

Response to the review of “Should radar precipitation depend on incident air temperature? A new estimation algorithm for cold climates”.

RC2: Responses to Anonymous Referee 2

The authors wish to thank the reviewer for his constructive comments and corrections to the discussion paper. In the following, we have responded to each of the comments from the reviewer. The comment from the reviewer (RC) is in italic font while the author comment (AC) and changes in the manuscript (CM) are in blue normal font. Revised Figures are added herewith as Appendices.

RC: Overview

In their paper entitled “Should radar precipitation depend on incident air temperature? A new estimation algorithm for cold climates”, Sivasubramaniam and colleagues present an attempt to improve radar-based quantitative estimation of precipitation by using observed near surface air temperature as a co-variate. The authors claim that the inclusion of air temperature in a non-parametric statistical estimation technique reduces the RMSE of precipitation estimates, particularly for cold climates.

The paper is fairly well written and fits the scope of HESS, although a discussion of hydrological implications is clearly missing. However, I have serious doubts about the methodological approach, the significance of the results, and the conclusions drawn by the authors, some of which I will explain in the following.

Understanding the cumulative and non-uniform effects of temperature

The authors state, on p. 2, ll. 9 ff. that “the scope of the present study is limited to radar precipitation estimation uncertainty during conversion from reflectivity to rain rate, with a focus on cold regions experiencing a mixture of solid and liquid precipitation.” I doubt that. Temperature not only affects the R(Z) relationship e.g. subject to the precipitation phase. Near surface air temperature either indirectly affects or correlates with different processes in the formation of precipitation in the atmosphere, but also along the radar observation and processing chain. That might lead to systematic effects on estimated precipitation intensities which accumulate over the entire estimation chain with R(Z) transformation only being the very last step. Near surface air temperature is e. g. indicative of different vertical reflectivity gradients (thus affecting observed reflectivity as a function of distance from radar), and also vertical air density gradients (affecting atmospheric refractivity and thus beam propagation / altitude). Higher precipitation intensities and thus path-integrated attenuation tend to increase with higher temperatures. The study misses is a systematic framework that takes into account different temperature effects that cause systematic precipitation estimation errors. The fact that the VPR is addressed in the met.no radar data processing chain does not mean that VPR effects (or the effects of correction) are not systematically present in the data anymore. A way to better understand these effects is to use polarimetric radar observations. That way, snow and rain can be discriminated (where radar actually measures them – not in a gauge on the ground!), so the quantitative effects could be investigated in order to understand the contribution of R(Z) uncertainty. The authors provide a general reasoning about some temperature-related effects (p. 3. ll. 25-30), but these effects are never really picked up again in the rest of the paper.

AC: We agree with the reviewer that there are effects of temperature with different processes in the formation of precipitation in the atmosphere and along the radar observation. However, in this study, we investigated the radar data product from met.no, which are converted by using single Z-R equation. The focus of this study is to improve the data for practical applications by using the available data sources. As discussed in the summary and

conclusion (ref. p. 14 ll. 20 – 22), if precipitation phase can be identified by alternative mechanism (e.g., polarimetric radar), the temperature effects can be better understood.

CM: The “Results” section will be renamed as “Results and Discussion” and hydrological implications will be discussed.

The following text will be added.

“A more accurate estimation of radar precipitation can provide better estimation of hydrological response. It is common practice to adjust radar precipitation against gauge observation. It should be noted though that this adjustment, while effective in an operational sense, can often mask the corrections that could be incorporated in the estimation procedure through use of additional data for the VPR and its relationship to snow. We recommend, when such data is readily available, use be made to infer the changes to the VPR for the low temperatures in focus here. This, however, is not attempted in our study as we are restricted to using available gauge and weather radar data collected in Norway.”

RC: Effects of systematic undercatch of snowfall

Still, one might argue: I am not so much interested in understanding the processes behind the phenomenon. I just want to produce a better precipitation estimate (i.e. decrease the systematic error). But is that really achieved? On p. 3, ll. 1-15, the authors illustrate the motivation of their study by showing different regression slopes between radar-based precipitation (predictor) and precipitation as observed by the gauge (independent variable), for assumed snowfall and rainfall conditions. That example clearly reveals a fundamental issue: Measurement of snowfall by precipitation gauges has consistently been shown to exhibit a systematic undercatch that is significantly more pronounced than the undercatch of rainfall (see e.g. Gross et al. 2017, Wolff et al.2015). So which effect do the authors observe: a “bias in the radar precipitation estimation for snow” (p. 3, l. 11), as assumed out by the authors, or a bias in the snow observation by gauges. Maybe a mix of both? In my opinion, it is almost impossible to reach any substantial conclusions based on the data and methods presented in the study.

