240 VERIFICATION OF THE MESH PRODUCT OVER THE CANADIAN PRAIRIES USING A HIGH-QUALITY SURFACE HAIL REPORT DATABASE SOURCED FROM SOCIAL MEDIA

Julian C. Brimelow[1] and Neil M. Taylor[1]

[1]National Lab-West, Environment and Climate Change Canada (ECCC), Edmonton, AB

1. INTRODUCTION

Timely and accurate severe weather reports support the issuance of warnings, warning verification, and research and development. Obtaining accurate severe reports for these purposes is, however, plagued by many uncertainties. Moreover, Blair et al. (2011) found that ~25% of hail reports in the U.S. *Storm Data* archives were not associated with strong radar echoes at the time of the report

Of all precipitation types, reports of hail are probably prone to the largest uncertainties and inaccuracies. The literature is replete with examples of uncertainties related to reporting hail (e.g., Witt et al. 1998; Schaefer et al. 2004; Blair and Leighton 2012; Allen and Tippet 2015). These uncertainties can be traced to several factors related to the physical characteristics of hailswaths (e.g., high spatiotemporal variability) and hail (e.g., irregular shapes). Additional sources of uncertainty arising from human-related errors are:

- Timing and location of reports.
- Whether the reports reflect the largest hail at a given location.
- Measuring the maximum hail size.
- Confusion surrounding reporting procedures—what axis dimension to report.

Other confounding factors that affect hail reports are:

- The hail size is often a guess-estimate or related to an object of unknown size.
- The size is estimated after stones have experienced substantial melting.
- Reports can be biased to severe thresholds and/or familiar objects,
- Population bias in remote areas.

* *Corresponding author address:* Julian C. Brimelow, National Lab-West, Environment and Climate Change Canada. 9250-49th Street Edmonton, AB, T6B 1K5; Email: julian.brimelow@canada.ca

Consequently, obtaining reliable hail reports, particularly for sparsely populated areas such as the Canadian prairies, is especially problematic.

In an attempt to address some of the above problems, three projects in the United States (US) over the past decade have conducted storm-scale sampling of hailstorms:

- SHAVE (Severe Hazards Analysis and Verification Experiment; Ortega et al. 2009)
- IBHS (Insurance Institute for Business and Home Safety, Heymsfield et al. 2014; Giammanco et al. 2015,)
- HailSTONE (Hail Spatial and Temporal Observing Network Effort; Blair et al. 2017)

Findings from HailSTONE highlight serious problems with hail reports in the U.S. *Storm Data*—problems that almost certainly apply to hail data in Canada. For example, Blair et al. found that the extent and magnitude of hail reported by dedicated chase teams was frequently much larger than that suggested by information in *Storm Data*. Moreover, their research found that more than 30% of the hail storms sampled during HailSTONE did not have an associated hail report in *Storm Data*. Blair et al.'s findings suggest that addressing report biases and uncertainties requires detailed observations at the storm scale. This is not feasible, nor practical in everyday operations (e.g., Hyvären and Saltikoff 2010).

The above limitations of human-sourced hail reports paint a rather dire picture. Could alternative data sources help mitigate the myriad of issues facing traditional hail reports? The answer to this is a cautious "Yes". For example, radar-derived products such as the Maximum Expected Size of Hail (MESH; Witt et al. 1998) could be used as a proxy for human reports, and also have the advantage that the location, timing and magnitude of hail are not subject to human-caused uncertainties and errors, or population biases. Moreover, radar-derived data are also available in near-real time, whereas there is oftentimes a delay before forecasters are alerted to a hail event via human reports. MESH has successfully been used to study the size and spatiotemporal distribution of hail in Europe (e.g., Betschart and Hering 2012; Nisi et al. 2016), Australia (e.g., Warren et al. 2017), and the US (e.g., Ortega et al. 2009; Cintineo et al. 2012). But before using MESH for this purpose, one needs to first verify its suitability and accuracy in Canada. That in turn requires addressing the inaccuracy and paucity of publicly sourced hail reports.

Over the past 10 years or so, social media has become an increasingly important source of information for precipitation type and severe weather reports (e.g., Ferree et al. 2009; Hyvären and Saltikoff 2010; Chen et al. 2016; Snook et al. 2016). A number of Apps have also been developed to facilitate the reporting of severe weather—for example, mPING (Elmore et al. 2014) and a photo report system developed by Longmore et al. (2015). Following the success of these initiatives, we set out to determine the utility of using reports from social media of hail on the data-sparse Canadian prairies.

