Dear Editor and Referees, thank you very much for the comprehensive, constructive and helpful feedback on our study. In this document, we address all the comments and questions raised by the two referees. For a better overview, we have retained the structure of the manuscript and subdivided the comments according to the corresponding sections. Below, comments of Referee #1 are highlighted in blue and comments of Referee #2 in orange. Our responses are provided in *black*. Line numbers of the referee comments refer to the original submission. Line numbers in the answers refer to the revised manuscript.

General comments

The authors combined two existing morphometric approaches in a single dune tracking tool and applied it in a Monte Carlo Simulation to investigate the influence of varying input parameter settings and quantify the (un)certainty of outputs. The methodology and results of this sensitivity analysis are described in much detail and supported by multiple illustrations. I am convinced that the study is worth publishing, yet the manuscript is relatively long and mainly descriptive. This is also reflected in the discussion section which describes new data rather than interpret the existing results and discuss implications, e.g. recommendations for future morphometric studies. Overall, the analyses appear scientifically sound and relevant, but I have some suggestions on how to adjust the focus and improve transferability. Please find these suggestions as comments in the attached PDF file.

Furthermore, I provide my responses to the reviewing questions as a guideline here:

1. Does the paper address relevant scientific questions within the scope of ESurf?

The paper addresses relevant scientific questions that fit ESurf's scope but its narrative and focus could be improved.

2. Does the paper present novel concepts, ideas, tools, or data?

Although tools and concepts are taken from previous studies, the presented analysis is new and worth discussing.

3. Are substantial conclusions reached?

The conclusions are relevant but mainly descriptive in their current form. The authors should try to integrate implications and recommendations for future studies.

4. Are the scientific methods and assumptions valid and clearly outlined?

The methods and assumptions are suitable and described clearly.

5. Are the results sufficient to support the interpretations and conclusions?

The results are well described and in much, if not too much, detail. But the manuscript falls short with respect to their interpretation and contextualization. As outlined before, this leads to conclusions which are mainly descriptive.

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?

The methodology is described in great detail and looks readily reproducible.

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?

The authors give proper credit to the original sources of their methodology but, in my view, should focus more on their own contribution which is the sensitivity analysis for an exemplary dune tracking algorithm (cf. next comment). Especially with respect to the current title, the authors fail to summarize the current state-of-the-art of bedform identification and previous studies on the subjectivity of their results.

8. Does the title clearly reflect the contents of the paper?

This is my central point: the title implies that the authors present "a new dunetracking tool" which I think i) they don't and ii) is not their main result. The authors even state that "There are many so-called dunetracking tools ..." So why present yet another one?

I would hence strongly recommend to clarify what is the novelty of this study, i.e. the quantification of uncertainty resulting from different (non-objective) input parameters and their weighting calculated for one exemplary dune tracking algorithm. Furthermore, this finding calls for recommendations to be followed in future morphometric analyses.

9. Does the abstract provide a concise and complete summary?

The abstract is overall concise and complete (notwithstanding the previously raised issues).

10. Is the overall presentation well-structured and clear?

The presentation of results is well-structured and clear, yet in my view too detailed and descriptive (the total number of figures is 18). The manuscript could dwell more on implications and recommendations and limit results to the essential.

11. Is the language fluent and precise?

The language is fluent and, apart from very few occasions, precise.

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?

Mathematical formulae, symbols etc are used correctly.

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?

As outlined in the provided PDF, some parts of the paper could be reduced or combined, e.g. Figures 11/13/14. Others, such as the discussion chapter, could be extended.

14. Are the number and quality of references appropriate?

Considering the research question of this study, I was surprised to see that only a few dune tracking tools were discussed and existing studies on their subjectivity were not references. I would suggest extending the literature research and improving the contextualization of results.

15. Is the amount and quality of supplementary material appropriate?

Currently, the Appendix contains a table of measurement pairs which is needed as context. As outlined in the PDF comments, it could be helpful to be more explicit about the essential differences of these pairs, i.e. mainly the time shift between measurements.

Thanks a lot for this feedback. We have revised the manuscript accordingly and also chose a new title. We have shortened the results section by combining figures and removing certain paragraphs as suggested. Further on, we have added a new subchapter to the discussion section. Here, we interpret the results, relate our findings to other studies and provide recommendations for future studies. We agree that the performance of the MCS and its results are the crucial aspects of the manuscript. Nevertheless, we consider a detailed description of the developed workflow - including already existing methods - important, because a) this is the only way we can show the reader in detail how the analyzed parameters work within the individual methods and the entire workflow and b) we would recommend other researchers to perform similar analyses for other methods and datasets.

