Review of doi.org/10.5194/amt-2024-2
"Double moment normalization of hail size number distributions over Switzerland"
by Alfonso Ferrone, Jérôme Kopp, Martin Lainer, Marco Gabella, Urs Germann, and Alexis Berne
Submitted to *Atmospheric Measurement Techniques*, April 2024.

Summary

In this work the authors apply a recently introduced –but for rain– method to describe the distribution of hailstone diameters, observed by the Swiss network of 80 automatic hail sensors, based on distribution "moments". In practice, for each of the 95 hail events studied (each one with a different time duration), they can compute the "p-moment", M_p, as the scalar quantity defined by:

$$M_p = \int_{5\,mm}^{D_{max}} D^p\, N(D)\, \mathrm{d}D \tag{1}$$

where $N(D)$ is the probability density function describing the distribution of hailstone diameters, D, of that individual event and D_{max} is the maximum hailstone diameter observed by that specific hail event ($D_{max} \leq 20\,\mathrm{mm}$, a part for very few cases). Following what done in literature for rain, the authors rescale the hailstone size distribution introducing the "normalized diameter":

$$x = \left(\frac{M_i}{M_j}\right)^{\frac{1}{j-i}} \cdot D \tag{2}$$

and obtaining the "normalized distribution":

$$h(x) = \left(\frac{M_j^{i+1}}{M_i^{j+1}}\right)^{\frac{1}{j-i}} \cdot N(x(D)) \tag{3}$$

All the event distributions are converted in their normalized counterpart because $h(x)$ vs x should "collapse" in a smaller area than $N(D)$ vs D (even if it is not explained *why* that should happen). Thanks to this property, the authors are then able to fit a single gamma distribution $\hat{h}(x)$ on the space $h(x)$ vs x to fit all the individual distributions with a single function. From this single function, they can derive the estimated $\hat{N}(D)$ distribution for each event and compare it with the observed one.

To verify their approach the authors divide all the 95 events in a training set (70% of cases used to fit $\hat{h}(x)$) and a test set (30%), using the distributions of BIAS, RMSE and R (Pearson correlation coefficient) as a measure of performance. Moreover, they use the same normalized distribution $\hat{h}(x)$ built from automatic hail sensors to fit a single hail event observed by drone on a large area. The last is characterized by a much higher number of hailstones and also larger diameters. The result is quite good, in particular as shape of the distribution, even if it tends underestimate $N(D)$. Lastly, since the hail sensor network is divided in three different subareas, the authors tried to compare the performance of one area relatively to the other two, finding that Ticino seems to be more different from Jura and Napf area.

In conclusion, I found this work very interesting and promising, but I have a long list of suggestions and hence kindly ask for a relatively long revision, as suggested below.

Major comments

- One of the most annoying part of the paper is the somewhat excessive/not-friendly notation used. The authors introduce HSD, HSND, $N(D)$, $N_u(D)$ probably to indicate only two different concepts. The term "normalized" is probably used with different meaning: they say that HSD (hail size distribution, same as $N(D)$?) is *"usually computed over a unit area and for a fixed duration of time"*, hence one can think that it is "normalized" with respect to time and area, while the distribution studied here, HSND (hail size number distribution, same as $N_u(D)$?), is "unnormalized" because *"it is dependent on the detection area of the instrument and the duration of the event."*. Here there is already a problem, since all the hail sensors have the same area and hence only the single event observed by drone has a different detection area. But, a part from that, then the main topic of the work is to transform $N_u(D)$ into the normalized distribution $h(x)$ (eq. 3). In which sense reshaping $N_u(D)$ with equation 3 (where $N(x(D))$ is simply multiplied by the scalar value $\left(\frac{M_j^{i+1}}{M_i^{j+1}}\right)^{\frac{1}{j-i}}$) "normalize" the distribution? Is $h(x)$ normalized by area and time in some way?

- In the manuscript there is a lot of math, while I'm missing simple equation that can make the concepts easier to be understood by the readers. For example, is it correct that the total number of hailstones, N_i, of the i-hail event is given by:

$$N_i = \int_{5mm}^{D_{max}} N_i(D)\,\mathrm{d}D\ ? \tag{4}$$

  If so, why such simple equation is never shown? BTW, if that is correct, it seems to me that $N_i(D)$ is more a probability *density* function that a number of hailstones, as written in the paper. Please clarify if $N_u(D)$ is a number distribution or better a density distribution, that gives "number" only when multiplied by dD (as written in line 52).
  Moreover, at the end the paper focus only on moments of order 2 and 4, thus all the complex equations above reduces (if I'm correct) to:

$$x = \sqrt{\frac{M_2}{M_4}} \cdot D \tag{5}$$

  and

$$h(x) = \sqrt{\frac{M_4^3}{M_2^5}} \cdot N(x(D)) \tag{6}$$

If these are the final equations used in this work, why they are never shown? Moreover, why the values of M_2 and M_4 are never discussed? For example, comparing Fig. 2 with Fig. 4 it seems that $x \cong D/20$. Please, could you show the distribution of $\sqrt{\frac{M_2}{M_4}}$ for your hail events? For example, I would be very interested in seeing a scatterplot of $\sqrt{\frac{M_2}{M_4}}$ versus D_{max} of each event, to see if there is any systematic dependence of the double-moment normalization from the maximum diameter observed the original distribution. The same can be done for $\sqrt{\frac{M_4^3}{M_2^5}}$.