Here we seek to answer the following questions:

1) *Can leveraging the crowd sourcing capability of social media be used to collect images of hail and obtain objective and accurate hail size measurements?*
2) *Can MESH be used as a reliable surrogate/proxy for hail reports?*

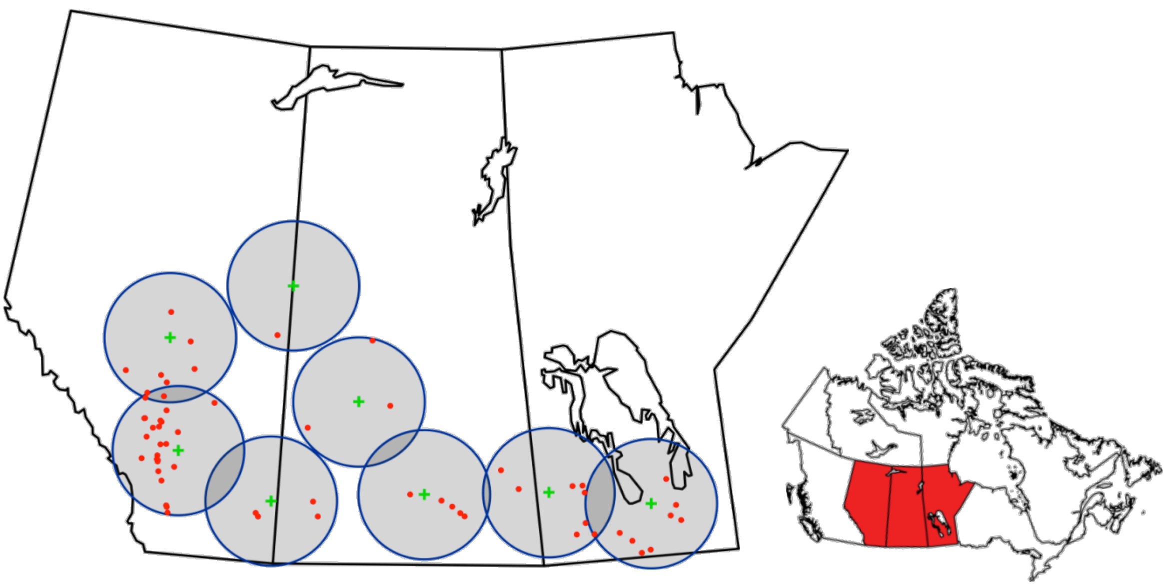

Fig. 1: Study area over the Canadian Prairie Provinces. Distribution of reports (red dots) used in analysis. Radars (green crosses) and 150 km range rings (blue and shaded) also shown.

2. DATA AND ANALYSIS TECHNIQUES

The study area we considered is the Canadian Prairies (see Fig. 1). This region is known for its relatively high incidence of severe summer weather (including hail), and with the exception of a few large urban centres, has a very low population density.

2.1 *Social Media Reports*

Starting around 2012, three hashtags were introduced that users of Twitter could use to report high-impact weather. Here we focus on hashtags used to report severe weather in the provinces of Alberta (#abstorm), Saskatchewan (#skstorm) and Manitoba (#mbstorm).

The aforementioned Twitter feeds were scanned for the summers of 2014 through 2016 for Tweets that included images of hail, or links to images of hail elsewhere (e.g., Facebook and Instagram). Almost 1700 images were acquired in this way.

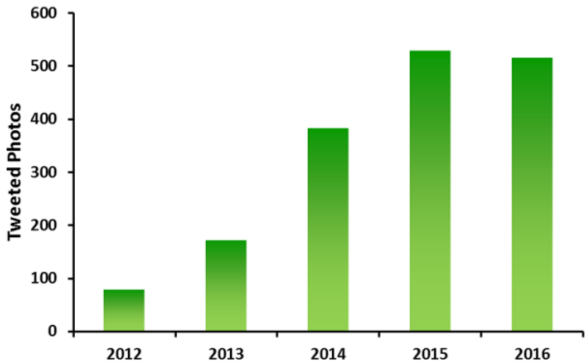

Fig. 2: Number of Tweeted hail photos by year.

The storm hashtags have become increasingly popular since they were first introduced (Fig. 2), with the majority of images collected in our database increasing each year. Most of the photos were from Alberta (Fig. 3).

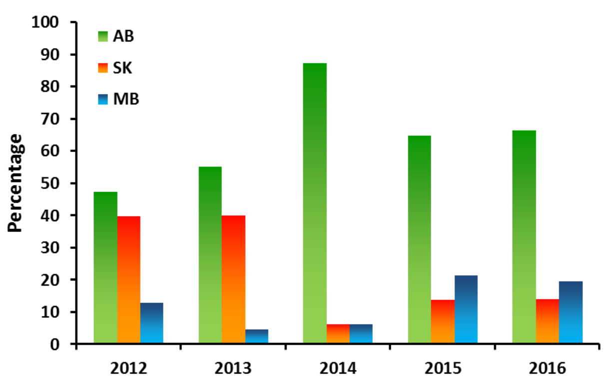

Fig. 3: Percentage of photos by year for Alberta (AB), Saskatchewan (SK) and Manitoba (MB).

To facilitate analysis of the images, they were classified into four groups (see Fig. 4):

1) Photos of damage only.
2) Photos of hail scenes (hail too far away to accurately infer size).
3) Photos of hail that did not include a scalable object.
4) Photos of hail that included a scalable object.

