Specific comments:

1) Starting with equations 3-4 the authors introduce an undefined quantity h that seems to play the role of vertical spacing. This quantity appears alongside the quantity h_α, which instead is defined on page 4. Later on, quantities l_α are introduced and employed without being defined. Whether these have the same meaning usually implied in the multilayer literature is never stated. This does not allow the reader to understand exactly how the proposed numerical method is formulated and makes the results described in the preprint impossible to reproduce. A mandatory revision is to define properly all quantities that are being introduced and explain clearly their relationship (if any) to analogous quantities introduced in the literature on multi-layer models. Furtermore, in the appendix (line 386) the authors seem to imply that in models for which $h_\alpha = l_\alpha H$ the mass transfer terms across the layers have to be zero. However, this is not true for the (Audusse et al. 2011) paper which apparently constitutes the reference for the present paper nor for the related discretization approach introduced in

   Fernandez-Nieto, E. D., Kone, E. H., Chacon Rebollo, T. (2014). A multilayer method for the hydrostatic Navier-Stokes equations: a particular weak solution. Journal of Scientific Computing, 60, 408-437.

   and employed in (Bonaventura et al. 2018). The authors should clearly specify to which multi-layer formulation they refer and remove any incorrect statements in this respect.

2) The authors devote a significant effort to the important issue of proving that what they call the 'tracer constancy condition'. They also refer to this condition as 'Geometric Conservation Laws', but no reference is given for either denomination. However, since the seminal paper

   Lin, S. J., Rood, R. B. (1996). Multidimensional flux form semi-Lagrangian transport schemes. Monthly Weather Review, 124(9), 2046-2070

   it has become customary to describe this condition as 'consistency with continuity' or 'compatibility with continuity', see e.g.

Gross, E. S., Bonaventura, L., Rosatti, G. (2002). Consistency with continuity in conservative advection schemes for freesurface models. International Journal for Numerical Methods in Fluids, 38(4), 307-327.

Fringer, O. B., Gerritsen, M., Street, R. L. (2006). An unstructured-grid, finite-volume, nonhydrostatic, parallel coastal ocean simulator. Ocean modelling, 14(3-4), 139-173.

Kuhnlein, C., Smolarkiewicz, P. K., Dornbrack, A. (2012). Modelling atmospheric flows with adaptive moving meshes. Journal of Computational Physics, 231(7), 2741-2763.

The authors might consider using (also) this terminology in the revised version. More importantly, as shown in (Gross et al. 2002), this consistency/compatibility must be guaranteed also for the time discretizations of the tracer and continuity equations. Apparently, this aspect is not discussed in the preprint, so that the consistency proof provided by the authors cannot be considered complete. This is especially important for semi-implicit discretizations, since using advecting velocities at different time levels might easily occur in this context. Completing the discussion on this aspect would definitely increase the value of the preprint. Furthermore, in the numerical experiments this property is only checked in a case with flat bottom, while a numerical check also for more complex bathymetry is necessary. A mandatory revision is to include similar checks of the preservation of constants also for the sloping channel and Venice lagoon benchmarks.

3) In many parts of the paper, the authors try to consider different $z-$coordinate formulations within the same multi-layer framework. While this is definitely a positive thing to do and a potentially important contribution of the preprint, often the way in which the different formulations are handled is confusing, also because of the related lack of specific definition of the l_α coefficients. The authors are strongly suggested to review all the parts of the text in which the different formulations are presented and make sure that all the quantities involved are properly defined and the specific steps to be taken for each formulation are described completely and in detail.

4) In the introduction (line 40) the authors claim that their remeshing strategy solves possible stability problems of approaches proposed earlier in the literature. This claim is repeated later in the preprint (page 9, line 202). However, no stability analysis is provided to support this claim. In the revised version, the authors should either provide a proof of stability for the proposed algorithm or remove/reformulate any claims of superior stability properties.

5) At the end of section 3.2 (line 175) the authors introduce a pseudo-time quantity τ which is then discretized in steps $\Delta\tau$ to proceed to the remapping of discrete quantities, see equation (15). However, there is no indication on how this pseudo-time step should be chosen and on whether any empirical or theoretical bounds should be respected to maintain stability. Inclusion of some criterion for the choice of $\Delta\tau$ (sufficiently small fraction of Δt?) is mandatory for the revised version.