The manuscript „A new dunetracking tool to support input parameter selection and uncertainty analyses using a Monte Carlo approach" deals with an interesting and relevant topic and is of interest for both the scientific community and practitioners. The proposed method to determine dune geometry and celerity is based on the combination of two approaches from the literature and is outlined in very much detail. However, it can, in my opinion, still be sharpened. The procedure is then applied to a field data set for which further details should be given. The obtained results are interesting and highlight the need for refined and harmonised methods to enable comparisons with results from other field sites (and hence studies). I do have some more comments on the manuscript (see below) so that I recommend returning the manuscript to the authors for revisions.

Thanks a lot for this feedback. We are pleased that our work is considered interesting and relevant for future studies. We have tried to sharpen and clarify some aspects in terms of methodology. We have also provided more information on the field data as suggested.

Abstract

L1: In my view, this tracking tool is neither needed nor your key result given that it is a mere combination of existing methods. However, the sensitivity analysis you conduct and the resulting quantification of uncertainty can be of great use for future studies. So I suggest focussing on these instead.

We have revised the whole manuscript and changed its title based on this perspective. We agree that the performance of the MCS and its results are the crucial aspects of the manuscript. Nevertheless, we consider a detailed description of the developed workflow - including already existing methods - important, because a) this is the only way we can show the reader in detail how the analyzed parameters work within the individual methods and within the entire workflow and b) we would recommend other researchers to performing similar analyses for other methods and datasets.

L11: How to set the input parameters is typically described in the respective publications. Their influence also depends on methodology ...

Agreed. However, extending this investigation to other methodologies would require considerable additional programming effort, which we would not be able to do in terms of time and resources. The length of the manuscript would also increase significantly. This is why we have selected and implemented the zerocrossing procedure, as it is a widely used and well-established method in bedform analyses and which is implemented in recently published tools. However, we would recommend performing similar analyses for other methods and datasets (see our conclusions). Our findings indicate the possible order of magnitude.

L17: More accurate than what? There are fully automated procedures available, which use a combination of wavelet transform and zero-crossing analyses (e.g. by Wang et al. 2020).

We show that the results from the wavelet analysis provide an orientation to specify the range of values entering the MCS. For values outside this specified range, divergence behavior could be observed for several resulting bedform parameters. So, accuracy and robustness of the procedure can be increased by adding this step. In the revised introduction section (L86 ff), we explicitly point out that a combination of wavelet transform and zerocrossing is used in the algorithm introduced by Wang et al. (2020). Concerning automation see answer to comment L131.

L21: What does this tell about the identified entities? Is this a measuring or a detection sensitivity?

The higher uncertainties found for bedform lengths compared to bedform heights are probably rather detection-related, although measuring sensitivity was not investigated in this study. However, from the measurement data DEMs with a spatial resolution of 10 cm were derived. For the detected bedforms with a length of several meters, a possible measurement uncertainty in length is therefore not significant. In the revised discussion section (L512 ff), we point out that the analysis of uncertainties could be influenced by the specific morphological situation for the test dataset from Lower Rhine.

L22: Not all dune tracking tools consider secondary dunes. What does this tell about the generalization of the investigated sensitivity?

See answer to comment L1 *and* L11.

L20: It could be mentioned that field data are used.

Done. See L20 in revised manuscript.

L23: Just a comment – this result is not really surprising and can be expected.

We analyzed the following three input parameters for the zerocrossing procedure: i) window size for the small-scale bedforms, ii) window size for the large-scale bedforms, iii) zc-threshold. All of them affect the determination of bedform geometries. Varying the window size for the small-scale bedforms during the sensitivity analysis showed the greatest influence on the estimation of bedform characteristics. This is a result that we were not aware of before carrying out the sensitivity analysis.

L27: See my last comment. In this context, since a specific time interval is given, I do recommend to indicate that field data collected at the river Rhine are analysed.

Done. See L20 in revised manuscript.

Introduction

L33: No reference?

We have added two references (L36).

L42: I don't ask for an exhaustive list here, but given the title of this manuscript, I would recommend reviewing as much literature on dune tracking as possible.

We extended the literature review and added more references (L45 ff).

L48: Agreed. But some efforts to standardize these inputs via benchmarking data sets are well documented.

