Data selection to assess bias in rainfall radar estimates: An entropy-based method

S. Spina, S. Sebastianelli, E. Ridolfi, F. Russo, L. Baldini, and L. Alfonso

Citation: AIP Conference Proceedings **1558**, 1665 (2013); doi: 10.1063/1.4825849
View online: http://dx.doi.org/10.1063/1.4825849
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1558?ver=pdfcov
Published by the AIP Publishing

Data Selection to Assess Bias in Rainfall Radar Estimates: An Entropy-based Method

Spina S.[a], Sebastianelli S.[b], Ridolfi E.[a,c], Russo F.[a], Baldini L.[b], and Alfonso L.[d]

[a]*Dipartimento di Ingegneria Civile, Edile e Ambientale, Sapienza University of Rome, Rome, Italy*
sandra.spina@uniroma1.it
[b]*Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, Rome, Italy*
[c]*H2CU-Honors Center of Italian Universities, Sapienza University of Rome, Rome, Italy*
[d]*Department of Integrated Water Systems and Governance - Hydroinformatics, UNESCO-IHE Institute for Water Education, Delft, The Netherlands*

Abstract. Miscalibration of radar determines a systematic error (i.e., bias) that is observed in radar estimates of rainfall. Although a rain gauge can provide a pointwise rainfall measurement, weather radar can cover an extended area. To compare the two measurements, it is necessary to individuate the weather radar measurements at the same location as the rain gauge. Bias is measured as the ratio between cumulative rain gauge measurements and the corresponding radar estimates. The rainfall is usually cumulated, taking into account all rainfall events registered in the target area. The contribution of this work is the determination of the optimal number of rainfall events that are necessary to calibrate rainfall radar. The proposed methodology is based on the entropy concept. In particular, the optimal number of events must fulfil two conditions, namely, maximisation of information content and minimisation of redundant information. To verify the methodology, the bias values are estimated with 1) a reduced number of events and 2) all available data. The proposed approach is tested on the Polar 55C weather radar located in the borough area of Rome (IT). The radar is calibrated against rainfall measurements of a couple of rain gauges placed in the Roman city centre. Analysing the information content of all data, it is found that it is possible to reduce the number of rainfall events without losing information in evaluating the bias.

INTRODUCTION

Rainfall radar estimates are affected by errors due to many causes that include, among others, radar miscalibration, range degradation (including beam broadening and sampling of precipitation at an increasing altitude), attenuation, ground clutter, variability of drop size distribution, vertical air motion, anomalous propagation and beam-blocking, variability of Z-R relationships, where Z is the radar reflectivity factor and R is the rainfall intensity [1]. Therefore, spatial and temporal averaging of radar and rain gauge data has always been used to reduce the measurement errors and the discrepancy between radar and rain gauge estimates. This work in particular focuses only on the calibration bias error, a systematic error affecting radar estimates of rainfall independently from both the instant of measurement and the location of the sampling volume. Calibration bias error is ascribed to a systematic error converting the power received by the radar receiver into radar reflectivity factor that requires either an a-priori knowledge or a difficult measurement on site of systems parameters, such as power transmission and antenna pattern that can vary, although slowly, with time. In the literature, several radar calibration methods can be combined to assess this bias such as the use of standard targets, the receiver static calibration, the use of the sunlight as a source of energy, the calibration using the self-consistency principle and the calibration with rain gauges [2, 3, 4]. The aim of this paper is to compare two methods to estimate the bias that affects radar estimates of rainfall: 1) a gauge-based radar calibration and 2) an information theory-based method. In the following section, the former method is described, followed by a description of the latter. Finally, results are presented and discussed and conclusions are drawn.

