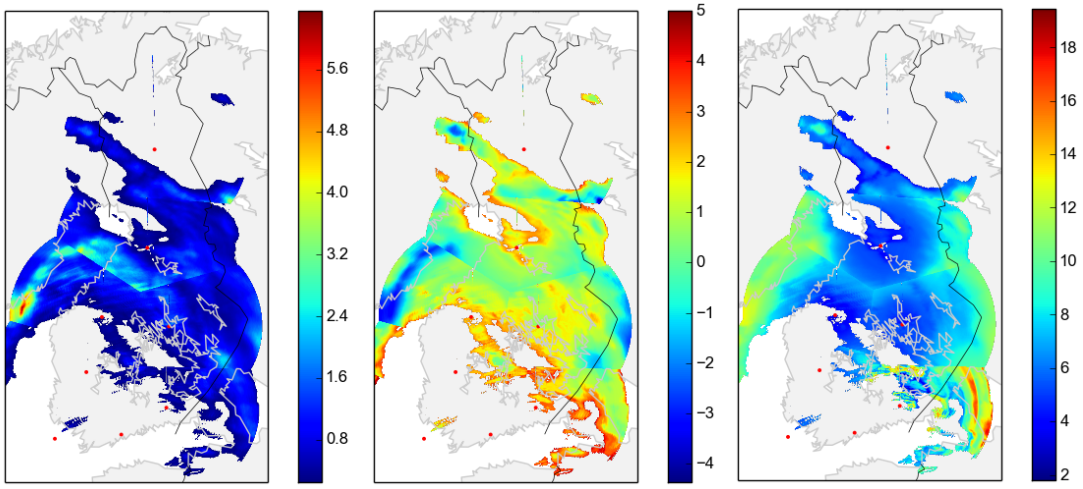
In this section we give a numerical verification of the model described in Section 2. This is done by using data obtained from the FMI radar and gauge networks. We also give an example where the model is applied to radar composites to produce probabilistic radar-based rainfall estimates.

The results of our validation experiments are listed in Tables 1-4. We have computed two validation measures for each radar and rain gauges within the radar range. The first measure is the mean of average F-values computed for each gauge location. The second validation measure is the mean of variances of per-gauge F-values. The corrected values obtained from the model have been computed for independent rain gauges (i.e. those not included in the model fitting) by using ten-fold cross validation. In all tests we have used the lowest elevation angle of each radar and applied a median filter and attenuation and VPR corrections. Radar and gauge observations outside the range 0.25-15 mm are not included in the fitting or validation data.

		regression variable(s)		
radar	uncorrected	R_r	distance	R_r & distance
Anjalankoski	1.035	-0.218	-0.091	-0.199
Ikaalinen	0.168	-0.131	-0.169	-0.112
Korppoo	0.318	0.004	0.056	0.003
Kuopio	0.315	-0.016	0.026	0.003
Luosto	0.403	-0.074	0.091	0.025
Utajärvi	0.465	-0.017	-0.016	-0.015
Vantaa	0.949	-0.138	-0.085	-0.146
Vimpeli	0.292	-0.108	0.076	-0.019

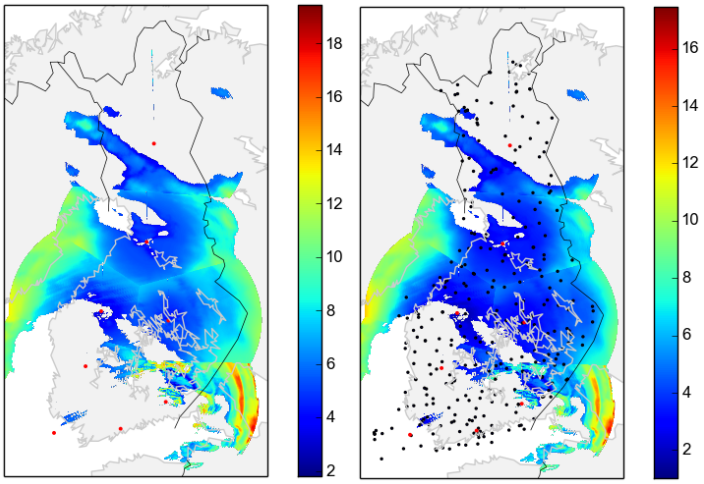
Table 1: F means for uncorrected and regression-corrected hourly radar rainfall accumulations with different regression variables (observation period June-August 2013).

