Dear Editor and Referee #1,

We thank you for the time spent to evaluate and review our manuscript and for the critical and constructive comments. The major and minor comments brought by Referee #1 were all addressed in a revised manuscript.

Here, we will repeat the reviewer's comments (italic font) and response directly below (standard font). After each response, the associated changes applied to the revised manuscript are included in a table box (coloured in red).

Sincerely,
G. Kesserwani and M. Shirvani

General comments

"The value of Agent-Based Modelling as a technique to understand the role of individual decision making in response to hazards, such as flooding, has increased in recent years, since the initial publication of ABM evacuation behaviour was published in 2011. This study is an interesting example of further development of this application of ABM and demonstrates a good synthesis of empirical research to formulate agent behaviour. Through introducing agent characteristics such as age, gender, weight, variable movement speeds and stability patterns, the model introduces an enhanced level of heterogeneity amongst the agent population from previous iterations of the model."

Major comments

"However, there lies two fundamental concerns related to the outputs of the model:"

The two fundamental concerns have been insightful for us to carry out the revision work necessary to ascertain the credibility of the model outputs based on which the results were analysed. The minor/technical comments raised were also found valuable to sharpen the quality of the manuscript. A point-by-point reply to each comment is provided below it, together with an explanation of the associated changes made to the manuscript.

"1. It would appear that uncertainty inherent within modelling stochastic agent behaviour is not acknowledged. The impression given is that model simulations were ran only once for each scenario (lines 331 – 333) and outputs were then analysed. There should be evidence of multiple simulations for each test scenario to account for variance in behaviour of agents. For example, confidence intervals and averaging over an appropriate number of simulations would provide a more representative set of outputs that account for stochasticity."

The reviewer is correct about this important point. As the outputs analysed in the initially submitted preprint were extracted from a single run, they do not account for the uncertainty inherent within stochastic agent behaviour. Therefore, the preprint has been revised to: first, acknowledge this uncertainty; then, to explore the sample size needed for a series of probabilistic runs to reach statistically significant outcomes for a range of confidence levels; and, finally, reproduce new outcomes averaged out from the probabilistic runs for the two case studies explored and each of their test scenario.

Section 2.3.1 was therefore extended to:
...As the motion of each pedestrian agent is governed by a stochastic (space-time) process, series of 10 and 20 simulation runs were conducted to average out a plausible outcome for each of the configuration Modes. The plausibility of the average outputs from both series of runs is evaluated, by estimating the margin of error (*MOE)* assuming confidence levels ranging between 90% and 99.9%. The following formula is used to evaluate the *MOE*: $$MOE = Z_{score} \times \sqrt{\frac{\sigma^2}{n}}, \quad (3)$$ where, Z_{score} is the critical value, which is equal to 1.65, 1.96, 2.17, 2.58 and 3.29, for confidence levels of 90 %, 95 %, 97 %, 99 % and 99.9 %, respectively (Hazra, 2017); σ is the standard deviation from the sample of outputs of size $n = \{10, 20\}$; and $\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$, with x_i representing the number of pedestrians with a particular HR-related flood risk or stability state extracted from the recorded outputs, and $\bar{x}$is the averaged value. Table 7 lists the maximum *MOE* evaluated for the different confidence levels, with respect to the average number of pedestrian agents under different HR-related flood risk and stability states for configuration Mode 0 to Mode 4.

Table 7: Maximum margin of error (*MOE*) for the average number of pedestrian agents with different HR-related flood risk or stability states that are extracted from the recorded outputs of all the configuration modes (Table 6) and across different confidence levels ranging from 90 % to 99.9%. Different ranges of the evaluated maximum *MOE* are highlighted with different colour shades: green, orange and red to indicate $MOE \leq \pm 5$, $6 \leq MOE \leq 9$ and $MOE \geq 10$, respectively.

