Dear Reviewer,

We appreciate your detailed and insightful comments to improve our manuscript. This reply is not a paper revision. Here, we would like to describe how we intend to address your primary suggestions and answer the most critical questions. Please see our individual responses below.

Revier's comments	Our Responses
The manuscript "Estimating return intervals for extreme climate conditions related to winter disasters and livestock mortality in Mongolia" analyzed the return periods/levels of the extreme events on summer drought and winter temperature, key factors of dzud (winter weather disaster) in Mongolia. The analysis was carried out by the Extreme Value Analysis framework (i.e. fitting to two types of models). In doing so, length of the period covered by the data is attempted to extend by using tree-ling and Siberian station records, which is the author insist helpful to let the results more reliable. Although I found potential of contribution by the current study, explanation of motivation and technical detail is not enough in the current manuscript. Before consideration for acceptance, I think substantial revision is needed.	Thank you very much for your detailed comments.
(1) Purpose of presenting the ARIMA model and the models with climate covariants is not clear. The authors say "establishing a relationship between drought conditions and climate variables, particularly precipitation and snow, is useful in understanding the dynamics that determine dry conditions". However, it is obvious that PDSI is dependent on precipitation, and for understanding the dynamics the conditions in the preceding some months should be analyzed. For winter temperature, the reason why the dependence of the return level on the Siberia's data should be considered is unclear. It may be better to first convert to the values in Mongolia (e.g., by regression	Thank you very much for pointing out important points in our paper. In our ARIMA (p, d, q) model, we analyzed the following parameters: p: the number of lag observations in the model d: the number of times that the raw observations are differenced q: the size of the moving average window; These parameters are shown in the selected model in Line 179-180 as the following: "The order of the best ARIMA models in each cluster is (3,0,0) for the Southwest, (1,0,2) for the Northwest, and (1,0,0) for the East." Please note that the tree-ring PDSI values used in this study are annual data. We used Siberia (Irkutsk) data since the gridded climate database covers Mongolia starts after 1901, but if we can use Siberia (Irkutsk) data, we can get back to 1820.

models) and then to input the converted values to GEV and GPD than to directly consider the Irkutsk's data in GEV and GPD.	The rationale for choosing our method to use the Siberia data is as follows. First, Siberia is a good proxy for Mongolia. Existing studies, such as Munkhjargal et al. (2020), Iijima et al. (2018), and He (2017), suggest the winter temperatures between Mongolia and Siberia are correlated spatially, which are driven by polar jet dynamics. Second, the relationship between these two variables is robust using bootstrap from Irkutsk data.
(2) Discussion on whether the reliability has been improved compared to the case where data only after 1940 is used is needed.	Our studies extended the time horizon by using tree-ring and proxy data. We can compare an analysis based on the 1940-data with the complete data in the revised version.
(3) Stationarity means there is no impact by the global warming?	Though we don't specify the influence of global warming on stationarity in our data, Batima (2006) analyzes climate trends and implications for impacts on livestock mortality in Mongolia. Batima (2006) reveals that the trend for increasing incidence of drought in the growing season is statistically significant at the 95% level and contributes to the higher mortality rates in 2000-2003.
(4) English should be checked by native editors.	Thank you very much for pointing out these parts. We will certainly address these parts in the revised version.
Page (P) 1 Lines (Ls) 23: In the manuscript, the usages of "return level" and "return period" are confusing. (see also Ls78, 101, 103, 116, 122, 124, 286 (and Fig 5), 297)	We will fix these in the revised version.
P1 L32: "thus": I could not understand how the previous sentence supports this sentence.	We will clarify our logical flow in the revised version.
P2 Ls33-34: "Here ---insurance" These are not treated effectively by the current study. If the authors want to argue this, reorganization of the manuscript is needed.	We will play down the contribution of our study to the insurance sector.
P2 Ls 38-40: A total ---in Mongolia: present the source (literature).	Accordingly, we will add the source, Middleton et al. (2015).
P2 L49: Tachiiri et al (2008) used snow amount (SWE), not temperature, isn't it?	Thank you very much for pointing out this. We meant that we referred Tachiiri et al. (2008) for "precipitation." But it might not be clear. We will accordingly cite the study.
P2 Ls63-66: I could not understand the advantages of the index-based insurance compared to the livestock-loss-based one.	The advantage of index-based insurance is that payment is faster than loss-based insurance because payment will be made once the predetermined index, such as precipitation amount or temperature, exceeds the threshold. In contrast to loss-based insurance, an insurance company must assess losses before making payment, which requires labor and time. We will explain this adequately.
P3 L69: "few studies" Do you mean there are some examples? If so it is better to briefly introduce them.	Thank you very much for pointing out this. We will cite a few relevant papers, such as Hessl et al. (2018).