AC: We agree with the reviewer that precipitation undercatch is more important for snowfall than for rain. However, for the gauges used in this study, there is a lack of data for performing wind corrections. Only 15 out of 88 gauges are equipped with wind speed measurement and thereby suitable for correction. By using the Nordic correction model (Wolff et al., 2015), we corrected precipitation at temperatures below +3 C for wind induced under catch for these 15 gauges. It was found that 14 gauges out of 15 showed less correlation between radar precipitation and gauge precipitation after wind correction. Further, the total radar precipitation volume is less than the uncorrected gauge precipitation volume computed from the data used in the study for all those 15 gauges. We think this analysis show that there is a winter underestimation of precipitation by the radar. Hence, the bias is not due to precipitation undercatch by gauges for snow in the data.

CM: Text will be included to the paragraph on p3, l1-14 and the precipitation under catch and catch correction will be discussed.

RC: Transparency of the cross-validation framework

The methods section does not elaborate on the leave-one-out-cross-validation (LOOCV) setup. Only in the results section, on p. 10, ll. 17-25, the application of LOOCV is pointed out. Still, the exact setup of the LOOCV remains unclear and leaves the reader with substantial doubts. If one gauge is left out to test the prediction, on which basis are the partial weights inferred for that prediction? From the nearest neighbour? From a weighted average of neighbours? For Fig. 3, as a result of what is described on p. 10, ll. 21-25, we do not know

that. For Fig. 4, where an average partial weight of the entire study area is used, we do not know whether LOOCV has been applied at all, and on what basis. Only for Fig. 5, the authors state that for each gauge, the partial weights had been derived from the five neighbouring gauges. Apart from that, p. 11, l. 16 – p. 12, ll. 1-2, casts serious doubts on the integrity of the LOOCV setup: “As mentioned in section 3, this study uses the gauge precipitation as the observed response for the regression estimation. So insufficient data points can also be the reason for lesser improvement in these locations because nonparametric k-nearest neighbour prediction (a data-based model) depends highly on the availability of sufficient data.” Does that mean that only the partial weights are independently computed from the validation target, while the observations of the target are still used in order to carry out the k-nearest neighbour regression? That could not be considered a valid LOOCV approach. The issue needs to be clarified in the minutest details, including a full disclosure of the data and the code used for the analysis. In this context, I have to emphasize that I cannot say anything about the nonparametric statistical techniques used in this study, which are outlined in section 3.1. I assume these techniques are fine, but I do not feel qualified to assess their applicability and the implications for the validity of the LOOCV, at least not from the present manuscript.

AC: We agree with the reviewer that LOOCV setup was not elaborated in the manuscript. We clarify as follows and we will add text to the manuscript.

In the study, two settings of LOOCV were used.

1) k-nn regression estimate of expected response by leave one out cross validation (“knnreg1cv” - tool in the NPRED package (Sharma et al., 2016)).

As described in the manuscript (p.10, ll. 17-18), k-nn regression estimate of expected response was calculated by leaving out that observed response value from the regression and then RMSE was calculated. For each gauge location, data from that gauge location was used (Results on Figure 3).

Partial weight for each gauge location was calculated independently using the data at each gauge location and then RMSE was estimated as described above by using entire data at that gauge location. However, a split sample test was done to verify the results, where two-thirds of the data were used to estimate partial weight and one-third of the data were used to estimate RMSE for each gauge location. The split sample test gave the same results as before.

2) Spatial cross validation

As described in the manuscript (p.12, ll.3-5), for each gauge location, partial weight was calculated by leaving that gauge out and from 5 nearest neighbours. Then for that gauge location (independent from partial weight computation), expected response was calculated by using tool “knnreglcv” in the NPRED and then RMSE was calculated (Results on Figure 5).

