Reponses to Reviewer #2

We greatly appreciate your comments and suggestions. We have make a careful plan to revise the manuscript according to what you have pointed out and believe that the quality of this manuscript will be improved as a result of the revision. We have included our detailed responses to each of your comments raised.

[General comments]

"Probabilistic risk assessment of livestock snow disasters in the Qinghai-Tibetan Plateau" by Ye et al. applies boosted regression tree and general additive modeling methods to the snow disasters in the Qinghai-Tibetan Plateau, as an event-based evaluation, and the results are basically consistent with existing studies. The research topic is within the scope of the journal, but there are some substantial flaws in the study, which should be addressed. The major concerns are:

1: The advantage of event-based evaluation is not clear. Rather than hay preparation based on event-based analysis of this study, that based on annual-based analysis will work well when the intervals between the two events are very short (as time to prepare for the next event is not enough, preparation for annual basis is better, particularly when modeled annual frequency > 1).

Page 19 Sect. 4.2: As "more than one snow disaster a year is unlikely", annual evaluation is enough, isn't it? For me it looks to prepare hay based on annual evaluation is OK.

2. The authors say probabilistic analysis is one of the advantages of the study. However I consider a year-by-year evaluation is more sophisticated, and the probabilistic analysis used here is not necessarily an advantage, but a result of taking a simple evaluation dealing with what should be separately treated as one set of data. An effective PRA would be a result from a probabilistic function, not a result from treating various conditions as one case.

RE: Above questions are inter-related and are therefore responded together.

We totally agree that annual based modeling has its own advantages, but the event-based approach is also very important. By nature, livestock snow disaster can occur multiple times in a winter, and the losses would accumulate event by event. Capturing the details of event occurrence and intensity are important for following aspects:

1) ln*LR* (natural logarithm of loss rate) shows a concave relationship with disaster duration (can be found later in details in response to your comment #4). Therefore, the total loss of one event lasting 30 days and two-event in one winter lasting 15 days each will be totally different. Knowing only the aggregate duration in a year cannot reflect such important details.

2) Although presently our data shows that historically it has been less likely to have more than 1 snow disaster in a winter for most parts of the Qinghai-Tibet. But if we are considering a method that can be generalized, then we need to consider the possibility of changing frequency and intensity, and the capacity to model it, particularly with the observed and

projected slight increase in precipitation on the Plateau (Kuang and Jiao, 2016; GUO, SUN and YU, 2018). Using the annual loss approach cannot capture such changes.

3) From the perspective of risk-informed action, annual evaluation (i.e. potential aggregate duration) would be temporarily sufficient for preparedness by the government and community. But it can hardly work for risk-transfer mechanisms as insurance schemes were mostly based on events. This is also the critical reason that catastrophe risk models are mostly built on event-basis (Michel-Kerjan *et al.*, 2013).

For the advantage of this study, we totally agree with you that a year-by-year evaluation is more sophisticated (than a simple statistical analysis). Our probabilistic analysis was built on event-by-event loss modeling, which is even more sophisticated than year-by-year evaluation. Benefited from your comment, we plan to add these discussions into the introduction section and discussion section to highlight the advantage of event-based modeling instead of probabilistic analysis.

3: To evaluate livestock number by carrying capacity, and to evaluate carrying capacity by grassland type is questionable. The appropriateness of them should be more carefully discussed.

Page 8 Lines 18-20: More careful discussion on the validity of this method is needed. I want to see the scatterplot of observed and estimated values.

Page 8 Lines 21-22: More careful discussion on the validity of this method is needed. I want to see the scatterplot of observed and estimated values.

Page 21 Lines 19-20: More careful evaluation is needed for the performance of herd size estimation.

RE: Thank you very much for your comment. Above comments are related and therefore responded together.

1) The appropriateness of evaluating livestock number by carrying capacity

We first need to clarify that we are trying to use carrying capacity to evaluate livestock number EXPOSED to snow disaster, but not the total livestock number. In the study area, only livestock grazing on grassland in pastoral and agro-pastoral regions (central to western part of Tibet Plateau) are exposed to snow disaster. Livestock kept in livestock farms in agricultural regions (mostly the eastern and low altitude parts of the Plateau) are not. When estimating exposure, we were trying to estimate the livestock number grazing on grassland.

The livestock number grazing on grassland is essentially determined by livestock carrying capacity, which exactly defines the maximum livestock number that can graze in a specific area of grassland, given local forage productivity and grazing style. Since 2011, the Tibet Autonomous region started the program of " subsidy and award policies for grassland ecological conservation efforts" to reduce the number of livestock and conserve grassland. In 2014-15, news articles intensively reported that "Tibet has basically reached forage-livestock balance" (government release, http://www.gov.cn/xinwen/2014-05/12/content_2677946.htm, in Chinese). In other words, livestock number grazing on grassland is highly consistent to the carrying capacity officially designated.