P3 L74: In Bayasgalan et al. (2009), I could understand the mechanism how drought becomes more frequent in the future, but why dzud will be more frequent is not clear.	In addition to Bayasgalan et al. (2009), Munkhjargal et al. (2020) discuss the increased frequency of dzud while analyzing the interdecadal variability of extreme cold air outbreak events. Also, Cohen et al. (2010) analyzed the 2009-2010 winter and found that the cold winter of 2009-2010 in Eurasia was more associated with an extreme Arctic Oscillation (AO) event than El Niño/Southern Oscillation (ENSO). We will elaborate more on this in the revised version by following Munkhjargal et al. (2020), Cohen et al. (2010), and Iijima et al. (2018).
P3 L76: For the future, models are used?	We refer to Bayasgalan et al. (2009) here to highlight that their results show that dzud is projected to increase in terms of frequency and magnitude using their model. In revising the paper, we will mention using CMIP 5 or 6 models for future assessments.
P3 L88: Does "distribution" here mean probabilistic distribution, or spatial distribution?	We meant a probabilistic distribution. We will clarify this in our revised manuscript.
P3 L100 Snowfall in winter is also considered to be important, isn't it?	As cited in our paper, Rao et al. 's 2015 model is based on winter temperature, summer drought, summer precipitation, and summer potential evaporanspiration. Their model explains 48.4% of the total variability in the mortality dataset.
P4 Ls108-9: "mortality assumes that the size of the population does not matter." I do not agree with that, as in my understanding it is possible that mortality depends on the size. Also, this does not explain why analyzing climate variables is more effective to moderate the damage of dzud than analyzing mortality or loss. If the authors consider mortality is not a good indicator, how about number of livestock killed?	We agree with you as we wrote in the following sentence below: "In fact, changes in livestock populations also matter since they can also be related to changes in socio-economic factors, such as shortage of food supply, which can be related to non-climate factors." What we meant here was that mortality "rate" assumes that the size of the population does not matter. We will clarify this in the revised version.
P4 Ls110-5: If livestock death depends not only on climatic factors, but also on socio-economic factors, index insurance is not reasonable, is it?	Index insurance requires a pre-determined index which other parties cannot manipulate. In this sense, climate factors are a good candidate for an index. Socio-economic factors would determine the target populations to be covered. Index insurance can be applied at many scales. Here, we consider the Government of Mongolia as the purchaser of the insurance at an aggregate level, and the amount of coverage they would seek would indeed be informed by their assessment of the socio-economic factors. If the index insurance targeted individual herders, then the coverage purchased per herder would relate to their socio-economic factors. However, we considered the country scale, since the government is expected to help during such an emergency, index insurance would provide rapidly accessible capital for relief programs, and the government planning process can account for socio-economic factors as part of a delivery mechanism for the relief of the population.

