

Reply to Referee #2

We thank Referee #2 for the detailed and constructive review and for the suggestions to strengthen the study. We have carefully considered each comment and describe below the corresponding clarifications and proposed revisions, as well as our rationale where we prefer to retain the present scope of the study.

5 Major revision

- 1. The motivation (the why) of this study is not completely clear from the Introduction only. Section 5.2 in Discussion helps to better understand the motivation. However, from Results in Section 4, I am left thinking “so what?”. To make the paper more concrete, I recommend adding at least one numerical experiment demonstrating an application of the Tweedie power- p to assess the statistical realism of precipitation products. For example, the Tweedie power- p for different spatial or temporal scales could be assessed for one or several statistical or artificial intelligence-based precipitation downscaling methods.*

We agree that the practical aspect of the motivation is developed more clearly in Sect. 5.2 than in the Introduction, and we will revise the Introduction to explain more clearly why characterizing the dependence of Tweedie power is useful. We have also considered the suggested numerical experiment carefully. However, applying the framework to a particular precipitation product would merit a separate model-evaluation study beyond the present objective.

The purpose of this study is to establish how p varies across the relevant dimensions using observations. These dependencies provide observational reference behaviour for subsequent applications, where a precipitation product can be assessed according to whether it reproduces the observed scale-dependent mean–variance structure, rather than matching precipitation statistics only at a single scale. Such an evaluation would need to account for the characteristics of the particular product and the observational reference used for comparison. We will therefore retain the present observational focus rather than introduce a model-specific numerical experiment, while strengthening the Introduction and the connection between the Results and Sect. 5.2 to make the intended application of the framework more concrete. We will also clarify that evaluating individual statistical, numerical, or machine-learning precipitation products against these observational references is an extension of the present work.

- 2. Results show that Tweedie power- p varies depending on accumulation length, radar VS gauge observations, locations and seasons, but it is not clear what causes p to vary. It is not clear whether the difference between the Tweedie power- p for radar and gauge is due to observation errors, to spatial aggregation (for the radar) or to representativeness error. A deeper analysis of the behavior of the Tweedie power- p using a different (not necessarily better) statistical*

model of precipitation would provide some insight. For example, using the Markov-Gamma statistical model and aggregating to different interval lengths, you could compute the Tweedie power- p and see if it is constant or varies with aggregation length. A more advanced spatio-temporal model of precipitation could also be used to study the effect of spatial aggregation and compare it with radar precipitation. Finally, a model of observation error could be used to assess the impact of ground clutter or radio-frequency interference on the Tweedie power- p .

By definition, p characterizes the mean–variance scaling of precipitation and can therefore vary when the statistical structure of accumulated precipitation changes. This may reflect differences in precipitation occurrence and intensity distributions, temporal and spatial aggregation, representativeness, and observing-system characteristics. While the present analysis demonstrates that p varies across the dimensions examined, these results alone cannot determine the individual contributions of these factors.

We have considered the suggested experiments using Markov–Gamma, spatiotemporal, and observation-error models. Such experiments could assess the sensitivity of p to prescribed mechanisms under controlled assumptions. For example, a Markov–Gamma process could be aggregated to different interval lengths to examine whether it reproduces similar scale dependence. However, such experiments would not uniquely attribute the observed differences to those mechanisms, and the resulting sensitivity would remain conditional on the assumed model, its parameterization, and associated uncertainties. A systematic attribution of the observed variations would therefore require a mechanistic modelling analysis. We will revise the interpretation to clarify this distinction and make the limitations more explicit regarding spatial aggregation, representativeness, and observational errors as potential contributors to the radar–gauge differences.

Minor comments

1. l. 27 : “systems” -> “weather systems”

We agree and will revise that at l. 27, as well as at l. 286.

2. l. 33: “rainfall statistics” -> “rainfall statistics at a particular location” (note that these statistics do not account for the spatial structure of precipitation)

We agree and will revise this. We will also clarify elsewhere in the manuscript that the rainfall statistics considered here are location-specific and do not characterize the spatial structure of precipitation.

65 3. *Eq. 1: why only show the wet case? “where the wet or dry probability only depends on the previous state $W_{t-1} = i$: $P(W_t = j | W_{t-1} = i) = p_{ij}$, where I and j are wet-dry indicators.*

Eq. (1) is intended as an example describing the conditional probability, but we agree that the dry probability should also be represented to state the model more clearly. We will revise Eq. (1) accordingly.

70

4. *Eq. 2: Define parameters k and θ .*

We agree and will define these Gamma parameters.