A single average partial weight was calculated by arithmetic mean value of partial weights of all gauge locations which were computed independently at each gauge location and presented in the Figure 2 and Table 1. The idea is to use a single regional value for partial weight. RMSE was calculated and presented (Results on Figure 4).

CM: Text will be updated in the manuscript (Results section) to clearly describe the two settings of LOOCV used in the study.

RC: Other issues (some of them major)

The entire section 2 (background) is far too extensive and provides a lot of information that is not pertinent to the study, and that does not play a role for discussing the results.

AC / CM: Background section provides additional information for those who are unfamiliar with the topic. We will update the text to make it more succinct where this can be accomplished without sacrificing clarity.

How is Eq. 1 an equation?

AC: Expression can be a better word.

CM: "Equation" will be replaced with "Expression".

Section 3.2: Would be helpful to additionally use an evaluation criterion that measures the systematic error (e.g. mean error).

AC: We estimated mean absolute error (MAE) and it showed similar patterns as RMSE. We feel use of additional metrics may not offer new insights to what we have reported here.

p. 8, ll. 27-29: It should be carefully analyzed whether simply choosing the nearest neighbour provides the best correspondence between radar-based QPE and rain gauge observations. Particularly at hourly intervals, the consistency depends on the neighbourhood definition due to representativeness issues.

AC: We agree that there are representativeness issues. However, in this study we assumed that, radar precipitation estimates at a gauge location are as same as the pixel location where the gauge located.

p. 8, ll. 33-25: Including such intensities as low as 0.05 – 0.5 mm might lead to the fact that insignificant precipitation dominates the results in terms of relative changes of the RMSE, as hourly rainfall follows a gamma-like distribution. Confining the analysis to significant precipitation should address that issue.

AC: Intensity of precipitation in the study area is low moreover; winter precipitation intensity is very low. Nearly 75 % of the data used in the study is within the intensities of 0.05 - 0.5 mm at temperatures below +3° C.

Most figures in section 5: I find the continuous colorbars very difficult to interpret.

Please use a discrete colorbar instead.

AC / CM: The colour ramp for the Figure 2 will be differentiated from Figure 3-5 (ref. Figure 2 Appendix). It can be noted that presented values are in the continuous domain, continuous colour bar could be more appropriate.

p. 10, ll. 4-6: How do you know the number of snow data pairs?

AC: Air temperature was used to classify the data pairs as snow or rain and hence number of snow pairs were counted in the preliminary investigation presented in the manuscript (ref. p.3, ll.5-6). As Figure 1 is referred, simplistic estimation of snow pairs was mentioned in this sentence.

p. 10, ll. 9-10: "This outcome is a result of sampling uncertainty due to which a minimum of 0.2 for the partial weight for radar rain rate has been used in the results." That explanation is not satisfactory at all. More generally, any gauge with a very high partial weight for temperature should be considered with great caution – why should temperature be a better predictor than radar rainfall?? Presumably only in case radar rainfall at that particular location is affected by serious artefacts. Here, a systematic analysis of the relation between radar and gauges on a per-gauge basis is required, together with a spatial analysis of systematic errors in the radar rainfall (e.g. due to partial beam blockage, residual clutter, ...).

AC / CM: Please note our response against this point to Reviewer 1. Essentially, there is sampling uncertainty in any statistic that is estimated. Similar uncertainty exists for the partial weight given it is a nonparametric measure. We argue that physical reasoning invalidates radar precipitation to be a function of temperature and not of ground precipitation. Hence we impose a prior belief that disallows this possibility. The threshold adopted reflects our prior belief and is consistent with any similar modelling applications that are used in hydrology. We emphasise here that our aim is to assess simply the importance of temperature on radar precipitation estimation, keeping other factors unchanged. We have approached this issue using extensive data allowing us to statistically meaningful assertions. It is for this reason we speculate on possible factors that may be impacting our results (in the lines of what the assessor has mentioned) but keep our results focussed on the temperature related impacts noticed. We will clarify this further and note other possible factors more clearly in the revised manuscript.

Table 1: Why not show a histogram instead?

AC: Histogram will be presented with Table 1. However, we keep Table 1 as it summarises the partial weight values which are used to describe the results in the manuscript (p.10, ll.13).