CASE STUDY

The rainfall monitoring system considered in this work is represented by the polarimetric Doppler radar Polar 55C located in the southeast of Rome and by a rain gauge network. Polar 55C is a C-band (5.6 GHz) Doppler dual polarised coherent weather radar with polarisation agility managed by the Institute of Atmospheric Sciences and Climate (ISAC), Rome (Italy). The radar is capable of transmitting and receiving horizontally and vertically polarised signals on alternate pulses, which measure the reflectivity factor (Zh), the differential reflectivity (Zdr) and

11th International Conference of Numerical Analysis and Applied Mathematics 2013
AIP Conf. Proc. 1558, 1665-1668 (2013); doi: 10.1063/1.4825849

the differential phase shift (dp). Radar measurements are obtained by averaging 64 pulses with a range-bin resolution of 75 m covering a 120 km radius from the radar site. This study considers a 1.5° elevation angle both to minimise the influence of ground clutter and to keep the radar beam low to sample precipitation close to the ground [5] .The elapsed time between two consecutive acquisitions with 1.5° elevation is approximately 5 min. Rain gauge data are provided by the Ostiense and the Via Marchi rain gauges located approximately 15 km away from the radar location inside the radar scanning area. These rain gauges are selected for the purpose of this work because radar errors at these gauge sites are due mainly to radar miscalibration (in addition to the variability of the Z-R relationship) as in [5, 1]. The influence of other types of errors (such as range degradation, urban clutter or beam-blocking) is negligible [5, 1]. Analyses are performed on a data set of 19 rainfall events aggregated at 30 min sampling time and collected by the two devices during the years 2008 and 2009 because differences between radar and gauge rainfall can depend on event characteristics (e.g., stratiform or convective).

CALIBRATION OF POLAR 55C RADAR WITH RAIN GAUGE

After removing background noise and ground clutter [5], radar reflectivity (Z_h) corresponding to meteorological returns is converted into rainfall intensity (R) by using the following relationship [6]:

$$R = 0.19055 \cdot 10^{0.5358 (Z_h/10)} \tag{1}$$

Polar coordinates are then transformed into Cartesian coordinates, and the 2 km x 2 km grid was overlaid with the map area. For each time interval, rainfall values are obtained at each pixel, and radar rainfall is then cumulated with a time resolution of 30 min to minimise the effects of mismatches in time and space due to sampling differences between radar and rain gauges [7, 8]. For each observed event, pairs of rainfall time series are obtained, coupling radar and rain gauge data. In particular, radar series refers only to the pixel of the Cartesian grid where the rain gauges are located. For each data set, the bias B is obtained as follows [5, 1]:

$$B = \frac{\sum_{j=1}^{M}\sum_{i=1}^{N} G_{ij}}{\sum_{j=1}^{M}\sum_{i=1}^{N} R_{ij}} \tag{2}$$

where Gij and Rij are rain gauge and radar rainfall amounts, respectively, M is the number of rainfall events, and N is the length of the events. B is a dimensionless multiplicative bias that is applied uniformly to each rainfall value estimated by the radar to obtain the corresponding measurement collected by the gauge. When the number N of events is that of the whole data set (i.e., 19), the value of B equals 0.333 and 0.326 for the Ostiense and the Via Marchi rain gauges, respectively.

THE ENTROPY-BASED METHOD

The so-called heuristic entropy was introduced by Shannon [9] as a measure of the uncertainty associated with a certain event. Heuristic entropy was used to solve a wide range of issues and applied to several fields ranging from mathematics to computer science, from ecology to economics. Heuristic entropy was also applied to hydrologic and water resources fields. For instance, entropy was used to define rainfall threshold values through the minimisation of a risk entropy-based function [10], to optimise location of water level monitors [11, 12], to design a water monitoring sensor network for flood monitoring purposes [13] and to find the maximum non-redundant information content of a rain gauge network [14]. An extensive review of other applications to hydraulic issues is reported by Singh [15]. In particular, the marginal entropy of a discrete random vector (RV) is defined as:

$$H(X) = -\sum_{i=1}^{N} p(x_i)\log_2[p(x_i)] \tag{3}$$

where $p(x_i)$ is the probability that the RV X assumes the value x. The joint entropy of M RVs is:

$$JH = H(X_1, X_2, ..., X_M) = -\sum_{i_1=1}^{N_1} ... \sum_{i_M=1}^{N_M} p_{1_1,...,i_M} \log_2\left(p_{1_1,...,i_M}\right) \tag{4}$$