P4 L119: "for the future": is the future discussed in the manuscript?	In our paper, this "the future" means the future design of index insurance, for example, as discussed in the other parts of the paper.
P4 L130: What do "many opportunities" mean here?	Many opportunities mean the following sentence, "This has the advantage of reducing the bias in the near-term projection, assessment of the return period, and recurrence interval associated with the event."
P5 L147: I could not understand what "because …" is for.	We will clarify this in our revised manuscript.
P5 L150: "Kaheil and Lall" should be "Lall and Kaheil (2011)"? I could not access Lall and Kaheil (2011). Add more information on this literature in the reference list so that the readers can access that when needed.	The suitable references are the following: Lall, U., Devineni, N., & Kaheil, Y. (2016). An empirical, nonparametric simulator for multivariate random variables with differing marginal densities and nonlinear dependence with hydroclimatic applications. Risk Analysis, 36(1), 57-73; and Y. Kaheil and U. Lall, 2011, Investigation of Climate Impact on Mongolian Livestock Mortality, Report to the World Bank, Project TF94002. We will cite them properly.
P5 Ls151-2: "which were adjusted ….in growing seasons": More information on how they were done is needed.	The clustering was initially made in Lall, U., Devineni, N., & Kaheil, Y. (2016). Their spatial clusters were identified using hierarchical clustering (Johnson, S. C. (1967). We will put a brief explanation about this.
P5 L154: "to improve risk analysis of Dzud and mortality of livestock in Mongolia": Is this consistent with Ls 89-91?	This is consistent. We meant that in Line 89 – 91, our risk analysis is not the one with a probability of an event x consequent losses/damages. We focus on the probability analysis as we write in Line 89-90.
P6 Ls161-2: The Mann-Kendall test is for each time series? If so each row of Table 1 is a pair of two clusters, to show the results of the Mann-Kendall test in the right-most column of Table 1 is not appropriate?	Thank you very much for pointing out this. These two columns should be separated. The right column should indicate Mann-Kendall values for Southwest, East, and Northwest, respectively. We will revise it accordingly.
Ps7-8 Ls190-191: how they are "scaled"?	We will adequately explain this in the note in the revised version.
P8 Table 2: how fine is "high resolution" of WMO and CRU data?	Both datasets has high-resolution of 0.5 x 0.5-degree resolution. We will add this in the revised version.
P9 Eq 1 (and Eq.3): What does" +" in the last term mean? Also, there is no y_+ in Eq 1 (may be z is so?).	y_+ means max {y, 0}, meaning that if y is negative, choose zero, otherwise choose y. Then, in equation 1, "+" indicates the same meaning. If the inside of the parentheses is negative, take zero.
P9 Eq. (5): Zt is equivalent to H(x) in Eq.3? (same in Eqs 6-7)	H(x) and G(z) are the distribution "functions". Equation (6) means that Zt is a random variable Z following the distribution GEV with parameters, μ(t), σ(t), ε(t), which are functions of time.
P10 L223: it is better to move Section 3.1 to Methodology section.	We put Section 3.1 here since it is relevant to only Section 3, not Section 4.

P10 L224: How "fit" was done?	Fitting was done following Section 2.
Ps10-11 L244-254: The oscillation shown in Fig 2 is not reflected in the models? Why? Also, I want to know how good fit was obtained by the models. (for other models too)	Time components are captured here by taking AR models. The fit was measured by BIC, as shown here.
P10 L246: “sigma = 0.95+0.002t”: Do you mean sigma=exp(0.95+0.002t)?	Yes, thank you for pointing out this. It was a typo.
P11 Table 3: AR(3): Why order (3) is selected as the best model? By what mechanism?	Autoregressive-Moving-Average (ARMA) models with different orders are evaluated based on the BIC.
P12 L260: "could be a real feature or an artifact": Which is more likely here?	The nonstationarity is in the scale parameter for the Southwest, with a mean coefficient of 0.002 relative to the constant value of approximately 0.9, so over 100 years the variability could increase from 0.9 to 1.05, if we take 0.9 to be a mid-period estimate, which is rather a modest change. This could reflect changes in the system dynamics, or it could reflect changes in the observational process.
P12 L266: Why not 1902?	Our analysis starts in 1903 because the AO index starts in 1903. We will revise the description adequately in the revised version.
P18 Table 7: Why the difference between the values for 10,50,100 years are so small? Also if PDSI of around 4 is returned every 10 years, the threshold of 1 is not too small as catastrophes?	The difference between the values for 10, 50, and 100 years is slight because the shape parameters estimated from the GEV for each case are negative. This means that the data is negatively skewed and this leads to an implicit upper bound for the process. As a result, each of the quantiles is restricted by that upper bound and ends up quite close to each other. For the second part, the threshold of 1 corresponds to 0.21 upper quantile for southwest and 0.26 for both northwest and east, corresponding to 4-5 years return levels. Setting these return levels as thresholds is of interest in terms of social concern.
P18 Ls346-7: "All the results show that…" What meaning does it have in light of the purpose of the study? (same for Ls419-20)	This sentence means that we identified which distributions the PDSI values follow. This information is necessary to estimate return periods of extreme drought and cold temperature, which are objectives of this study in Line 122 – 125.
P19 L378: Why seasonality should be removed?	If both series data have seasonality or periodicity, the correlation between them will be high just for that reason. Removing seasonality helps us identify if the anomalies from the periodic behavior are correlated, or namely if they share similar dynamics in effects induced by atmospheric circulation beyond the seasonal cycle.
P19 L380: "correlation" with what should be clarified.	Here, the correlation is between the temperature data in Mongolia and Siberia.
P20 Ls391-3: The text is not consistent with Table 8.	We apologize for the typo. The correct one should be the following: "Models with Siberia data both in the location and scale parameter are the lowest BIC for the Northwest. For the Southwest and East, the one with Siberia data in the

