Response to reviewers

We thank the reviewer for their constructive criticisms and suggestions. We have taken these on board to improve the manuscript. We hope the paper is now acceptable to both reviewers.

In the following, we respond to each of the reviewer's comments in turn. Reviewer's comments are in Italic font, with our response both indented and in Roman font.

Overall, this work represents an interesting development in the field of Earth Surface dynamics and, in particular, the study of bedform interactions. The finding that dune interactions are probabilistic represents a novel result and may come as something of a surprise to many researchers in the community. I welcome this finding and believe that this work represents a sufficient advancement to warrant publication. The authors should also be commended for the style of the report which is generally well written, organised, and presented.

I do, however, have some general problems with the study. These issues centre primarily around the limitations of 2-dimensional regimes and the question of how well these results scale to realistic 3-dimensional systems. More specifically, I find the attention paid to exactly determining the form of the observed stochasticity rather irrelevant since any real-world system will behave differently because of the increased dimensionality. Additionally, the comparison with the experiments presented in Jarvis et al. (2022) is perhaps an odd choice since, as the authors point out, those experiments involved a "train" of interacting dunes rather than the binary system considered in the simulations presented in this work. Although the authors mitigate some of the issues this might present by considering only interactions where there was no physical overlap with the additional bedforms, there may be wake effects such as those identified in Bacik et al. (2020) which are not accounted for in this work. I believe that the discussion of the limitations of this work should be made more detailed as, in my opinion, these problems are more substantial than they are made to seem in the current manuscript. Nevertheless, the essential finding that collisions are stochastic rather than deterministic is an important one and the authors should be congratulated for their work.

> We thank the reviewer for this comment. We agree that there are limitations to this work (as there is to all studies) and have made an effort to be more detailed in this, as described in our response to specific comments below. Regarding the specific wake effects described by Bacik et al. (2020), the author is correct that these are not accounted for in the numerical model. Indeed, they cannot be reproduced due to the pure 2D geometry of the numerical domain. We now try to better acknowledge this throughout the paper (see our responses to specific comments below for exact line numbers).

I will now provide some more specific comments.

Specific Comments:

Line 25 - "pattern coarsening, whereby a larger number of smaller dunes transition to become a smaller number of larger dunes…"

Coarsening has been observed in many experimental and numerical studies of dune dynamics. However, many natural dune systems have been shown to be homogeneous rather than coarsening. This point should be explicitly made here.

We have now added the line "This coarsening process occurs during the development of a dune field from a flat bed whereas more mature dune fields can attain a steady state, where their mean wavelength and amplitude remain relatively constant." Lines 28-30.

Line 41-42 -"*Turbulence, however, is an inherently 3D phenomenon…*"

Interesting claim given that the water tank experiments of Bacik et al. (2020) reported that induced turbulence led to repulsion even in their quasi-2D setup. Furthermore, the claim that only the size ratio controls the collision outcome is likely only true for sufficiently large bedforms where the assumption of scale invariance applies.

We now clarify that the size ratio is the only controlling parameter if the dunes are sufficiently large to be scale invariant (Lines 46-47). We also explicitly state that the dunes in our simulations are scale invariant (Lines 291-292). Regarding the turbulence, we note that the quasi-2D experiments of Bacik et al. (2020) are inherently 3D. Thus, turbulence can be, and is, present in those experiments, as we now note on Line 45. Our simulations, however, are pure 2D. Consequently, vortices, such as the recirculation zone behind the crest of dunes, cannot decay into a 3D turbulent field.

Line 59 - "However this means we can exhaustively…"

The merit of such an exhaustive study is severely limited however by the fact that 3D and 2D systems are inherently different. Although you may be able to fully understand the problem in 2D, one must recognise that this is still a toy model and that a study in 3D is going to have more real-world impact even if it cannot be quite as exhaustive.

We fully agree that we are not capturing phenomena that will be present in a 3D system. However, 3D systems contain so much complexity, the fundamental result which we observe here, namely the fact dune collisions can be modelled probabilistically, would be much harder to identify in the wider parameter space. To address this, we have added the sentence "Although our study is strictly only valid for 2D systems, our results should motivate further research to test if the fundamental characteristics of collisions that we observe can be preserved in 3D environments, where turbulence and flow perturbations also have an influence." (Lines 69-71)

Line 60 -"By performing large numbers…"

This makes it sound as though thousands of simulations have been performed but we later find out that it was only ~50 and that, in fact, the number of simulations that could be performed was a limiting factor on the uncertainty of the findings.

To be clear, we have performed many more than 50 simulations. In fact, each data point in Fig. 3D represents 50 simulations alone. Fig. 3D contains 32 data points. Thus, we actually performed $32 \times 50 = 1600$ simulations. We now state clearly that we performed a large number of simulations (line 65) as well as re-emphasise this in the results (line 198).

Line 126 -"Two distinct types of coalescence…"

This phrasing makes it sound like they are very different processes, however the authors themselves describe that the intermediate stages are only "slightly different" and later (line 151) state explicitly that "...close to the transition it is very difficult to distinguish the two types of behaviour". This suggests that the types of coalescence are not really "distinct" as claimed here but two regimes of a single coalescence process between which there exists a continuous transition.

Our results show that the two types of coalescence, upstream-dominant and downstream-dominant, are distinct. This is most clearly demonstrated in Fig. 1. It can be shown that, for downstream-dominant coalescence, the sediment from the upstream dune (red) is preserved in a slip slope-parallel layer in the final dune whereas, for upstream-dominant coalescence, the red sediment remains preserved in a mixed region above an unmixed basal layer preserved from the downstream dune (blue). These very distinct morphologies clearly demonstrate that the two types of coalescence are distinct processes and not end-members of a single phenomenon. We have now reworded the manuscript to make this clear (lines 153 and 189).

