Response to review AR1

The reviewer's comments are mentioned in black and author responses are in blue.

We thank the reviewer for her/his time and efforts in highlighting parts of the manuscript that require changes and clarification.

The paper "The Effects of Late Cenozoic Climate Change on the Global Distribution of Frost Cracking" by Sharma et al. presents the results of three frost cracking models applied on global scale. The authors used published temperature reconstruction in an 80 to 80 km resolution for four time slices: Pre-Industrial (~1850 CE, PI), Mid-Holocene (~6 ka, MH), Last Glacial Maximum (~21 ka, LGM) and Pliocene (~3 Ma, PLIO) times. These temperatures were used to calculate days spent in the frost cracking window (Anderson, 1998) and two existing frost cracking models (Andersen et al., 2015; Hales and Roering, 2007). The authors analyzed spatial variation of frost cracking intensity (FCI) at individual time slices and observed large deviations of FCI between warmer (PI) and colder (LGM) climate. The paper addresses an important topic about spatial and temporal variation of frost weathering that shapes the Earth's surface. My major concern is that there is no independent data set that enables the validation of the model results. The authors provide three model results but there is no possibility to review the model results for a location where data exist or for different time slices.

The reviewer raised a very important concern regarding the unavailability of data for paleoclimate time-slices to validate the model results. We addressed this in section 5.3 of the revised manuscript by comparing the trends of our global FCI estimates with results from previous studies. Due to the scarcity of regional data, we focus on the global and general trends in these comparisons and provide a few suggestions (e.g. in section 5.5) for future regional studies that will result in a better basis for detailed regional estimates and their validation.

Lastly, we briefly mention here a, perhaps, philosophical difference with the reviewer in how models should be used. Throughout geoscience literature, numerical models play a role in the formulation of hypotheses that can be tested in observational studies. Models are useful for hypothesis formulation because the underlying physics/processes producing different observables are known/controlled for. Observations, in contrast, are useful for telling us what has happened, but they often (not always) do not tell you the process(es) responsible for the observations recorded. There is no shortage of observational studies that lack a process based (modeling) framework to be interpreted, and instead infer (qualitatively) the underlying process. The two approaches (models, observations) go hand in hand and science progresses by development in both approaches. This was our motivation for conducting a global scale modeling analysis of frost cracking (and for setting up the simulations in Mutz et al., 2018; and Mutz and Ehlers, 2019, ESurf).

Concerning paleoclimate model validation – the standard approach used in this community is to conduct present-day simulations that are compared to observations (usually reanalysis data). Our previous published work does this with the ECHAM5 model (for the same time slices presented in this study). If the comparison is satisfactory, then the modeling approach that follows is to modify boundary conditions and model inputs (e.g., land cover, sea surface temperatures, etc.) for paleo conditions. Our experience in our previous publications is that the differences between a climate model results and reanalysis data (e.g., ERA40) for a region is typically less than the differences between different reanalysis data sets. This has repeatedly been shown for the ECHAM5 model (it has been around a while). In short, observed climate is not as well resolved as many people think, and model performance (at for ECHAM5 at least) is reliable for (paleo)climate studies. The assumption then made is that if the physics in the model worked correctly for the present-day, then they also work for the paleoconditions and meaningful comparisons to the modern can be made by differencing modern and paleo simulation results. This is the underlying approach of this community, and also what we follow in this manuscript.

Concerning frost cracking observations available to test model predictions: Later in this review we address the reviewers concern regarding which observational studies provide a meaningful comparison to the model results. We now focus our comparison to regional observations/studies only.

1. The use of paleodata

The study uses paleo-temperatures, which are air temperatures according to the papers by Mutz & Ehlers (2019) and Mutz et al. (2018) and not land surface temperatures as indicated in this paper. Snow cover and vegetation will result in temperature offsets between air and surface temperatures, which will cause large difference in the frost cracking results and are not addresses in this manuscript at all.

The reviewer raises a good point here, however her/his statement is due to our lack of clarity in the text (which we've now fixed). Our analysis was actually conducted on the ground temperatures. The above-mentioned studies i.e., Mutz & Ehlers (2019) and Mutz et al. (2018) used 2m air temperature in their analyses (as the reviewer states). However, the palaeoclimate experiments are conducted with a GCM (ECHAM5-wiso) that simulates more aspects and variables of the climate system. Snow and vegetation cover are included in the GCM's simulation of an equilibrium climate, and (land) surface temperature are computed and saved in the GCM's standard output along with various other variables. These have been available for download since their first publication in 2018 along with 14 other variables because they were not analyzed those studies. Not all were presented in the papers, but some have already been used in other studies (e.g. Werner et al. 2018, ESurfD). These can be downloaded as part of the p001 package from:

https://esdynamics.geo.uni-tuebingen.de/wiki/index.php/all-category/2-main/117-echam5-palaeoclimate-simulations-mutz-et-al-2018

Hence, we confirm that we did in fact use land surface temperature data in the study and obtained them from the Mutz et. al. (2018) simulations. We have mentioned this contribution in Acknowledgements section (lines 592-593). We have also added this point in the revised manuscript in Data section (line 83-85).

