Comment on egusphere-2022-770
Anonymous Referee #2

Referee comment on "Effects of complex terrain on the shortwave radiative balance: A sub–grid scale parameterization for the GFDL Land Model version 4.2" by Enrico Zorzetto et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-770-RC2, 2022

Summary and general comments

In this study, a parameterisation for the effects of sub-grid topography on surface shortwave radiation is presented. In a first step, the authors apply Monte Carlo ray tracing to simulate surface shortwave radiation for 3 geographic domains with complex terrain. These experiments serve as a reference to develop the (sub-grid) parametrisation. In a next step, terrain properties (μ, sky view factor and terrain configuration) are linked to modulated radiation fluxes with two statistical models – a Multiple Linear and a Random Forest Regression. Finally, sub-grid effects are considered by merging land units within a grid cell with similar terrain properties by means of hierarchical clustering.

The aim of this study is very interesting and relevant – namely improving the representation of surface shortwave radiation fluxes in an Earth System Model. Due to the plane parallel radiative transfer schemes applied in such models, surface radiation is typically simulated rather inaccurately in areas with complex terrain. The implementation of parameterisations, particularly on a sub-grid scale, has the potential to strongly reduce such biases. The approach presented by the authors is very interesting and the manuscript is well written and structured. However, I struggled to understand certain sections in detail – for instance the hierarchical clustering section in the methods and some passages in the Results and Discussion. Furthermore, the Results and Discussion section is sometimes incomplete in my opinion and should be extended (see the following comments for more details).
Major comments

Section about hierarchical clustering (2.4)

Until section 2.4, the methodology is very well described. However, I struggled to follow section 2.4. For instance, why do you want to partition land in hydrologically coherent units? From a “terrain-radiation-perspective” – this is not obvious. Has this approach been chosen due to an already existing tile classification in the GFDL Land Model?

I’m also confused why the clustering is performed twice (first in $k$ hillslopes, then in $p$ sub-units). I think a detailed flow diagram (e.g. with an example of the step-wise classification of sub-units of a geographic domain) would help the reader to understand these steps. Furthermore, it is also not obvious to me why lakes and glaciers represent separate classes. And are glaciers and lake classes further divided into sub-classes according to their terrain properties? Finally, some parts of section 3.4 (e.g. starting from line 354 could also be moved to the method section).

Analysis and results – improve consistency and completeness

- I’m missing the third domain (Nepal) in Fig. 7. I guess you used one domain to train the model and the other two domains for cross-validation – right?
- I think a performance comparison of the sub-grid to a grid-scale parameterisation would be very interesting to show. With this, you could emphasize the additional benefit of the sub-grid scale scheme.
- The discussion of certain findings should be extended. From the results, it seems that a tile number of ~100 captures the sub-grid characteristics already very well. Do you agree? And would such a number be feasible in an online ESM simulation?
Minor comments

Content-related (text)

Line 42: what does the abbreviation “WLH” stand for?

L139: “uniform” albedo -> how realistic is this assumption?

L139: I appreciate such clear definitions, it simplifies the comprehensibility of the subsequent text greatly!

L157: I’m not sure if I understand this sentence correctly. Do you mean that radiation fluxes significantly departure locally from areal-average fluxes?

L 162: “represents the fraction of the sky dome visible from a target site” -> technically, this is incorrect. Compare e.g. with Helbig et al. (2009) (text next to Eq. 8) and Zakšek et al. (2011). The sky view factor definition of Dozier and Frew (1990) yields the fraction of hemispherical radiation received under the assumption of isotropic radiation. The same is valid for the subsequent explanation of the terrain configuration factor $C_t$.

L171: Could you explain why you use this terrain configuration definition and not simply $C_t = 1.0 - V_d$ (compare e.g. with Chu et al., 2021)?

L174: I would briefly introduce and explain the parameters $\mu_i$ and $\mu_0$ here.

L194: I’m a bit confused by these lines. It seems that you perform the clustering only for soil elements (also according to line 207; $kp$ and $kp + 2$) and not for glaciers and lakes. What is the reason behind this? I guess glaciated areas and lakes can also have very variable topographic parameters (like e.g. sky view factor).

L204: It’s not obvious to me why you apply the clustering a second time. Generally, to increase the comprehensibility of this section, it might be worth to extend the workflow
diagram displayed in Fig. 4. One could show the classification of a certain domain
(resolved for every single step).

L207: I’m still a bit puzzled – what is the motivation behind categorizing land surface
based on hydrological properties? I don’t see the connection to topography-radiation-
processes.

L298: “reflected components are quite linear” -> for \( f_{\text{dir}} \), the deviations between MLR and
RFR are quite substantial...

L 305: “case in which...” -> I don’t understand this part; there is probably something
missing.

L316: First of all, I’m confused about which region (East Alps vs. Peru) is the
(in-)dependent domain. The caption of Fig. 7 does not agree with the statement here.
Furthermore, I’m not convinced that results from RFR are not location dependent. Looking
at Fig. 7, the RFR method consistently indicates a worse performance for the cross-
validation domain than the MLR method. For me, this is an indication that obtained
relations from the RFR simulation are very location-dependent and not easily transferable
to other terrain geometries (i.e. the model is overfitted).

L388: It would be interesting to see the results for these tests too. Maybe you could show
them in the supplementary material.

Typos, phrasing and stylistic comments

L124: references not correctly rendered

L153: I was a bit confused by this line, it might be better to write something like: “The
MC calculations were performed for three independent domains (Nepal, Peru, East Alps)...”
(if that is what you mean)

L157: “determines” -> “determine”
"represent" -> "represents"

"in order to compute the sky view factor"

"eq. 6" -> "Eq. 6"

"if these are present in a given grid cell."

"the is the indicator" -> "is the indicator"

"eqns. (1)" -> "Eq. (1)"

"angles compute based" -> "angles computed based"

"simulation (5)" -> "simulation (Fig. 5)"

I would rewrite this to e.g.: "...larger than approximately 5 km the effect disappears."

"case in which" -> "a case in which"

**Figures and Tables**

**Figure 2:** The colorbar labelling is erroneous – I guess it should be “Elevation [m a.s.l.]”. The same is true for the upper-left panel in figure 3. Furthermore, the degree symbol is missing for the cardinal directions.

**Figure 4:** It seems from these panels (x/y-coordinates) that the MC model was run on a
map projection. Could you specify the projection somewhere?

Figure 7: $\mu_0$ not correctly rendered in caption

New references

