Review of: **Improving object-oriented radar based nowcast by a nearest neighbor approach** by Bora Shehu and Uwe Haberlandt

Ruben Imhoff

Ruben.Imhoff@deltares.nl

May 27, 2021

Summary

The authors present a novel method to improve object-oriented radar rainfall nowcasts by making use of a k-nearest neighbor approach. The method uses already existing methods and on top of that, it replaces the generally used Lagrangian persistence for storm cell movement and tracking with a k-nearest neighbor algorithm. The results are promising and clearly outperform the benchmark (an object-oriented radar rainfall nowcasting algorithm that uses Lagrangian persistence for storm advection), especially for storms of short durations. These are the more convective storms, that often lead to flooding, and that are also harder to predict by other nowcasting methods. Hence, I think that this work is relevant and an interesting new approach. What I particularly have liked about the approach is the more physical look on using machine learning for a particular part in the nowcasting chain. The choice for a k-nearest neighbor approach makes sense from that perspective, and the results remain explainable (and thus also open for improvements).

After going through the manuscript, I have quite some suggestions and questions left, but I hope that these suggestions will help in further improving the manuscript. I would like to thank the authors for this interesting work and I look forward to seeing the revised version of it.

General comments

Methods

The unmatched storm cells are left out of the analyses. I do agree with this choice, but I wonder what the effect is on the algorithm performance once the ‘normal’ radar data, with these unmatched storm cells / artefacts, are fed into the algorithm. Can the authors comment on this?

What is the algorithm performance for extreme events or events that were not part of the training data? In addition, can the authors say some more about the size (memory) of the training dataset, the computation times and how big the training set should be for adequate use? From both an operational and reproducibility perspective, this would be very relevant information.

The authors often mention timestep of nowcast or nowcast time. Up to the end of the manuscript, I have found the terminology and meaning confusing (it may just be me..). Do the authors mean the issue time of the nowcast since the start/evolution of the storm with this? If so, I would recommend changing this for clarity.

I appreciate the use of an adequate benchmark (another object-oriented nowcasting method using Lagrangian persistence for the movement of storm cells) in this study. However, can the authors comment on other works that take next to Lagrangian persistence also other processes into account, e.g. splits and mergers (e.g. Dixon & Wiener, 1993; Han et al., 2009), and the rate of growth and dissipation (e.g. Pulkkinen et al., 2020 – although not necessarily constructed for object-oriented nowcasts)? Thus, how would this work relate to or even improve such methods? An analysis comparing such a method to this method is of course outside the scope of this paper, but it would be great if the authors can at least comment on it.

Ensemble approach and analysis

The construction of the ensembles is interesting, but the description in the methodology was not always clear to me (see the comments mentioned later in this document). Can I ask the authors to describe a little more elaborately how the ensemble members are constructed and how different weights are applied for example?

The ensemble validation (e.g. lines 306 – 311) is useful and is clearly tailored toward showing that the target value is present in the training dataset. Besides the focus on separate ensemble members, it may be interesting to also plot the ensemble spread vs. the error to get a more statistical indication of whether the observation falls within the ensemble spread or not (for the full ensemble). Note that this is to a certain extent already present in e.g. Fig. 13. Hence, see this rather as a suggestion than a must. For inspiration, see for instance Fig. 9 in Foresti et al. (2016) or supplement Fig. S6 in Imhoff et al. (2020) as spread vs error examples for ensemble nowcasting.

Results

Starting with the figures, I did like the analyses chosen by the authors, but the labels, legend and text in the figures was often hard to read. Zooming in is luckily possible, but can I ask the authors to make the font sizes of the figures bigger? Besides, the schematic figures describing the method are clear and are very nice figures (no changes needed there).

One figure that I did miss at the start of the results, is an example nowcast for several lead times with the k-NN methods, the Lagrangian persistence and the observations (the radar images, I suppose). This can directly show what to expect and visualize why we see certain results in the subsequent figures. Can I ask the authors to make such a figure, possibly one for each class (duration), so likely a small-scale convective event, a mesoscale convective event and perhaps a stratiform event?