In the revised discussion section, we refer to studies on benchmarking datasets that deal with the influence of the use of different methods and different definitions of bedform attributes on calculation results (L531 ff).

L51: Please outline in brief, what this means.

We have added a short explanation (L58 ff).

L52: But only for one dunetracking approach.

Yes, that's true. However, we have chosen an approach, which is widely used and which is also included in recently published tools. In the revised introduction section, we clarify what the objective was when developing the method and which aspects were particularly important to us (L68 ff).

L65: Highlighted text

Due to our revision of the introduction section this sentence is no longer included.

L67: I am almost certain that this has been done before.

This is true. In the revised introduction section (L86 ff), we explicitly point out that a similar combination is used in the algorithm introduced by Wang et al. (2020).

L69: Highlighted text

We have checked and corrected this throughout the entire manuscript.

L75: So this sensitivity analysis only holds for this specific procedure... I would therefore recommend stressing that not the dune tracking approach is the central novelty, but the sensitivity estimates (for an exemplary approach) as well as the resulting weighting of uncertainty for future analses.

See answer to comment L1.

L43: Please indicate the label 'a' also in the references (L684/L686).

Done.

L44: The reference Henning (2013) is only cited with the title in German. Please provide more information where this was published.

We have added the DOI in the references.

L49: I fully agree and it is good to see that this is explicitly mentioned.

We are pleased that this aspect is considered important.

L57: I recommend deleting "highly".

Due to our revision of the introduction section this sentence is no longer included.

L61: (Guitierrez, 2018) must be (Guitierrez et al., 2018).

Done.

L65: “analog” or “analogue”?

Due to our revision of the introduction section this sentence is no longer included.

L71: But the proposed procedure does not allow for an evaluation of 3D-dunes – correct?

No, there is no module included to explicitly analyze three-dimensionality of bedforms. However, by creating and analyzing bed elevation profiles for different angles, it is theoretically possible to derive statements about three-dimensionality.

Method

L85: How is the longitudinal axis (in a 3D MBES data set) defined?

The longitudinal axis is defined by the user by specifying XY-coordinates. We have added a few additional explanations to the method (L130 ff) and to the dataset section (L349 ff). However, the creation of the longitudinal profiles is not part of the workflow presented here, but more details can be found in Lorenz el al. (2021a).

L107: Please give on sentence what this geometric modeling comprised.

We have added a sentence for clarification (L130 ff). As mentioned above the details can be found in Lorenz et al. (2021a).

L130: How?

We have added a sentence for clarification (L154 ff).

L131: You have a point here, but this statement also undermines your demand for objectivity and automation.

This is correct. However, this is the only manual and subjective step in the entire process, which - at the same time - has the most significant impact on further analyses. This is also described in the recently published meta-analysis by Scheiber et al. (2024), in which the assumption of a second bedform scale was identified as the most significant cause of deviations between the results of different algorithms. In our opinion, a manual plausibility check is crucial here and allows more control for the user by deliberately including or excluding ambiguous cases as shown in the example (e.g. a peak marginally above or below the significance level). Although a full automation could be implemented technically, implausibilities can occur, particularly in such ambiguous cases.

L137: see comment before

See answer to previous comment.

L141: see comment above

Agreed. The decision on performing detrending and its impact on the results could also be included in a sensitivity analysis. However, this would also apply to other aspects such as the influence of hydrographic preprocessing. However, here, the MCS focuses on the direct input parameters of the zerocrossing procedure. We have added a paragraph concerning other sources of uncertainty to the discussion section (L530 ff).

L146: Please use (a) and (b) to distinguish between the two panels. Also, the labels "layer 1" and "layer 2" are somewhat unspecific.

We updated the figure captions by adding a/b labels throughout the entire manuscript. We also replaced the labels “layer 1” and “layer 2” by “small-scale” and “large-scale” bedforms.

L150: Wang et al. 2019 even combine it with a Wavelet transformation, very similar to what is presented here. But they presented a fully automated approach.

See answer to comment L17/ L131.

L156: Which is defined how?

As explained in Sect. 2.6, for zc-threshold also a range of values has to be defined. There is no objective criteria for this, but globally, it must not be greater than the minimum expected bedform height. We have added a sentence to Sect. 2.3 (L189).

L176: Is this from Wesseling & Wilbers (2000) as well? Otherwise, please explain and consider adding this to your figure.

It is based on the publication by Ten Brinkem Wesseling and Wilbers (1999). We have added the reference and moved the corresponding equation from section 2.5 here (L213 ff).