75 5. *l. 50-54: clarify if μ_t , p and ϕ are constants over time*

We agree that this should be clarified. We will state that μ_t is interval-specific, whereas p and ϕ are constant within a given fit. Separate values of p and ϕ are estimated for different accumulation lengths, seasons, locations, and observing systems.

80

6. *Define zero-inflated, positively skewed totals...*

We will clarify that this refers to precipitation accumulation totals characterized by a high frequency of zero values and a positively skewed distribution of non-zero amounts.

85

7. *How do we find if the total precipitation amounts are zero given a mean μ_t and a variance $\text{Var}(Y_t)$, i.e. can you sample the probability distribution? I would suggest to present the Poisson-Gamma distribution for $1 < p < 2$ first and mention that the Tweedie distribution generalizes it for other p 's.*

90 The compound Poisson–Gamma representation in Eq. (4) already shows how zeros occur, but we agree that this can be made clearer. We will include the corresponding zero probability using the Poisson representation and emphasize that the Poisson–Gamma representation corresponds to the Tweedie family.

8. *l. 57: The parameter λ_t is not defined, what range of values can it take?*

95

We will define the Poisson parameter and clarify its positive range.

9. *l. 57: What represents $X_{t,i}$, what are the parameters, are they all the same for all t and i ?*

100 We will define $X_{t,i}$ and clarify that these Gamma parameters are fixed within a given model fit.

105 *10. (Introduction) Why do we care about that the Tweedie power p varies “across space, time, and observing systems”? What are the concrete implications? Can you make a stronger point for applications (see major revision 1)? I don’t see how this study will be useful for “the evaluation of precipitation statistics in radar products, nowcasting systems, and other high-resolution earth system models”. I suggest a slight rewrite by bringing some of the points presented in Discussion (5.2) as well as introducing the application example from major revision 1.*

110 As noted in our response to Major Comment 1, we agree that the practical motivation should be made more explicit in the Introduction and will bring forward relevant points from Sect. 5.2 to clarify the intended applications of the Tweedie framework.

11. l. 113: Can you explain why the 10 minutes and the hourly time series are not consistent, are they measured with independent instruments?

115 The 10 min and 1 h series are obtained from the same precipitation gauges rather than independent instruments. However, they are supplied as separate processed series, and the 1 h product is not necessarily identical to hourly sums formed from the 10 min series. Differences in quality control or processing procedures may contribute to this discrepancy, which we are clarifying with the data provider. We will revise Sect. 2.2 to clarify the possible source of this difference accordingly.

120 *12. How is the block length L selected in Table 1? How are the results sensitive to different choices?*

125 The block lengths were selected to balance the number of accumulation windows available within each block against the number of blocks retained for the regression from the dataset, particularly after seasonal stratification and accounting for missing data. The choices were also guided by block lengths used in previous Tweedie precipitation studies and adjusted to the temporal resolution and data availability of the present datasets. Exploratory tests with nearby block lengths produced similar p estimates and did not alter the main qualitative patterns. We will clarify the rationale for the selected block lengths in Sect. 3.1.

130 *13. Figure 3: Explain in Methods what the Poisson and Gamma limit represents/how they are computed.*

The lines in Fig. 3 are theoretical reference values corresponding to the Poisson case at $p = 1$ and the Gamma case at $p = 2$, rather than values computed from the data. We will emphasize this in the Methods and Fig. 3 caption.

135 *14. Figure 4: the x-axis is logarithmic, are the Gamma and Tweedie distribution also using a logarithmic x-axis? The match between the histogram and the distributions is not very convincing...*

The histogram and fitted distributions are shown on the same logarithmic x-axis, which is intended to display the strongly skewed precipitation totals more clearly. As stated in the manuscript, the Gamma and Tweedie curves are shown as reference densities rather than as a direct goodness-of-fit comparison with the histogram. In particular, the histogram represents relative frequencies within bins, whereas the curves represent probability densities. Therefore, their heights and detailed distributional shapes are not directly comparable. We will emphasize this distinction further in the text and Fig. 4 caption.

145 *15. Figure 5: suggest to also add Gamma and Tweedie distribution fit for each season*

We appreciate the suggestion. Fig. 4 is intended to illustrate the Gamma and Tweedie reference densities, while Fig. 5 primarily compares the empirical DJF and JJA precipitation distributions directly. Adding multiple fitted curves to Fig. 5 would reduce the clarity of the seasonal comparison, so we will retain the current presentation.