The results listed in Table 1 show that the multivariate model with two regression variables gives the average F having the smallest absolute value in most cases. Furthermore, the variances listed in Table 2 show that removal of large systematic errors (such as those seen from Figure 1 for small rainfall accumulations) also decreases the variance of F. Again the multivariate



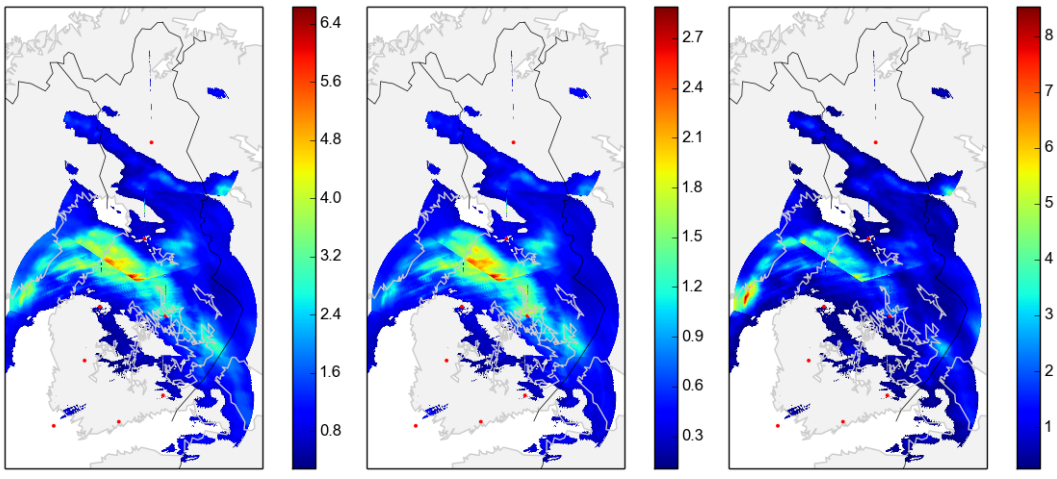
(a) uncorrected R_r (mm/h) (b) F mean ($\hat{\mu}_{\boldsymbol{H}}$) (c) F variance ($\hat{\sigma}^2_{\boldsymbol{H}}$)

Figure 3: Uncorrected radar composite and mean and variance estimates of F computed by using the regression model. Radar locations are plotted with red dots.



(a) regression only ($\hat{\sigma}^2_{\boldsymbol{H}}$) (b) regression and Kriging ($\hat{\sigma}^2$)

Figure 4: Variance of F estimated from the regression model and the combined regression and Kriging model. Gauge locations are plotted with black dots.



(a) expected rainfall ($\hat{R}_g$) (b) lower bound (c) upper bound

Figure 5: Expected rainfall estimated from the combined regression and Kriging model and 90% probability intervals.

model has the best performance. Here we have used only two variables, but including additional variables in the regression model is expected to further improve the results.

		regression variable(s)		
radar	uncorrected	R_r	distance	R_r & distance
Anjalankoski	12.032	10.309	12.032	10.379
Ikaalinen	11.055	9.571	11.055	9.573
Korppoo	11.537	9.508	11.537	9.271
Kuopio	11.359	9.547	11.359	9.268
Luosto	8.322	7.629	8.323	7.841
Utajärvi	10.905	9.595	10.905	9.436
Vantaa	11.827	10.115	11.827	9.810
Vimpeli	10.617	9.145	10.617	9.035

Table 2: F variances for uncorrected and regression-corrected hourly radar rainfall accumulations with different regression variables (observation period June-August 2013).

The validation results for the full model with Kriging interpolation are listed in Tables 3 and 4. In these tests the regression variables are chosen as the hourly radar rainfall accumulation and distance from radar. For variogram fitting and Kriging interpolation we have used the gstat package implemented for the R software. As seen from Table 4, interpolation by using nearby observations in space and time clearly reduces the variance of the observed F-values. The results show a clear improvement over a static regression model that does not take into account the time-dependent behaviour of gauge-radar errors.

radar	uncorrected	regression	regression+Kriging
Anjalankoski	1.614	-0.031	-0.024
Ikaalinen	0.719	-0.005	-0.010
Korppoo	1.142	0.020	0.016
Kuopio	0.411	-0.012	0.002
Luosto	0.400	0.001	0.055
Utajärvi	0.675	-0.032	-0.010
Vantaa	1.483	-0.046	-0.039
Vimpeli	0.627	0.003	0.006