Line 149 -"...so many simulations would be required to gain meaningful outcomes."

But the authors claimed in the introduction that they had performed "large numbers" of simulations and were able to "exhaustively study" the phenomena. This is a direct contradiction and makes it seem as though the claims made in the introduction were unwarranted.

The line in question refers specifically to the transition between downstream- and upstream-dominant coalescence. The key focus of our manuscript is to quantify the transition between coalescence and ejection. We have now edited the manuscript to make this clearer (line 60).

Line 160 -"...creating multiple small bedforms"

Would it not make sense to define the cases where different numbers of bedforms were generated as different types of collisions, particularly as in 3D it may be possible for these new bedforms to escape from between the dunes? This would also be more consistent with the distinction between the types of coalescence identified by the authors for which the intermediary stages were key.

We prefer to maintain the current classification, whereby the events which result in more than two bedforms being created are classified as ejection. This is because, ultimately, the underlying mechanism is the same regardless of whether one or more downstream dunes are ejected. However, in order to quantify the effect of initial size ratio on the number of dunes created would require many more simulations than we currently perform, since there would be many more possible outcomes. We have now added some text to expand on this (lines 170-173).

Line 188 -"Performing further simulations…"

Again, this ought to be mentioned earlier as the introduction makes it seem that the study was not restricted in this manner.

Please see our above responses to your comments on the number of simulations, and our focus on quantifying the coalescence-ejection transition

Line 194 -"finding $a = 14 \pm 2$ and $b = 0.509 \pm 0.005$."

Given that the results of 2D experiments are not likely to be fully scalable to 3D I do not believe that exact determinations of these constants are particularly relevant to real-world systems. As such, I think these values could easily be removed to an appendix.

We agree that these values are specific to the 2D case and, given the larger dimensionality of the parameter space and the greater number of possible outcomes, are not directly applicable to 3D systems. However, these results are a key result for the 2D system we study here. Additionally, we rely on these values to perfom the comparison to experiments in Section 4. We therefore chose to include them. However, we now acknowledge in the manuscript that these fitted values are not appropriate for 3D systems (lines 208-209).

Figure 3 caption -“...is due to only 50 simulations…”

Same point made previously, this is a major shift in tone from the introduction.

Please see our responses to the above comments. In particular, we note that each data point in Figure 3d corresponds to 50 simulations and we have actually performed 1600 simulations in total.

Line 204 -“...only simulated interactions between two discrete dunes… train of interacting dunes”

I think this is a more important caveat than the authors make it seem. Other similar studies (e.g. Bacik et al. (2020) have found that in these systems wake induced turbulence plays a critical role. The turbulence generated by multiple interacting bedforms in these experiments is likely to be greatly affecting the outcomes. Whereas, this is not the case in the simulations where only two dunes were present.

We agree with the reviewer that this is an important caveat and we do not wish to underplay this in the manuscript. As we already state in the manuscript (lines 226-228), we attempt to mitigate this by selecting interactions a) between discrete dunes and b) during times when there was no physical overlap with neighbouring dunes. However, we now explicitly acknowledge that the presence of a dune train will lead to fluid-transmitted interactions which we cannot remove (lines 228-229).

Line 205 -“...the number of simulations… is relatively small”

Same point made previously!

Please see our our above responses on this point.

Line 209 -“However, additional simulations…”

If these simulations have been performed already and the authors wish to compare with the experimental results of Jarvis et al. (2022) then why not simply present the results from the experiments where θ = 18â¦ rather than those where θ = 35â¦ ?

The text in the previous version of the manuscript was erroneous and misleading. Although we have performed some simulations with a leeside slope angle of θ = 18 º, they are only indicative, showing that ejection and coalescence occur as observed in the simulations where θ = 35 º. We have now clarified in the manuscript this caveat (lines 224-226).

Line 225 -“There is reasonable agreement…”

I would like to see a table containing these data and ideally some statistical tests about whether the experimental observations are statistically similar to the stochastic rules defined in this work.

Following Greenhill et al. (2011) and Davidson-Pilon (2015) (https://nbviewer.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter2_MorePyMC/Ch2_MorePyMC_PyMC3.ipynb) we now present a separation plot comparing the predicted and observed number of ejection events. We chose to use a visual comparison rather than a quantitative statistical test since, such tests often rely on arbitrary thresholds, e.g., the expected Percentage of Correct Predictions (Heron, 1999), or lack statistical interpretations, e.g., Brier Scores (Brier, 1950).

As can be seen in our new Fig. 4D, and as described in the text (lines XX – XX), the separation plot shows that the model suggested in equation 5 does a reasonably good job of predicting the experimental data, since the majority of ejection events occur on the right hand side of the plot.

Finally, as suggested by the reviewer, we have now inlcuded a table (Table 1) detailing the dune collisions selected from the experiments and the modelled probability of ejection for each event.

Line 253 -"...agree well with numerical determined…"

Again I would like to see these data.

Please see our response to the above comment.

References (if not in paper)

Brier, G. W. (1950). Verification of Forecasts Expressed in Terms of Probabilities. *Bull. Am. Meteorol. Soc.*, 78, 1-3.

Herron, M. C. (1999). Postestimation Uncertainty in Limited Dependent Variable Models. *Polit. Anal.*, 8(1), 83-98.