Summary and Recommendations:

In this manuscript the authors show how small scale variability of the topographic gradient in the Amazon rainforest influences the simulation results of a land surface model (INLAND). By showing the differences between a simulation at a plateau and a valley they highlight that a too coarse grid resolution which neglects this small scale variability can lead to significant simulation errors. To address this question the authors show in great detail that their model is capable to mimic different observed fluxes and state variables.

I think the general idea and methods of this manuscript is within the scope of HESS, however I believe some major revisions are needed beforehand. The main reason is a missing clear hypothesis and the overall structure of the manuscript.

General comments:

1) Your manuscript could be improved upon significantly if you would focus your manuscript around a clear hypothesis. If your goal is to show that the small scale variability of the topographic gradient is important and a dominant control for different fluxes in your landscape you might consider building your story around that hypothesis. Compare the two models and show that the measurements and the simulations differ significantly at each site.

2) Your abstract needs to be re-written. It should clearly state what you have done and what the take home message is. At the moment there are a lot of minor details.

3) Section “f. experimental design”. While your overall manuscript is rather long this section is a little short. Make clear why you do different simulation experiments and consider an extra section where you explain 1) your model setup and 2) your model evaluation. I would recommend not mixing them.

4) What happens if you remove your ground water model? Is your model performance decreasing? Why do you add a conceptual groundwater model instead of switching to statured conditions with the Darcy equation? Wouldn’t this be much more consistent?

5) In your results and discussion you go into great detail. However for an improved readability consider moving some of the details to an appendix or consider doing a separated result and discussion section. You have nice results and discussion points it is just difficult to find them between all the model details.

6) In your conclusion you again explain in great detail that your model is capable to mimic different fluxes and states variables. You state that this confirm your initial hypothesis I would explain a little more why do you think this is true.

Technical comments:

1) Line 274: Where is equation 3 and 4?

2.) Line 346-347: Why do you pick the bias, R^2 and the RMSE as error statistics? Why do you think these three are a good choice in your case?

3) Line 309: Up to a grid resolution of 4 m with Darcy Richards. I know that this is common practice in the land surface community but you might want to read the papers of Lehman, Or and Vogel about the basic assumptions of the Darcy Richards equation and if they hold within a 4m large grid.

4.) Line 353: Maybe I missed that part. How do you know that your vegetation and soil parameters are optimal?

5) You write that your model reproduces several fluxes in the different parts of the result section reasonable well. I would consider sharing what you mean by reasonable well. Is the difference above or below the measurement uncertainty? Does it show in comparison to other models or model studies much better or similar performance?

One example is Line 476 to 480. Here you write that your model is doing well because it simulates different ET values for the plateau and the valley in good accordance with the observed data of 2006. However, your model simulates 3.7 mm day^{-1} for the plateau and 3.2 mm day^{-1} for the valley. Your observations are 3.0 mm day^{-1} and 2.9 mm day^{-1}. First of all is the differences between the observations with 0.1 mm day^{-1} above your measurement uncertainties? Is it possible to say that the two sites are different on an annual scale based on these measurements? Secondly if this is the case why do you think your model is doing well here? It overestimates the differences and the plateau observation is closer to the valley simulation.