Responses to Review #1

We thank the reviewer for taking the time to thoroughly evaluate our work. Our detailed responses and the resulting changes to the manuscript are noted below. We have numbered the various comments to allow cross referencing. Our responses are in italics and we show text from the revised manuscript in quotation marks. We also append the new abstract to the bottom of this response.

R1.01: I am not convinced that the word "secular" used in the title and in other parts of the text is good to represent the type and temporal span covered by the national inventory. Is this the same type of inventory that Malamud and co-workers (2004) [cited] have called "historical", and Guzzetti et al (2012) [cited] "geomorphological"? Please clarify.

We prefer to retain the use of secular in the sense of a background level of activity over an indefinite amount of geological time; this term is not new in the Earth sciences. The term historical in this context is not good, because it refers more generally to a time period that is recorded by humans, and therefore a historical dataset can be interpreted as one that is recorded via written records, in contrast to one that is recorded by the landscape.

R1.02: The second paragraph of the Introduction (page 115, lines 5-16) is out of context, really. Consider deleting the paragraph. The section 4.4 Implications for hazard assessment does not justify the paragraph.

We feel that the second paragraph offers an important lead in to the subject matter, namely that landslides are both important in a socio-economic sense as well as geomorphic agents (paragraph 3). As such we have kept it.

R1.03: (A) Some confusion exists because in the literature authors have used "cumulative" (e.g., Dussauge et al. 2003 [cited]) and "non-cumulative" (e.g., Malamud et al. (2004) [cited]) statistics. Comparison of the results of the different studies is therefore problematic. The authors should make this clear, and specify which statistics were obtained from cumulative distributions, and which from non-cumulative distributions.

We concur. Clear statements that our study is only analysing non-cumulative distributions have been added to the manuscript. Moreover as you further state in R1.0, we have mixed comparison of landslide area and rock fall volume statistics, which are not directly comparable. This has been corrected.

R1.04: (B) In the text, the author mix (confuse?) statistics of landslide area (e.g., Hovius et al. 1997, Pelletier et al. 1997, Stark & Hovius 2001, Guzzetti et al. 2002 [all cited]) and landslide volume (Dussauge et al. 2003 [cited]). Dussauge et al. (2003) studied rock falls, and all the other authors studied landslides of the slide or flow types. The difference is relevant, and the statistics may not be comparable. For landslides of the "slide" type authors have found that the relationship linking landslide area and volume is non-linear (e.g., Parker et al. 2011 [cited], Guzzetti et al. 2009, doi:10.1016/j.epsl.2009.01.005, Larsen et al. 2010, doi:10.1038/NGEO776). I am not aware of studies linking the area and volume of rock falls.

We concur. The data from Dussauge et al. (2003) are also related to the cumulative frequency statistics of rock falls (see R1.04), and so there are now two reasons to remove this comparison from the manuscript.. We now write:

"Several studies have proposed that the non-cumulative size-frequency distribution for landslides (i.e. the number of slides of a give size occurring over a given length of time or within a given area) follows a negative power-law relationship for medium to large landslides (sensu lato) (e.g. Guzzetti et al., 2002; Hovius et al., 1997; Pelletier et al., 1997;

Stark and Hovius, 2001; Turcotte et al., 2002). Estimates of the exponent α for power-law scaling of large events vary from α =1.4 up to α = 3.3 (Van Den Eeckhaut et al., 2007). Van Den Eeckhaut et al. (2007) report an average value of α = 2.3 ± 0.6, and Malamud et al. (2004) suggested α = 2.4 might be universally applicable to event-triggered inventories based on consensus between three contrasting event-driven datasets. Larsen et al. (2010) caution that estimates of volume of material transported by landslides may be very sensitive to this scaling exponent, resulting in prediction errors of over an order of magnitude. The scaling exponents may vary with underlying geology (e.g. Frattini and Crosta, 2013; Guzzetti et al., 2008), and the type of failure event (e.g. Brunetti et al., 2009; by analysis of landslide volume rather than area statistics)."

R1.05: (C) Brunetti et al. 2009 [doi:10.5194/npg-16-179-2009] have re-examined the dataset of rock fall volumes compiled by Dussauge et al. (2003), and have determined a different (and larger) scaling exponent for the power law distribution that describes the empirical data. These authors have also found a difference on the scaling of the power law describing the volume of "falls" and "slides".