L184: Briefly explain, for which reason/purpose. And why is it called T90 and not H90?

We have added an explanation to the text (L222 ff). T90 measures the accumulated bedforms layer thickness (T) along the entire BEP instead of the height (H) of individual bedforms.

L191: A second sub-chapter with this number?

We have updated this.

L161: Please use descriptors and captions for each panel.

We updated the figure captions by adding a/b labels throughout the entire manuscript.

L199: "in a analysis" or "in analyses" Please check Sg./Pl. throughout the manuscript.

We have checked and corrected this throughout the entire manuscript.

L202: Please elaborate on the window size used in these cross-correlations.

We have added an explanation to the manuscript (L238 ff). In fact, cross-correlation analyses are also part of the MCS. This means, that all the different solutions generated in the zerocrossing procedure (due to different window sizes) are taken into account.

L203: Is this the correct term if you speak of correlations?

We have rephrased the sentence (L247).

L214: What does this mean and how did you determine the threshold of 25%? Is this threshold referring to length AND area as written in brackets or referring to the aforementioned ratio?

The assumption here is, that for longer time offsets between two measurements, the deformation of bedforms increases due to splitting and merging processes. This makes it considerably more difficult to detect the same bedform in two consecutive measurements. This is why we implemented a threshold to exclude those bedforms whose shape has already changed too much between two measurements to enable a reliable detection. The threshold is determined and checked for both length (L(t2)/L(t1)) and area (A(t2)/A(t1)). The value of 25% was chosen to be "on the safe side" (i.e. to exclude one pair of bedforms too many rather than to make a wrong assignment) and resulted from preliminary investigations (not shown here). We have added an explanation to the text (L258 ff).

L222: At this point, this sounds like a part of the Discussion section to me.

Agreed. We have moved the paragraph accordingly (L563 ff).

L228: Please introduce (a) and (b) as suggested above.

We updated the figure captions by adding a/b labels throughout the entire manuscript.

L233: I like this figure a lot. You may consider adding the total result (weighted average) for

Done.

L243: Isn't this the sum of all bedform lengths?

Yes, that's correct.

L270: In my view, this is the essential objective of this study. It should be highlighted in the title.

See answer to comment L1.

L80: Personally, I don't like starting a paragraph with a figure.

In this case, we considered it helpful to visualize the rather complicated workflow directly at the start of the paragraph in order to allow the reader to easily find the figure during the entire reading process, making it easier to jump back and forth. But this is certainly a matter of taste.

L85: Is this procedure also applicable in curved river sections or should it only be applied to straight sections?

Technically, the procedure is also applicable in a curved section. However, due to occurring transverse flow, we would prefer not to apply 2D dunetracking methods in curved sections. Just as 3D-models are used in hydrodynamic numerical modelling to investigate complex flows, three-dimensionality must also be considered when investigating the migration of bedforms under such conditions.

L94: Strictly speaking, an analysis cannot "look at" something – I recommend rephrasing.

Agreed. We rephrased the sentence (L118).

L104: What is the scale of the subsections? Multiple bedform lengths (and if so, how many)?

This is user-definable, as it depends on occurring bedform lengths and also on the research question. In case of the test dataset, we have chosen a length of 100 m. However, this only affects the forming of the resulting parameters (e.g. median bedform heights and lengths) at the end of the workflow and not the construction of the baselines. These are always constructed along the entire length of the BEP.

L105: It is a bit confusing that Figure 6 is already referenced here.

Agreed. We now refer to the relevant section instead (L129).

L106: Can you give some examples what kind of steps are meant here?

E.g. plausibility checks, geometric modelling, generation of BEPs. We have added a few additional explanations to the method and to the dataset section. However, more details can be found in Lorenz et al. (2021a).

L107: What kind of geometric modeling?

We have added a sentence for clarification (L131 ff). As mentioned above the details can be found in Lorenz et al. (2021a).

L108: What about curved river sections (see my comment @L85) – how to generate meaningful BEPs for such cases?

Due to occurring transverse flow, we would prefer not to apply 2D dunetracking methods in curved sections. Just as 3D-models are used in hydrodynamic numerical modelling to investigate complex flows, three-dimensionality must also be considered when investigating the migration of bedforms under such conditions.

However, in case of curved dune fields, dune slip faces can be computed, e.g. by using the algorithm described by Lorenz et al. (2021b), and profile axes can be defined parallel to the dune field orientation.

