Final response in the interactive discussion

Dear Referees,

We would like to thank you for your positive comments and constructive suggestions for the improvement to our manuscript "Optical flow models as an open benchmark for radar-based precipitation nowcasting (rainymotion v0.1)". In this document, we would like to provide our responses to the comments of each of the three referees in one single document.

The referee comments turned out to be very helpful. Based on these comments, we suggest several changes to the manuscript and the rainymotion library which we will outline in detail on the following pages.

For that purpose, we will show the referee comments in **black** font, and our responses in **blue**. For the sake of clarity, we have also not reproduced some introductory parts of the referee comments in this comment. Parts that were not reproduced, are marked as [...].

We hope that the suggested changes sufficiently address the referees' concerns, so that we can, given the approval of the editor, finalize the revision of our manuscript.

Sincerely,
Georgy (on behalf of authors)

Referee comment #1 (by Loris Foresti)

Main comments

1. The forecast verification is well done, but in my opinion it should include a verification of the statistical properties of the advected rainfall fields to understand the degree of numerical diffusion, which can be a major problem in precipitation nowcasting if not properly handled. Such effect usually leads to an undesired smoothing of the precipitation fields, which reduces the more interesting high rainfall intensities and complicates the inter-comparison of models.

We entirely agree that it would be interesting to verify the statistical properties of the advected rainfall fields. It will be done as suggested in comment #11: using a periodogram of rainfall intensities of advected precipitation fields for different lead times.

2. As the paper presents new optical flow and advection techniques, it must include some additional figures showing examples of motion fields and precipitation nowcasts, e.g.:
   2.1. A multi-panel figure with vector plots of the motion fields retrieved by the different methods overlaid on top of radar images (for example one "rotational" precipitation event).
   2.2. A multi-panel figure showing examples of observed and nowcasted precipitation fields at different lead times, e.g. 30 or 60 minutes. This would be very useful to understand the quality and realism of the advected rainfall fields, and check

whether there are any artefacts due to numerical diffusion and interpolation processes.

We will add requested figures in the revised version of the manuscript.

3. Some statements in the literature review are a bit imprecise and could be improved.

We will revise the literature review in the introductory section based on several referee comments (comments #5, #6, #7, and #24 by Dr. Foresti, #1.2 and #2 by Dr. Pulkkinen, #1, #2, and #3 by Dr. Uijlenhoet).

Specific comments

4. Page 1, line 3, Page 2, line 14. "extrapolate the motion" -> "extrapolate the radar echoes". The motion field is usually kept fixed and only the radar echoes are extrapolated, although in some cases it may be beneficial to extrapolate the motion field together with the precipitation echoes.

We suggest to rephrase this to "[...] and then to displace the precipitation field to the imminent future (minutes to hours) based on that motion, [...]".

5. Page 2, line 4. The cited approaches (analogue, local Lagrangian and stochastic) were mentioned in the context of probabilistic precipitation nowcasting. They all provide empirical estimates of the probability density function in different ways. Please update accordingly.

Please see response to comment #7.

6. Page 2, lines 5-6. Foresti et al. (2015) did not use the correlation coefficient as a measure of similarity to retrieve the analogues (as done e.g. by Atencia et al., 2015), but rather the Euclidian distance in the space of principal components. Please adjust the statement.

Please see response to comment #7.

7. Page 2, lines 8-10. I think there is some confusion about the definition of "local Lagrangian method". The cited paper (Foresti et al., 2015) follows the definition of Germann and Zawadzki (2004), which defines the "local Lagrangian" as one possible method to derive a probabilistic nowcast. This is achieved by collecting the precipitation values upstream in a local neighbourhood, whose size is increased as a function of lead time.