Fig. 4: Four classes into which photographs of hail posted on social media were binned.

It is clear from Fig. 5 that photos with no scalable objects or photos of hail scenes made up the majority (~73%) of the photos. Photos of damage were in the minority (~4%). Only about 22% of the photos included a scalable object next to the hail. By far the biggest issue with hail photos is that people tend to hold the hail in their open hand, which precludes accurately estimating the hail size. Consequently, over 80% of the 1700 photos were of little or no use for our research purposes.

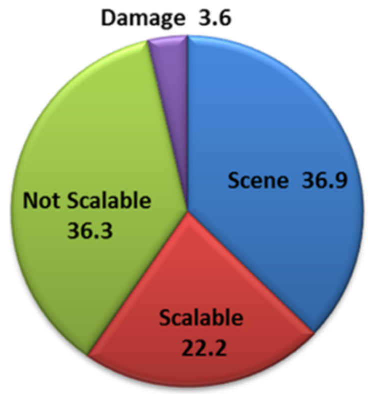

Fig. 5: Breakdown (in %) of hail photo classes.

Photos were also classified as to their likely location:

- Events observed at home or summer cabin.
- Events away from home (e.g., at work, shopping, chasing, camping).
- Events where the location was unclear.

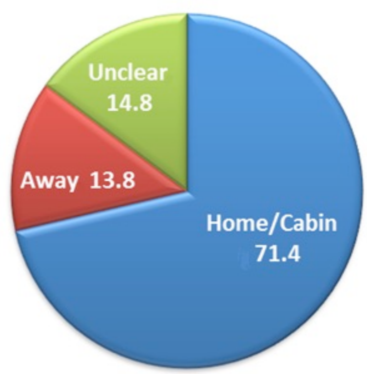

Fig. 6: Location where photos appeared to have been taken in 2016.

Data shown in Fig. 6 indicate that most photos in 2016, for example, were taken at peoples homes or cabins, with an even split (~14% each) for photos taken away from home or for which the location is unclear/uncertain. Implications of this finding are discussed in Section 4.

The dimensions of the primary (D_x) and secondary axes (D_s) were calculated by scaling the dimensions in the photo using an object of known diameter included in the photo (See Fig. 7). These data were then used to calculate the aspect ratio (α) and finally the equivalent spherical diameter (D_e) using the following formula:

$$D_e = \alpha^{0.333} D_x \qquad (1)$$

The D_e of the largest stone in each photo was recorded. In reality, most stones are tri-axial ellipsoids, with a mean axis ratio of D_m (minimum axis) to D_x of ~0.77 (Matson and Huggins 1980).

Fig. 7: Image showing how each hailstone was measured using a scalable object. Emphasis was placed on measuring the largest hailstones.

Using the observed axis ratio from the photos to calculate D_e yields equivalent diameters within a few mm of the diameter estimated assuming the stones are spheroids. Additional sensitivity tests for the expected errors in measuring the scaling objects and in fitting an ellipsoid around the stones typically produced uncertainty in D_e of 1 mm to 3 mm (i.e., within ~10% of D_e). Photos of hailstones having highly irregular shapes, such as shown in Fig. 8, were not included in our dataset because of the large uncertainty in estimating a value of D_e.

2.2 *Radar Data and Analysis*

Eight C-band radars provide coverage for most of the agricultural zone and urban centres in the study area, although there are significant gaps in radar coverage as evident in Fig. 1.

For each convective cell identified and tracked by the Universal Radar Processor (URP) software, numerous key radar-derived metrics are archived for the purpose of quantifying the severity of the cell. Specifically, we are concerned with hail-size estimates at the surface determined using empirically derived algorithms of Treloar (1998) for Australian hailstorms (later implemented by Joe et al. 2004), and the MESH algorithm of Witt et al. (1998).

Fig. 8: Photo of hail with an irregular shape because of surface lobes.

Here we compare hail-size estimates from these two hail algorithms with hail-size estimates calculated from photographs as described in Section 2.1. For the radar data, we used the largest hail-size estimate from the two algorithms for each cell—that is, the pixel having the largest estimated diameter. We recognize that a method utilizing a cell-representative MESH value may be more appropriate, but here adopted the existing methodology used to produce post-processed products in URP.

To avoid data-sampling issues associated with the cone of silence around the radars and at large distances from the radars, we considered only those events (and cells) located between 30 km and 150 km from the radars. Because we were using data from an experimental server, there were sometimes dropout issues with the archived data that reduced the number of possible events.

Fig. 9 below summarizes the methodology adopted for this study. We used a search radius of 10 km around each report location, which is similar to the radius used by Warren at al. (2017). Reasons for using 10 km are as follows:

- Canadian radar products are updated every 10 minutes, which introduces uncertainty in the storms' locations relative to the report, especially when cells are moving rapidly.

