This manuscript describes the procedure used for the successful prediction of the detachment of a hanging glacier. Impressively, through a combination of skill (and maybe a little luck) the authors have managed to obtain a record of displacements that extends right up to failure. This provides them with an extremely detailed dataset that they can analyze and use to make forecasts. There are two interrelated themes threaded throughout this manuscript. The first is the practical task of forecasting a likely time when the hanging glacier will detach based on measurements of surface displacement/velocity. This is no small task and is vitally important to those living in imminent danger of being crushed or at least inconvenienced by one of these events. Here, unlike many glaciological studies the authors were not only able to make a prediction that was verified, but this prediction had consequences. The second theme is more theoretical and in it the authors argue that surface velocities exhibit log-periodic oscillations and that this behavior is universal and the result of the discrete scale invariance. These two themes mesh together as the authors show that using the additional information associated with the log-periodic oscillations enables them to make more accurate predictions. I have some comments and questions (see below), but overall I enjoyed reading this manuscript and feel like it has the potential to make a significant contribution to avalanche forecasting and our understanding of rupture processes in general.

Thank you

One of the reviewers argued that this manuscript is not hypothesis driven and as such may not be appropriate for Cryosphere. I disagree. If we were forced to eliminate all manuscripts that aren’t more clearly hypothesis driven half the manuscripts on the Discussion page would be eliminated. However, I do agree with most of the other comments posted and will echo some of them in my more detailed comments below.

Detailed comments: My detailed comments focus individual on the two main themes of the paper.

1. Time to rupture forecast. As the other reviewers noted, I think it is important to provide a more detailed description of the survey requirements and data processing algorithms. I have never been involved in avalanche forecasting so take my comments with some skepticism, but I would imagine that other practitioners would be interested in knowing more about the quality and quantity of data required for accurate time to failure forecasts. What kind of data collection rates are needed? How many/few stations are needed? What measurement precision/accuracy are needed to make accurate forecasts?

See new paragraph 5.5

How far in advance can forecasts be made?

See paragraph 5.4

The authors also make several claims about the unique quality of the data. That argues to me that the authors should make sure the data are archived somewhere they are publicly available so that others can examine their method and test their own prediction algorithms on this well studied case. (Although I make this claim in incomplete ignorance of standard practices in thefield.) I a little more interested in knowing more about the fitting algorithms. For example, Equation (1) has fewer parameters than Equation (2). Because of this Equation (2) can't (or at least shouldn't) generate larger residuals than Equation (1). A more statistically significant question is where Equation (2) performs in a statistically significant way better than Equation (1) **when the additional degrees of freedom are included**. That is a quantitative answer that the authors can provide. Similarly, assuming the authors are doing a standard least squares/maximum likelihood estimate of parameters, then it is possible to compute the uncertainty in each fitted parameter. This may be an especially useful quantify for the estimated rupture time. This type of analysis would allow the authors to address questions like: Does increasing the number of model parameters decrease the uncertainty in the predicted time to failure in a statistically meaningful way? Does increasing/decreasing the amount of time used in the time series used for fitting parameters affect the uncertainty of the predicted time to failure? I have some other minor questions, like what is the uniform time step that the data was interpolated to? How did the authors account for the effect of interpolating the data on the spectral analysis (or is it too small to bother with.)

Finally, I think the authors should provide more detail on their fitting algorithm, especially in regards to how they determine the power-law exponents. There is a long history in the geosciences of calculating exponents by fitting a straight line in log-space. We know now that this procedure can be dangerously inaccurate and much better procedures are available (see, for example: A. Clauset, C.R. Shalizi, and M.E.J. Newman, "Power-law distributions in empirical data" SIAM Review 51(4), 661-703 (2009). (arXiv:0706.1062, doi:10.1137/070710111.)

We described in more details the fitting procedures line 175-181. Associated errors are now included in Table 1. The errors in estimated rupture time for both regression are contained in Fig. 8b. We also checked the Degree-of-freedom adjusted coefficient of determination (dfa) for both fits: with eq. 1: dfa=1-1.7 10-3 and with Eq 2: dfa=1-1.1 10-4

See also new paragraph 5.1

Clauset et al 2009 deals with power law distributions, not with power law acceleration (when t_c is not known, without possibility to plot displacement as a function of tc-t).