Mode	HR-related flood risk and stability states	Maximum *MOE*									
		$n = 10$					$n = 20$				
		90 %	95 %	97 %	99 %	99.9 %	90 %	95 %	97 %	99 %	99.9 %
0	HR < 0.75	± 5	± 6	± 6	± 8	± 10	± 3	± 4	± 4	± 5	± 6
	0.75 < HR <1.5	± 4	± 5	± 6	± 7	± 9	± 3	± 3	± 4	± 4	± 5
	1.5 < HR <2.5	± 3	± 3	± 4	± 5	± 6	± 2	± 3	± 3	± 4	± 5
	HR > 2.5	± 1	± 1	± 1	± 2	± 2	± 1	± 1	± 1	± 1	± 1
	Toppling-only	± 5	± 6	± 7	± 8	± 10	± 4	± 4	± 5	± 6	± 8
	Toppling-and-sliding	± 4	± 5	± 5	± 6	± 8	± 3	± 4	± 4	± 5	± 9
1	HR < 0.75	± 6	± 7	± 8	± 9	± 12	± 4	± 4	± 5	± 6	± 7
	0.75 < HR <1.5	± 6	± 7	± 8	± 10	± 12	± 4	± 4	± 5	± 6	± 7
	1.5 < HR <2.5	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1
	HR > 2.5	± 0	± 0	± 0	± 0	± 1	± 0	± 0	± 0	± 0	± 0
	Toppling-only	± 6	± 8	± 8	± 10	± 13	± 4	± 4	± 5	± 6	± 7
	Toppling-and-sliding	± 5	± 6	± 7	± 8	± 10	± 3	± 4	± 5	± 5	± 7
2	HR < 0.75	± 6	± 7	± 7	± 9	± 11	± 4	± 5	± 5	± 6	± 8
	0.75 < HR <1.5	± 7	± 8	± 9	± 10	± 13	± 5	± 6	± 6	± 7	± 9
	1.5 < HR <2.5	± 1	± 1	± 1	± 2	± 2	± 1	± 1	± 1	± 1	± 1
	HR > 2.5	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1	± 1
	Toppling-only	± 6	± 7	± 8	± 9	± 12	± 4	± 4	± 5	± 6	± 7
	Toppling-and-sliding	± 6	± 7	± 7	± 9	± 11	± 4	± 5	± 5	± 6	± 8
3	HR < 0.75	± 6	± 7	± 8	± 9	± 12	± 4	± 4	± 5	± 6	± 7
	0.75 < HR <1.5	± 6	± 7	± 8	± 9	± 12	± 4	± 5	± 5	± 6	± 8
	1.5 < HR <2.5	± 1	± 1	± 1	± 1	± 1	± 0	± 1	± 1	± 1	± 1
	HR > 2.5	± 0	± 0	± 0	± 0	± 0	± 0	± 0	± 0	± 0	± 0
	Toppling-only	± 6	± 7	± 8	± 9	± 12	± 4	± 5	± 5	± 6	± 8
	Toppling-and-sliding	± 6	± 7	± 8	± 9	± 12	± 4	± 4	± 5	± 6	± 7
4	HR < 0.75	± 5	± 6	± 7	± 9	± 11	± 4	± 4	± 5	± 6	± 7
	0.75 < HR <1.5	± 7	± 9	± 10	± 12	± 15	± 5	± 6	± 6	± 7	± 10
	1.5 < HR <2.5	± 1	± 1	± 1	± 1	± 2	± 1	± 1	± 1	± 1	± 1
	HR > 2.5	± 1	± 1	± 1	± 1	± 1	± 0	± 0	± 1	± 1	± 1
	Toppling-only	± 6	± 7	± 8	± 9	± 12	± 4	± 4	± 5	± 6	± 7
	Toppling-and-sliding	± 6	± 7	± 8	± 9	± 12	± 4	± 5	± 5	± 6	± 7