150 *16. Figure 6: Radials artifacts (radio-frequency interference?) can be seen in the radar images...*

- *A comment about that should be added in the radar data description.*
- *The high values of these artifacts might alter the analysis of radar data; it would be best to exclude these radials when computing statistics.*

155 As mentioned in the Data section, we have already applied standard quality control to the radar composite, including capping, filtering and de-speckling. We have noticed the significance of radial artefacts during early inspection, and the processing therefore also includes a specific procedure to identify bright azimuths and reduce anomalous values using the surrounding azimuths, which substantially suppresses such artefacts.

160 Nevertheless, some radial structures remain visible in Fig. 6, because occasional anomalous values can affect the mean–variance relationship even after their amplitudes have been reduced. We acknowledge that these may locally influence the estimated p , as already noted in the Results and Limitations. Complete removal of radial artefacts cannot be guaranteed without also risking the removal or modification of valid meteorological echoes, and we therefore do not plan to introduce further processing or masking specifically for this analysis. Optimization and comparison of

alternative radar quality-control techniques would constitute a separate methodological study beyond the objective of the present study. We will describe the radial artefacts more clearly and clarify their potential impacts on the study.

17. *Figure 6: The range-dependence of the Tweedie- p parameters should be commented on.*

- *Are ground clutter artifacts affecting the precipitation values at lower range?*
- *Do you have under-detection of shallow precipitation (e.g. snow) at longer ranges?*
- *How the larger binning size at the longest range affects the results?*

Estonia is characterized by flat terrain and low surrounding structures, while the radars are situated on open and high towers, which limits ground clutter artefacts, although wind farms may remain. Such clutter would be expected to produce localized artefacts rather than explain the range dependence. Nevertheless, we agree that the range dependence should be discussed in more detail. At longer ranges, increasing beam height can lead to under-detection of shallow precipitation, particularly during winter in Estonia, while attenuation and increasing spatial averaging may also affect the retrieved precipitation statistics. These effects are currently described only briefly as beam-height effects and attenuation in the Limitations, and we agree that they should be explained more clearly.

18. *Figure 6: The sub-title over each sub-figure is too small. It should be removed.*

We agree and will remove the sub-titles in Fig. 6.

19. *Figure 7: How are the patterns robust to finite sampling? E.g. are the patterns similar from year to year?*

We agree that finite sampling may affect the estimated spatial patterns. However, the available radar record is relatively short and estimating p separately for individual years would substantially reduce the number of valid blocks, particularly at longer accumulation lengths and after seasonal subdivision. Uneven periods of missing radar data could introduce biases and further affect year-to-year comparisons, especially when certain seasons have fewer available observations, while annual differences would also reflect both sampling variability and underlying interannual variability in precipitation.

As an exploratory check, we extended the radar dataset by an additional year and did not find noticeable differences in the spatial patterns or p estimates. Similar behaviour was found when extending the gauge analysis using the longer available observational record. These tests provide some indication that the main patterns are not strongly dependent on the exact study period, although they do not constitute a formal assessment of interannual variability. A climatological assessment using the Tweedie framework is also beyond the scope of the present study. We will

emphasize the limitation associated with the short radar record and note that assessment of interannual variability would require a substantially longer observational record.

20. Figure 7: I don't really get what insight someone would get by looking at a map of Tweedie- p exponents. Figure 7 could be condensed by selecting only one accumulation length.

We appreciate the comment. The purpose of Fig. 7 is to examine whether the spatial distribution of p changes with season and temporal aggregation. The all-year fields in Fig. 6 are relatively smooth, whereas the seasonal maps in Fig. 7 reveal clearer spatial contrasts and how these patterns vary seasonally. Moreover, the comparison between 6 h and 24 h shows that these contrasts generally become more apparent at longer accumulation lengths. We therefore prefer to retain both accumulation lengths, since selecting only one would remove the scale comparison. We will revise Sect. 4.3.2 to make the intended insight from Fig. 7 clearer.

21. Table 2 and 3: Definition of Δp should be included in the caption. Actually, I would recommend to remove section 4.4 altogether and replace it with another application (see major revision 1).

The definition of Δp is already included in the captions of both Tables 2 and 3. We prefer to retain Sect. 4.4 and not replace it with another application, as discussed in our response to Major Comment 1, since the station-wise radar–gauge comparison directly quantifies how p varies with observing systems. We will further strengthen this comparison by adding statistical tests, as suggested by Referee #1.

22. Figure A1: I suggest to also add Tweedie distribution fit for each instrument

We agree that the Tweedie fits would provide useful additional context for radar–gauge comparison here. We will add the corresponding fitted Tweedie distributions with the histogram frequencies.