Thank you for highlighting this study, which we were unaware of. Whilst we cannot directly compare the dataset presented in this study since we are analysing area rather than volume, we add to the discussion concerning the exponents and landslide type since the interpretations by Brunetti et al. (2009) are pertinent to our own analysis and interpretations, and we cite this paper where appropriate (see also R1.04):

"Magnitude-frequency scaling of landslides classified by the type of mass movement have power-law scaling exponents (α = -0.57 to -1.5) lower than the dataset as a whole (α = -1.76). Lower exponents suggests the subset of data may be biased towards larger events, and indeed it seems likely (and reasonable) that detailed field studies to determine the style of failure are preferentially carried out for larger failure events. Unfortunately there are few observations of landslide type below areas of 10^3 m^2 (Figure 4b) yet there are a large number of landslides in the NLD of this magnitude (Figure 3b). However, whilst the sample sizes are small, there is a suggestion that the gradient of the most likely power-law decreases with landslide type, for falls/topples is low (exponent α = -0.57). This is broadly consistent with the results of Brunetti et al. (2009) who performed analysis of landslide volume distributions rather than area as in the present study. Brunetti et al. (2009) performed magnitude-frequency analysis of 19 landslide inventories and found that the scaling exponents for landslide volume were lower for rock falls and rock slides than for slides and soil slides. More data are needed to provide an empirical test that more resistant lithologies preferentially yield deep-seated landslides whilst weaker materials preferentially yield shallow landslides, as found theoretically by Frattini and Crosta (2013)."

R1.06: (D) The "double Pareto" (Stark & Hovius, 2001 [cited]) and the "inverse Gamma" (Malamud et al., 2004 [cited]) are the most common distributions used to model the probability/frequency-area distribution of landslides. With a few exceptions (e.g., Guzzetti et al. 2008 [cited]), they are not used to describe the probability/frequency-volume of landslides (of the "slide" or the "flow" types), although this is certainly possible. For the investigation of the probability distribution of rock fall volumes see e.g., Brunetti et al. (2009) [doi:10.5194/npg-16-179-2009]. Recently, Chen et al. (2011) [doi:10.1209/0295-5075/95/49001], based on the application of non-extensive statistics, have proposed an analytical distribution function to describe the frequency-area distribution of landslides.

Our analysis is concerned only with landslide area, and not landslide volume statistics. We thank the reviewer for highlighting that we may be misleading the reader by erroneously comparing volume and area statistics, and have amended the manuscript accordingly (see R1.04).

R1.07: I have some concern on the method used to assemble the landslide database that was used for the analysis. The authors are clear in explaining the steps they have taken. However, inspection of Fig. 2 makes me think that landslides were not mapped very accurately, at least for part of the database. This has influence on the frequency-area statistics, and on the conclusions drawn. As an example, the shape of the landslide polygon centred on coordinates E412000,N384000 in the map shown in Fig. 2 is indicative of possibly multiple coalescing landslides, and not of a single (and larger) landslide. This may be due to the scale of the mapping, which is relatively small (1:50K). The authors should comment a bit more on the quality (and diversity in quality) of their inventory.

We are pleased that the steps we have taken to compile the dataset are clear. We were particularly anxious to communicate explicitly to the reader that there are concerns over data quality and we required total transparency to the data processing which was carried out. There may be issues with the quality of the landslide data we analyse, and we tried to be explicit about this from the outset, stating that mapping has taken place over a long timescale by numerous individuals. During quality assurance the source of the mapped deposit is revisited where available (usually a field map), and the polygon is compared to and verified against aerial photography and high resolution (5 m) topographic data. There is the possibility that landslides in our dataset may consist of several amalgamated events when only a single event exists in the National Landslide Database. Our policy of linking only single recorded landslide event to single mapped landslide deposits goes someway to dealing with these problems, in that we retain only data that has undergone quality assurance procedures. However, with a large incomplete record, and with the scale at which mapping took place there may be some issues still. In the case of the specific landslide you refer to, we cannot observe any convincing evidence from surface features observed using topographic data and aerial photography that there are multiple failure tongues, though we acknowledge that the polygon shape could suggest this. We note from the underlying OS map that there are streams draining the surface of the landslide inbetween the potential tongues and therefore the morphology may just be the result of reworking of the deposit. At this stage, we prefer to provide the reader with the clear caveat that the data are not perfect and that our conclusions are inevitably limited in this context. Nonetheless, we also note that the data are in general well-behaved, and that our interpretations are not offering a radical new model of landslide behaviour. In other words, we do not believe that we have over-interpreted a dataset that is inevitably limited.

"These mapped deposits have been recorded by a multitude of geologists whose primary concern may not have been the precise recording of landslide deposits, thus the quality of the dataset is likely to be highly variable. The NLD is not considered to be a complete inventory of landslide occurrence in the UK. The data were collated from a multitude of sources and therefore approaches to mapping are unlikely to have been consistent. Additionally the spatial coverage of landslide mapping is unlikely to have been consistent. As part of the continuing collection, updating and verification of landslide information by the BGS, existing landslide data are subjected to a standardized QA procedure designed to improve the consistency and reliability of the these datasets, and areas with poor coverage are identified for detailed resurveying (Foster et al., 2012)."