2) The appropriateness of evaluating carrying capacity by grassland type

Evaluating carrying capacity by grassland type is the way that adopted by both official calculation technical manual (*Ministry Standard of Calculation of Rangeland Carrying Capacity issued by Ministry of Agriculture of China*, NY/T 635-2015) and academic studies regarding forage-livestock balance, i.e. (Xin *et al.*, 2011). Grassland type essentially determines the key values of evaluating carrying capacity, i.e. forage regrowth percentage, proper utilization rate of rangeland (of different grazing seasons), conversion coefficient of standard hay. Using grassland type as a identifier of different carrying capacity is also the widely adopted way of presenting the evaluation results.

3) Verification of livestock number estimated

We obtained official release of carrying capacity by Tibet Autonomous Region from *Statistical Materials of Grassland Resource and Ecology of Tibet Autonomous Region* published by the Department of Agricultural and Pastoral of Tibet Autonomous Region in year 2011. We used zonal statistics to derive county-level carrying capacity estimated by our method, and compared with the official release (Figure 1).

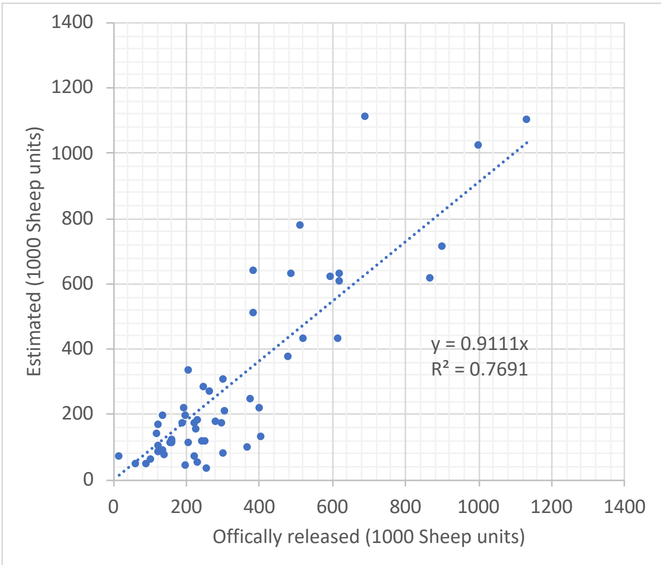

Figure 1 Estimated and officially released livestock carrying capacity in Tibet Autonomous Region

As the above figure shows, the estimated and officially released livestock carrying capacity basically agree with each other.

In the revision, we plan to 1) clarify that we are estimating the herd size exposed to snow disaster (livestock number on free grazing) rather than total herd size; and 2) further explain the appropriateness of estimating livestock number of carrying capacity, and estimating carrying capacity by grassland type.

4. Explanation and discussion on Eq. 1 is not sufficient. The form of functions for each term in the right hand side, and the performance of the equation should be clearly presented.
Page 4 Line 10 (eq 1): Present the detailed forms of s(Duration), s(Wind), s(P) and s(GDP). (all parameters of spline curves)

Page 4 Line 14: More detailed performance check is needed (not for lnLR, but for LR). How large is RMSE? Is the error random or systematic? I want to see the scatterplot of observed and modelled values.

RE: Thank you for these questions and comments. In light of the suggestion from Reviewer #1 (please refer to Reviewer#1 Comments 5,6, and 9), we have updated the model by using value added of animal husbandry of the underlying county instead of GDP as the indicator of prevention capacity. We have updated Eq(1) accordingly to: $\ln LR = s(Duration) + s(Wind) + s(P) + s(Value\_Add)$. Details of the updated model are provided below.

(1) Model-fitting statistic:

```
Family: gaussian
Link function: identity

Formula:
lnLR ~ s(Duration) + s(Value_Add, k = 4) + s(maxWind) + s(P)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.5300     0.1708   3.103  0.00287 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
               edf Ref.df     F  p-value
s(Duration)  5.626  6.705 9.706 1.96e-08 ***
s(Value_Add) 1.000  1.000 8.878  0.00407 **
s(maxWind)   2.732  3.440 2.747  0.04463 *
s(P)         1.000  1.000 2.809  0.09864 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.563   Deviance explained = 62.5%
GCV = 2.5508  Scale est. = 2.1593    n = 74
```

Figure 2 Fitting statistic of the updated GAM model

(2) The response curves (spline curves)