For $n = 10$, there is a considerable increase in the maximum *MOE* with Mode 1 to Mode 4 compared to Mode 0. This is particularly seen for the number of pedestrian agents in low and medium flood risk states (HR < 0.75 and 0.75 < HR <1.5, respectively) and with toppling-only and toppling-and-sliding stability states. This suggests that the more sophisticated the pedestrian agent characteristics and rules, more discrepancies would appear in the simulator's outcomes. The maximum *MOE* identified suggests a deviation of around ± 15 from the averaged outcomes. However, when the sample size is increased to $n = 20$, the maximum margin of error does not exceed ± 10 for all the modes and confidence levels. Therefore, the simulation results analysed next are averaged out from a sample of 20 simulation runs, subject to ± 10 maximum *MOE* for a population of 1000 pedestrians in the flooded walkable area, which corresponds to 1 % per 1000 pedestrian population.

With the changes above, Sect. 2.3.2 has been revised too in order to incorporate new figures and accommodate their respective changes in the text.

Section 3.2.3 has also been revised to confirm the appropriateness of using a series of 20 simulation runs to average out outputs for the real-case study:

3.2.3 Simulation runs

Series of 20 simulations were run under configuration Mode 2, which was deemed suited for the scope of this case study (Sect. 2.3.2). Three series were performed associated with the 20 %, 30 % and 40 % thresholds for the 'autonomous change of direction' condition, respectively (visualisation of a simulation can be found in the video supplement in Shirvani (2021)). Outputs averaged from each series of simulations included spatial and temporal information, at each time step, about the pedestrian agents as they evacuate ($t > 0$ min). The averaged outputs include the position, HR-related flood risk state, stability state (with a toppling-only condition and a toppling-and-sliding condition), and the choice for the destination selected by the pedestrian agents during the evacuation process. Considering the uncertainties associated with the motion of pedestrian agents, the plausibility of the averaged outputs relevant to the number of pedestrian agents was evaluated. The evaluation was based on the *MOE*, using Eq. (3), for 99.9 % confidence level only, informed by the results of the analysis in Sec. 2.3.1. Table 8 shows the maximum *MOE*s found for the number of pedestrians predicted to be in the considered HR-related flood risk states and the stability states, obtained from the series of runs using each of the 20 %, 30 % and 40 % threshold, respectively. It can be seen that the maximum *MOE* increases as the risk perception level decreases, suggesting a notable increase in uncertainty from the incorporation of the perception component on the pedestrian behaviours.

Table 8: Maximum margin of error (*MOE*) for the average number of pedestrian agents with different HR-related flood risk or stability states that are extracted from the recorded outputs throughout the simulations for each 20 %, 30 % and 40 % threshold.

HR-related flood risk and stability states	Maximum *MOE*		
	20 % threshold	30 % threshold	40 % threshold
HR < 0.75	± 16	± 16	± 19
0.75 < HR <1.5	± 2	± 8	± 15
HR > 1.5	± 0	± 1	± 2
Toppling-only	± 2	± 5	± 13
Toppling-and-sliding	± 1	± 4	± 7

Next, the averaged outputs are analysed for each of the 20 %, 30 % and 40 % thresholds, considering the popularity of the destination selected by the pedestrian agents (among south, east and north) together with their flood risk state and stability state.

With the changes above, Sect. 3.3 have been revised too in order to incorporate new figures and accommodate their respective changes in the text.

"2. The number of agents included in the Hillsborough case study is 4,080. The lowest recorded attendance for Sheffield Wednesday in 2019 was counted at 21,485, meaning that the agent population is only 18% of what is considered 'real'. Given that the start of the paper states that the FLAMEGPU platform can handle "as large population size as needed (line 79)" and that throughout the paper congestion is frequently stated as a factor that influences flow dynamics and evacuation time, a question emerges as to why such a significantly low agent population has been adopted. The paper states that "using a bigger population size would lead to extreme pedestrian congestion that impacts the movements of individual pedestrians (lines 489 – 490)".Surely, in the context of a football match, this is an important factor that must be represented as accurately as possible. There is no justification for the chosen figure of 4,080 and it seems arbitrary. I am not wholly convinced that subsequent outputs reflect evacuation times that would be representative of a football match. Either a more realistic agent population in line with actual attendance rates is necessary, or justification that the emergent behaviour is not dramatically impacted by the chosen number of spectators."