- To account for hydrometeor drift. This is especially relevant for the MESH and Teloar algorithms because they relate data above the freezing level to the expected hail size at the surface.
- To allow for a margin of error with the report location (we did not have an exact latitude and longitude for each photo, rather the location was estimated using information provided about road intersections, neighbourhood names, or distances from the nearest town, for example). Unless provided, the longitude and latitude for each event were determined using Google Earth.

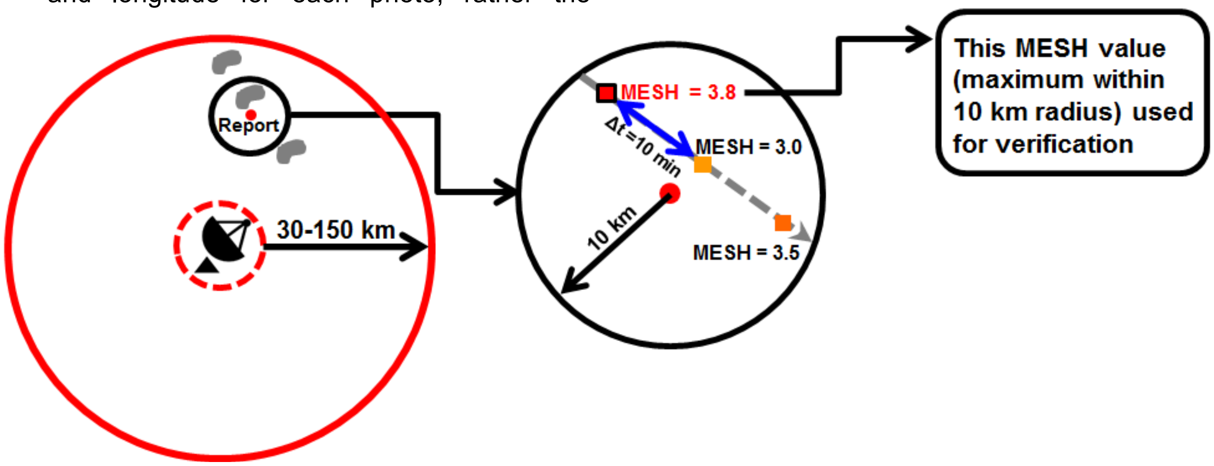

Fig. 9: Schematic showing method used to identify the relevant MESH value for each hail event in the database

Both radar and hail events were subjected to rigorous quality control (QC). For hail, only reports which met the following criteria were included:

- Reports for which we were highly confident that the size was representative of the largest hail within 10 km.
- Reports for which we had high confidence in the location.
- Reports for which we had high confidence in the scaled D_e, and for which the hail had experienced minimal melting.

Archived storm cell data were also subjected to rigorous QC. We excluded hail events if the storm associated with the report was attenuated or affected by beam blocking. Sometimes individual scans indicated MESH values that were much higher (one standard deviation) than preceding and subsequent scans. In such instances, the next largest MESH value within 10 km was selected.

Despite the challenges associated with using photos submitted on social media, the subset of viable images (374) is still relatively large due to the large numbers of photos submitted using this crowd sourcing mechanism.

Ultimately, we identified 65 "golden" cases which could be used to verify the hail-size estimates from the MESH and Treloar algorithms. The locations of the reports are shown in Fig. 1, and the size distribution of the hail sizes used in our analysis is shown in Fig. 10.

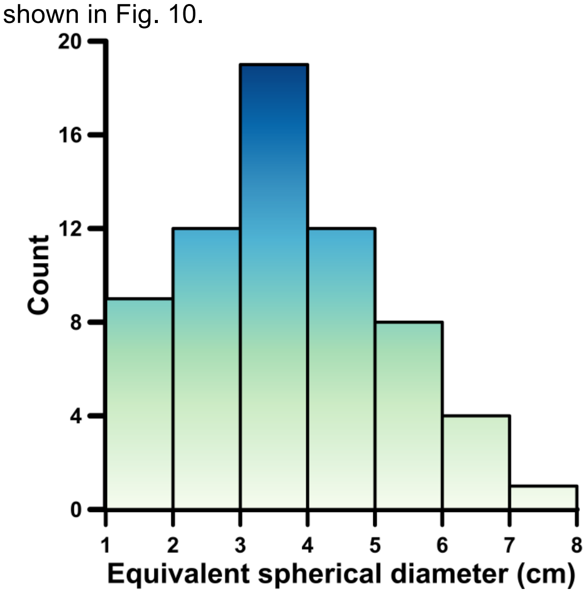

Fig. 10: Histogram of measured equivalent spherical diameters for the 65 hail cases used in our analysis.

The mean D_e of the 65 reports was 3.8 cm, with D_e ranging between 1.1 cm and 7.7 cm. The mean aspect ratio of the stones was 0.88. Fig. 10 suggests that events for which D_e was < 3 cm were underrepresented in our dataset (see also section 3.1).