L118: Please improve the caption of Figure 2, also by adding some more explanations (e.g., labels a/b should be used; which exemplary profile is shown (or is it a created one); why does x-axis start at 390 m; why is the minimum wavelength not limited to 0).

We updated the figure captions by adding a/b labels throughout the entire manuscript. Although the BEPs are all taken from the test dataset, they are not yet used to demonstrate any site-specific results in this context. In the method section we use different excerpts of BEPs in order to illustrate the explained aspects as clear as possible. This is why, x-axis does not start at 0. The minimum wavelength is actually 0.

L128: How are these layers specified in the analysis (i.e., how to get the data for layer 1 / 2 from the total data set)?

The layers are separated by the calculated baseline of the upper layer of bedforms. At the same time, the baseline of the upper layer represents the top of the lower layer. The baselines are calculated by the zerocrossing procedure.

L130: Can you please indicate how outliers were removed?

We have added a sentence for clarification (L154 ff).

L133: Is it possible to define more than two layers? Hoe exactly are these layers determined? Some more guidance would be helpful for the reader.

By making minor adjustments to the algorithm, it would be possible to define a third layer, which would make sense in the case of underlying bars, for example. However, the workflow is currently designed for two layers. We have therefore adapted the corresponding text passages in the manuscript. We have also revised figure 2, in order to clarify how the layers are identified.

L134: Isn't this partly subjective contradicting the statement regarding automatization?

See answer to comment L131.

L135: I am not sure that I can follow this argumentation – I can three lines for the different significance levels and I see a total number of three detected peaks, but is not clear to me for which significance level (and how the different significance levels indicate a different number of layers). Some more explanations would be helpful here for the reader.

We have revised the figure and have added a few sentences for clarification to the text (L160 ff).

L138: See my comment @L134.

See answer to comment L131.

L154: What is the span of the moving average?

The span (= window size) of the moving average depends on the detected wavelengths in the previous step (wavelet analysis). The values are varied within the MCS. So, for each iteration a different window size is selected within the previously defined range.

L157: Why does the lowest baseline separate the bedforms from the immobile bed? What if the bed below would also be moving with a constant speed (I know this is hypothetical, but so is the assumption that the bed below layer 2 is immobile)?

For clarification we replaced the term "immobile bed" by "non-active layer" according to the definition of Kleinhans (2005), in which the active layer is described as "the layer of moving dunes (represented by their average dimension) in which sediment is sorted through the time of dune passage" (see L192). However, we consider it unlikely that there is significant movement of the bed below the lowest baseline, as the actual mobile layer (or the "dynamical active layer" according to the definition of Church and Haschenburger, 2016) is usually described to be only several grains deep with decreasing speed as depth increases (e.g. duBoys, 1897).

L159: "Several" indicates again that more than two layers can be defined (see my comment above).

Agreed. Corresponding to the answer on comment L133 *we replaced "several" by "two".*

Caption Figure 3: "For layer two the procedure is based on baseline 1 instead of the BEP." This could be indicated in the text. It seems that a moving average was used for the layer two analysis (based on baseline 1 – what was the span)? How would one continue if there would be more than two layers?

Yes, the moving average for layer 2 (now referred to as "large-scale") is based on baseline 1. We have added this to the text (L195 ff). The span, again, depends on the detected wavelengths in the previous step (wavelet analysis). In this case the previously defined range for layer 2 is used. As mentioned above, the workflow is currently designed for two layers. For a third layer, however, the procedure would be similar to layer 2. The only difference would be that the moving average would be based on baseline 2 instead of baseline 1.

On another note - which profile is shown in this plot (it seems to be different from the one shown in Figure 2)?

As explained above, in the method section we use different excerpts of BEPs in order to illustrate the explained aspects as clear as possible. If we used the same excerpt for all figures, some details could not be shown properly. Although all BEPs are taken from the test dataset, they are not yet used to demonstrate any site-specific results in this context.

L175: Figure 4: Why using again a different profile than the ones shown in Figure 3?

See answer to comment above.

L176: There is no L_total defined in Figure 4.

L_total is equal to L2. We have added this for clarification.

L191: Check numbering of heading.

Thanks. Done.

L200: See my comment @L65.

We have updated this (L247).

L206: See my comment @L43.

Done.

L214: What is a short measurement interval?

This depends on the morphological and hydrological conditions. For very rapidly migrating smaller bedforms even a measurement interval of only a few hours might be too long in order to track them reliably. Due to the ambiguity, we have rephrased the sentence (L260).