Comments #5, #6, and #7 are related to one paragraph (Page 2, lines 4-9). We will rewrite the whole paragraph in accordance with the referee's suggestions and try to make the main message of this paragraph (classification of methods used for radar-based precipitation nowcasting) clearer.
We suggest to rephrase the corresponding paragraph to:
"A variety of radar-based precipitation nowcasting techniques can be classified on three major groups based on assumptions we make regarding precipitation field characteristics (Germann and Zawadski, 2002). The first group -- climatological persistence -- provides nowcasts by using

climatological values (mean or median). The second group -- Eulerian persistence -- is based on using the latest available observation as a prediction, and is thus independent from the forecast lead time. The third group -- Lagrangian persistence -- allows the extrapolation of the most recent observed precipitation field under the assumption that the motion field is persistent (Germann and Zawadzki, 2002; Woo and Wong, 2017). In addition, we can classify nowcasting methods based on introduced prediction uncertainty: In contrast to deterministic approaches, ensemble nowcast attempt to account for predictive uncertainty by including different realizations of the motion field and the evolution of rainfall intensity itself. In this study, we focus our model development around the group of Lagrangian persistence models which provide deterministic precipitation nowcasts. Yet, the unified availability of different tracking and extrapolation techniques in the rainymotion library could directly be used to construct ensembles that account for the uncertainty of rainfield displacement."

8. Page 3, line 1. I fully agree that optical flow libraries have been around for long, but they cannot be directly applied for the retrieval of radar echo motion without important adaptations and tests. For example, they must be tuned to represent the typical range of advection speeds of real precipitation fields, they must be spatially dense and extrapolate well also in regions without precipitation, etc. This is why papers like yours are important contributions to make the necessary adaptations and tests.

We agree that the original manuscript does not sufficiently address how, on the one hand, parameters of different optical flow techniques affect the specific problem of precipitation field tracking, and, on the other hand, how the results of different optical flow techniques might need further post-processing in order to enhance their usefulness for the extrapolation step (e.g. filling or interpolating zero velocities that might occur in regions of zero rainfall). Given that we also introduce further optical flow/tracking as well as extrapolation techniques (comments #9 and #12 by Dr. Foresti, #4-7 by Dr. Pulkkinen) in the revised version of the manuscript and the rainymotion library, the revised manuscript will address these requirements more precisely and comprehensively.

9. Page 3, line 8. Page 4, line 24. It would be interesting to know why you decided not to include in the list of benchmark extrapolation techniques the backward-in-time semi-Lagrangian scheme, which is generally accepted to be the most appropriate method (Germann and Zawadzki, 2002). The forward scheme is known to produce holes in the precipitation field in presence of divergent vectors, which need to be interpolated. This inevitably leads to additional numerical diffusion.

We originally implemented the forward scheme because it is more intuitive to advect the precipitation field "forward in time" and "downstream in space". Based on the referee's comment, though, we decided to complement the revised version of the rainymotion library with a backward method for the optical flow calculation.
On that basis, we repeated our benchmark experiments by using the backward scheme both for the Dense (constant-vector) and DenseRotation (semi-Lagrangian) models. Results show that the backward scheme performs slightly better for low rainfall intensity rates (under 0.5 mm/h) and longer lead times (from 30 minutes). For rainfall intensity rates over 0.5 mm/h and shorter lead times (up to 30 minutes) there are no significant differences between both schemes. Based on the new results we decided to implement the backward scheme as a default option for precipitation motion field calculation in the revised version of the rainymotion library. We will

update the revised version of the manuscript and the supplementary material in accordance with the new results.

However, we also want to note that the intercomparison of different advection schemes provided in Germann and Zawadzki (2002) cannot, in our opinion, be interpreted in a way that "backward-in-time semi-Lagrangian scheme [...] is generally accepted to be the most appropriate method". In the corresponding paper, the forward-in-time scheme is concerted with a gaussian redistribution of advected rainfall in contrast to the interpolation used for backward-in-time scheme. In our library, we adapt the same "interpolate only once" idea of Germann and Zawadzki (2002) -- regardless of the direction used for velocity field (optical flow) calculation that allows intercomparison of forward and backward schemes in a similar setting. Although the new results of our intercomparison are consistent with the referee's statement on the backward scheme being superior (for low rainfall intensities and longer lead times), in our opinion further research is needed to compare the efficiency of different implementations in detail.

10. Page 3, line 26. I cannot understand properly why you mention the concept of scale-dependence in the context of local LK methods. Please explain how local optical flow techniques account for scale-dependence.