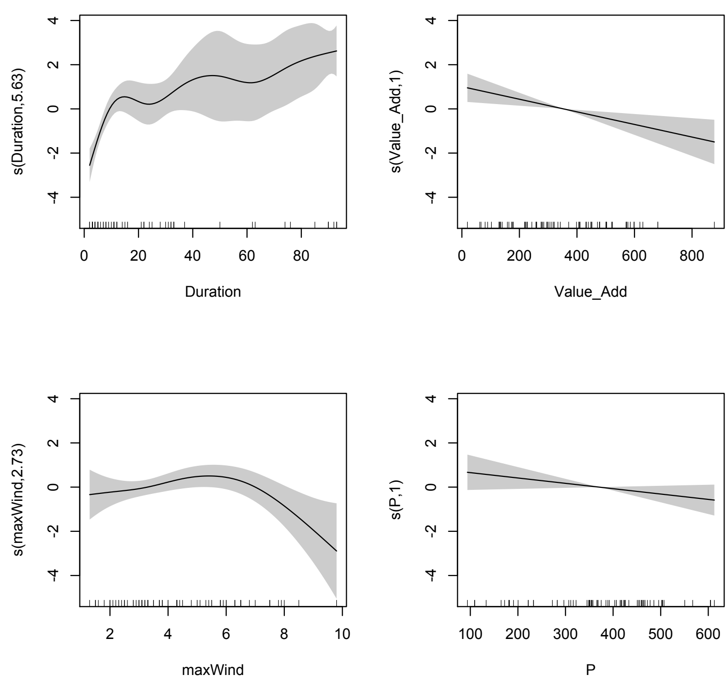

Figure 3 Response curves of the updated GAM model

Its response curves (Figure 3) indicate that: (1) ln*LR* is increasing with snow disaster duration. Duration up to 15-18 d is a critical period that mortality will increase rapidly. (2) ln*LR* decreases with value added of animal husbandry (*Value_Add*), indicating the effect of stronger prevention capacity in reducing mortality, i.e. government expenditure in reserving hay for preparedness, and subsidy to herders to build/enlarge warm sheds. (3) An inverted-U shaped relationship between daily maximum wind speed and ln*LR*. The up-slope part indicates the increasing stress of stronger wind on livestock, but the down-slope part (beyond 5-6 m/s) indicate herder's reaction to stop free-grazing and keep herds in shelters (Wu *et al.*, 2007). (4) ln*LR* decreases with growing season precipitation. Larger growing season precipitation indicates more abundant food for livestock in summer, and therefore better body-condition in resisting low temperature and lack of food in snow disaster times.

(3) Model performance diagnostics

We performed 10-fold cross validation, and found RMSE, MAE and ME for the model were 1.747, 1.325, and -0.002, respectively.

The performance diagnostics charts of the model (Figure 4) indicate that 1) QQ plot is very close to a straight line, suggesting our distributional assumption of normality about ln*LR* is reasonable. 2) The variance is approximately constant as the mean increases. 3) The histogram of residuals appears approximately consistent with normality. 4) the response against fitted values show a positive linear relationship in the scatter plot.

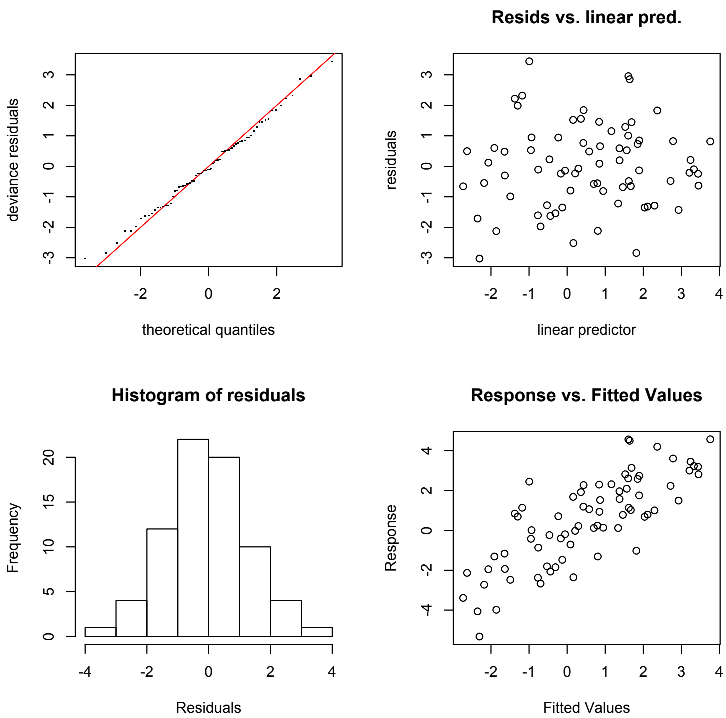

Figure 4 Performance diagnostics charts for the updated GAM model

In the revision, we shall supply more details about the model fitted in to the manuscript. In addition, model details, including the fitting statistics, response curves, and performance diagnostics charts (Figure 2~Figure 4) will be provided in supplementary document for reference.

[Specific comments]

Page 3 Line 1: Do you mean that the final metrics should always be mortality, not mortality rate? It looks the study considered mortality rate as the final metrics, and I think each study can use its own final metric.

RE: Thank you for your comment. The sentence was misleading. In the revision, we plan to delete the latter part of this sentence and kept on "*Compared to earlier works, they successfully extended the framework to future climate change analysis*".

Page 3 Line 13: Provide the rationale of "some of the highest livestock snow disasters". What is "highest", by the way? Largest damage? Highest frequency?

RE: We intended to say "one of the regions with the highest risk" in terms of both high frequency and large damage. In the revision, we plan to this sentence to "*…the QTP is one of the regions that has suffered the most from livestock snow disasters due to its large area of snow cover area, long-lasting snow cover days, and nomadic grazing.* "

Page 3 Line 14: Provide literature for "This region is also a hot spot in climate change".