L216: There may be gaps in the data if too many outliers would be present, so that the migration rates may be biased?

Yes, we show this in figure 11: With increasing dt more outliers are identified, which may lead to biased / non-reliable results. We also discuss this aspect in the discussion section (L563 ff).

L225: Does this mean bedforms corresponding to layer 2?

Yes. However, we replaced the labels "layer 1" and "layer 2" by "small-scale" and "large-scale" bedforms.

L228: Figure 5: Please define the data used for this plot (see my comments above) and a/b labels should be used. What was the time interval between the measurements?

The time difference was 0.4 h. We have added this information as well as a/b labels to the figure.

Dataset

L277: Coming back to my previous comment, this - I think - is not the main purpose of this study, esp. because you only combine two existing algorithms.

As mentioned above, we have revised the whole manuscript based on this perspective, which is also reflected in the title now. We have rephrased this sentence accordingly (L316).

L278: Given the study site location, I assume, it's the German Federal ...

Yes. Updated.

L282: Figure 6?

Yes. Updated.

L284: It could be helpful to know the long-term mean discharge.

Agreed. Added to the text (L324).

L286: What measurements? Please be clear about the utilized instrumentation and expectable measurement accuracy.

We have added some information about used equipment and measurement accuracy (L328 ff).

L297: The box in overview (a) (!) looks larger than the actual study area.

We have updated this.

L305: Am I correct that these profiles are not straight but bent? How did you define them?

Yes, that's true. We have added a short explanation to the text (L349 ff).

L311: "To previous" or "to first" measurement?

To previous is correct.

L312: Consider using these descriptors (1-16) in your overview in Fig 6 as well.

Done.

L282: Fig.5 should be Fig. 6? Please be consistent in using Fig. or Figure in brackets (e.g., L306).

Yes. Done.

L288: Please provide more information on the data (especially spatial resolution; what kind of MBES was used and how was it deployed etc.).

We have added a corresponding paragraph to the manuscript (L328 ff).

L300: No. 5-14: are these the detailed measurements in Figure 6? How good was the spatial match of these profiles?

The detailed measurements are MBES-data as well, but they were carried out along a single measurement swath in the center of the bedform field at shorter intervals. To derive the BEPs from the MBES-data identical tracks were used in each case, so there is no spatial offset.

L306: See my comment before – what is the "raw resolution" of the data and what is the area of the MBS-footprint? What kind of interpolation algorithm was used to create the high-resolution grid with a point spacing of 10 cm? I assume that a total of four DEMs were derived (for No. 1-4 in Table 3)? How were the BEPs derived (there is some curvature of the channel)? How accurate were the

We have added a corresponding paragraph to the manuscript (L328 ff).

L306: What is meant by "normalized"? Is it meant that the mean bed elevation in this figure corresponds to zero? However, this value is different from the z-range shown in the previous figures. Please explain. Can you also comment more on the lateral differences in bed elevation profiles?

Yes, we divided the z-values by the mean z-value. We have revised some figures so that this is consistent throughout the manuscript. The lateral differences are 10 m each.

L307: In this context – in my opinion, the largest bedforms are present in BEP 1-4. Please provide some more explanations. Can you also please indicate BEPs 1-16 in Figure 6 (i.e. where is BEP 1 and where is BEP 16 – that will be helpful for the reader).

You are probably referring to the macro structure between Rh-km 860.1 and 860.3. It extends from the head of the groyne on the left bank to the middle of the channel. The length of the structure exceeds that of the longest other bedforms by an order of magnitude. By analyzing the collected MBES-data no migration of the structure could be identified within 48 hours. We therefore assume that this is not a migrating bedform, but that sediments were washed out between the groynes during the previous flood event. The geometry, with a steep stoss side and a flat lee side, is also contrary to the conditions typically found for bedforms. Further on, even if the structure was migrating, the time intervals between the measurements would not be sufficient to analyze its migration.

This is why we decided, not to include it in further analyses. We have added this explanation to the manuscript as well (L353).

We have added the numeration of BEPs to Figure 6.

Results

L323: ?

We rephrased the sentence.

L334: I suggest using inverted commas or italics to highlight the name of your input parameters.

Done.

L336: +/-

Done.

L367: I can tell that I am having a hard time differentiating between these two labels throughout the text. Maybe you can find more descriptive names, such as small-scale vs large-scale, primary vs secondary or underlying vs superimposed.