where $p_{i1...iM}$ is the joint probability of the M variables. As in this work, if the logarithm is base 2, entropy is measured in bits. In this paper, an entropy-based method is used to choose the optimal set of events to estimate the bias. In this case, X_1, X_2,…X_M are the decision variables that represent the difference between rainfall radar estimates and rain gauge measurements or the difference between rainfall events measured by radar and by the Ostiense rain gauge. This set, in particular, should have maximum joint entropy (as a measure of the information content) and minimum total correlation (as a measure of redundancy). The latter is evaluated as:

$$C(X_1, X_2,...X_M) = \sum_{i=1}^{M} H(X_i) - H(X_1, X_2,..X_M) \tag{5}$$

The optimisation issue is posed as a Multi-Objective Optimization Method (MOOP) [16], and the issue is solved by the Non-dominated Sorting Genetic Algorithm (NSGA-II) [17]. The two objective functions are:

$$\begin{aligned} Min(C) &= Min\{C(X_1, X_2,..., X_M)\} \\ Max(JH) &= Max\{H(X_1, X_2,..., X_M)\} \end{aligned} \tag{6}$$

where C and JH are the total correlation and the joint entropy, respectively, of the M variables. To evaluate the total correlation of discrete random vectors, the grouping property of mutual information is adopted [18]. A further explanation of this method can be found in [19]. The optimal solutions are plotted on a Pareto front (Fig. 1) considering a number of events ranging from 2 to 8. Each point in the graph represents a potential solution, consisting of a set of rainfall events to calibrate the radar estimates.

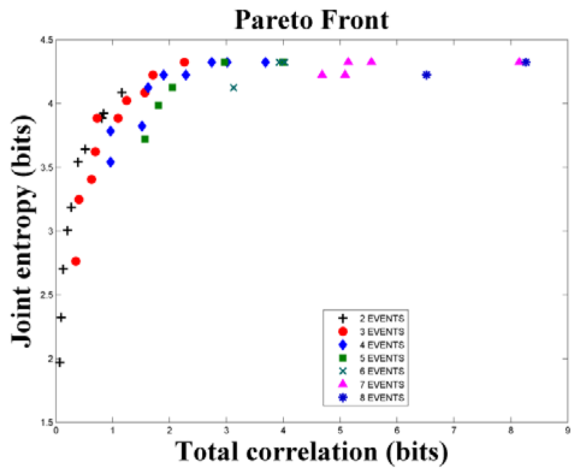

FIGURE 1. Pareto front of optimal set of solutions, considering an increasing number of events.

Fig. 1 shows that the curves converge to the maximum joint entropy value (i.e., log2N, where N is the RVs length) as soon as the number of events is equal to 5. Fig. 1 demonstrates that it is possible to choose a subset of events that are as much informative as the complete data set. The bias value is then estimated for all solutions through Eq. (2). In particular, results highlight that the set of 5 events is characterised by a bias value equal to 0,331, close to the value computed with the complete data set. To compare the two bias values, the Fractional Standard Error (FSE) is used. FSE is defined as follows:

$$FSE = \left[\frac{1}{N}\sum_{i=1}^{N}(R_i - G_i)^2\right]^{0.5} \cdot \left(\frac{1}{N}\sum_{i=1}^{N} G_i\right)^{-1} \tag{7}$$

The mean FSE values considering 19 and five events are, respectively, equal to 1.840 and 1.839. To validate the events selection procedure through entropy, radar data are calibrated against the Via Marchi measurements using the same subset of events. Results show the same performance as the previous case: the bias value equals 0.310. The entropy can therefore be considered a valid support in radar calibration, detecting a significant sample of events (i.e., only informative ones). The procedure leads to minimisation of the computational time and the amount of data required for the process.

CONCLUSIONS

This paper evaluates the information content of rainfall events registered by weather radar and by the corresponding rain gauge for validation. First, the radar is calibrated considering all rainfall events, and then the optimal number of events is defined through an entropy-based approach. In particular, the selected subset should fulfil the two conditions of maximising joint information and minimising total redundancy. The best solutions are plotted on a Pareto front. Analyses show that a number of five events satisfies the two objective functions. Bias values estimated using these five events and the 19 events available are slightly different, but the average error values, computed through the FSE, are the same for the two cases. Results are confirmed by repeating the procedure using data provided by another rain gauge. Results show that radar calibration, through bias estimation, can be performed by lowering the number of rainfall events without any loss of information content.