The reviewer raised fair points regarding the lack of justification of the selected 4,080 agent population and of how effects of pedestrian congestion on the evacuation times were handled. In terms of agent population, the average recorded attendance should be around 24,000 in normal weather conditions. As the case study assumes severe weather conditions and flood warnings were issued before the event, it makes sense to consider around 20% of the expected spectators; this is also representative of the percentage of people who would ignore the warnings and attend the match (Fielding et al., 2007).

As the statement: *"using a bigger population size would lead to extreme pedestrian congestion that impacts the movements of individual pedestrians"* was a major oversight, it has been removed from the manuscript. Instead, new clarifications are provided on how a dischach rate was implemented to the simulator for this case study to comply with a safe evacuation measure, in other words ensure that *"the emergent behaviour is not dramatically impacted by the chosen number of spectators"*.

End of Sect. 3.1 has been extended to more clearly justify the reduced number of spectator selected and to also justify the outcomes of the simulator won't be significantly impacted for a higher population size.
In this study, it is assumed that the site in Fig. 7 is hit by a flood, similar to the one that had happened in 2019, during a football match where the spectators are caught unaware of the rainfall accumulation around the stadium. ... The evacuated spectators gradually enter the walkable area where they get in direct contact with the propagating floodwater along their ways to any of the south, east or north destinations. A population of 4,080 spectators was assumed, which is lower than normal due to the severe weather condition and flood warnings issued prior to the event. This population is around 20 % of the spectators expected, and represents the relative number of people who would ignore the warnings and attend the match (Fielding et al., 2007). For this case study, a dispatch measure was introduced to the simulator to release the evacuees into the walkable area during the flooding. The dispatch measure limits the influx rate to person-per-second per width unit to comply with guidance methods for controlling the density of large crowds outside the stadiums for safe evacuation (Minegishi and Takeichi, 2018; Still, 2019). For a gate that is around 4 m wide, four pedestrians per second are dispatched from the stadium to the walkable area. Using the simulator with this dispatch rate limits the overall number of pedestrians that would be present in the walkable area at a time. Therefore, running the simulator to analyse the evacuation of a larger number of spectators is expected to lead to similar evacuation patterns and flood risk trends that would be prolonged over a larger evacuation time. The flood-pedestrian simulator is applied to simulate this scenario for analysing the pedestrian evacuation patterns including their preference for the destination during flood evacuation, by activating the 'autonomous change of direction' condition (Sect. 2.2.3).

Technical/Minor Comments:

- *"In section 2.3.2, it states "...the more crowding of pedestrians the more energy loss in the floodwater dynamics for low risk floodwaters, which in turn enables the pedestrians located behind to take a faster moving speed (lines 346 – 348)".Whilst it is well noted that crowding would disperse the floodwater, making pedestrian movement easier, I can't help but wonder if collective moving speed would in turn be slowed down by the congestion itself, therefore this could be acknowledged."*

While revising Sect. 2.3.2, the above-raised aspect has been addressed to acknowledge the influence of pedestrian congestions in the floodwater on the collective moving speed. Our findings suggest that, with the two-way interaction condition, crowding of pedestrians disperse the floodwater, in turn allowing more spreading of the moving pedestrians and those in the front to move faster (see revised Fig. 6 below). However, the collective moving speed does not exhibit any significant difference due to congestion, as can also be noted in the overall trends of HR-related flood risk states of pedestrians (addressed in the discussion of the revised Fig. 4, not shown in this letter).