Our intention was to highlight that for the Sparse group of rainymotion's tracking models we use distinct "corners" instead of storm cells -- this eliminates the need to specify arbitrary and scale dependent characteristics of "precipitation features" while the identification of "corners" depends only on the gradient sharpness in a cell's neighborhood. Of course this will not solve the issue of scale-dependence of the average motion itself. We will clarify these aspects in the revised manuscript.

11. Page 4, line 5. I am a bit worried that the use of warping and interpolation of discontinuities in the advected radar field can lead to serious numerical diffusion effects. The most appropriate method to test this issue is to compute the Fourier spectrum of the original and advected fields to check whether there is loss of power at the high spatial frequencies (see Fig. 10 in Germann and Zawadzki, 2002). A simpler approach would be to compare the histogram of nowcasted rainfall fields at different lead times with the one of the last observed radar image. The variance and histogram should be conserved during the extrapolation.

We will update the revised version of the manuscript with figures that describe the level of numerical diffusion for the different models by using the nowcasts' spectral power density for different lead times.

12. Page 4, line 26. I agree that the constant-vector approach does not explicitly allow to account for rotation. However, if the advection is applied recursively in short time steps the rotation can be approximated by a set of short straight lines (at the cost of stronger diffusion). Despite this fact, I believe that a good implementation of the semi-Lagrangian scheme should consistently give better (or comparable) results than the constant-vector approach.

We agree that an accurate implementation of the semi-Lagrangian scheme should yield a skill that is at least equivalent to the constant-vector approach. We have found two possible reasons

why our original implementation did not achieve that: 1. Errors in the estimation of motion fields (e.g. with anomalies, artefacts etc.) could affect the forecast in the semi-Lagrangian advection scheme more than in the constant-vector scheme, since displacement vectors from regions of higher uncertainty might be "activated" more frequently; 2. Higher complexity of semi-Lagrangian scheme implementation which involves interpolation on two levels: when we advect each pixel and try to find the new velocity vector for any new pixel location, and during the final interpolation of intensities.

We attempted to address the estimation of field motion using the (local) Farnebäck optical flow method by implementing a variational refinement procedure to smooth the velocity field, and to get rid of spurious velocities, and by implementing different global optical flow methods that usually provide more smooth and robust motion fields (see also comment #4.1 and #7 by Dr. Pulkkinen). As a result, we included both the variational refinement and different global methods for the tracking step in the rainymotion library, and included these approaches in our benchmarking experiments.

According to these new results, the implementation of the Dense Inverse Search (DIS) global optical flow method (Kroeger et al., 2016) provides better results than the Farnebäck method with variational refinement and other global methods such as DeepFlow (Weinzaepfel et al., 2013) and PCAFlow (Wulff and Black, 2015). Based on these new findings, we decided to use the DIS method as a default method for the precipitation motion field calculation in the revised version of the rainymotion library. We also found that using the DIS method, our results show no significant difference between Dense and DenseRotation models. That confirms the strong influence of motion field estimation on the performance of the DenseRotation model.

We will update manuscripts accordingly and show the intercomparison of of different optical flow methods in the supplementary material.

Kroeger, T., Timofte, R., Dai, D., & Van Gool, L. (2016, October). Fast optical flow using dense inverse search. In *European Conference on Computer Vision* (pp. 471-488). Springer, Cham.
Weinzaepfel, P., Revaud, J., Harchaoui, Z., & Schmid, C. (2013). DeepFlow: Large displacement optical flow with deep matching. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1385-1392).
Wulff, J., & Black, M. J. (2015). Efficient sparse-to-dense optical flow estimation using a learned basis and layers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 120-130).

13. Page 4, line 29. Also here I would study the effect of numerical diffusion caused by the interpolation. Numerical diffusion can also have undesired consequences when comparing (benchmarking) different nowcasting models. In fact, a precipitation nowcast that loses power at the high spatial frequencies will be generally smoother. This behavior will be rewarded in terms of some verification scores (in particular the MAE/RMSE), which affects the comparison with other models. A fair comparison of different nowcast systems should be done at similar spatial scales, for example using Fourier or wavelet decompositions.