ACKNOWLEDGMENTS

The authors are grateful to Ufficio Idrografico e Mareografico of Lazio Region and to Institute of Atmospheric Sciences and Climate (ISAC), Rome (Italy) for providing data series.

REFERENCES

1. S. Sebastianelli, F. Russo, F. Napolitano and L. Baldini, Assessment of an adjustment factor to model radar range dependent error, *AIP Conf. Proc*. **1479**, 1784-178, doi: 10.1063/1.4756523 (2012).
2. Atlas D.: Radar calibration: some simple approaches, Bull. Amer. Meteor. Soc 83, 1313-1316 (2002) .
3. G. Scarchilli, E. Gorgucci, V. Chandrasekar and A. Dobaie, Self-consistency of polarization diversity measurement of rainfall, *IEEE Trans. Geosci. Remote Sens*., **34**, 22-26 (1996).
4. J. Koistinen and T. Puhakka, Can we calibrate radar with raingauges, *Geophysica*, **22 (1-2)**, 119-129 (1986).
5. S. Sebastianelli, F. Russo, F. Napolitano and L. Baldini, On precipitation measurements collected by a weather radar and a rain gauge network, *Nat. Hazards Earth Syst. Sci*., **13**, 605-623, doi:10.5194/nhess-13-605-2013 (2013).
6. E. Gorgucci and G. Baldini, Radar An examination of the validity of the mean raindrop-shape model for dual-polarization radar rainfall retrievals, *IEEE T. Geosci. Remote,* **47(8)**, 2752-2761 (2009).
7. G. Villarini and W.F Krajewski, Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall, *Surv. Geophys*., **31(1)**, 107-129 (2010).
8. K. Ozturk and A.U. Yilmazer, Improving the accuracy of the radar rainfall estimates using gage adjustment techniques: Case study for west Anatolia, Turkey, *Atmos. Res*., **86**, 139-148 (2007).
9. C.E Shannon, A mathematical theory of communication, *Bell Syst. Tech. J*., **27**, 379-423, 623-656 (1948).
10. V. Montesarchio, E. Ridolfi, F. Russo and F. Napolitano, Rainfall threshold definition using an entropy decision approach and radar data, *Nat. Hazards Earth Sys*., **11**, 2061–2074 (2011).
11. L. Alfonso, A. Lobbrecht, and R. Price, Information theory-based approach for location of monitoring water level gauges in polders, *Water Resour. Res*., **46**, W03528, doi: 10.1029/2009WR008101 (2010).
12. L. Alfonso, A. Lobbrecht, and R. Price, Optimization of water level monitoring network in polder systems using information theory, *Water Resour. Res.,* **46**, W12553, doi: 10.1029/2009WR008953 (2010).
13. E. Ridolfi, K. Yan, L. Alfonso, G. Di Baldassarre, F. Napolitano, F. Russo, and P.D. Bates, An Entropy Method for Floodplain Monitoring Network Design, *AIP Conf. Proc.* **1479**, 1780-1783, doi: 10.1063/1.4756522 (2012).
14. E. Ridolfi, V. Montesarchio, F. Russo and F. Napolitano, An entropy approach for evaluating the maximum information content achievable by an urban rainfall network, *Nat. Hazard. Earth Sys*., **11**, 2075-2083 (2011).
15. V.P. Singh, The use of entropy in hydrology and water resources, *Hydrol. Process*., **11**, 587-626 (1997).
16. E. Ridolfi,, L. Alfonso, G. Di Baldassarre, F. Dottori, F. Russo, and F. Napolitano, An entropy approach for the optimization of cross-section spacing for river modelling. *Hydrological Sciences Journal*, doi: 10.1080/02626667.2013.822640 (2013).
17. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, *Evol. Comput*., **6**, 182-197 (2002).
18. A. Kraskov, H. Stoegbauer, R.G. Andrzejak and P. Grassberger, Hierarichical clustering based on mutual information, *Europhys. Let*t., **70**, 278 (2005).
19. L. Alfonso, L. He, A. Lobbrecht and R. Price, Information theory applied to evaluate the discharge monitoring network of the Magdalena River. *J. Hydroinform*, **15(1)**, 211–228, doi:10.2166/hydro.2012.066 (201