Dear authors,

Thanks for submitting this interesting article. I enjoyed reading and studying it. When data is challenging to measure, combining these two techniques, SFM and CFD, seems to be a promising and reliable way to expand our knowledge. I believe that until we can measure these turbulent flows, this will be the best way to obtain good spatially distributed data in such complex systems.

The article presents a "Hybrid" approach to gain insight into the characteristics of a turbulent flow in a step-pool system. The Hybrid part consists of a digital elevation model, built based on structure from motion data, which is used to define the bathymetry in a CFD 3D model, including a free-surface capturing component. By doing so, the authors show the complex flow structure at the different parts of the test channel and comment on the importance of different hydraulics variables and their effects on the stability of the structures. The base experiments are well designed and contain enough information for this exercise. Also, they account for different conditions, ultimately reaching the destruction of the step. This is key for characterizing the complete range of flow conditions.

The general structure of the article contains a description of the experiments, CFD simulations, and data analysis. A long series of appendixes and additional figures are given at the end to support the observations. After reading it several times, I felt that it was really long and could be reduced in length without losing any of the information provided.

General comments:
1) The article is well written but too long. Lots of repetition of ideas are found through the text, while at the same time, some details are missing, as will be detailed in the "detailed comments" section. Some ideas are broken between consecutive sentences but do not affect the delivery of the message.

2) To me is not clear why use the term "Hybrid" modeling because only one model is being used, the 3D CFD. I know that using a model of the bed surface elevation is the second model here. Still, when calling this hybrid modeling, I was expecting a dynamically linked set of models. For instance, I initially thought about an Hybrid RANS-LES type of model or something similar. This may arise from my CFD background, but it can be a little bit misleading.

3) Some more robust validation is required for the CFD model. In general, the validation is based on one measure of the error of water surface elevation but is not clear in terms of water depth or individual velocity. For example, velocities at the pool could have been provided and compared to the model. Zhang's 2020 paper described some PIV measurements.

4) All CFD model descriptions are based on the developer's descriptions, in this case, Flow Science, but more information from peer-reviewed or independent tests are required.

5) After reading the article is still unclear to me what the "insights for the stability and failure of step-pool units are also provided." The article seems to focus more on describing the results of the modes instead of describing a mechanism of stability and failure.

6) In general, references are too few. The use of "e.g.," in several cases throughout the text reduces the list of essential studies used. For example, CFD simulation with irregular channel boundaries (line 62) only cites one study. There are dozens of relevant studies here. I am not saying that you should cite all of them. But using, e.g., should not reduce the number of citations to just one or two examples.

Based on these general comments and what is provided below, I decided that probably after a major revision the article will be suitable for publication. I genuinely believe that there is immense potential when using the approach that the authors show here and that this article, once refined, will be highly cited and a reference in the field.

Detailed comments:
- The title could be misleading. Maybe in my case, for a particular background, the "hybrid" part of the title sounds like a hybrid CFD turbulence model. If the authors have another understanding please discard this comment.

- Is the @163.com a good email for future contact with the author? What about a couple of years from now? Maybe one from an institution provides a better way to connect with the author.

- line 34: "The high-resolution information of both topography and hydraulics for step-pool features is the key to fully reveal and describe these characteristics" It is not clear what you mean here for these characteristics.

- Some references are missing. Please check that all are included. One example is Golly et al 2017, which is not listed.

- line 39: "Although detailed topographic information has been available" Where is that information available? Please provide references (hereinafter ref. needed)

line 54: "The PTV method managed to visualize" Is not really the method that managed to visualize but what you do with the method. I understand what you are trying to say, but it is technically incorrect. It should be something like, by using the PTV method, we can visualize. So, it is not the method that visualizes things but who provides the data.

lines 62 and 63: Many more studies can be cited here. Please provide some more examples.

line 66: The CFD approach has been applied in some numerical studies. This is redundant. Of course, that CFD has to be applied to numerical studies. Suggest to change to "The CFD approach has been applied in studies containing step-pool features" or a variation of this.

line 69: Please define what a sub-unit is. Also, previously it was used a unit-scale. Please define it too.

line 78: "In the flume experiment of Zhang et al" In that article there is more than one, so using "the" is incorrect here.
This is critical when using this approach. Why use only water surface elevation (WSE) and not other variables? What about the PIV data of Zheng et al 2020? I understand how difficult it is to measure velocity in flume experiments like this one, but this velocity data seems to be available for your team.