3. RESULTS

3.1 Hailstone Characteristics

Here we discuss the size distribution of all 600 stones for which we calculated D_e using photos from social media. Importantly, Fig. 11a,b shows that using photos from social media instead of public reports avoids the minimum in hail sizes between 3 cm and 4 cm and also reduces the spike in occurrence of golfball-sized stones.

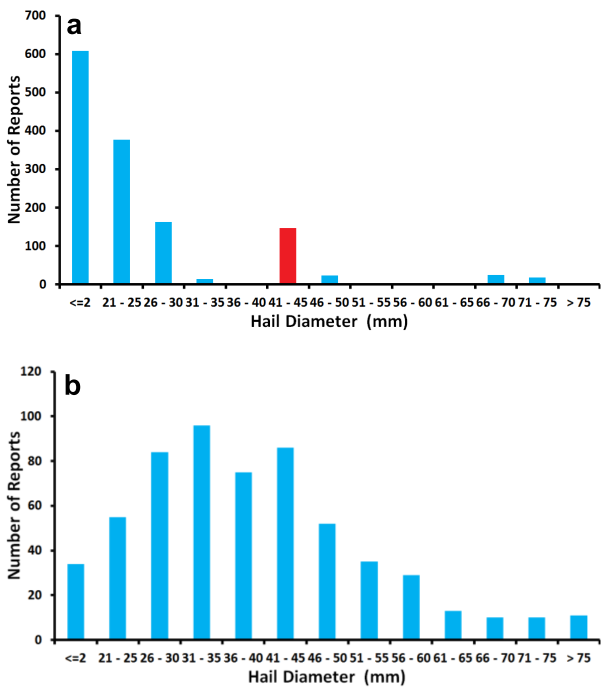

Fig. 11: Histograms of hail sizes from (a) public reports for 2014-2016 and (b) from social media (all measured stones). Red bar indicates spike in reports of golfball-sized hail in public reports.

Problems with the hail-size distribution of public hail reports has been noted elsewhere (e.g., Schaefer et al. 2004; Jewell and Brimelow 2009). One limitation of our database is that smaller hail events (i.e., D_e < 3 cm) are underrepresented. Inspecting the collection of photographs we obtained from social media reveals a possible explanation for this. Specifically, it appears that people are much less likely to measure small hail, or place an object next to small hail when taking photos to post on social media.
The large database of ~1700 photos permitted us to identify the most frequent issues that were encountered when selecting suitable photographs of hail for our analysis:

1) No location provided or only a vague reference to location provided.
2) Tweet posted tens of minutes to hours after the storm.
3) Uncertainty in event time (disagreement between time stamp on Tweet and time provided in Tweet).
4) Using objects that do not have a fixed size or indeterminable size.
5) Photos of hail held in someone's hand, with no scalable object.
6) Photos taken after hail experienced significant melting.
7) Photos taken at oblique angles.
8) Photos showing the "wrong" side of a coin (i.e., heads).
9) Photos of the scalable object(s) and hailstones lying in different planes.
10) Combinations of the above.

The above list indicates where we should be focussing our efforts when educating the public on how to report hail in general, but also when educating them on using social media to post useful photographs of hail.

3.2 Verification of MESH

A scatterplot of MESH against D_e (Fig .12a) shows that the MESH algorithm does a fairly good job of capturing the variability in hail size in our dataset. Overall the correlation coefficient is 0.70. In contrast, Fig. 12b shows that the Treloar algorithm has a much lower correlation (r = 0.54) with D_e, as evidenced by the larger scatter of the data.

The scatterplot in Fig. 12a indicates that there is a tendency for MESH to overestimate smaller hail, whilst underestimating giant (D > 5 cm) hail events. In comparison, the Treloar algorithm (Fig. 12b) tends to overestimate D_e across the entire size spectrum, but especially for hail reports with D < 4 cm.