The paleo-data is available at 80 to 80 km resolution which is much too coarse to apply these to high-topographic environments as the European Alps, Andes or Tibetan Plateau. The coarse resolution is not integrating topographic effects in is not applicable to mountains. The authors should downscale their data, which is a standard procedure in alpine studies (e.g. Fiddes and Gruber, 2014).

As a bit of background: Typically when we publish our climate model results the climate community modelers/reviewers complain that our resolution for the simulations is excessively high (not too low)!. The reasons for them saying this is that lower resolution simulations run significantly faster and capture them sample climatology as the higher resolution models. We have been insistent to the climate modeling community in running our simulations at a high resolution (80x80km) globally because they are more useful for comparison to geomorphic studies (such as this manuscript). Thus, by climate community standards, - the simulations presented in this study are very high resolution. They also took about 2 months each to run on a large cluster, running them at an even higher resolution (e.g., T256) would likely take 9-12 months (each).

In this study, we attempt to present a first order (global) approximation of frost cracking intensity subjected to different Cenozoic time-slices. We state this clearly in the text, and our interpretations (e.g., all the figures) are conducted at regional and global scales only. Dynamical downscaling of a palaeoclimate simulation (e.g. Wang et al. 2021, JGR-Atmospheres) would be ideal in such a case and preferable over simple topographic corrections that do not adequately consider the impact of topography-related atmospheric dynamics leading to further changes in climate elements (incl. surface temperature). However, this is only feasible for studies focusing on one region, as regional climate simulations for only one orogen would likely take several months to a year, depending on the researcher’s level of expertise and challenges encountered along the way. The application of such regional climate models (or even simpler downscaling approaches suggested by the reviewer) would require additional expertise/training and significant computation time that is beyond the scope of this study. Furthermore, Pliocene downscaling is complicated further by uncertainties in topographic reconstruction on a much finer scale. We therefore decided to not pursue the idea of global-scale climate downscaling in orogens and instead clarify our focus on global first order trends in the introduction section 1 (lines 68-69). Additionally, we make recommendations for future regional/local studies (e.g., in high-topographic environments) that warrant the intermediate step of climate downscaling (lines 528-533 in the limitation section 5.5), but please note that each of these studies would likely be a paper/study in itself and could not present a global analysis as we've done here.

The data is available at daily time steps and could be used directly to calculated frost cracking. However, the authors calculate a mean annual temperature and half amplitude of annual temperature. They used sinusoidal daily temperatures but it remains unclear if these temperatures are from the paleo-temperatures or assumed values. A more direct use of paleo-temperatures would be better suited.

We prefer using sinusoidal daily temperatures over paleo-temperatures at daily time-steps for following reasons: Daily temperature variations predicted by GCMs are difficult to validate. Consequently, we do not expect any added value to forcing the frost-cracking models with GCM temperatures instead. Also, using daily temperatures instead of sinusoidal variations based on MAT and Ta would require relatively longer computation time for a global analysis like ours. Using the GCM temperature cycle is thus unlikely to produce significantly different results and we opt for the computationally more efficient option. However, we agree that in future local studies, using paleoclimate data at daily time-steps could provide higher accuracy to the model results with respect to time and depth of occurrence of frost cracking if daily variations are demonstrated to reflect realistic conditions in the study area at a specific time in the past. To our knowledge, the latter has not been demonstrated for the regions and time slices treated here. We have added the main reasons for our choice in the methods section 3.1 (lines 148-151).

2. Frost cracking models

The authors used three proxies or models for frost cracking but only focus on model 3 in their paper. The days spent in the frost cracking window is only a poor proxy for frost cracking (Anderson et al., 2013). The model by Hales and Roering (2007) is out-dated and not including any lithological differences. Both models are barely used in the results and discussion section, therefore they could be omitted from the manuscript.

The first two proxies for frost cracking used in our study are simpler models, which require fewer input parameters and may still be used when a first order approximation is needed, especially when all required inputs for a more complex model are not available. Hence, we include the results of simpler models i.e., model 1 (Anderson et al., 1998) and model 2 (Hales and Roering, 2007), in the main text, but still include them entirely in the supplement for reference. These results are still of potential use to the broader community and have merit when insufficient data is available to constrain a more complex model.