Please see response to comment #11.

14. Page 6, line 10. "programmatic realization" is a strange expression.

We will rephrase to "In this study we used the RV product data as an operational baseline and did not reimplement the underlying algorithm itself."

15. Page 6, line 31. "rainfall depth product". Is it the instantaneous intensity in mm/hr or an accumulation?

The RY product represents rainfall depth in mm for a five minute interval which is however derived from an instantaneous intensity considered representative for that interval.

16. Page 6, line 23. It would be very interesting to move the CSI verification at a threshold of 5 mm/hr from the supplementary material to the actual paper. These rainrates are the ones that are relevant to trigger warnings for severe weather.

We will update Figure 6 to represent the CSI for the threshold of 1 mm/h and Figure 7 to represent the CSI for the threshold of 5 mm/h.

17. Page 8, lines 5-10. You are correct. Detailed motion fields provide better skill at short lead times, while smoother motion fields are more adapted for longer lead times. Similarly to precipitation fields, the motion fields also have an intrinsic predictability (persistence). This can be exploited by gradually smoothing the motion field in a way that is consistent with its predictability.

We agree. In fact, users can use the library to implement such ideas.

18. Page 8, line 14. All the proposed solutions to the problem of low predictability at convective scales are based on the optical flow and are all valid options. However, precipitation, and in particular the one of convective nature, has a large unpredictable component that we will likely never be able to predict. Therefore, the nowcasting community needs to admit the incapability of providing accurate deterministic precipitation forecasts and find ways to estimate and communicate the inherent uncertainty. I am glad that you presented this issue in the conclusion at page 9, lines 22-25, but it would be a good idea to make this point stronger.

We will try to emphasize this point in the revised version of the manuscript.

19. Page 9, line 20-21. I also believe that we should not discard the Sparse models. One possibility is to make them "dense" by interpolating the motion vectors before applying the advection scheme (hopefully semi-Lagrangian).

We thank referee for this comment. More generally, future research should analyse in more detail which steps in our Sparse model chain contribute the most uncertainty. We still think the combination of Sparse optical flow and warping is very efficient and promising, but should be understood better.

20. Figures 1 and 2. These are extremely clean and nice presentations of the methods.

We thank referee for this comment.

21. Figure 3. You may add in the caption that the figure shows the forward-in-time semi-Lagrangian method.

Figure 3 caption will be updated accordingly.

22. Figure 5. You may consider writing a more descriptive figure caption, e.g. "Verification of the different optical flow based nowcasts in terms of MAE for 11 precipitation events over Germany".

Figure 5 caption will be updated accordingly.

23. Page 7 line 15, Figures 6-7. Is there an explanation on why the RADVOR nowcasting method performs poorly in the first 5-10 minutes? The effect seems quite systematic and I have a hard time explaining it with the faster movement of precipitation fields.

We briefly described the possible reasons of this RADVOR behavior on Page 7 lines 14-15 and Page 8 lines 3-8. In our opinion, the use of smoothed displacement fields that focus on a large scale motion patterns particularly cause a loss of skill in the RV product for the first 5-10 minutes. We will update the corresponding paragraph of Section 5.1 (Model comparison) to make that point clearer.

24. Page 9, lines 27-30. With respect to the use of open source libraries to promote the developments in the field of nowcasting, you could also mention how you would imagine the contribution from rainymotion to the developments of other projects, as for example the probabilistic nowcasting library pysteps (https://pysteps.github.io/). In my opinion, any improvement in optical flow methods, e.g. using the rainymotion library, will also have a positive impact on the quality of probabilistic nowcasts. This could represent an interesting synergy between the two libraries, in line with the open source philosophy.

At the moment of paper submission (6th July), no reference to pySTEPS was known to us (first commit on GitHub from 9th July). We will update the introductory section to add a reference to pySTEPS, but we will also include the perspectives mentioned by the referee in the "Summary and conclusions".