	location parameter and constant in the scale parameter shows the lowest BIC (Table 8)"
P21 L408: How the threshold 20 was determined should also be explained.	As we explained in supplement S1, selecting a threshold is made by looking at the threshold plot of the distribution, similar to Fig S3 and FigS4. We will add these figures in the supplemental in the revised version.
P22 Fig 10: Why the uncertainty is very small for East and NW?	Thank you very much for raising this point. In the revised version, we will add a description more on these uncertainties, including high uncertainty for SW.
P22 Ls425-8: "We use the GEV model because the winter minimum temperature data is a single extreme value and that the GEV model is suitable for maxima and minima of block data." then why you also presented the result of the GPD?	Thank you very much for raising this point. Since it may cause confusion, we will remove the sentence in the revised version.
P23 Fig 11: What is the cause of the variations? Siberian data?	The variations come from both the statistical properties of the Mongolia data itself and variations attributed to Siberia data.
P23 L440: what does "self-calibrated" mean here?	Self-calibrated means that the index automatically calibrates the behavior of the index at any location by replacing empirically constants in the index computation with dynamically calculated values.
P23 L449: "can be used to improve the risk calculations for livestock index insurance in Mongolia" How? (also in P 24 L 463)	Insurance is priced considering three main factors. The first is the probability of exceeding the threshold at which the payout occurs. This is called the fair premium, and it is equal to the payout*annual probability of exceedance of the threshold. The second term is the company's transaction costs and profit. The third term is related to the uncertainty associated with the estimation of the probability. This term reflects the estimation of the uncertainty, which is reduced as the length of record (in this case from the paleoclimate extension) increases, and it also incorporates some consideration of systematic non-random variations that may lead to clustering in payouts, e.g., 7 years wet followed by 7 years dry. An assessment of these variations also contributes to the price associated with uncertainty. Typically, for the record lengths available, the uncertainty term can contribute more to the premium than the other two terms. So far, no one in the industry is using paleoclimatic information to extend and reduce coverage costs, but there is interest in using it to understand the clustering of payouts.
P23 Ls453-4: Is it the case where rainfall is included as a covariate?	Yes, if you mean by this sentence, "The GEV model also shows that the return levels of drought conditions are changing over time and variability is increasing for all the regions." As shown in table 4, we use models with location parameters, including precipitation data.
P24 L462: "through early warning systems" how the lead time is considered?	We meant here that our results could provide implications for designing early warning systems. The lead time of the early warning systems is out of the scope of our paper.
P24 Ls462-3: "the estimation of extreme value distributions and return levels has the potential to	Our study estimates the return intervals and underlying probabilistic characteristics of the climate variables.