The discussion of Fig. 4a in Sect. 2.3.2 was revised to acknowledge this aspect:
Figure 4 shows the trends in the number of evacuating pedestrians with different HR-related flood risk states predicted by the simulator after 20 runs using all the configuration modes (Table 6). Figure 4a represents how the number of pedestrians with a low flood risk state (HR < 0.75) change during 20 minutes of flood time. Figure 4a-left includes the trends predicted after enabling the walking condition for the age-related moving speeds (Mode 1) versus those predicted by further enabling the two-way interaction condition (Mode 2). In Mode 1, the trend is in good

agreement with the baseline predictions (Mode 0, with non-age related moving speeds) at flooding times when there are less than 100 pedestrians in the walkable area with a low flood risk state, during 3.5 min to 7 min. A considerable difference among the predictions starts to appear when more than 150 pedestrians are present, around 2.5 min and 8.5 min. This difference seems to impact the overall trend, suggesting a 6 min longer duration with a higher number of pedestrians being predicted to be under this flood risk state, during 8 min to 18 min. In Mode 2, compared to Mode 1, the number of evacuating pedestrians is seen to reduce further at flooding times involving more than 150 pedestrians, around 2.5 min and 10 min. This is expected as crowding of pedestrians in low risk floodwaters is expected to disperse the floodwater dynamics, which in turn help pedestrians evacuating ahead to pick up faster moving speed (Shirvani et al., 2021). This does not seem to influence the collective moving speed of pedestrians, for example by generating additional congestions (as shown later in Fig. 6), as the overall flood risk trends with Mode 1 and Mode 2 are very close. Figure 4a-right contrasts the trends predicted after activating the running condition for the age-related moving speeds (Mode 3) to those predicted by also enabling the two-way interaction condition (Mode 4). In Mode 3 and Mode 4, the trends show a considerably faster moving speed of pedestrians (than with Mode 1 and Mode 2), significantly reducing the duration when pedestrians fall under a low risk state, suggesting a flood risk trend that is close to the baseline predictions (Mode 0). With Mode 3, discrepancies (compared with Mode 0) only occur between 2.5 and 3.5 min and after 8 min of flooding, when there are more than 150 pedestrians moving under the running condition. In Mode 4, with further enabling the two-way interaction condition, the trends remain close to those predicted under Mode 3, except at 2.5 min flooding time that involves more than 200 pedestrians under low flood risk state. This suggests that activating the two-way interaction condition with the running condition may only temporarily influence the pedestrians' collective moving speed, namely when more than 200 pedestrians are caught under a low flood risk state. Overall, there is a major difference in the collective moving speeds of pedestrians when age-related walking vs. running speeds are deployed, leading to prolonged vs. shortened evacuation times compared to the baseline predictions (Mode 0), respectively. Also, using the two-way interaction condition seems to be a sensible choice for simulating mass pedestrian evacuations in low risk floodwater.

- *"In section 2.3.2, it also states "enabling it with the two-way condition (Mode 2) increases slightly the evacuation time as crowding is more likely under the walking condition (lines 435 – 436)". It would be reasonable to assume that the 'running condition' would be more likely to cause crowding as it represents a more 'excited' response, causing further bottlenecking in the evacuation process compared to more organised walking agents. Some consideration of this would be insightful."*

As shown in the revised Fig. 6, the reasonable assumption raised above is retrieved by the simulator when running with the two-way interaction condition and contrasting pedestrian distribution at 6 min when largest population of pedestrians are at risk of toppling and the medium risk floodwater (0.75 < HR < 1.5) affect the majority of the population.