- Figure 1 is never cited (actually, there is a wrong reference to it). RNG-VOF model is not described at the point where Figure 1 is cited.

"Two side cameras were used to capture the longitudinal profiles of the bed and water surface near the flume walls" After reading Zhang et al 2020 and Zhang et al 2018 it is still not clear how the measurements of WSE are obtained. It is not described in these two papers and I think it should be better described in this article because it is the base of the model validation.

- Details on the grain size distribution (GSD) are required in the article. Some general information about the D50, grain sizes in the pool, etc will make the article more robust.

- Line 110 The terminology used at this point in the article is not clear. What is KS (I know that means keystone, but that is described several pages later)? What means T2 or a T run?

- Line 113 : Any comments about the effect of stopping the water circulation on the bed response would be appreciated.

- line 126: What other facilities? please list them.

- line 126: DMS at different flow rates? Is that "for" different flow rates? Please re-write this sentence because it is confusing.

- Line 128: What means relatively poor? Compared to what? Can you provide a measure of good or poor?


- Lines 132 and 133: If these WSE were captured every 2 s, what is then provided? Is it an average over X seconds? Please describe it in detail. Also, how are longitudinal profiles
of the bed used in this study?

- Line 137: This "high enough" is the 3.3 - 3.9 mm? Can you provide a measure to understand why is high enough? For example, the D50 was 1 cm, therefore, with 3.3 mm all grains are captured.

- Line 137: " geometric feature " Please list them. It is not clear what geometric feature means in this context.

- Line 138: " reduced the requirements for computing resources of the numerical simulation " How does the elevation model help reduce the computing resources used? That most likely is the only function of the numerical mesh, but not on the DSM. Please explain. This could be helpful for other researchers.

- Line 140: Commercial solution is difficult to understand. Could it be commercial software or commercial model, or a commercial solver?

- Line 141: This part is critical. What do we understand as "shown good performance"? Please provide more information here, some peer-reviewed examples using flow3D.

- Line 142: Please provide references for the TruVOF. The article of Hirt and Nichols is well known and one of the fundamental studies in VOF, but it is not about TruVOF.

- Line 143: Same for FAVOR, please provide more information here, some peer-reviewed examples using FAVOR

- Line 146: Please provide more references to justify the model selection. Using the vendor's information is not enough. I believe that it is a good model, but using Flow science, 2016 as a reference is suitable just for description.

- Line 148: How was the FAVOR technique tested here?

- Line 154: What does it mean by "leaking" - "When leaks emerged between the DSMs of bed surface ..."
- Line 168: Please describe in more detail the Renormalized Group (RNG) k-ε turbulence model and its implementation in the solver. No references to the turbulence modes are given. Please provide references to literature and now the software developer.

Line 169: "The VOF technique ..." This has been said already.

Lines 171 to 175: Why there is always ranges of mesh properties? 2-3 structured mesh blocks, 24-37 cm, 6.5-9.4 million units, etc. Are those different for different flows?

Line 176: It is unclear how uniform and non-uniform mesh differ when using them in the model.

General comments on the simulations: Before resuming the line-by-line analysis, I will summarize some missing numerical details in the article. These are critical because they give the reader an idea about how the model was configured. After analyzing the results and looking at the figures, I think first order numerical schemes were used. I can't know this for sure, but it is my impression because velocity profiles and wse are super smooth.

- Please provide details about how each term was treated (Numerical schemes) For example, did you use MULES for the water phase? If so, was it limited? with artificial compression? second-order in velocity? etc.

- What convergence criteria were used for Vel and Pressure? What about other variables?

- The results we are seeing, are they time-averaged? Over how long?