CM: Histogram will be added to the manuscript (ref. Figure *new nr.*, Appendix)

p. 10, ll. 27-28: "Further, it can be noted that all the gauge locations with an associated partial weight of air temperature ($\beta_T > 0$) shows an improvement in radar precipitation estimation." What does that mean in the context of a crossvalidation where you do not know the partial weight at the target location?

AC: As described above for the response to "RC: Transparency of the cross-validation framework", two LOOCV settings used in this study. Here the "knnreg1cv", tool in the NPRED was used for each gauge location. Partial weight was estimated independently for each gauge location and that partial weight was used in k-nn regression.

p. 10, ll. 1-2: "raingauge locations with a minimum partial weight for radar rain rate ($\beta_R = 0.2$ $\beta_T = 0.8$), did not show improvement in RMSE." How can you compare the two settings – one only based on radar rainfall and the other with β_R arbitrarily set to 0.2?

AC: Partial weight computation resulted in a zero partial weight for radar rain rate on those locations. Presumably, partial weight of radar precipitation rate could not be 0 and hence we set a minimum partial weight of 0.2 for radar precipitation rate. However, in order to inform the reader that whether such less partial weight for radar rain rate (0.2) can or cannot improve RMSE, RMSE was estimated and presented.

For all investigated changes in RMSE, please investigate the statistical significance of that change (e.g. by using bootstrapping). E.g. on p. 11, ll. 7-8, it is stated that "over 80 percent of the gauge locations in the study area show more than 3 percent improvement in RMSE". Which portion of these changes is statistically significant?

AC: All changes denoted in the manuscript are assessed using cross-validation. As a result, any differences, when present, represent statistical significance. There is no need of bootstrapping in this scenario as model complexity is not relevant in the results obtained. In fact, if the model were overly complex (for instance using additional predictor variables or a smaller value of 'k'), cross-validation performance will deteriorate, with poorer results than have been presented.

p. 11, ll. 3-11: What is the implication of the fact that using an average partial weight over the entire study area produces better results? Was that result also achieved from an LOOCV? It

is important to understand how the partial weights were assigned in the analysis that assigns a specific partial weight to each station, on order to understand the implications of this experiment.

AC: The implication is that a single regional partial weight can be used at any ungauged location within the study area. A single average partial weight is an arithmetic mean value of the partial weights of gauge locations in the study area. It can be noted that results (improvement in RMSE) from single average resembles with the results from 5 nearest neighbours (ref. p.12, ll. 6-8) which was achieved by LOOCV.

p. 11, ll. 3-11: One implication of this paragraph is that the spatial pattern of partial weights is meaningless (unless proved otherwise), since a simple average provides better results. As a consequence, I recommend to drop the maps in section 5, and show e.g. histograms instead, or any other visualisation that allows the reader to better understand the quantitative implications of the results.

AC: Spatial plots shows also the reader that the temperature dependency is spread throughout the study area. Further, it can be noted that there are spatial variation of partial weight and RMSE, e.g., the most southerly locations showed lesser improvement in RMSE (ref. p. 12, ll. 9-13).

CM: Histogram will be added to show the estimated partial weights at gauge locations (ref. Figure *new nr.*, Appendix).

p. 12, ll. 11-13: Meaning remains unclear.

AC: The most southerly locations showed less dependency on air temperature. The mean negative temperature is lower (almost zero degree Celsius) for these locations than other locations. There can be no many snow events at these locations.

CM: The following sentence is added on p.12, ll.12. "Hence there can be no many snow events at these locations."

p. 13, ll. 1-5: "[...] resulted in a maximum of 20 percent total improvement [...]" – where has that been shown? "[...] The nonparametric k-nn predictive model with radar rain rate as a single predictor improves the prediction." – that has not been shown in the results, either – no benchmark based on a direct rainfall estimation from reflectivity is shown.

AC: As main focus of this paper is to investigate the dependency of radar precipitation on air temperature, results from the k-nn predictive model with radar rain rate as a single predictor was not presented for all stations. To compare and discuss the results with those in the literature, the maximum improvement was reported in the results and then discussed.

CM: Text will be updated.

p. 13, l. 12 ff.: That paragraph basically shows that the predictive model does not contribute to an understanding of temperature effects. It rather feeds the suspicion that the effect of temperature merely balances different observational biases of the precipitation gauges with regard to rain and snow.