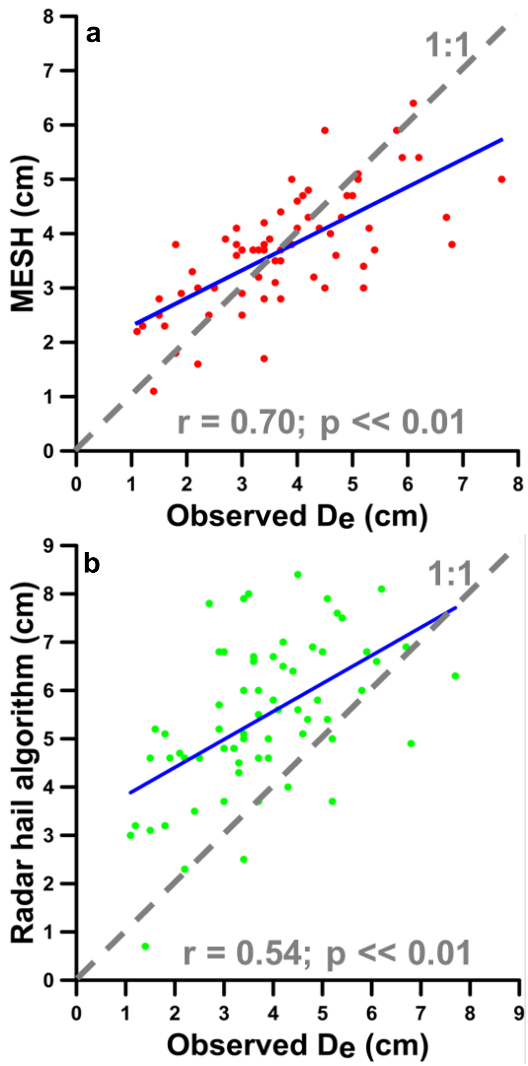

Fig. 12: Scatter plots and linear best fits for (a) MESH and (b) existing hail algorithm (Treloar 1998). Dashed grey line represents 1:1 line. "r" represents the correlation coefficient and "p" the p-value.

The above comments are reflected in whisker-box plots showing the distribution of errors for all observed sizes and by observed size category (Fig. 13 and Fig 14a). For all sizes, the interquartile range for MESH errors is between -0.5 cm and 0.7 cm, with a median error of zero (i.e., no bias). Also, almost all the MESH errors are within +/- 2 cm. The mean absolute error (MAE) for MESH is 0.79 cm (21% of the mean observed diameter). In contrast, the Treloar algorithm displays a distinct positive bias, with an interquartile range between 0.7 cm and 2.6 cm, and a median error of 1.8 cm (MAE of 1.67 cm). Moreover, almost 50% of the Treloar algorithm errors exceed 2 cm.

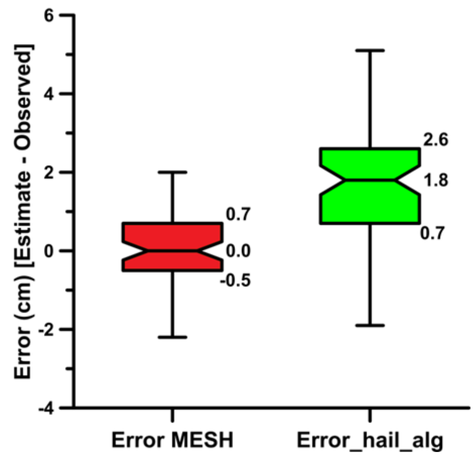

Fig. 13: Errors for the MESH (red) and the existing Treloar (green) hail algorithms in URP. Whiskers represent the inter-quartile range multiplied by 1.5.

The tendency for MESH to overestimate D_e for smaller hail events and overestimate D_e for larger hail events is indicated in Fig. 14a. Specifically, the MESH error for observed hail less than 3 cm in diameter typically ranges between 0.1 cm and 1.2 cm, with a median of 0.8 cm.

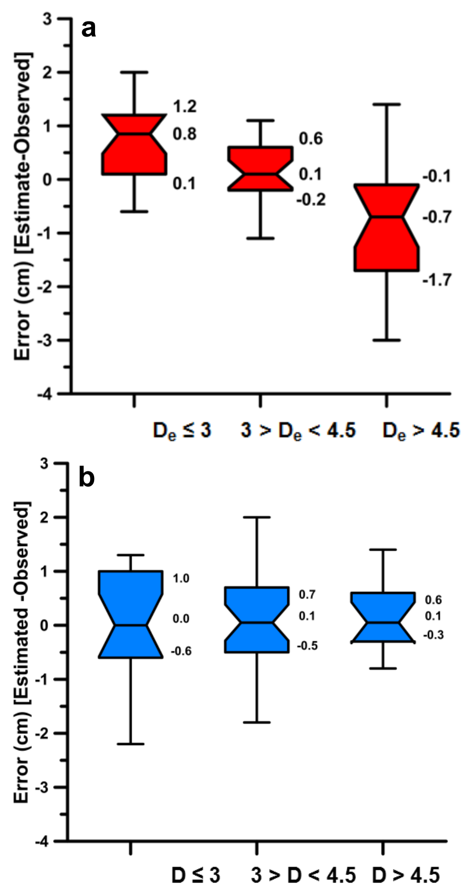

Fig. 14: Errors ranked by observed diameter (D_e) in (a) and by MESH diameter (D) in (b).

In contrast, MESH errors for larger than golfball hail range between -1.7 cm and -0.1 cm, with a median error of -0.7 cm. For observed hail between 3.0 and 4.5 cm, MESH displays a small bias of near 0.1 cm, indicating only a slight tendency to overestimate the hail size over this size range.

The errors shown in Fig. 14a are, however, not necessarily of much use operationally because forecasters do not know *a priori* what the observed hail size is going to be. Consequently, it is more instructive to look at the errors as a function of the estimated MESH hail size (D; Fig. 14b).