improve livestock index insurance": How?	Index insurance requires a proper threshold and the understanding of underlying distributions of risk events. Thus, our study would improve the design of index insurance.
SI P 1 L7; Whereabouts is the "inflection point" in the Figure?	The inflection point is shown around 0-1 in Figure S.4. We subjectively select 1 as explained in Gilleland and Katz (2016).
SI P 1 L14:" Therefore, it is reasonable to use a threshold of 1.0." I could not understand why you could say so from the previous sentence.	We try to capture the upper quantile by this threshold. Our selection of this threshold leads to 0.21 and 0.26 quantile of distributions, which corresponds to 4-5 years return levels. Looking at events above these return levels are of our interest.
SI P1 Fig S1: What causes high ACF for x > 20 (particularly for East). Also, add explanation of the dotted line near y=0.1.	We don't know the reason for this.
SI P2 Fig S3: Add a bit more explanation on what the figure shows. What is reparameterization? Why the sign was changed by that?	Fig S3 is used to choose a threshold for GPD following Gilleland and Katz (2016). Fig S3 repeatedly fits the GP distribution function to the data to plot a sequence of threshold choices with some variability information. A subjective selection of 1 as a threshold appears to yield estimates that will not change much. Also, this selection is made based on Fig S4 and the theoretical and practical justification (which means that a threshold of 1 corresponds to a 4- 5 year return level, which is of social interests.) Reparametarization means here the scale parameter is adjusted so that it is not a function of threshold (Gilleland and Katz (2016)). We will provide a brief note of the technical details in the revised version.
SI Ps3-4; Add more explanation on Figs S5 and S6 (on the red marks, red curves, numbers at the center of the boxes etc).	Red marks show statistical significance, while red curves show the smoothed curve.
SI P6 Table S4: Add information on the statistical significance.	We will add them adequately in the revised version.
Typos etc	
P3 L84: extraordinarily -> extraordinary?	Yes. Thank you very much for pointing out this.
P7 L179: ARIMA is not spelled out. Also, a brief description of (p, d, q) (like (3,0,0)) would be helpful.	We will adequately correct the typo.
P9 L210: explanation after "where" is confusing. P16: Figure 6. Add unit for the x-axis.	We will properly correct the typo.
P16 L320: chosen -> chose	We will adequately correct the typo.
P17 L328: PDSI -> for PDSI?	We will properly correct the typo.

P23 L448: Add "for GPD" after "the upper-bound Beta distribution".	We will adequately correct the typo.
SI P1 L5: delete one "Figure S.4"	We will properly correct the typo.

Reference

Batima, P. (2006). Climate change vulnerability and adaptation in the livestock sector of Mongolia. International START Secretariat: Washington DC, 204-209.

Cohen, J., Foster, J., Barlow, M., Saito, K., & Jones, J. (2010). Winter 2009–2010: A case study of an extreme Arctic Oscillation event. Geophysical Research Letters, 37(17).

Gilleland, E., & Katz, R. W. (2016). extRemes 2.0: An extreme value analysis package in R. Journal of Statistical Software, 72(1), 1-39.

Hessl, A. E., Anchukaitis, K. J., Jelsema, C., Cook, B., Byambasuren, O., Leland, C., ... & Hayles, L. A. (2018). Past and future drought in Mongolia. Science Advances, 4(3), e1701832.

He, S., Gao, Y., Li, F., Wang, H., & He, Y. (2017). Impact of Arctic Oscillation on the East Asian climate: A review. Earth-Science Reviews, 164, 48-62.

Iijima, Y., & Hori, M. E. (2018). Cold air formation and advection over Eurasia during “dzud” cold disaster winters in Mongolia. Natural Hazards, 92(1), 45-56.

Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32(3), 241-254

Middleton, N., Rueff, H., Sternberg, T., Batbuyan, B., & Thomas, D. (2015). Explaining spatial variations in climate hazard impacts in western Mongolia. Landscape Ecology, 30(1), 91-107.

Y. Kaheil and U. Lall, 2011, Investigation of Climate Impact on Mongolian Livestock Mortality, Report to the World Bank, Project TF94002.

Munkhjargal, E., Shinoda, M., Iijima, Y., & Nandintseteseg, B. (2020). Recently increased cold air outbreaks over Mongolia and their specific synoptic pattern. International Journal of Climatology, 40(13), 5502-5514.