RE: References will be provided into the text as suggested in the revision, i.e. (Diffenbaugh and Giorgi, 2012; Gu *et al.*, 2014).

Diffenbaugh, N. S. and Giorgi, F.: Climate change hotspots in the CMIP5 global climate model ensemble., Clim. Change, 114(3–4), 813–822, doi:10.1007/s10584-012-0570-x, 2012.

Gu, H., Yu, Z., Wang, J., Ju, Q., Yang, C. and Fan, C.: Climate change hotspots identification in China through the CMIP5 global climate model ensemble, Adv. Meteorol., 2014, doi:10.1155/2014/963196, 2014.

Page 4 Fig 1: It is difficult to understand the relationship between Timing, and Duration and Wind speed from the figure.

RE: The figure presented in the online document was not complete due to unknown technical reasons when generating .pdf files. The correct figure should be:

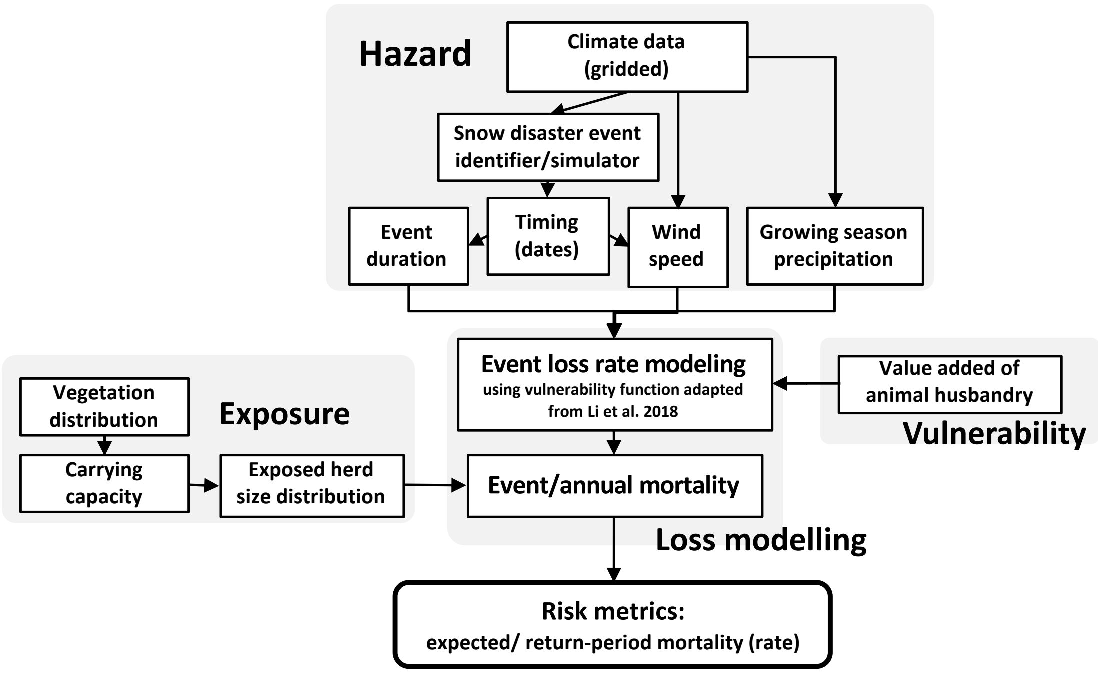

Figure 5 Corrected technical flow of the study (Fig 1 in the manuscript)

In the modeling process, we derive timing (both starting and ending dates) from the "snow disaster event identifier/ simulator". Then the duration of the event (days between the starting and ending dates) and the wind speed during this period can be derived (as we have daily wind speed data). In the revision, we will correct the figure and add necessary explanation about the flow.

Page 4 Sect. 2.1: Fig 1 indicates GDP part is the vulnerability function, while Sect. 2.1 reads Eq. 1 (to derive mortality rate from GDP as well as hazard indicators). Which is true?

RE: GDP was one out of the four inputs of the vulnerability function (others were hazard indicators), which was used to indicate prevention capacity. In the revision, we have updated Fig. 1 in the manuscript by moving the vulnerability function into box "Event loss rate modeling" to keep it consistent with Eq. 1.

Page 5 Line 19: How the data of the previous days are used?
Page 6 Line 14: Why since the last snow fall day?
RE: The two comments are related and are therefore responded together.

We used average daily precipitation (snowfall) since the last snowfall day (the last day with effective precipitation) to represent the diminishing impact/pressure from snowfall as time elapses. In general, the larger amount of snowfall, the longer the impact of the snowfall would be given identical temperature conditions – it would take longer time to thaw. Similarly, the closer a day to the snowfall day, the more likely it would be identified as a "snow disaster" day. It worked especially for deciding whether two snow fall days should be regarded as parts of one single event or two independent events.

Page 6 Lines 12-13: How do you compromise when the two standards are different with each other?
RE: Thank you for your comment. The two standards, although designed for different regions, are not conflicting with each other on the key indicators (Table 1), but shares similar indicators (i.e. continuous days of perpetual snow cover, or number of consecutive snow fall days since the last snow fall day).