Finally, the second reviewer of this manuscript appreciated having all three models and requested a more detailed comparison between them (which we've done – see response to reviewer AR2).

The model by Andersen et al. (2015) is applied using soil thickness to constrain a soil layer with an assumed porosity of 30% which is located above a bedrock layer of 2% porosity. The soil thickness is derived from a global database with 5 km resolution and used for every time slice, however, it is unrealistic that soil thickness is a constant over Cenozoic time scales.

We agree that it is an unrealistic assumption to have a constant soil thickness over Cenozoic time-scales. However, there is no available dataset for past soil thickness, so we have used the present-day data, and clearly stated that it is a limitation. To address this concern, we have added additional text (lines 87-93) in the Data description (section. 2) and revisit it again in the model limitations section (section. 5.5, lines 508-514), to explain this limitation more prominently. We have also changed the caption of Fig. 1 to support our argument for the above assumption.

The substrate classification into soil and bedrock changes water flow in the subsurface within the frost cracking model. For alpine regions the database provides relative high soil depths, however, rockwalls with 30 % porosity are not existing, which highlights the problem of spatial solution and applicability of this model in this way to alpine conditions using a soil map. In addition, the model by Andersen et al. (2015) uses a fixed frost cracking widow between -8 and -3 °C that is not supported by laboratory data (e.g. Murton et al., 2006), field data (Girard et al., 2013) or physical models (Walder and Hallet, 1985). As lithology and rock strength show variations across the Earth, lithology will control weathering, which could be incorporated to include more realistic results.

We agree that the models are based on simplistic assumptions with coarse spatial resolution in paleoclimate and soil data. The model results cannot replace in-depth regional studies using high spatial resolution lithological distribution and downscaled (topographically corrected) paleoclimate data. We thank you for raising a very important concern. We have modified the model limitations section 3.5 (lines 526-531) to include highlight these simplifications and avoid overinterpretation of our results.

We would like to thank the reviewer for raising the concern about the assumption of fixed frost cracking window. We have added the above argument in Introduction section 1 (lines 55-60) in the revised manuscript.

3. Glaciation

The authors provide a glacier mask in the supplementary and compare this mask to FCI. The glacier mask is not including any glaciations in the European Alps during LGM or 1850 (Little Ice Age). On which scientific basis is the map derived? Why are the authors comparing the spatial distribution of FCI with their glacier mask? When a glacier is there, then there is no frost cracking as the rock is disconnected to atmospheric processes (Grämiger et al., 2018). By not including a glacial cover, the authors are overestimating the FCI by far.

The paleoclimate time-slice specific glacier masks (ice cover reconstructions) are the same as those that serve as input for the Mutz et al. (2018) simulations. In turn, Mutz et al. (2018) follow the experiment protocols of the Paleoclimate Modelling Intercomparison Project (PMIP) (Bracannot et al. 2012), also described on the PMIP3 wiki (wiki.lsce.ipsl.fr/pmip3/doku.php/pmip3:index). This mask describes a synthesis of significant ice sheet cover, but does not account for different reconstructions or all regions covered by glaciers at the time. While PMIP4 has moved to the use of several different ice sheet and mask reconstructions for the LGM (Kageyama et al., 2021), PMIP3 use one reconstruction that comprises a synthesis of several reconstructions of significant ice cover (that would significantly impact climate). This is described in detail in Abe-Ouchi et al. (2015), which we list as the source for our boundary conditions in Table. 1 in the manuscript. We have also mentioned the source of glacier masks in the captions of Fig. 6-10 in revised manuscript.

Furthermore, in the revised manuscript, the glaciated regions in different time-slices were masked from the FCI results in different time-slices in all the three models. Also, in Pliocene simulations, regions which experienced glaciation during the Pleistocene, have been removed from the results. We do this, as the assumption of constant soil cover is heavily violated in regions that experienced Pleistocene glaciation (as mentioned in RC2.1). The glacier covers for respective Cenozoic time-slices have been masked and highlighted with distinct color (violet) in Fig. 6-10 and supplement Fig. 1-2.

Note that we were comparing the FCI difference maps to the glacier mask, as a comparison of FCI inside and outside the ice sheets in different time-slices creates confusion. This has been rectified in the revised manuscript, where we have applied the glacier mask to the FCI difference maps too (Fig. 8-10). Hence, we do not compare glacier mask and FCI difference maps in the revised manuscript. Hopefully these modifications and explanations meet with the reviewer's approval.