- What are the boundary conditions for k and epsilon?

How does the model treat the variables when it has a discontinuity in permeability (bed/water transition)? I am wondering if wall boundary conditions are required.

- Why this particular turbulence model was used? Changes in pressure would suggest using a k-omega sst type of closure. I don't think it is wrong to use k-e, but it must be justified.
- In the upstream end, the flow enters the study reach. The images show that distance is too short to be a fully developed flow. Is there any consideration used to ensure that the numerical flow is a good representation of the actual flow once it reaches the step?

- How was the flow initialized? There is a mention that the pool had water initially, but no other details are provided.

- How was defined the location of the WSE. In VOF, the fractions are a continuum, so depending on the flow type, some studies use 0.5 (others 0.1 or 0.9) to define the boundary between water and air.

- How much space was left from the WSE to the upper boundary. What are boundary conditions used to model the air entering/leaving the domain?

- Was the velocity specified at the inlet uniform? a logarithmic profile?

All these comments point towards a better understanding of the model configuration.

Now I will continue with line-by-line comments.

- Line 180: For the no-slip condition, what is the cell size on the walls? What about the other variables, for example, pressure, k or epsilon

- Line 185: 'which efficiently accelerated " Compared to what? compared to starting with an empty (only air) domain? How do you know that it was faster? This could be helpful for other researchers because when using VOF obtaining good initial conditions is a time-consuming process.

- Line 190: "were collected after the solution was steady, with the variation from the mean less than 0.5% at each flow rate" This is an essential aspect of the numerical experiment design. In a turbulent environment like the one used in this case, we would expect high fluctuations in every single variable, this can be seen in the images of Zheng et al 2018 and 2020. A variation of 0.5% from the mean seems to indicate the use of a highly diffusive numerical scheme. The good thing is that it is stable for this complex simulation but not really accurate. Please provide more details. Maybe a plot of velocity in time at some locations could be helpful.
- Line 191: A frequency of 2 Hz could be too distant in time to capture some important turbulence properties. This is only 2 observations every second. For example, you mention later that you use velocity fluctuations for some calculations. Please comment.

- Line 195: Please explain how it was decided that grid independence was reached. Please provide a metric.

- Line 200: Although there are several metrics for evaluating the error in WSE, a comparison to water depth will also be helpful. It is difficult to judge if 2-3 cm is a small quantity or not because there is no information related to water depth.

- Line 214: How are velocity fluctuations evaluated?

  Line 219: Are all these terms in equation 3 calculated by post-processing the information or given as a result of the model? Do you have to calculate all the gradients or there is a function within Flow3D that does the job for the user?

- Line 221: Then, obtaining all these results, what is the typical time step in the simulations? Any CFL criteria were used?

- Line 221: How is the shear stress calculated? If it is a result from the solver, what equation is used?

- Line 222: Please explain why the dynamic pressure is the one analyzed here.

- Line 225: Please explain how it is obtained in a given model. Is a result of manually post-processed from VOF?

- Line 225: Was water density equal to 1000 kg/m3? or any correction for temperature was used? What about air density and viscosity?

- Line 231, Is Using measured at the inlet?

- Line 234: Please describe the threshold method.
- In general, the results are well described. They could be reduced a little bit, but they are well explained. Good job!

- Line 297: What is here a contraction? This term may be used with different meanings in different fields. Please provide a definition.

- Figure 6. At the end it says, Figure 3. Probably it is Figure 5.

- Section 3.2.2 What is the role of shear stress in this type of structure?

- In the discussion section, I believe some paragraphs connecting the observations from the results are required. How is everything connected to stability?

- Line 553 - Air entrainment is possible to include, at least partially, in VOF. Especially in this type of model when there is a jump. Please comment on why it is not considered. I said partially because it depends on the grid size. Required more post-processing because air and water fractions or their boundaries are more difficult to isolate, but it is possible. Please check the literature on bubbles (OpenFOAM has been used widely for this).

- As a limitation, please mention all those that may arise by selecting the numerical schemes and model configuration, also, by using a k-e turbulence model.