Table 1 Variables considered in national and industrial standards for snow disaster

Snow Disaster Grades in Grazing Regions of China (GB/T20482-2017)	Meteorological Grades of Urban Snow Hazards (QX/T 178-2013)
Snow depth (cm) Grass height (cm) Continuous days of snow cover (d) % of grassland covered by snow (%)	Cumulative snowfall (mm) Maximum daily snowfall (mm) Snow depth (cm) Number of consecutive snowfall days (since the last snowfall day) Daily lowest temperature (degree C) Windspeed (m/s) Minimum relative humidity *(%)

Page 7 Lines 1-2: This sentence dose not explain why satellite is not used here.

RE: Thank you for your comment. Data need from satellite imagery is mainly the daily snow cover ("yes/no" data) and snow cover rate (% area covered by snow). One of the major purpose of this study is to develop a probabilistic risk assessment framework for risk assessment future climate change scenarios. For such a purpose, our input variables used in the framework must be available in climate projections. As far as we know, there are no projections of daily snow cover for the future. This is one of the main reasons that we did not consider it.

In the revision, we will try to make this point clear to the readers.

Page 7 Lines 4-6: Are there no bias between the two data?

RE: A similar question has been raised by Reviewer #1 (comment #7). According to (Wang *et al.*, 2013), the data for period of 1980-2007 were obtained from the yearbooks of meteorological disasters. Therefore, these data were originally recorded officially by provincial meteorological administrations, and published as a collection in books. The data for 2008-2015 were directly obtained form China Meteorological Administration in digital format. Therefore they are both from official records from meteorological administration, and the standards in identifying snow disasters are the same. However, we cannot perform a bias check as the data from two different sources do not share any overlapping period.

In the revision, we will supply information to clarify that the potenail bias between two data is very limited.

Page 7 Line 11: Explain the meanings of lr and tc.

RE: Boosted Regression Trees have two important parameters that need to be specified by the user. Tree complexity (*tc*) controls the number of splits in each tree. A *tc* value of 1 results in trees with only 1 split, and means that the model does not take into account interactions between environmental variables. A *tc* value of 2 results in two splits and so on. Learning rate (*lr*) determines the contribution of each tree to the growing model. As small value of *lr* results in many trees to be built. These two parameters together determine the number of trees that is required for optimal prediction. The aim is to find the combination of parameters that results in the minimum error for predictions.

In the revision, we will supply these information into the text.

Page 7 Lines 13-14: How the number of the variables and prediction power is weighted? Any kind of criteria like AIC or BIC is used?

Page 10 Line 4: Why SD, minWind and Pre were excluded?

RE: SD, minWInd and Pre were excluded due to their modest contribution in explaining the response variable. BRT uses a process of variable selection analogous to backward selection in regression. It drops the least important (in terms of relative influence) predictor, then re-fits the model and sequentially repeating the process (Elith, Leathwick and Hastie, 2008). In each step, after the removal of one predictor, the change in predictive deviance is computed relative to that obtained when using all predictors. Finally, a list containing the mean change in deviance and its standard error as a function of the number of variables removed will be returned

(Hijmans *et al.*, 2011). From the list, the optimal number of variables to drop can be identified, i.e. the number of variables that yield the minimum predictive deviance.

In the revision, we will supply more information about the mechanism of variable selection and model simplification.

Page 7 Line 14: The cross validation is how many fold?
RE: We used a 10-fold cross-validation. This information will be supplied in the revision.

Page 7 Line 20: Is "prediction error" random? If they are systematic, to take average may not be a good solution.
RE: The CV estimates of prediction error (predictive deviance in BRT models, Elith et al. 2008) of our BRT model indicates that the error is random.

Page 9 Line 9: How good is the performance of the equation?
RE: This equation is derived using logic reasoning: if a share of livestock died in one event, then the actual herd size exposed to the next event should be reduced correspondingly. As we rarely have two or more subsequent events in one year and one place in our historical records, we cannot evaluate the performance of this equation. But it is logically correct.

Page 10 Line 7: How relative contribution is calculated?
RE: In BRT, the relative importance are calculated based on the number of times a variable is selected for splitting, weighted by the squared improvement to the model as a result of each split, and averaged over all trees. The relative importance for each variable is scaled so that the sum adds to 100 (Elith, Leathwick and Hastie, 2008).