Reproducibility

Are any of the data, scripts, etc. publicly available? That would increase the reproducibility, but also the impact of this very interesting work.

Specific comments

Title: Perhaps good to mention here radar-based rainfall nowcasts, to clarify the focus on rainfall forecasts.

Lines 33 – 43: Could you explain why the focus is on object-oriented nowcasting of primarily convective events here (so, why object-oriented and why convective events as focus)? I suppose this is because the authors focus on mostly convective events that can or have resulted in flooding.

Line 118: Is the C-band radar single or dual-pol?

Lines 122 – 123: Can I ask the authors to elaborate a bit on this merging method. It can be brief, the rest is mentioned in the paper of course.

Line 124 – "110 events": What is the definition of an event in this study and did you systematically look for certain event characteristics? Seeing the following lines, the authors have chosen to focus on mostly convective events. How were those selected?

Line 146 – "unmatched storm cells": What have you done with these unmatched storm cells? Are they left out of the method and analysis or not? I see the answer now in Sec. 3.2.3 Perhaps good to very briefly mention this (or point to Sec. 3.2.3) here.

Lines 146 – 152: How is the storm duration defined?

Lines 158 – 161: I agree! Nowcasting, especially Lagrangian persistence, either in an object-oriented or intermittent field based approach, works quite well for these stratiform events. It may be worth mentioning that this is not where the k-NN method will provide a lot of improvement (at least, that is what I expect) compared to already existing methods. Hence, it is no problem that the sample size is leaning more heavily on the convective events.

Line 164 / Eq. 1: What if one radar image is missing or another problem occurs for a certain time step? How will this be treated? In case of a missing radar image, would it be safer to divide by the number of used predictors in equation 1 (max. 6, but possibly less) instead of by 6? In addition, there are seven 5-min steps from t=0 to t=t-30, assuming that t=0 and t=t-30 are included. Should you divide by seven or is step t-30 not included?

Lines 242 – 243 and lines 253 – 255: How are the weights determined? I see that the weight of eq. 5 is determined from the results (Fig. 6), perhaps good to refer to this.

Line 253 / Eq. 6: In the estimated response of the to-be-nowcasted storm, are the total lifetime, area and intensity simply the weighted average of the k nearest neighbors? What about the location, are V_{X+LT} and V_{Y+LT} used to displace the location of the current state of the storm at t_0? It may be clear for future readers to specify this a bit in more detail.

Lines 255 – 261: I'm a bit lost with the way the ensemble nowcast is constructed. Are 30 individual members issued by taking every time a different selection of neighbors? Or just by one neighbor out the top 30?

Lines 277 – 298: Is there a difference in the training between VS1 and VS2, mainly w.r.t. the error per target or for all targets together? Or is the training identical for both approaches?

Line 304 – "Lagrangian persistence": Just to be sure, Lagrangian persistence in object-oriented nowcasting, right?

Lines 314 – 316: Do you have an indication how much less this was (in the worst-case scenarios)?

Lines 367 – 369: Do you have any idea why the result is different for the Total Lifetime?

Lines 380 – 385 and Fig. 8: How large is the spread in the results (the optimal k-value) per class? I can imagine that this does not make the figure clearer when added (as IQR for example), but perhaps the authors have an indication of this.

Lines 392 – 401: Is the decreasing error with increasing lead time (mainly visible in the top row of Fig. 9) a result of storms dissipating sooner than 30 min, which is then forecast well? Could the authors comment on that? It would otherwise be unexpected to see the performance improve with increasing lead time (you would expect the opposite).