Referee comment #2 (by Seppo Pulkkinen)

Relation to previous work and literature review

1. There are two important classes of optical flow methods that are only briefly mentioned or not mentioned at all:

    1.1.   In the variational methods, a smoothness constraint is added to the optical flow equations and they are solved "globally" over the whole domain. The key practical difference to the "local" methods, such as Farnebäck and Lucas-Kanade is that the motion field is automatically filled to areas of no precipitation.

    1.2.   In the spectral methods, the Fourier transform is applied to the inputs and the optical flow equations are solved in the spectral domain. The authors could add a citation to [3].

[3] E. Ruzanski, V. Chandrasekar and Y. Wang, The CASA Nowcasting System, Journal of Atmospheric and Oceanic Technology, 28(5), 640-655, 2011.

We thank referee for the clarification. We will add the corresponding methods and references to the introduction of the revised version of the manuscript, particularly since we added a variational approach and several global methods to the rainymotion library and our benchmark experiment - please see our response to comment #12 of Dr. Foresti.

2. There are several widely used optical flow algorithms developed in the machine vision literature. The authors could cite the Brox and CLG algorithms ([1] and [2]). These have also publicly available C implementations (see the IPOL journal).

[1] T. Brox, A. Bruhn, N. Papenberg and J. Weickert, High Accuracy Optical Flow Estimation Based on a Theory for Warping, ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV, 25-35, 2004.
[2] A. Bruhn, J. Weickert and C. Schnörr, Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods, International Journal of Computer Vision, 61(3), 211-231, 2005.

We will add those references to the paragraph where we mention only openCV as an open software library with optical flow algorithms implementation (Page 3 line 1 of the discussion paper).

3. The paper cites to a large number of references where more advanced probabilistic nowcasting methods are described. Therefore, in the third paragraph of Section 6 the authors should be more concrete about future plans to include such features into rainymotion, and not just present ideas of potential improvements. Or will rainymotion be restricted only to deterministic extrapolation nowcasting based on Lagrangian persistence?

We would like to thank the referee for this suggestion, yet we are hesitant whether more detailed perspectives on future developments should be elaborated in the paper. Based on its current design, rainymotion's focus is to track the motion of rainfields and to extrapolate future rainfall on that basis. At the same time, the rather low-level implementation easily allows for the flexibility to manipulate the displaced precipitation fields in order to represent -- stochastically or

deterministically -- the dynamics of precipitation intensity. Yet, there are no specific plans to implement such features to the rainymotion, but we will, in the revised version of the manuscript, highlight more explicitly the possibility to include such developments.

Methodology

4. Precisely speaking the Farnebäck optical flow algorithm is not global or dense and should not be called such. This misuse of terminology originates from the OpenCV library.

    4.1. The Farnebäck method is dense only in the sense that it produces gridded output instead of motion vectors for sparse feature points as Lucas-Kanade does. If you look at the paper of Farnebäck, the method is formulated as local feature matching, where the solution of the optical flow equations is done by using a polynomial approximation. As a result, the method produces zero motion velocities to areas of no precipitation. You can verify this by plotting motion fields produced by the Farnebäck method.

We thank the referee for pointing that misuse in terminology which we have not been aware of so far. In our paper we use the term "local" and "sparse" in the sense that these methods provide motion vectors at specific locations only. In contrast, we use "global" and "dense" for pointing out that the motion vectors are calculated for an every radar image pixel.
We will revise the paper in a way that terminology is both accurate and easy to understand. As a possible solution we propose to change "local" and "global" to "sparse" and "dense" in the revised version of the manuscript and provide a more detailed description of what we consider as "sparse" and "dense" models in the Section 2 (Model). Furthermore, we have actually added global optical flow techniques to the set of tracking models (please see our response to comment #12 of Loris Foresti), so the revised manuscript version will explicitly address the issue of global vs. local optical flow.

    4.2. It follows from the above that when a pixel is advected into area of no precipitation and a new motion vector is taken at that location (as in the DenseRotation method), it's motion to stops at the boundary. This could explain why Dense has in many cases better performance than DenseRotation.