6) The truncation error analysis presented in the appendix uses in an essential way the linearized equation

$$\partial_t \zeta + H_0 \partial_x u = 0.$$

This form is consistent with the assumption that the linearization has been performed around the constant state $U = 0, H = H_0$ and that the velocity field u is a first order perturbation. This seems however inconsistent with the assumption of an $O(1)$ tidal amplitude A. The correct linearized equation would be in this case

$$\partial_t \zeta + U \partial_x \zeta + H_0 \partial_x u = 0$$

with $U = O(1)$. As a consequence, the upper bound on the divergence would also depend on the free surface gradients, which would seem physically reasonable. The whole derivation in the appendix would have to be reformulated taking into account the correct linearized equation. Furthermore, the following numerical experiments seem to consider a constant laminar diffusivity, which is rather different from the turbulent profiles that would typically arise in a realistic situation, making the whole discussion somewhat academic. Either the authors find a way to address this major shortcomings of the analysis presented in the appendix, or they would be strongly suggested to remove this analysis which is only marginally related to the main topic of the paper.

7) The model equations are written in dimensional form, so measure units should be introduced for all the quantities reported when describing the numerical experiments (they are missing in section 5.1)

Technical corrections:

1) line 13: replace 'that follow the materials' with 'that are material surfaces'
2) line 21: the reference to (Cheng et al. 1993) could be complemented with a reference to the unstructured UNTRIM-3D model, such as e.g.

   Casulli, V., Walters, R. A. (2000). An unstructured grid, threedimensional model based on the shallow water equations. International journal for numerical methods in fluids, 32(3), 331-348.

   Casulli, V., Zanolli, P. (2002). Semi-implicit numerical modeling of nonhydrostatic free-surface flows for environmental problems. Mathematical and computer modelling, 36(9-10), 1131-1149.
3) line 87: the explicit definition of the term IPG_α should be introduced
4) line 107: in formula (8), the argument of the flux limiter should be specified; this should also be done in all the other points where this quantity is introduced, most of the time without specification of the argument and of the location at which it is computed; also in formula (9), a ϕ_α appears that is not defined anywhere
5) line 114: in formula (9), the second derivatives on the right hand side have no reason to be positive, so that this kind of upper bound should be performed considering only absolute values of the quantities involved
6) line 135-139: what the authors denote as the space discrete and fully discrete variables, respectively, are indeed (see e.g. formula (17)) the P^1 finite element approximation of the solution, which is a piecewise polynomial continuous function; the text should be changed to avoid this confusion; even though the notation $u_h(x)$ is customary in the finite element literature, it is a bit confusing in a context where h has a different meaning (the finite element 'h' would correspond to what in

the preprint is called Δx_E.)

7) line 146: the first sentence of section 3.1 is superfluous, any time discretization method will update the free surface based on equation (3)...the sentence should either be removed or reformulated if something else was meant

8) line 158: coefficients $l_{\alpha,i}$ are introduced without having been previously defined; this is related to point 1) in the specific comments above; clear definition of these quantities is essential, since otherwise the proposed methods are not completely defined nor reproducible

9) line 188: it is unclear what do the authors mean by 'z-layer depth at rest Δz_α^0; if this is the depth at the initial time, it is better to say so because what 'at rest' means in a hydrodynamical simulation is very unclear

10) line 217: the expression 'hanging interfaces' is probably derived from the 'hanging nodes' used in the literature on numerical methods for non conformal meshes; however, while hanging nodes makes sense (the quadrature nodes on one side do not have a counterpart on the other side and numerical fluxes or mortar procedures must be employed), hanging interfaces does not make much sense in my opinion, since the interface is a perfectly well defined geometrical object; the authors are strongly suggested to modify this terminology and use instead e.g. non-conformal boxes, as they do in the following

11) line 221: change 'sophisticate' into 'complicate'

12) line 240: change 'kernel' into 'basis function'

13) line 244: the formula below this line is obtained according to the authors by integration by parts, but contains no boundary terms, the authors should explain whether these terms are zero and why or correct the formula

14) line 286: the rule used to define the mesh size should be explicitly reported, e.g. h_K as the maximum length of the triangle sides

15) in the caption of Figure 8, the quantities T, T_0 used to define the relative tracer conservation error are not defined; if they are meant to be the total tracer mass at the end and at the beginning of the simulation, then $|T - T_0|$ (absolute value is missing in the text!) is the absolute

error, not the relative error; the authors are suggested to display values of $|T - T_0|/|T_0|$

16) line 333: replace 'summerized' with 'summarized'

17) line 437: replace 'its' with 'her' !!! it would also be appropriate to specify better the direct or indirect contribution of Dr. Bellafiore to this work

18) general comment: personally I think it is graphically better to write $z-$coordinate than z-coordinate; this is not a required change but I think that it would be appropriate