4. Scale issues

The authors use a simple bottom-up approach to model frost cracking for different time slices. They have no independent data that they could use to validate their models. Consequently, they have a problem to discuss their own results and put them into a perspective. They compare a 80 x 80 km model for North America and Alaska for PI, MG and PLIO to a frost cracking studies at Jungfraujoch that measured frost weathering using acoustic emissions on one rockwall at 3500 m for 4 days (Amitrano et al., 2012) or one year (Girard et al., 2013). I cannot see how these studies support the author's results on much larger scale at different time steps in the past in completely different environments. Furthermore, the author states that their model results at higher Asia and Alaska during LGM are consistent to periglacial processes observed in Oregon (Marshall et al., 2015; Marshall et al., 2017). I cannot see the context between periglacial conditions and landforms in Oregon and the author's observed FCI in other areas of the Earth. These are just a few examples but the whole discussion shows no argumentation. Model results will be compared to models from Hales and Roering (2007) or Andersen et al. (2015), which are used to derive the same model results.

The reviewer makes a very valid point regarding scale differences. We have revised the data comparison section 5.3 to exclude the short-term studies and focus the revised section on discussion of broader trends and their comparison to more comparable findings by previous studies. Additionally, we highlight the need for scale-bridging for regional studies (lines 531-533 in the limitation section).

References cited in our response:

Anderson, R. S.: Near-surface Thermal Profiles in Alpine Bedrock: Implications for the Frost Weathering of Rock, Arctic and Alpine Research, 30, 362–372, https://doi.org/10.1080/00040851.1998.12002911, 1998.

Abe-Ouchi, A., Saito, F., Kageyama, M., Braconnot, P., Harrison, S. P., Lambeck, K., Otto-Bliesner, B. L., Peltier, W. R., Tarasov, L., Peterschmitt, J.-Y., and Takahashi, K.: Ice-sheet configuration in the CMIP5/PMIP3 Last Glacial Maximum experiments, Geosci. Model Dev., 8, 3621–3637, https://doi.org/10.5194/gmd-8-3621-2015, 2015.

Braconnot et al, Evaluation of climate models using palaeoclimatic data, Nature Climate Change 2, 417-424 (2012), doi:10.1038/nclimate1456

Hales, T. C. and Roering, J. J.: Climatic controls on frost cracking and implications for the evolution of bedrock landscapes, J. Geophys. Res., 112, F02033, https://doi.org/10.1029/2006JF000616, 2007.

Kageyama, M., Harrison, S. P., Kapsch, M.-L., Lofverstrom, M., Lora, J. M., Mikolajewicz, U., Sherriff-Tadano, S., Vadsaria, T., Abe-Ouchi, A., Bouttes, N., Chandan, D., Gregoire, L. J., Ivanovic, R. F., Izumi, K., LeGrande, A. N., Lhardy, F., Lohmann, G., Morozova, P. A., Ohgaito, R., Paul, A., Peltier, W. R., Poulsen, C. J., Quiquet, A., Roche, D. M., Shi, X., Tierney, J. E., Valdes, P. J., Volodin, E., and Zhu, J.: The PMIP4 Last Glacial Maximum experiments: preliminary results and comparison with the PMIP3 simulations, Clim. Past, 17, 1065–1089, https://doi.org/10.5194/cp-17-1065-2021, 2021.

Mutz, S. G. and Ehlers, T. A.: Detection and explanation of spatiotemporal patterns in Late Cenozoic palaeoclimate change relevant to Earth surface processes, Earth Surf. Dynam., 7, 663–679, https://doi.org/10.5194/esurf-7-663-2019, 2019.

Mutz, S. G., Ehlers, T. A., Werner, M., Lohmann, G., Stepanek, C., and Li, J.: Estimates of late Cenozoic climate change relevant to Earth surface processes in tectonically active orogens, Earth Surf. Dynam., 6, 271–301, https://doi.org/10.5194/esurf-6-271-2018, 2018.

Taylor, K.E., R.J. Stouffer, G.A. Meehl: An Overview of CMIP5 and the experiment design." Bull. Amer. Meteor. Soc., 93, 485-498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Wang, X., Schmidt, B., Otto, M., Ehlers, T. A., Mutz, S. G., Botsyun, S., and Scherer, D. Sensitivity of water balance in the Qaidam Basin to the mid-Pliocene climate. Journal of Geophysical Research: Atmospheres, 126, e2020JD033965. https://doi.org/10.1029/2020JD033965, 2021

Werner, C., Schmid, M., Ehlers, T. A., Fuentes-Espoz, J. P., Steinkamp, J., Forrest, M., Liakka, J., Maldonado, A., and Hickler, T.: Effect of changing vegetation and precipitation on denudation – Part 1: Predicted vegetation composition and cover over the last 21 thousand years along the Coastal Cordillera of Chile, Earth Surf. Dynam., 6, 829–858, https://doi.org/10.5194/esurf-6-829-2018, 2018.