We have replaced the labels "layer 1" and "layer 2" by "small-scale" and "large-scale" bedforms.

L372: In my view, these results show that there is a larger difference in tracking layer 1 or layer 2, respectively. Giving a median therefore seems pointless.

Figure 10 was slightly misleading in this regard, so we have revised it. The black dots do not represent the median values of layer 1 and 2 but the results for the total BEP (without considering two separate layers). This is also the reason why there is only one result per measurement (black dots). Only if we consider separate layers of bedforms and determine them by means of MCS we obtain a range of results per measurement (red and blue polygons show the respective range of results).

L376: Should readers know, what this means in terms of parameter combinations? You should be more explicit about what distinguishes measurement pairs 1-12 (or 1-17?) and 12-45.

We have revised the relevant figures to clarify which measurement pair corresponds to which time difference.

L393: Why not compare tracking methods 1 and 2 (Figure 14) first?

Agreed. We have revised the results section, accordingly.

L398: To be honest, I don't see the value of this (figure and) part in your analysis.

The intention of this figure was to show the variation of bedform geometry and migration rate of individual bedforms. We agree that this is an aspect that can be omitted in favor of shortening the results section. It is therefore no longer included in the revised version.

L407: Discussion?

Agreed. We have moved the paragraph accordingly (L563 ff).

L421: I would recommend merging Figures 11 and 14 into one illustration, which juxtaposes the performance of both tracking methods. This figure should also highlight the major differences in the 45 measurement pairs, i.e. the dt or at least the dt gap between No 17/18. You may consider doing the same for Fig. 13.

Agreed. We have revised the figures accordingly.

L431: Highlighted text

Done.

L439: "... depending on the methods' input variables". I would stress this a bit more explicitly.

Agreed.

L330: It is interesting to note that the max. wavelength is about 25 m. Note that Lokin et al. (2022) (https://doi.org/10.1029/2021GL097127) identified wavelengths of up to 140 m in the downstream Waal river (although for lower discharges) – nonetheless, for higher flows (Q = 4000 m^3/s) the wavelengths were still > 60m. This is quite a difference that may be explained in the manuscript (I guess Lokin et al. 2022 should also be referenced as it deals with a similar topic – showing the need to adjust the literature review).

Bedform lengths depend on hydrological and morphological conditions and can vary a lot along the Rhine (see e.g. study by Wilbers and Ten Brinke (2003), https://doi.org/10.1046/j.1365-3091.2003.00585.x). We therefore see no contradiction in the fact that different bedform lengths were determined in a different river section with a different grain size distribution (in case of the mentioned dataset from river Waal, the river bed has a higher sand content).

In any case, the test dataset is used to conduct a sensitivity analysis and to quantify resulting uncertainty, which is the main objective of this study. It could also be taken from any other river, if a comparable dataset (in terms of spatial and temporal resolution) is available there. A field study to characterize and compare occurring bedform dimensions along the Rhine is not the aim of this manuscript.

L336: I am not sure that the “sigma”-sign was appropriately defined in the text.

Agreed. We now introduce it in the method section (L155).

L344: Why reference to Figure 8 showing wavelengts? Should be Figure 9?

Agreed.

L349: If this statement refers to Figure 9 (what I assume), I cannot see these 10 m.

You are right. It is BEP 14. We have updated this.

L394: Please give a reference for Lowess smoothing.

We have removed this paragraph from the manuscript in favor of shortening the results section.

L431: A closing bracket is missing.

Done.

Discussion

L446: I would again stress that this is the range of UNCERTAINTY under consideration of different input parameters in your MCS, which I see as an essential result.

See answer to comment L1.

L449: Although this is a very interesting aspect, the section is currently limited to a mere description similar to the previous Results chapter.

We have added a subchapter containing an interpretation and contextualization of the key findings of the results section.

L464: Can you explain, why?

We have added an explanation to the text (603 ff): The reason for this is that the T90-parameter is independent of the number of identified bedforms. It is not based on measuring individual bedform heights but on measuring the accumulated bedforms layer thickness (T) in every x-position along the entire BEP. So, always the same number of input values are used for the calculation.

L468: Why?

It can be assumed that convergence will eventually occur with increasing values, as the calculated moving average value successively approaches a horizontal line. We have added this explanation to the text (L610).

L484: Please quantify.

Done (L626 ff).

L491: If you give this recommendation here, can you please specify the general character of this uncertainty (e.g. order of magnitude, random vs systematic error etc) based on your findings? Or is it your recommendation that every future study includes your methdology to quantify uncertainties?