Page 10 Line 19 "well captured": p-values of 0.118 and 0.189 are not necessarily good (not statistically significant with 10% level). To check the model's representability more carefully, let us see not cumulative but probability density function before accumulation.
RE: We agree with your that we may not use "well captured". But we want to double confirm about the meaning of "not statistically significant with 10% level". The null hypothesis of the Two-sample Kolmogorov–Smirnov test is: "The two samples come from a common distribution", and the alternative hypothesis is: "The two samples do not come from a common distribution" (https://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ks2samp.htm). Our test statistics indicated "not statistically significant at 10% level", and therefore we failed to reject the alternative hypothesis and had to believe that the two samples (observed and predicted) were from a common distribution. It said that our prediction had captured the statistical feature of the observed duration (both event and annual). We posted the probability densities of these variables below, and actually they show similar thing with the cumulative density function:

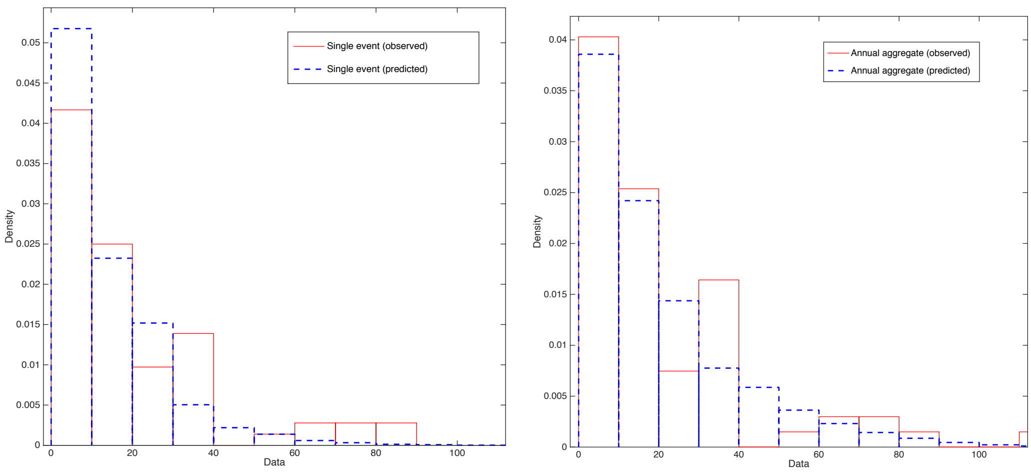

Figure 6 probability densities of observed and predicted snow disaster durations

In the revision, we will change "well captured" to "captured" following your suggestion.

Page 12 Fig. 5: Better to present the annual total number of SDDs.
Page 13 Line 13: Fig 5b shows annual aggregate snow disaster duration? But the caption says "mean event duration".

RE: Two questions above are related and responded together.

We believe that the reviewer has been discussing Fig. 5 (b). In the revision, we will follow your suggestion and present the annual aggregate duration ("annual total number of SDDs"). Correspondingly, the description on Page 13 will be updated to keep consistency with the figure.

Page 13 Line 11 (also for Page 18 Line 1): It is obvious when the Gaussian approximation is used.
Page 13 Lines 18-19: If "The distribution of annual average mortality rate is extremely positively skewed", the Gaussian kernel function (Page 9 lines 21-22) is not appropriate, is it? BTW, is it related to the dependent variable in Eq .1 is lnLR, not LR?

RE: Thank you for your question. The two comments are related and are responded together. There are three points to clarify:

1) We did not use Gaussian approximation over the simulated annual loss rates, which is a parametric and symmetric distribution function. In stead, we used a non-parametric approach called kernel density function (Page 9 lines 21-24). The kernel density approach does not specify any functional form and so it is flexible to capture probability densities of different degrees of skewness.
2) Page 13 Lines 18-19 was describing the distribution of annual average mortality rate of different grids (spatial locations), but not the distribution of annual mortality rate for any specific location. "Extremely positively skewed" means that the grids/regions of high mortality risk take only a small portion of all places on the Plateau, but their annual average mortality rates were much higher than those of other grids. It is not related to the distributional assumption of dependent variable LR.
3) LR is also positively skewed. Only its natural logarithm (lnLR) exhibits normality.

Page 18 Table 1: What is the trend of actual herd size in QTP? To consider a static herd size is reasonable?

RE: The trend of herd size as a total of Qinghai and Tibet is provided below in Figure 7. As shown in the figure, number of cattle has not changed much for both regions. Number of sheep in Qinghai has not changed much since 2006, but that of Tibet has been decreasing for recent years, mainly due to the forage-livestock balance policy. In terms of aggregate size (1 cattle = 5 sheep units), Qinghai remains quite stable since 2006, and Tibet keeps dropping rapidly up to year 2014. Once it drops to the upper limit of carrying-capacity, it is very likely to keep nearly constant (please also refer to the discussion on your comment #3).