New Fig. 6 has been produced and discussed in Sect. 2.3.2 to include insights on this aspect:

Figure 6 compares the spatial distributions of the evacuating pedestrians over flood HR map at flood time 6 min, obtained from simulator runs under Mode 1 to Mode 4. In each of the sub-plots, the framed 50×50 m^2 before the emergency exit includes the number of pedestrians in that area, where the congestion of pedestrians is assessed for the different modes. With all the modes, the simulator predicted a dominance of medium risk floodwaters (0.75 < HR < 1.5) over the walkable area, causing the majority of the pedestrians to fall into a toppling-only condition (purple dots) and a minority to have a stable condition (green dots) in front of the emergency exit and from the left side of the crowd. By contrasting the spatial distribution of pedestrians obtained from Mode 1 and Mode 2 (upper panels), there seems to be a considerable increase in the number of pedestrians with a stable condition when the two-way interaction condition is enabled with the walking condition (Mode 2). The same pattern is observed with Mode 3 and Mode 4 (lower panels), but this is accompanied by a shift in the position of pedestrians towards the front, as expected for the running condition. On the other hand, by contrasting the number of pedestrians in the small square obtained from Mode 1 and Mode 3 (left panels), it can be observed that enabling the running condition results in a decrease in the congestion of pedestrians in front of the emergency exit. The opposite pattern is observed when enabling the two-way interaction condition in Mode 2 and Mode 4, showing an increase in the congestion of pedestrians under a running condition compared to the walking condition. Hence, using the two-way interaction condition with the simulator may be useful to more realistically evaluate bottlenecking impacts of an evacuation process.