2. Universality and log-periodic oscillations. This is the part of the manuscript I was most interested, but it was also the part that left me with the most questions. Some of these questions may not be too esoteric for this manuscript and I leave this decision with the authors and editor. Nonetheless, the universality claim is very interesting. However, I'm not sure that I understand the sense in which the authors are using the term "universal". The concept of universality (and criticality) that I'm familiar with originates in the statistical physics of phase transitions. Phase transitions were found to exhibit power-law fluctuations near a critical point in phase space and the exponents of the power-laws were found to be "universal" in the sense that many seemingly different systems exhibited the exact same exponent. The renormalization group was eventually used to show that all of these disparate systems with the same exponents lived in a broad "universality group". The crucial result being that details of small-scale interactions between components were unimportant and the universality group was largely controlled by factors like dimensionality.

This history (and perhaps my own misinterpretation of the authors use of universality) is what feeds my confusion. As far as I can tell the authors find different critical exponents for each stake (Table 1). This would seem to indicate that either the critical exponents have large uncertainties and these differences are not statistically meaningful (does this have implications for forecasts?) or universality (at least as I'm used to using it) isn't supported by the data. I have similar questions about the critical exponents obtained for the Weisshorn glacier. If this really is universality then surely these two glaciers should surely belong to the same universality group and we should find the same exponents for both glaciers? Shouldn't there also be other systems that aren't hanging glaciers that live in the same universality class and exhibit the same critical exponents as for hanging glaciers? I apologize to the authors if I have misunderstood their intended usage through my own ignorance.

You are right concerning universality. In our case, the critical exponents are not the same for both glaciers, and therefore universality cannot be evoke in this study. However, we mentioned only once the term universal (in the last sentence: the present methods exploiting the log-periodic oscillating behavior are universal), but it was in the sense "general". Our mistake. We removed "universal" line 330-332.

A second, but more technical comment on the assumption of criticality and scale invariance is that scale invariance is a concept that breaks down at scales comparable to the fundamental scale of the interacting components of the system (actually much before). Presumably, the fundamental components here are interacting micro fractures within the glacier and then critically is obtained by taking the thermodynamic limit in which the system size tends to infinity (as compared to the fundamental scale of microcracks). Is the system actually large enough that you can have a large enough number of micro cracks interacting for the thermodynamic limit to be a valid approximation? What does this tell us about the size of micro cracks? Similarly-and this is what I'm more interested in-traditional assumptions break down as the scale of fluctuations approach the system scale and this gives rise to finite size effects. I would expect that as fractures start to penetrate a significant portion of the glacier that these finite size effects would begin to manifest themselves in deviations from criticality. But I don't see any evidence in the data for this. Why is this? Is it possible that much of the observations are corrupted by finite size effects and this is why the critical exponents fail to converge to a single value? Or is it possible that much of what is being observed is a consequence of finite size effects and not universal critical behavior? Is there a renormalization calculation that can be used to estimate exponents and/or finite size effects?

Such a behavior was investigated in the 2005 Weisshorn event (Faillettaz et al. 2011) with seismic measurements. It appears that just before failure a change in the size frequency distribution of icequake energies was detected: larger icequakes occurred as expected (signature of overcritical system?). Therefore a deviation from criticality was also observed in seismic as the reviewer conjectured, but no finite size effect was detected. We doubt that finite size effect manifest itself in this case.

Technical comments:

Section 3 results, 1st paragraph. I think the authors are saying that they interpolated the unevenly spaced measurements onto measurements that are evenly spaced in time. This can probably be rephrased to say this more clearly. What evenly spaced interval was used?

OK line 160-164

Page 4926: unique material -> single material?

OK

Page 4931: missing word? “Downstream this crevasse” ->Downstream **of** this crevasse

OK

Page 4932: Paragraph starting with Moreover should be joined with previous paragraph or introduced with a separate topic sentence.

OK

Section 3.2: Why apply the fitting procedure to 1 month of data? Does the analysis fail if, say 2 months of data are used or if 2 weeks are data are used? I would not be surprised to find that considering a longer time series doesn’t help. However, it is modestly interesting to ask how little data you need to make a predicting given that many glaciers may not be as consistently monitored.

We found that taking a longer time serie does not really help. Therefore we just used the last month of data to perform our analysis.

Page 4934: I suggest removing the exclamation point. This is indeed impressive and impressively large oscillation. I don’t think you need the exclamation to call readers attention to this.

OK

Section 4.2: Again, are these differences statistically significant?

See new section 5.1

Section 4.4 is what I was looking for earlier. You might want to point readers towards this section when you first describe the data analysis.

OK line 195-197