At the time of submission of this manuscript and thus when the results for this paper had been produced, the implementation of the extrapolation algorithm of the dense optical flow models did in fact not account for the case that pixels are advected into regions of zero velocities. In the meantime, however, we have revised the algorithm so that zero velocities are discarded and replaced by interpolation, and the results will be updated accordingly. Yet, the hypothesis that the insufficient treatment of zero velocities was responsible for the Dense model outperforming the DenseRotation model could not yet be corroborated (please also refer to our response to comment #12 of Dr. Foresti).

5. In Germann and Zawadzki (2002), the authors conclude that the backward semi-Lagrangian has better performance than the forward method. In fact, a majority of existing nowcasting methods use the former that is widely regarded as the best approach. However, here the authors use only the latter. If possible, the authors could also implement the backward method and include it in the performance comparison.

Dr. Foresti in his comment #9 also raised this issue. Please refer to our answer there.

6. Using the backward method would require filling the gaps in the motion field on areas of no precipitation. Otherwise, no precipitation would be advected into areas where it does not exist at the nowcast start time. A simple distance-weighted interpolation should be sufficient for this purpose. For the above reason, using gap-filling would also improve the performance of the forward semi-Lagrangian method.

In the revised version of the rainymotion library we implemented the referee's suggestion of "[..] filling the gaps in the motion field on areas of no precipitation" by utilizing inverse distance weighted interpolation to fill zero-gaps in the motion field. However, the benefit of this implementation on the performance of the forward semi-Lagrangian method (the DenseRotation model) is not so distinct probably because of the reasons we highlighted in the response on the comment #12 from Dr. Foresti (motion field estimation errors by the Farnebäck algorithm and additional interpolation).

7. Note that the gap-filling is automatically done in the variational methods without the need for separate post-processing of the motion field. Therefore, such methods are truly dense and global. The authors could consider implementing a variational method and include it in the performance comparison.

We thank the referee for his recommendations regarding the implementation of variational optical flow models in the rainymotion library. We incorporated global optical flow methods which are available in opencv library as additional options for motion field calculation in the rainymotion library (see also our response to the comment #12 from Dr. Foresti), verified their skill for nowcasting and have to conclude that using more advanced global optical flow methods advances an efficiency of a semi-Lagrangian advection scheme. Based on the new obtained results we decided to replace the Farnebäck method by the global Dense Inverse Search (DIS, Kroeger et al., 2016) as a default tracking option. We will also update the supplementary material with intercomparison results of different optical flow methods.

Kroeger, T., Timofte, R., Dai, D., & Van Gool, L. (2016, October). Fast optical flow using dense inverse search. In *European Conference on Computer Vision* (pp. 471-488). Springer, Cham.

Software library

8. Sections 2.4 and 3: Is the library restricted only to using the DWD data? Please add discussion about how to use the library with other file formats? For instance, by using wradlib this should be easily done because it supports a large number of different formats.

There is no restriction in using different data formats because of rainymotion works directly with numpy arrays, and the data preprocessing routine is fully on the user-side. There is a set of available open software libraries for radar data reading and preprocessing (the list available on https://openradarscience.org/). We will add the corresponding information to the Section 2.4 (rainymotion Python library).

Verification

9. Section 2.6: MAE could be computed conditionally over those pixels where both the nowcast and the verifying observation exceed the detection threshold. Otherwise, there would be overlap with the CSI statistic as both penalize incorrect forecasts of precipitation/no precipitation.

In our study we decided to use MAE as a score from a continuous category and implement it directly without making specific thresholds (like we do for categorical category of verification scores). In our opinion, this admittedly arbitrary decision of using different verification score categories helps to represent a diversity of obtained results.

10. A large number of CSI and MAE statistics are shown for different lead times. There could be more analysis of the results.
    10.1. There is no indication about what can be considered as a good CSI or MAE value for the nowcast to be usable. Can you give some thresholds?

At the best of our knowledge, there is no convention regarding what to consider as "good" or "bad" for any verification metric commonly used in radar-based QPN. For our benchmarking experiment, the focus is on the differences of scores between the different models, not on their absolute values.

    10.2. The differences between the methods (excluding Persistence) are relatively small in terms of CSI and MAE statistics. Based on such differences, the authors should be more careful when claiming that some method is better than another. For instance the maximum mean difference between Dense and DenseRotation is only 0.01 according to Table 3.