R1.08: It is not entirely clear to me how the frequency/probability density was obtained, and showed in the Figures 3, 4 and 5. The authors first state that they have (page 121) "calculated FD and PD for the NLD dataset by sorting the data into logarithmically-spaced bins in A". It is unclear how this was done. Were bins all of the same size, in log coordinates, or not? Logarithmic binning has problems when the number of points is reduced. This may be the case for the very large landslides. Then the authors state (page 123) that the double Pareto and the inverse Gamma distributions were estimated using "maximum likelihood estimates to find the best fit parameters". I presume MLE was applied to the raw data (and not to the log-binned data), but this is not clear. More information is required on the technique(s) used to determine the frequency/probability densities.

We have improved our statistical approach in order to provide greater confidence in our interpretations. To do so, we now include a new section in the "Data and Methods" section to document the analysis we now perform, which is different to the analysis presented in our submitted manuscript.

Previously we were performing MLE to the binned data rather than the raw data (note, we are analysing the SLI dataset, not the NLD dataset).We now perform MLE on the raw data when finding the most likely parameter combinations for the various models. We obtain uncertainties on our parameter fits by performing a bootstrap method by replacement and taking our estimated parameters as the mean bootstrapped value ±1 standard error. For power-law fitting this is done following the analytical solutions of Clauset et al. (2009). We subsequently still bin the data into bins spaced evenly in logarithmic space for visualisation purposes only.

The new statistical methods section reads:

"To visualize the size distribution of the SLI we calculated F_D and P_D following section 2.3.1 for the dataset as a whole and for subsets grouped by lithology, landslides type (see Figures 3, 4 and 5). We use a maximum likelihood estimation (MLE) approach to fit statistical models in equations 2-4 to the data and various subsets but apply this to the raw data rather than the binned frequencies and probabilities shown in Figures 3, 4 and 5. To do so we calculate the log-likelihood L according to:

$$L = \ln P_D\left(A \mid \theta\right) = \sum_{i=1}^{n} \ln P_D\left(A_i \mid \theta\right) \qquad [5]$$

where P_D is a probability density model (e.g. equations 3-4), and θ is the parameter to optimize. The log-likelihood of a particular set of parameters θ is therefore the sum of probabilities for all landslides in the dataset or subset (of size n). Finding the combination of parameters which optimizes L gives the MLE of parameters for a given model.

To constrain the uncertainty on MLE parameters we perform bootstrap analyses in which we repeat the MLE method on 10,000 datasets sampled by replacement from the SLI (and subsets). We therefore generate 10,000 estimates of the most likely parameter combinations for the respective models and use the mean and standard deviation to report our most likely parameter combinations. For fitting of power-law distributions we use the MLE solutions for α provided by Clauset et al. (2009). Testing their solutions against our bootstrapping approach for fitting power-laws yields identical parameter estimates but with larger standard errors. We use these analytical error estimates for power-law MLE, and report standard deviations about bootstrapped parameter means when performing MLE for double pareto and inverse gamma functions (equations 3 and 4 respectively)."

R1.09: Some of the discussion is based on the qualitative (visual) or quantitative (using the parameters controlling the equation) comparison of the different probability density models obtained for the different subdivisions of the landslide dataset. However, it is not clear what is the uncertainty associated with the different probability densities. Depending on the uncertainty, in Fig. 4a the densities for the surficial, mudstones, interbedded and coarse clastic deposits may be indistinguishable, or may be statistically different. The same is [true] for the density for carbonates, metamorphic and igneous rocks. In Fig. 4b, the four frequency densities for rotational, planar, flow and fall type landslides may also be statistically indistinguishable, or not, depending on the associated uncertainties. This is a crucial point that needs to be resolved. The density models were determined using MLE. It should therefore be possible to determine the confidence levels for the individual density models (e.g., using a bootstrap method), and to compare them. My doubt is that for some of the datasets the number of samples may be too limited to constrain sufficiently the density models.

See R1.08. We thank the reviewer for their helpful suggestions. We have revised our methods accordingly. As a result there have been some changes to the absolute parameter values we estimate but the general distributions observed remain consistent with many of our interpretations (to be expected now that we are analysing the raw dataset rather than binned data).

R1.10: The authors observe a paucity of very large landslides in their distributions. The explanations given for this finding are plausible. I have two suggestions on this topic. First, clearly the size of the very large landslides is somewhat controlled by the size of the slopes where the large landslides occur. Is it possible that the lack of large landslides is related to the lack of very large slopes? Second, is it possible to segment the landslide database for the UK on time, and use the (relatively) recent landslides (e.g., those occurred in the last 50 or 100 years) to investigate the extent to which the power law scaling for large landslides holds?