We would recommend to investigate the sensitivity of input parameters for other methods as well. The shown results provide an indication of the possible order of magnitude. We therefore quantified the found uncertainties. In the revised discussion section (L512 ff), we also point out that the analysis of uncertainties could be influenced by the specific morphological situation for the test dataset from Lower Rhine. We therefore recommend to continue the analysis by using datasets with different morphological conditions in order to validate the findings.

L547: Highlighted text

Done.

L551: Highlighted text

Done.

L565: More generally speaking, what is the order of uncertainty inherent to alternative bedload transport estimations?

We wanted to point out that there are other sources of uncertainty that need to be taken into account for a comprehensive estimation. However, our focus was on the influence of the input parameter settings. There are multiple other sources of uncertainty when determining bedform geometries (L530 ff). When calculating the bedload transport, even more are added. We have tried to clarify this aspect in the text (L575 ff). A comparison of uncertainties between dunetracking and direct bedload sampling is addressed in another publication, which is currently in preparation.

L459: “sizes” should be “size”

Done.

L460: Are BEPs 8-14 meant?

Yes. We have clarified this in the text (L598).

L462: Do you mean it is sensitive for window-sizes < 5 m?

Yes, exactly.

L464: Is this because individual high bedforms are not as significant in the “T” analysis compared to the “H”-analysis?

Yes. We have clarified this in the text (603 ff): The reason for this is that the T90-parameter is independent of the number of identified bedforms. It is not based on measuring individual bedform heights but on measuring the accumulated bedforms layer thickness (T) in every x-position along the entire BEP. So, always the same number of input values are used for the calculation.

L475: Are there other studies available that have reported bedform geometries in the Rhine river (in regions nearby) that can be used to validate the findings (see, e.g. Lokin 2022).

For determining bedform geometries we use a method (zerocrossing), which is widely used. However, in the revised manuscript we have added a short discussion about the calculated geometries, the possible influence of the specific hydrological conditions during the measurements and relations to other studies (L494 ff).

In any case, the test dataset is used to conduct a sensitivity analysis and to quantify resulting uncertainty, which is the main objective of this study. It could also be taken from any other river, if a comparable dataset (in terms of spatial and temporal resolution) is available there. A field study to characterize and compare occurring bedform dimensions along the Rhine is not the aim of this manuscript.

L499: See my comment @L475.

See answer above.

L500: Figure 16 should be Figure 17.

Done.

L534: Figure 18: Are there bed-load measurements available to verify the determined total transport rates?

Yes, there actually are bedload measurements available from a station nearby (2.5 km upstream) for even similar discharge (the samples were taken three days earlier). Median values are in the same order of magnitude (about 170 g/sm for direct sampling and about 120 g/sm for dunetracking). However, we have decided not to include this aspect, in order to focus on the developed workflow and the obtained uncertainties due to input parameter selection. There is another publication in preparation dealing with the comparability of dunetracking and direct sampling, in which this aspect will be discussed in detail.

L538: Section 5.2: In my opinion, this paragraph is not really a discussion. It is rather a summary and repetition of the contents described before.

We have revised the discussion section and added a subchapter including an interpretation and contextualization of the key findings of the results section.

L547: delete “shortly”.

Done.

L551: “repeat” should be “repeated”.

Done.

Conclusions

L588: I think, you call them secondary bedforms for the first time here. It should also be highlighted what influence the time interval has on larger primary dunes.

Agreed. We have replaced ‘secondary’ by ‘small-scale’. We have also added the influence the time interval has on large-scale bedforms to the text (L708).

L591: For readers who just read your conclusions, this term might be irritating.

We replaced the labels “layer 1” and “layer 2” by “small-scale” and “large-scale” bedforms.

L597: So, every morphometric analysis in the future should conduct your MCS approach?

We have specified this conclusion (L720 ff): Overall, it was shown that varying input parameter settings can have a large influence on the determination of bedform parameters. But at the same time, we have introduced a workflow that can provide proof of robust estimates of these parameters. We therefore recommend carrying out similar investigations for other bedform analysis methods and datasets, in order to assess the robustness of derived results. However, in this study we focused on the uncertainties resulting from varying input parameter settings. There are multiple sources of uncertainties in bedform analysis like the choice of a method or tool in general or the geometric definitions of bedform attributes. All these uncertainties must be considered together in field studies characterizing prevailing bedform conditions derived from measurement data, in order to obtain the full picture.