Table 3 represents statistics which are averaged over all the analyzed events and two lead time periods (5--30, and 35--60) and primarily highlight the difference between the Dense group of rainymotion models (Dense and DenseRotation) and and the RV product (as mentioned on the Page 7, lines 27--31) -- which is more distinct than the difference between Dense and DenseRotation models themselves. For the verification procedure we also carried out the Student's independent two-sample *t*-test to find whether differences between mean CSI and MAE values for the specific lead times are significant or not (not shown in the manuscript). We found that the results of the visual inspection of the verification plots are well consistent with the formal statistical evaluation: if there is a clear difference in the plots, it is typically significant in a statistical sense.
We will update Table 3 with the new results and adjust our statements about considering one model better/worse than another correspondingly.

11. Figures 5-7 and p. 9, lines 19-21. The authors should indeed take a closer look on why the performance of the sparse methods is poor. Some comments about this:
    11.1. The relevant parameter here is the number of features used in the tracking and nowcasting. If this number is too small, the motion vectors of the features are not representative of the large-scale motion field. Can you check this by adjusting the thresholds in the feature detector?

11.2.   In addition, can you specify somewhere how many feature points are used with the sparse methods because this is a key parameter?

11.3.   Another point missed in the paper is that the corner detector tends to pick features that have high intensities and gradients. Therefore, a very careful quality control is needed to ensure that the features are precipitation and not some random artefacts in the radar data. Can you be sure that the quality control is sufficient?

11.4.   Even if the features are precipitation, they represent small-scale phenomena that can have very different motion from the large-scale advection field. Thus, the representativity of such features can be very poor.

We agree with the referee that the sensitivity of the Sparse group of models to specific key parameters needs to be investigated more closely. Yet, we consider such an analysis beyond the scope of this study. Another study is underway that specifically and systematically focuses on the error of the forecast location of detected features based on a vast set of tracking and extrapolation techniques, and including different parameterisations as mentioned by the referee (such as the maximum number of features detected, or different approaches to filter spurious or non-representative velocities at small spatiotemporal scales). In the present manuscript under discussion, however, our aim is to present two basic and open architectures of nowcasting models based on optical flow which can serve as a baseline for future developments - as part of the rainymotion library itself or in combination with the library, and to demonstrate that these are skillful. Still, the parameters of the Shi-Tomasi corner detector provide us a possibility to control the maximum number of features, their quality (which is based on the minimal eigenvalue) and a minimum euclidean distance between the nearest identified points. A calibration of these parameters had been performed on different events and the most robust values had been set up as default parameters as follows: maximum number of features -- 200; quality level -- 0.2 (the corners with the quality measure less than the product of quality level and minimal eigenvalue will be rejected); minimum euclidean distance -- 7 pixels (the corners which have stronger neighbors in a neighborhood less than 7 pixels will be rejected). As for quality control of the actual radar data, we rely on the DWD's processing workflow that produces the RY product and which eliminates vast parts of spurious echoes. Yet, even in the presence of residual static or dynamic clutter, the tracking algorithm has proven to be robust against producing zero velocities.

12.   Forecasting the occurrence of precipitation/no precipitation for high intensities is highly relevant for practical applications. Therefore, I would suggest moving the results with the 5 mm/h threshold from the supplementary material to the main paper.

We support referee's recommendation (see also comment #16 by Loris Foresti) and will transfer the corresponding figure from the supplementary to the main paper.

Figures

13.   Since the motion field determination plays a key role in the paper, the authors should show at least one figure with an observed precipitation field and the computed motion field plotted on the same figure. Even better would be a figure showing motion vectors of features and motion fields computed by using different methods.

We agree with the referees' recommendation (see also comment #2 by Loris Foresti) and will add the requested figures to the revised version of the manuscript.

14.   Figure 4: Are names of individual functions relevant here? Consider removing them.

In our opinion, it is informative to show the key functions that we used from various libraries in order to put together the main functionality of rainymotion. It illustrates that the combination is, from a technical perspective, not too complex.