Partitioning the errors according to the predicted hail sizes yields a very different picture. The data in Fig. 14b indicate that when the MESH errors are classified in this way, the nature of the errors is such that applying a bias correction to the MESH data is not feasible. Instead, it is probably more beneficial to determine a MESH threshold for severe hail ($D_e \geq 2$ cm) as has been done by Cintineo et al. (2012) and Warren et al. (2017), who determined thresholds of 2.9 cm and 3.5 cm, respectively.

4. DISCUSSION AND CONCLUSIONS

Obtaining timely and accurate reports of hail continues to be a challenge. Herein we investigated the potential to use the radar-derived MESH product as a surrogate for human-sourced hail reports, and whether social media could be used to collect accurate measurements of hail for three summers (2014–2016) on the Canadian Prairies.

We find that while social media has the potential to be a useful source of hail data to supplement traditional reports, many issues need to be overcome in order to realize its full potential. Issues with how hail is currently reported using Twitter mean that ~80% of the posted photographs are of little or no value for quantitative hail-size verification.

To maximize the value of social media reports, weather agencies need to be proactive and reach out to educate social media users on best practices for posting reports. The most frequent problems with social media reports could be avoided, or mitigated, by encouraging users to follow some key guidelines:

1) Include a scalable object of known size when taking photos of hail.
2) Take photos looking vertically down on the hail and scalable object (see Fig. 7).
3) Photograph hail as soon as it is safe, and search for a sample of the largest stones.
4) Include the time and location in the message accompanying the photo.

We found that most (~70%) of the photos were likely sent from people's homes or summer cabins. This suggests that an alternative method for reporting hail size is possible. Namely, weighing the hailstone using a kitchen scale after photographing it with a scalable object and then including the weight in the message. This method avoids many of the uncertainties and measurement errors associated with traditional reporting methods. Additionally, it would be relatively easy to implement this approach, because many CoCoRaHS (Cifelli et al. 2005) volunteers already have a scale with which to weigh snow for determining the snow water equivalent.

Despite the aforementioned limitations of using photographs posted on social media, the large number of photographs collected for this study (~1700) ensured that even after applying rigorous quality control and limits, we still had a relatively large sample of 65 “golden” cases that could be used to verify the MESH product. Hailstones in our analysis spanned a wide range between 1.1 cm and 7.7 cm, with a median diameter of 3.8 cm.

Our verification metrics indicate that the Treloar (1998) hail-size algorithm shows little skill in estimating hail size and displays a large positive bias, especially for $D_e < 4$ cm. In contrast, we find that the MESH algorithm holds promise as a proxy hail-size product because of its small overall bias and mean absolute error (~0.8 cm). We find that the MESH errors change with the observed size, overestimating hail reports with $D_e < 3$ cm, while underestimating reports with $D_e > 4.5$ cm. Our findings for MESH are also broadly consistent with those of Betschart and Hering (2012). For example, the same bias behaviour displayed by MESH in our analysis was also noted in their work.

While the data suggest that it may be possible to correct the bias in the MESH data across all observed sizes, partitioning the errors according to the predicted MESH sizes indicates that the nature of the errors is such that applying a bias correction to the MESH size estimates is not feasible.

Of note is that our finding that MESH could be used as a surrogate for hail size is in conflict with findings made by Ortega et al. (2009) using data from SHAVE. This discrepancy requires closer examination, but we can think of two possible reasons:

1) We used a relatively large search radius of 10 km to allow for uncertainties in report location and hydrometeor drift. In contrast, Ortega et al. (2009) used a search radius of only 3 km.

2) We calculated the equivalent spherical diameter of the observed hail size from photos that included a scalable object. Ortega et al. (2009) based their hail sizes on reports solicited from phone calls, and the reported sizes were not necessarily obtained from directly measuring the hail.

Future work will focus on adding non-severe and null cases to our dataset with the intention of identifying MESH thresholds for hail and severe hail. Our dataset has other applications such as examining storm environments associated with different hail-size categories, verification of other hail forecast products and for investigation of lightning-hail relationships.

ACKNOWLEDGEMENTS

The authors would like to thank Jason Knight for providing the 2016 radar data, and summer students and forecasters who drafted the daily chronologies.

REFERENCES

Allen, J.T., and M.K. Tippett, 2015: The characteristics of United States hail reports: 1955–2014. *Electron. J. Severe Storms Meteorol.*, **10**, 1–31.

Betschart, M., and A. Hering, 2012: Automatic hail detection at MeteoSwiss. Arbeitsberichte der MeteoSchweitz 238: 61. (https://issuu.com/mariobetschart/docs/ab238).

Blair, S.F., D.R. Deroche, J.M. Boustead, J.W. Leighton, B.L. Barjenbruch, and W.P. Gargan, 2011: A radar-based assessment of the detectability of giant hail. *Electron. J. Severe Storms Meteorol.*, **6**, 1–30.

Blair, S.F., and J.W. Leighton, 2012: Creating high-resolution hail datasets using social media and post-storm ground surveys. *Electron. J. Severe Storms Meteorol.*, **13**, 32–45.