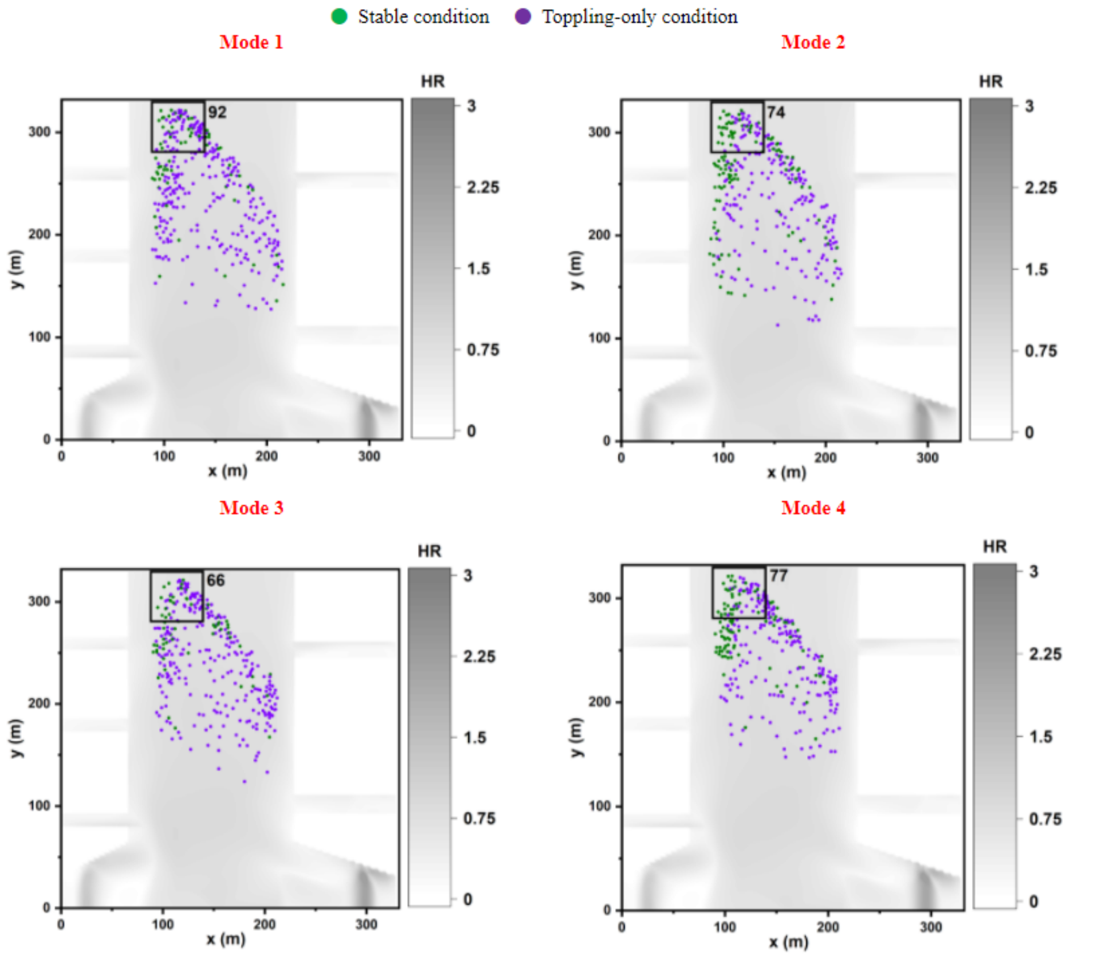

Figure 6: Spatial distribution of pedestrian agents, represented by coloured dots, predicted by the simulator under Mode 1 to Mode 4 at 6 min after flooding. The grey colour represents the floodwater extent based on the flood HR quantity and the back square before the emergency exit represents an area of 50 × 50 m² with a number printed alongside it representing the number of pedestrians in that area.

- *"The model demonstrates well a fundamental concept inherent within complexity sciences and Agent-Based Modelling; emergence. The adoption of risk perception thresholds seem to have influenced the favouring of destination selection by agents with varying levels of risk perception. As is the case for risk perception and adaptive behaviour in response to flooding (i.e. risk perceptions thresholds increase after an individual's property is flooded, which increases the likelihood of adopting protective measures in future) as an active area of interest within flood risk management. Some further comments on the overlap between risk perception and evacuation behaviour may be insightful to lay foundations for future research."*

In the revised paper, comments have been added to elaborate on the overlap between risk perception and evacuation behaviour, namely supported by further citing and discussing recently published papers who studied and reviewed the impact of past flooding experience of individuals on behaviours. Also, the uncertainty analysis added to Sect. 3.2.3 (to address major comment #1 above) offers new insights on the impact of adding the risk perception component on the variance in the simulator outcomes.

Changes made to Sect. 2.2.3:
2.2.3 Autonomous change of direction condition Each pedestrian agent is also featured with two extra rules to enable it to autonomously navigate into new pathways while moving within a flooded zone, where it encounters a non-zero floodwater depth from the navigation agent at its specific time and location. The first rule makes a pedestrian agent detect and choose another destination if the floodwater depth along its way becomes higher than a threshold of a floodwater depth to body height. The choice for the threshold is case-dependent and exploring different thresholds may be necessary (Sect. 3.3.2) as an individual's flood risk perception is dependent on different factors, including past flooding experiences (Hamilton et., 2020; Abebe et al., 2020). This affects the modelling of decisions of when and where people enter the floodwater or make a move into another destination (Becker et al., 2015; Netzel et al., 2021).
Also, the following changes are made in the 3rd paragraph of Sect. 3.2.2 .
As people's behaviour in floodwater is uncertain depending on their individual risk perceptions (Sec. 2.3.2), they may take the risks and enter the floodwater. To accommodate this uncertainty in the simulator, three thresholds of floodwater depth to body height were investigated (Fig. 11), inspired by the experiments in Dias et al. (2021). The '20 % threshold' was defined to represent people with high-risk perception, who previously experienced a critical flooding incident, and decided not to enter floodwater with a depth that is more than 20 % of their body height. This threshold, for this case study, is estimated based on the ratio of the dominant minimum value for the depth of floodwater that can occur over the walkable area (0.3 m) to the height of the shortest pedestrian agent available (1.4 m). With this threshold, the likelihood of the entire population to be in a condition to change their direction is ensured. The '40 % threshold' was defined to represent people with low-risk perception, who previously experienced a minor flooding incident, and decided to enter a floodwater with a depth that is even more than 40 % of their body height. This threshold, for this case study, is estimated based on the ratio of the dominant maximum depth of floodwater (0.9 m) to the height of the tallest population of pedestrian agents available (2.1 m). This threshold enables the entire population to have the freedom to keep moving even within the deepest floodwater in the walkable area (0.9 m). The '30 % threshold' was defined based on an average-risk perception, for those who previously experienced a major flooding incident. Pedestrians with average-risk perception would decide to enter floodwater up to their knees, constituting 30 % of human body height (Teichtahl et al., 2012).
Recall also the above-reported revisions applied to Sect. 3.2.3, addressing major comment #1, which demonstrates the impact of the risk perception on the variability in the simulator outcomes.