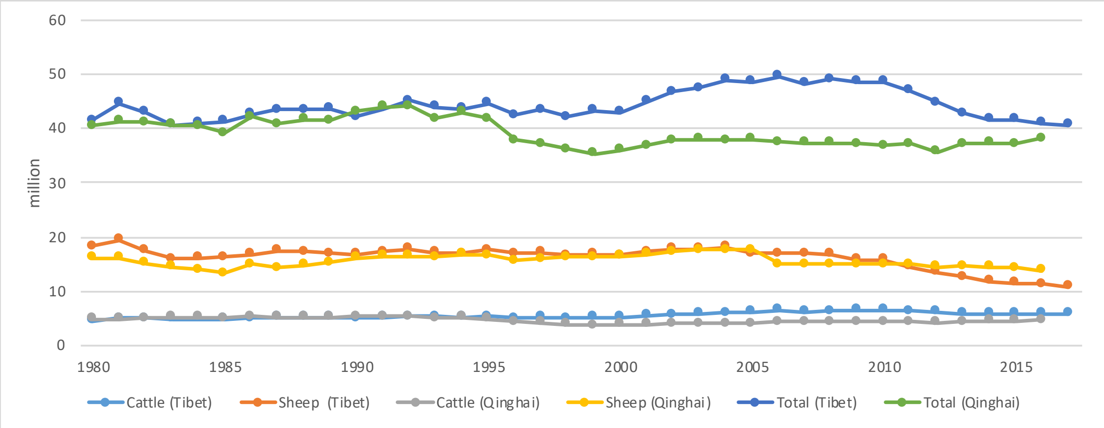

Figure 7 Change of herd size summed from Qinghai and Tibet

In our results, we will add discussion to this point.

Page 18 Line 8: Why mortality becomes small by the constraint of herd size by carrying capacity?

RE: This sentence was misleading. We intended to say that the mortality (sheep units) was small (mostly below 10 per grid), and the main reason was that the estimated carrying capacity was small. In the revision, we will change this sentence to: "*Mortality appears small in Fig. 7, generally several sheep units/km^2. However, when aggregated at the prefecture level, mortality remained considerable (Table 1).*"

Page 19 Lines 15-16: It is better to compare the modeled mortality with observed (historical) ones.

Page 19 Lines 27-28: Is there no possibility that this study overestimates?

RE: The two questions above are related to each other and are responded together.

To further test the performance of the model, we have re-run the model using historical value added of animal husbandry. For the two specific cases as mentioned in our previous manuscript, our modeled aggregate livestock loss for event (a) in 1996 in Yushu and Guoluo prefectures was approximately 1.20 million heads/units (turned from 2.64 million sheep units modeled given the cattle-to-sheep ratio in Qinghai Province), and the historical record was 1.48

million[1] livestock recorded), and for event (b) in 1998 in Naqu Prefecture was 0.72 million head/unit (turned from 1.59 million sheep units), compared to 0.82 million livestock. Therefore, our model result did capture the loss of major events in specific regions, although it still suffered from uncertainty.

In our revision, we will put the validation result into discussion. In addition, model results uncertainty will be added into the limitation section.

Page 21 Line 8 "two critical indices": Is this presented in the Result section?

RE: The "two critical indices" were referring to disaster duration and growing season aggregate precipitation, two variables that critically determines livestock mortality rate in our vulnerability function.

In the revision, we have deleted this part to make it clear. "*A more favorable choice is to adopt an index-based structure using livestock mortality rate as predicted by our model.*"

Page 22 Line 18: How the study can be applied for future? I consider that the method used here is not suitable when the climate is changing.

RE: Thank you for your question. Reviewer 1 has raised similar comment (comment#3 of reviewer#1). So we will refrain from claiming so in the current manuscript and leave the discussion for the next study.

Some discussion might answer your question here. In our modeling process, we have been trying to make our model framework capable of incorporating the changing climate and socioeconomic development. Following the existing work of climate change risk assessment (Tachiiri and Shinoda, 2012; Carleton and Hsiang, 2016; Winsemius *et al.*, 2016; Kinoshita *et al.*, 2018), our model consists of a set of response relationships: 1) a hazard module defines the relationship between daily weather condition and the occurrence (identification) of snow disasters; 2) a vulnerability function defines the relationship between even mortality rate and hazard intensity (duration, wind speed, growing season precipitation) and prevention capacity (as proxied by socioeconomic variable); 3) exposure in terms of herd size is used only in a multiplicative way to derive the final risk metrics in terms of sheep units.

In such a structure, climate condition and socioeconomic condition are merely inputs to our model, rather than a part of the model. Climate change will lead to changes in model input, and correspondingly model output. And that is what we want the model the have the capability to capture: in the short-term future, will the changing climate lead to more or less frequent snow disasters, with shorter or longer duration? Together with the enhanced prevention capacity, will the mortality risk increase or decrease correspondingly? For the first question, our hazard module is capable of identifying/simulating snow disaster days based on climate inputs mimicking meteorological observers' decision using machine learning algorithm: given daily maximum , mean and min temperatures, precipitation, and maximum and mean wind speed, the module can exactly derive corresponding snow disaster event set. Applying it to future climate scenario can then generate future event set and investigate the change of disaster event

[1] The figure 1.08 million in the previous manuscript was from a literature, but unfortunately it is not consistent with those reported in the books. We have corrected it.

frequency and intensity (duration) in the future. For instance, in a warmer climate we may expect snow disaster events with less frequency and shorter duration.

In summary, climate change will not influence the model structure, but certainly it will change the model results. That is how the climate changing is taken into account in our model.

[Technical corrections]

Page 1 Lines 20,22: 1/20a -> 20 years (also for all similar expressions).

RE: Thanks for pointing this out. We have corrected it through out the manuscript and in the figures.