Blair, S.F. and co-authors, 2017: High-resolution hail observations: Implications for NWS warning operations. *Wea. Forecasting*, **32**, 1101–1119.

Chen, S. J.J. Gourley, Y. Hong, Q. Cao, N. Carr, P-E. Kirstetter, J. Zhang, and Z. Flamig, 2016: Using citizen science reports to evaluate estimates of surface precipitation type. *Bull. Amer. Meteor. Soc.*, **97**, 187–193.

Cifelli, R. N. Doeksen, P. Kennedy, L.D. Carey, S. Rutledge, C. Gimmestad, and T. Depue, 2005: The community collaborative rain, hail, and snow network. *Bull. Amer. Meteor. Soc.*, **86**, 1069–1077.

Cintineo, J.L., T.M. Smith, V. Lakshmanan, H.E. Brooks, and K.L. Ortega, 2012: An objective high resolution hail climatology of the contiguous United States. *Wea. Forecasting*, **27**, 1235–1248.

Elmore, K.L., Z.L. Flamig, V. Lakshmanan, B.T. Kaney, V. Farmer, H.D. Reeves, and L.P. Rothfusz, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342.

Ferree, J.T. J. Demuth, G.M. Eosco, andN.S. Johnson, 2009: The increasing role of social media during high impact weather events. Preprints, *37th Conf. on Broadcast Meteorology*, Portland, OR, Amer. Meteor. Soc. 4.1A.

Giammanco, I.M., T.M. Brown, R.G. Grant, D.L. Dewey, J.D. Hodel, and R. Stumpf, 2015: Evaluating the hardness characteristics of hail though compressive strength measurements. *J. Atmos. and Oceanic Technol.*, **32**, 2100–2113.

Heymsfield, A.J., I.M. Giammanco, and R. Wright, 2014: Terminal velocities and kinetic energies of natural hailstones. *Geophys. Res. Lett.*, **41**, 8666–8672.

Hyvären, O., and E. Saltikoff, 2010: Social media asa source of meteorological observations. *Mon. Wea. Rev.*, **138**, 3175–3184.

Jewell, R. and J.C. Brimelow, 2009: Evaluation of Alberta hail growth model using severe hail proximity soundings from the United .States. *Wea. Forecasting*, **24**, 1592–1609.

Joe, P., D. Burgess, R. Potts, T. Keenan, G. Stumpf, and A. Treloar, 2004: The S2K severe weather detection algorithms and their performance. *Wea. Forecasting*, **19**, 43–63.

Longmore, S., S. Miller, D. Bikos, D. Lindsey, E. Szoke, D. Monenar, D. Hillger, R. Brummer, and J. Kaff, 2015: An automated mobile phone photo relay and display concept applicable to operational severe weather monitoring. *J. Atmos. and Oceanic Technol.*, **32**, 1356–1363.

Matson, R.J., and A.W. Huggins, 1980. The direct measurement of the size, shapes and kinematics of falling hailstones. *J. Atmos. Sci.*, **37**, 1107–1125.

Nisi, L., O. Martius, A. Hering, M. Kunz, and U. Germann, 2016. Spatial and temporal distribution of hailstorms in the Alpine region: a long–term, high resolution, radar-based analysis. *Q.J.R. Meteorol. Soc*., **142**, 1590–1604.

Ortega, K.L., T.M. Smith, L. Manross, K.A. Scharfenberg, A. Witt, A.G. Kolodziej, and J.J. Gourley, 2009: The severe hazards and analysis and verification experiment. *Bull. Amer. Meteor. Soc*., **90**, 1519–1530.

Schaefer, J.T., J.J. Levit, S.J. Weiss, and D.W. McCarthy, 2004: The frequency of large hail over the contiguous United States. Preprints, *14th Conf.. Applied Climatology*, Seattle, WA, Amer. Meteor. Soc., 3.3.

Snook, N., Y. Jung, J. Brotzge, B. Putnam, and M. Xie, 2016: Prediction and ensemble forecast verification of hial in the supercell storms of 20 Mat 2013. *Wea. Forecasting*, **31**, 811–825.

Treloar, A.B., 1998: Vertically integrated radar reflectivity as an indicator of hail size in the greater Sydney region of Australia. In Proceedings of *19th Conference on Severe Local Storms*, Minneapolis, MN, Amer. Meteor. Soc., 48–51.

Warren, R., H. Ramsay, S. Siems, M. Manton, J. Peter, R. Stone, and A. Protat, 2017: Hail climatologies for Sydney and Brisbane, Australia, derived from single-polariozation radar and insurance claim data. 8th *Conf. on Radar Meteorology*, Chicago, IL. Amer. Meteor. Soc., 232.

Witt, A., M.D. Eilts, G.F. Stumpf, J.T. Johnson, E.D. Mitchell, and K.W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*., **13**, 286–303.