- *"Some more clarity on what solvers are used to dictate the flow of water in the hydrodynamic model should be provided in Section 3.2.1."*

Clarification has been provided in the second paragraph of Sect. 2.1 instead, as it is the most appropriate location to include this general information about the simulator's make-up.

Changes made to Sect. 2.1:
The environment layout encodes force vector fields providing navigation to key destinations. These fields are stored within a grid of fixed discrete agents, forming a navigation map (Karmakharm et al., 2010). The navigation map is necessary for pedestrians' way-finding decisions while they are directed to reach their key destinations.

The hydrodynamic model is formulated based on a non-sequential implementation of a finite volume solver of the depth-averaged shallow water equations on a two-dimensional grid on FLAMEGPU, which was validated previously in Shirvani et al. (2021). The hydrodynamic model was applied on another fixed grid of discrete agents, flood agents, which is coincident with the grid of navigation agents...

- *"Calibration/validation evidence of the hydrodynamic model should be provided in more detail to highlight the robustness of hydrodynamic outputs. Two separate figures are provided of flood extents; an Environment Agency figure (Figure 8) and your own (Figure 10). These could be combined into one figure, by overlaying and analysing to provide an FSTAT value to indicate accurate representation of simulated outputs versus simulated. Similarly (at a minimum), a figure showing simulated and observed hydrographs should be provided to demonstrate that flood depths are within an acceptable tolerance."*

To the best of our knowledge, there is no record of an observed hydrograph at a gauge point located in the selected study site, and that's why we resorted to the flood risk maps of the Environment Agency (EA). These flood risk maps inform the range of possibilities of what the flood depth and extent would be for different return periods (Fig. 8), whereas the hydrodynamic model predicts maps and depth extents (Fig. 10) caused by the rainfall runoff from the events of 19th of Nov. 2019 (driven by the hydrograph in Fig. 9). Therefore, we had to keep them separated. Still, the range of water depth and extent predicted by the hydrodynamic model for this event agree well with the expected ranges sampled in the EA flood risk maps should a flooding occur under any of the reported scenarios, which is a good sign that the hydrodynamic model is fairly calibrated for the intended purpose: demonstrating the simulator's capability for modelling collective trends of flood risk states of pedestrians inferred from the heterogeneity in their characteristics and behaviours.

The following footnote text has been added as a footnote to Sect. 3.2.1 to provide clarification:
To the best of the authors' knowledge, there is no record of an observed hydrograph sampled at a gauge point located in the selected study site, though resorting to the EA's flood risk maps can be deemed sufficient to support the scope of this investigation that is mainly focused demonstrating the simulator's capability for modelling collective evacuation trends inferred from the heterogeneity in pedestrian characteristics and behaviours.

- *"A supplementary table providing a summary of all agent variables and values upon initialisation (flooding, navigation and pedestrian agents) would be useful. This would promote replicability and reproducibility."*

The *'Flood-pedestrian simulator user guide'* file on Zenodo (10.5281/zenodo.4672125) includes a summary of all agent variables and values upon initialisation, and many other informative steps to promote replicability and reproducibility.