Pages 14-15 Fig 6: To be multi-colored like Fig 7 would be more reader-friendly.

RE: Thanks for the suggestion. We have updated Fig. 6 to use multi-color in the map.

Page 9 Line 7: "although unlikely" should be rephrased with better expression.

RE: Thanks for the suggestion. We have removed the phrase from the text.

Page 5 Lines 4-6: Hard to understand. Too many "and"s. "its needs" -> "it needs"? Delete one of the two "provide"s?

Page 10 Line 14: Fig. 3 -> Fig. 4?

Page 13 Line 2: topology -> topography?

Page 19 Line 22: There is no Table 2.

Page 19 Line 26 (also in Page 21 Line 28): higher -> longer.

Page 20 Line 8: Fig A2 -> A3?

RE: Above comments are related to typos. We have revised/corrected them as suggested.

References

Carleton, T. A. and Hsiang, S. M. (2016) ‘Social and economic impacts of climate’, *Science*, 353(6304). doi: 10.1126/science.aad9837.

Diffenbaugh, N. S. and Giorgi, F. (2012) ‘Climate change hotspots in the CMIP5 global climate model ensemble.’, *Climatic change*. Springer, 114(3–4), pp. 813–822. doi: 10.1007/s10584-012-0570-x.

Elith, J., Leathwick, J. R. and Hastie, T. (2008) ‘A working guide to boosted regression trees’, *Journal of Animal Ecology*. Wiley/Blackwell (10.1111), 77(4), pp. 802–813. doi: 10.1111/j.1365-2656.2008.01390.x.

Gu, H., Yu, Z., Wang, J., Ju, Q., Yang, C. and Fan, C. (2014) ‘Climate change hotspots identification in China through the CMIP5 global climate model ensemble’, *Advances in Meteorology*, 2014. doi: 10.1155/2014/963196.

GUO, D.-L., SUN, J.-Q. and YU, E.-T. (2018) ‘Evaluation of CORDEX regional climate models in simulating temperature and precipitation over the Tibetan Plateau’, *Atmospheric and Oceanic Science Letters*. Taylor & Francis, 11(3), pp. 219–227. doi: 10.1080/16742834.2018.1451725.

Hijmans, R. J., Phillips, S., Leathwick, J. R. and Elith, J. (2011) ‘Package “dismo”’, *R Package*, p. 55. doi: 10.1016/j.jhydrol.2011.07.022.

Kinoshita, Y., Tanoue, M., Watanabe, S. and Hirabayashi, Y. (2018) ‘Quantifying the effect of autonomous adaptation to global river flood projections: application to future flood risk assessments’, *Environmental Research Letters*. IOP Publishing, 13(1), p. 014006. doi: 10.1088/1748-9326/aa9401.

Kuang, X. and Jiao, J. J. (2016) ‘Review on climate change on the Tibetan Plateau during the last half century’, *Journal of Geophysical Research: Atmospheres*. Wiley-Blackwell, 121(8), pp. 3979–4007. doi: 10.1002/2015JD024728.

Michel-Kerjan, E., Hochrainer-Stigler, S., Kunreuther, H., Linnerooth-Bayer, J., Mechler, R., Muir-Wood, R., Ranger, N., Vaziri, P. and Young, M. (2013) ‘Catastrophe risk models for evaluating disaster risk reduction investments in developing countries.’, *Risk analysis : an official publication of the Society for Risk Analysis*, 33(6), pp. 984–99. doi: 10.1111/j.1539-6924.2012.01928.x.

Tachiiri, K. and Shinoda, M. (2012) ‘Quantitative risk assessment for future meteorological disasters: Reduced livestock mortality in Mongolia’, *Climatic Change*, 113(3–4), pp. 867–882. doi: 10.1007/s10584-011-0365-5.

Wang, W., Liang, T., Huang, X., Feng, Q., Xie, H., Liu, X., Chen, M. and Wang, X. (2013) ‘Early warning of snow-caused disasters in pastoral areas on the Tibetan Plateau’, *Natural Hazards and Earth System Science*, 13(6), pp. 1411–1425. doi: 10.5194/nhess-13-1411-2013.

Winsemius, H. C., Aerts, J. C. J. H., Van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., Kwadijk, J. C. J., Ligtvoet, W., Lucas, P. L., Van Vuuren, D. P. and Ward, P. J. (2016) ‘Global drivers of future river flood risk’, *Nature Climate Change*, 6(4), pp. 381–385. doi: 10.1038/nclimate2893.

Wu, J., Li, N., Yang, H. and Li, C. (2007) ‘Risk evaluation of heavy snow disasters using BP artificial neural network: the case of Xilingol in Inner Mongolia’, *Stochastic Environmental Research and Risk Assessment*, 22(6), pp. 719–725. doi: 10.1007/s00477-007-0181-7.

Xin, Y., Du, T., Xin, Y., Wu, A. and Lu, F. (2011) ‘The Evaluation of Carrying Capacity of Grassland in Qinghai’, *Qinghai Prataculture*, 20(4), pp. 13–22.