Lines 402 – 412: I agree with the reasoning for the 36th time step. However, the y-axis of Fig. 9 (middle row) is clearly scaled to this time step, which makes it hard to distinguish what happens for the other two lines. Could you change the axis scale and just describe why the 36th time step falls outside this scale (which is already described now)? Another question regarding the 36th time step, because I'm not sure I understand the times of nowcast well here: isn't the time of nowcasts the same as the issue time, so the nowcast starts 3h after the evolution of the storm. In the class here, you are only considering storms that last maximum 3h. Hence, aren't we looking at storms that should have died already?

Lines 452 – 453: I think this conclusion needs the nuance that this is the case for the shorter storm durations, whereas for longer durations this is not or only to a lesser extent the case.

Line 492 – "Overall the ensemble results are better than the single 4-NN nowcast": Based on the results and the shown figure, I think you can only state that the best ensemble member performs better than the single 4-NN nowcast.

Lines 574 – 576: There should be some nuance here. Although the results are very promising and often outcompete Lagrangian persistence, these high improvement numbers are generally reached for short-living storms, while the improvement is less (sometimes even worse) for longer-living storms.

Line 586 – "additional predictors": For the interested reader(s), can you say more about which predictors from those sources you think would be feasible for this?

Figure 4 caption: Perhaps refer to Table 1 for the meaning of the symbols (the predictors) in the figure.

Technical corrections

Figure 2: As I hope that this interesting paper will be read by people from all over the world, it may be good to add a small subfigure indicating where this region is located in Germany or even in NW-Europe. Besides, the indicated coordinates seem to be in a local coordinate system. Can I ask the authors to mention this in the caption or, perhaps even better, to place lat-lon coordinates on the map?

Figure 3: What is the last group (duration) on the x-axis?

Figures 7, 8 and 9: The light grey color for the 5-min class is not easy to distinguish (especially on a colored background). Could you make the grey a little darker or use a different color?

Line 30 – "short-term rainfall nowcast": perhaps say short-term rainfall forecasting?

Lines 32 – 33: A minor detail about the storm and intermittent field references: the list of references can be almost endless here. It is not necessary to cite all of them, but perhaps you could say (e.g. + references) to indicate that this is just a sample of all studies to these topics.

Line 84 – "show for instance Hou & Wang (2017)": is for instance shown by Hou & Wang (2017).

Line 211 – "predictor important analysis": should this be importance instead of important?

Line 391 – “event-based 4-NN”: For consistency, did you mean storm-based 4-NN?

Line 463 – “same results”: similar results.

Line 561 – “Person”: Pearson.

Line 562 – “two measurement”: Two measurements.

Line 570 – “the death processes”: Perhaps better to use dissipation processes of storms?

References

Dixon, M., & Wiener, G. (1993). TITAN: Thunderstorm identification, tracking, analysis, and nowcasting—A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10(6), 785–797. https://doi.org/10.1175/1520-0426(1993)010<0785:TTITAA>2.0.CO;2

Foresti, L., Reyniers, M., Seed, A., & Delobbe, L. (2016). Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in Belgium. *Hydrology and Earth System Sciences*, 20(1), 505–527. https://doi.org/10.5194/hess-20-505-2016

Han, L., Fu., S., Zhao, L., Zheng, Y., Wang, H. & Lin, Y. (2009). 3D convective storm identification, tracking and forecasting – An enhanced TITAN algorithm. *Journal of Atmospheric and Oceanic Technology*, 26(4), 719 – 732, https://doi.org/10.1175/2008JTECHA1084.1

Imhoff, R. O., Brauer, C. C., Overeem, A., Weerts, A. H., & Uijlenhoet, R. (2020). Spatial and temporal evaluation of radar rainfall nowcasting techniques on 1,533 events. *Water Resources Research*, 56, e2019WR026723. https://doi.org/10.1029/2019WR026723

Pulkkinen, S., Chandrasekar, V., Von Lerber, A. & Harri, A-M (2020). Nowcasting of convective rainfall using volumetric radar observations. *IEEEE Transactions on Geoscience and Remote Sensing*, 1 – 15, https://doi.org/10.1109/TGRS.2020.2984594