AC: We repeated the computation for 15-gauge locations where wind speed measurements are available for wind induced precipitation catch correction. Partial weight and RMSE were estimated with corrected precipitation as observed response for these gauges. After we corrected the gauges for catch correction, we still see a partial weight for air temperature and an improvement in RMSE for the corrected gauges. We do believe that this shows a temperature effect and the viability of the method presented.

References

- SHARMA, A., MEHROTRA, R., LI, J. & JHA, S. 2016. A programming tool for nonparametric system prediction using Partial Informational Correlation and Partial Weights. *Environmental Modelling & Software*, 83, 271-275.
- WOLFF, M. A., ISAKSEN, K., PETERSEN-ØVERLEIR, A., ØDEMARK, K., REITAN, T. & BRÆKKAN, R. 2015. Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrol. Earth Syst. Sci.*, 19, 951-967.

Appendix – Revised Figures

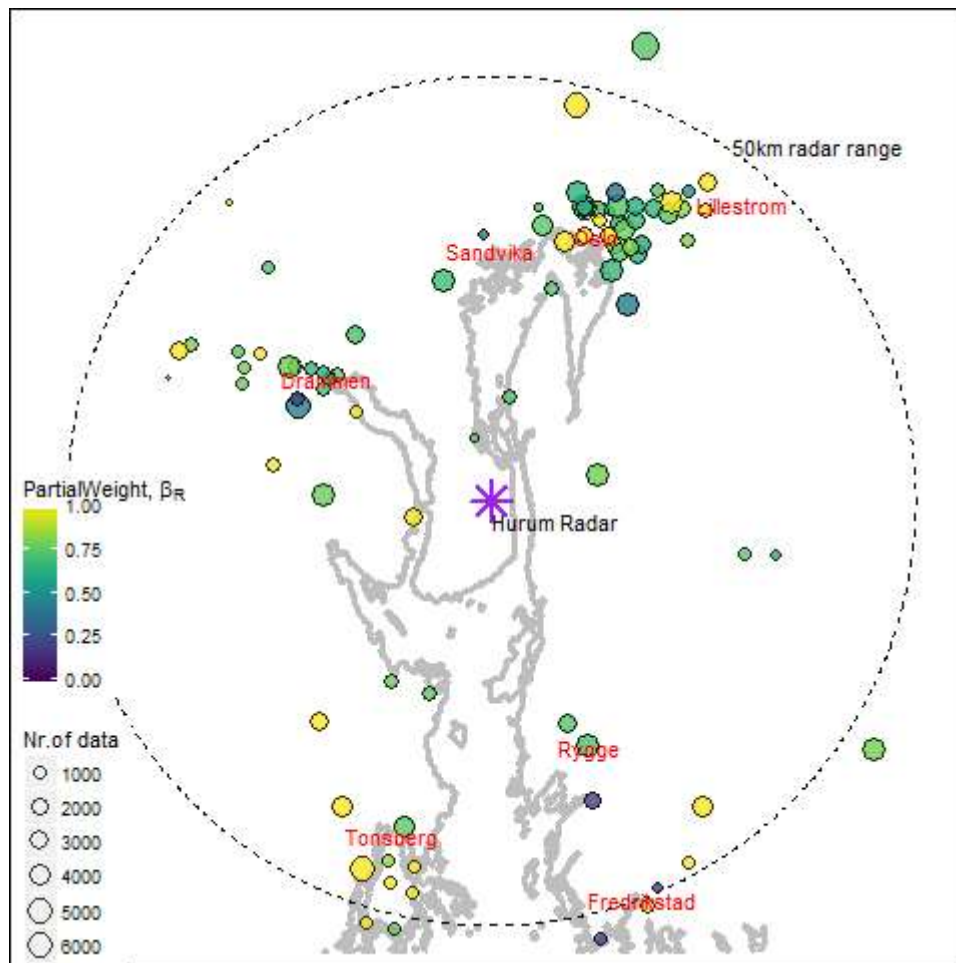


Figure 2. Partial weights of radar precipitation rate (β_R) for precipitation gauge locations (colour scale) and number of data pairs (circle size), which are used to estimate partial weight at each gauge location, overlaid on the coastline of the study area. Partial weights provide a measure of relative importance of predictor variables on the response and the summation of partial weights ($\beta_R + \beta_T = 1$) is equal to 1.