Minor details

15.   Page 4, lines 3-6 and Figure 1. How exactly is the affine transformation matrix calculated. In particular, is a single matrix estimated for all features or is this done separately for each feature?

The transformation matrix is calculated on the basis of all identified features. We will add this clarification to Section 2.1 (Local optical flow models).

16.   Page 5, line 24. Why the HDF5 file format was chosen? Please add some justification for this.

For all internal projects we use HDF5 database and corresponding file format as an efficient data storage with powerful set of archiving options (i.e. compression rate, chunk size) instead of using default binary files provided by the DWD. However, we propose to remove the reference to HDF5 file format and h5py library because of it is neither integral part of our analysis, nor the rainymotion library, but just a subjective choice we made regarding our research workflow. We will update the Section 3 correspondingly.

17.   Page 9, lines 7-9. I don't understand what this means. Can you clarify?

The statement "It might also be considered to combine the warping procedure for the extrapolation step with the Dense optical flow procedure for the tracking step in order to dramatically enhance computational performance" describes the idea to detect corners, then predict the future locations of these corners using the motion field from dense optical flow, and then construct the Affine Transformation Matrix for the warping based on the corner locations at forecast time and lead time t_n. That way, we would combine the robustness of the dense optical flow technique with the computational efficiency of the warping technique. We will clarify that idea in the revised manuscript.

18.   Page 9, line 24. Stochastic accounting <- stochastic modeling?

We thank referee for pointing out that mistake which will be corrected in the revised version of the manuscript.

Referee comment #3 (by Remko Uijlenhoet)

1. References to important papers from Marc Berenguer, Daniel Sempere-Torres and Geoff Pegram are missing (SBMcast, etc.). These are very relevant papers in the context of this manuscript, which discuss the issue of spectral decomposition of precipitation fields and scale-dependent radar nowcasting.
2. Reference to Berne et al. (2004; JoH) is missing. This is a (by now) classical paper on space-time scales of rainfall fields required for (urban) hydrological applications.
3. Reference to pySTEPS appears to be missing (https://github.com/pySTEPS). This is the open source Python version of STEPS. Highly relevant given the topic and focus of this manuscript.

We will include the suggested references in the introductory section. As for the missing reference to pySTEPS, we refer to our response to comment #24 of Loris Foresti.

4. Please provide some more detailed background information concerning: Shi–Tomasi corner detector (Shi and Tomasi, 1994); Lucas–Kanade optical flow algorithm (Lucas and Kanade, 1981); affine transformation matrix (Schneider and Eberly, 2003); warping and interpolation (Wolberg, 1990).

We will try to illustrate in more detail the main features of these techniques in the revised version of the manuscript.

5. Corrections, grammar and typos
   5.1. "Supplementary" –> "Supplementary Information" (several times in the manuscript).
   5.2. P.4, l.8: "24 recent radar images" –> "24 most recent radar images".
   5.3. P.5, l.20: "models' description" –> "model description".
   5.4. P.6, l.25–26: "rainfall rates prediction" –> "rainfall rate prediction".
   5.5. P.8, l.6: Insert comma before "which".

Will be fixed.

6. Is "RV" the same ad "RadVor"?

RADVOR is the entire nowcasting workflow used by the DWD. RV is a main product along that processing chain which is the forecast precipitation depth in five minute intervals over a lead time of two hours. The official main product of RADVOR, though, is the RQ product which is the precipitation depth accumulated over an interval of one hour for a lead time of two hours. It is basically obtained from the RV product, but includes an additional adjustment of the distribution function. In summary, the RV product is the part of DWD's nowcasting chain that is best comparable to our nowcasting products and the best "end product" that is available at an interval of five minutes.

7. General: (much) more detailed captions; figures + captions should be as self-contained as possible.

We will update the figure captions to make them more self-contained.

8.   Journal (Nature), issue, page numbers missing from reference to Bauer et al. (2015).

We will update the corresponding reference to the Bauer et al. (2015) paper as following:
Bauer, P., Thorpe, A., Brunet G.: The quiet revolution of numerical weather prediction, Nature, 525, 47–55, https://doi.org/10.1038/nature14956, https://www.nature.com/articles/nature14956, 2015.