- Speaking about the verification of the spatial invariance, I understand that Section 5.4 should clarify well the point, but it is not very clear how the events used from one subarea are compared with the events in the other two regions: is that done fitting all the events of one area to build a new $\hat{h}(x)$, which is then used to reconstruct $\hat{N}(D)$ in the other two regions? Or only the training subsample in one region?
  A part from that and considering also the fact that Ticino seems to have different characteristics from the other two regions, I think that a more robust approach would be to redo all the main work using as training set all the events from two regions and as test set all the events from the third region (e.g. Ticino). In fact, using sensor locations from all three regions in the training set (chosen randomly) will give less robustness to the spatial invariance result, in my opinion.

- The verification if done using 4 different metrics, but since there is a value for each event, at the end only the verification metric distributions are shown. It may happen that these metrics are not independent each other and that an event having, for example, a better BIAS could have a worse correlation, while in other cases these metrics are more independent (see Section 5.3.1. of Manzato et al. 2016). There is a well known tool to show many of the most used regression verification metrics together in a single diagram, that is, the Taylor diagram (Taylor 2001, 2005). I invite the authors to consider to plot all the event verifications in a single Taylor diagram, instead of using four different diagrams showing distributions. If there are too many points, they could use a sort of "density plot" in the Taylor diagram space. Of course, mine is just a suggestion.

Minor comments

41-42 *have a surface of similar scale.*
Maybe similar order of magnitude?

44 *northern Italy*
northeastern Italy

55 equation $N(D)$
Please add the equation number and $[units]$.

68 *and period of time for which it is used.*
Here is a good place to explain something more on why the double moment normalization should work, i.e. why it should collapse the single event distriutions in a closer space.

91 *the quantity that we model is not normalized over a unit surface*
OK, but all sensors have the same dection area, so it is a constant. Ar you sure that you can't simplify the notation $N_u(D)$?

136-139 *Therefore,... HSNDs.*
Please, could you clarify better?

141-143 *Each of them has been defined as a period in which hailstones are recorded with a gap between consecutive impacts of less than 20 minutes, corresponding to the largest blank period studied in Kopp et al. (2023a).*
20 min seems to me a little too long period, in particular reading the recommendation of Kopp et al. (2023a), which states *"we suggest using a 10 to 15 min Tmb in further studies"* Do you have a specific justification to take the longest blank period studied?

157 *94 events*
In the Conclusions you write 95 (line 526). Please double check.

161 *minimum values,..., reaching 9 additional impacts*
Please explain better if you speak of hail events or hailstone impacts per single event or what else.

165 *the scarcity of measurements for the largest diameters (e.g. above 20 mm),*
Please write somewhere how many cases do you have above 20 mm and which is the maximum diameter used in this study. I suspect that you are missing the tail of large hailstones because of a too short database and that could undermine the general value of your $\hat{h}(x)$.

Equation 3
Written in this way, $h(x)$ does not seem a function of x.

Equation 4
At the end of the work, could you write an analytical version of eq 4 with the values that you have found, so that if some reader would like to apply to his dataset your findings it will be easy to be done? In general, it would be nice if this work will give some more "practical" results and a little less math.

Section 3.2
Consider to use the Taylor diagram if possible.

267 *PCC*
Please replace here and elsewhere PCC with R.

274-275 *the randomness of the impacts on the relatively small surface of the hail sensor*
Here you can quote Grieser and Hill (2019).

275 *has the potential to be noticeable in the upper tail*
Please reformulate.

317-319 *The main benefit of computing a normalized distribution is the reduced spread of the h(x) values at each x when compared to the considerable variability of Nu(D) at each D. This effect, sometimes referred to as the "collapse" of the normalized distributions, is what allows us to fit $\hat{h}(x)$, as described in section 3.1.*
This part can be moved and further elaborated in the Introduction.

Section 5.2
It would be possible to consider a Pareto distribution?

394-395 *The Pearson correlation coefficient is the only metric in which the fitted $\hat{h}(x)$ has better performances for the drone event than for most of the hail sensor ones.*
Any idea why? Is it possible because the drone hail event has a broader hail size distribution?

Section 5.4
Not clear which data exactly are used to train on one region and which to test on the other two.

458 *represents a middle ground*
Please reformulate.

474 *is linked with Nu(D) being an integer number*
Not yet clear if $N_u(D)$ is a number of hailstones or a density that must be multiplied by dD to get an integer number.

572-574 *By finding a link between the empirical moments of the HSND and the radar measurements, it would be possible to use the formula and parameters of $\hat{h}(x)$ defined in this study to estimate the full distribution of hail diameters expected at the ground.*
That sounds very interesting. Could you elaborate further?

Appendix
It seems to me that there are not much differences (or using the words of the authors: *"Overall, the comparison between the various combinations of moments does not show any of the pairs clearly outperforming the others."*) thus the choice of moments 2 and 4 will never be very convincing. That is very evident already in Fig. A1, which I suggest to include in the main text.

Figures:
Plase make them larger. A page like that showing Fig. 2 is not acceptable.

Figure 3
The caption seems really too long.

Fig. 6 caption *The comparison is always performed between pairs of events,*
Not clear how the pair are chosen.

Fig. 9 colorbar
Please use more different colors, not only two.

Suggested References

I believe that this work could benefit considering at least part of the following references:

- Grieser, J. and Hill, M. (2019) How to express hail intensity modeling the hailstone size distribution. Journal of Applied Meteorology and Climatology, 58, 2329-2345. https://doi.org/10.1175/JAMC D-18-0334.1.
- Manzato, A., Cicogna, A. and A. Pucillo. 2016. 6-hour maximum rain in Friuli Venezia Giulia: Climatology and ECMWF-based forecasts, Atmospheric Research, 169B, 465-484.
- Taylor, K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106: 7183-7192. Bibcode:2001JGR...106.7183T. doi:10.1029/2000JD900719
- Taylor, K.E. Taylor diagram primer (2005), PDF available at https://pcmdi.llnl.gov/staff/taylor/CV/Taylor_diagram_primer.pdf.

Best Regards.

Tino