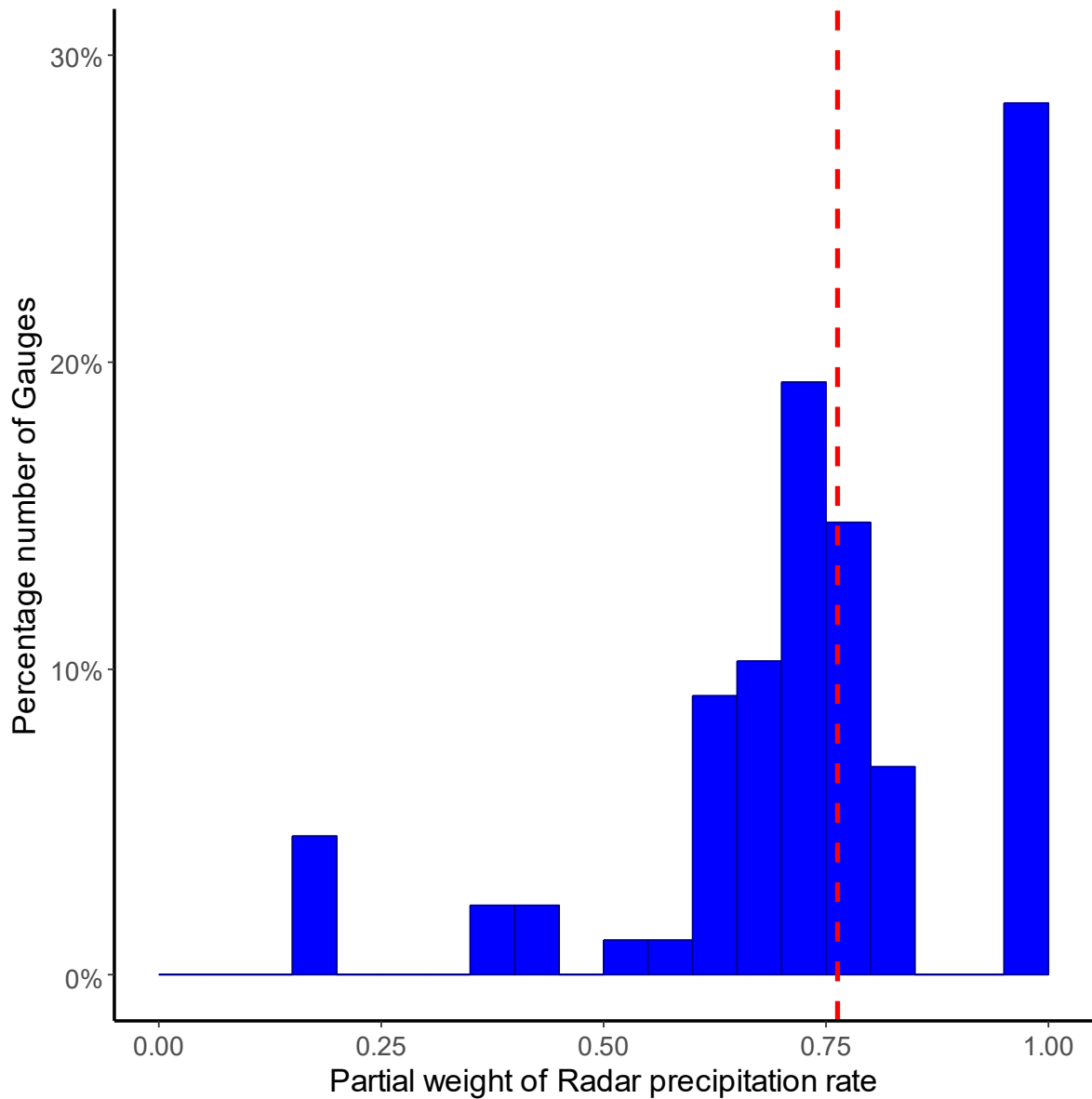


Figure *new nr*. Percentage number of precipitation gauges against estimated partial weight of radar precipitation rate (β_R) for those gauge locations and the mean partial weight (red dash line) for the study area. Total number of gauges used in the study = 88.