Table 3: F means for uncorrected, regression-corrected and regression+Kriging -corrected hourly radar rainfall accumulations (observation period June-October 2013).

radar	uncorrected	regression	regression+Kriging
Anjalankoski	12.361	9.654	8.741
Ikaalinen	11.552	9.129	8.151
Korppoo	13.642	9.564	8.340
Kuopio	10.652	7.822	7.290
Luosto	8.183	6.182	5.623
Utajärvi	10.102	7.873	7.400
Vantaa	11.742	9.039	8.237
Vimpeli	10.691	8.218	7.340

Table 4: F variances for uncorrected, regression-corrected and regression+Kriging -corrected hourly radar rainfall accumulations (observation period June-October 2013).

The radar composites shown in Figures 3-5 demonstrate applicability of the model to obtaining gridded radar rainfall and uncertainty estimates. Here we have fitted the regression and variogram models individually for each radar from observation period June-October 2013. As above, the regression variables are chosen as hourly radar rainfall accumulation and distance from radar. The accumulation period shown in Figures 3-5 is 23 October 2013 03:00-04:00 UTC. The composites are constructed by taking the nearest radar for each map grid point.

Figure 3 shows the mean and variance estimates $\hat{\mu}_{\boldsymbol{H}}$ and $\hat{\sigma}^2_{\boldsymbol{H}}$ computed from the regression model. As also observed from Figure 1, here we can see that the model gives large mean values for F for small rainfall accumulations and small mean values for large accumulations far from the nearest radar. We can also clearly observe that the model gives a large variance at faraway locations. Figure 4 shows the variance estimates obtained by using only the regression model and by using the full model with Kriging. The observation is that the including the Kriging interpolation in the model yields significantly smaller variance near rain gauges. The variances at non-gauge locations can be considered as credible uncertainty estimates due to the correlation structure illustrated in Figure 2.

Finally, Figure 5 shows a gridded estimate of the expected ground rainfall obtained from the full regression-Kriging model by using equation (2.3) and the intervals in which the rainfall lies with 90% probability. The intervals are computed by assuming

that F is normally distributed with the estimated mean $\hat{\mu}$ and variance $\hat{\sigma}^2$. Assuming this, a symmetric interval around $\hat{\mu}$ is taken so that a 90% probability is obtained by integrating the normal density over this interval. The bounds of this interval are then transformed to the lognormal scale by using equation (2.3). Comparing Figure 5a to Figure 3a shows a desired result. That is, the "seams" between different radars are mostly removed, suggesting that the error models for individual radars yield correct adjustments.

4. Conclusions and Future Research

We have developed a novel multivariate kernel regression model for gauge-radar errors. The model is capable of explaining multiple factors contributing to the error distribution. Preliminary cross-validation results show clear improvement when compared to simple univariate regression models. The regression model is combined with spatiotemporal Kriging of the residual error component. Our results show that by using Kriging, the error variance can further be decreased near rain gauges. This is due to the strong spatiotemporal correlation structure of the residual error. The improved accuracy of the model is beneficial in many applications. Those include computation of probabilistic rainfall maps and testing the reliability of radar and gauge measurements.

We are implementing a four-dimensional spatiotemporal covariance model that includes altitude of radar bins as a third spatial coordinate. This approach allows using measurements from multiple elevation angles, thus increasing the size of available samples for variogram model fitting and Kriging interpolation. Our preliminary cross-validation results show a significant improvement to a model that uses only one elevation angle.

References

G. J. Ciach, W. F. Krajewski, and G. Villarini, "Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data," *Journal of Hydrometeorology*, vol. 8, no. 6, pp. 1325–1347, 2007.

E. Gregow, E. Saltikoff, S. Albers, and H. Hohti, "Precipitation accumulation analysis - assimilation of radar-gauge measurements and validation of different methods," *Hydrology and Earth System Sciences*, vol. 17, no. 10, pp. 4109–4120, 2013.

D. B. Michelson and J. Koistinen, "Gauge-radar network adjustment for the baltic sea experiment," *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, vol. 25, no. 10-12, pp. 915–920, 2000, first European Conference on Radar Meteorology.

G. Villarini, W. F. Krajewski, G. J. Ciach, and D. L. Zimmerman, "Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates," *Water Resources Research*, vol. 45, no. 1, 2009.