The suggestion by the reviewer that a deficit of large landslides may be related to the lack of large slopes is entirely consistent with our interpretation that conditions have changed since the emergence of the UK from the last glacial maximum. One of those conditions is the initial exposure of relatively steep slopes that were previously ice-supported (whether by the ice-sheet or by permafrost): these slopes are able to yield relatively large landslides, but because there is no tectonic regeneration of such slopes, the landscape is gradually weaned of steep slopes and associated large landslides. The reviewer has articulated this better than we had, and the revised manuscript now clarifies this interpretation.

R1.11: In the text and the Table, do not use e.g., 15.3 x 10ˆ3 kmˆ2, but instead 1.53 x 10ˆ4 kmˆ2.

OK.

R1.12: Fig. 1. Add geographical reticule to the map.

OK

R1.13: Fig. 2. Add scale bar, or clarify in the map that coordinates are in metres (m).

OK

R1.14: Fig. 3. The journal accepts colour figures at no extra cost. The authors should exploit this opportunity and make the Figure in colours. This will improve the readability of the Figure. Indicate the number of samples in the landslide dataset. A suggestion: do not use dashed lines in the box plot, and provide a legend for the box plot (different criteria can be used to prepared the box plots, and without a legend it is impossible to tell what the different elements of the box plot (rectangle, central line, range) represent.

We have made the MLE for the double pareto and inverse gamma functions contrasting colours and line thicknesses so that they can be distinguished. We no longer used dashed *lines for the 'Malamud' model or the box plot. Description of the box plot is provided in the figure caption:*

"Figure 3: (a) Probability distribution of landslide deposit area for n = 8453 *landslides in the UK organized into bins spaced evenly in logarithmic space (open diamonds). Solid red and blue lines show MLE of a double Pareto function (α_p = 1.01 ± 0.01; β = 1.71 ± 0.07; A_{peak} = 8.09 ± 0.6 × 10^{-3} km^2) and inverse Gamma function (α_g = 0.95 ± 0.02; r = 10.9 ± 0.4 × 10^{-3} km^2; s = -1.91 ± 0.08 × 10^{-3} km^2) respectively (error ranges based on one standard deviation of bootstrapped MLE parameters) . The grey line is a proposed general distribution for landslides put forward by Malamud et al. (2004). Box plot shows the median (central line), upper/lower quartiles (extent of rectangle) and 5th and 95th percentiles (whiskers) of area data with a median value of 1.53×10^{-2} km^2 (b) Frequency density distribution for landslides in the UK. Solid lines represent the general distribution proposed by Malamud et al. (2004) for varying total number of landslides N_T."*

R1.15: Fig. 4. Suggestion: Indicate the number of samples in the different landslide datasets. Do not use dashed lines in the box plot, and provide a legend for the box plot (different criteria can be used to prepared the box plots, and without a legend it is impossible to tell what the different elements of the box plot (rectangle, central line, range) represent.

The number of landslides in each dataset is included in table 1, which we now refer to in the caption. Dashed lines have been removed from box plots and the box plots are now clearly explained in the figure caption as per R1.14.

R1.16: Fig. 5. The journal accepts colour figures at no extra cost. The authors should exploit this opportunity and make the Figure in colours. This will improve the readability of the Figure. Indicate the number of samples in the landslide dataset. Do not use dashed lines.

We have made the MLE for the double pareto and inverse gamma functions contrasting colours and line thicknesses so that they can be distinguished. The number of landslides in each dataset is included in table 1, which we now refer to in the caption. Dashed lines have been removed from box plots and the box plots are now clearly explained in the figure caption as per R1.14.

The new abstract:

Linking landslide size and frequency is important at both human and geological time-scales for quantifying both landslide hazards and the effectiveness of landslides in the removal of sediment from evolving landscapes. The statistical behaviour of the magnitude-frequency of landslide inventories is are usually compiled following a particular triggering event such as an earthquake or storm, and their statistical behavior is often characterized by a power-law relationship with a small landslide roll-over . The occurrence of landslides is expected to be influenced by the material properties of rock and/or regolith in which failure occurs. Here we explore the statistical behavior and the controls of a secular landslide inventory (SLI) (i.e. events occurring over an indefinite geological time period) consisting of mapped landslide deposits and their underlying lithology (bedrock or superficial) across the United Kingdom. The magnitude-frequency distribution of this secular inventory exhibits an inflected power law relationship, well approximated by either an inverse Gamma or double Pareto model. The scaling exponent for the power-law scaling of medium to large landslides is α = -1.71 ± 0.02. The small-event rollover occurs at a significantly higher magnitude ($1.0\text{-}7.0 \times 10^{-3}$ km^2) than observed in single-event landslide records ($\sim 4 \times 10^{-3}$ km^2). We interpret this as evidence of landscape annealing, from which we infer that the SLI underestimates the frequency of small landslides. This is supported by a subset of data where a complete landslide inventory was recently mapped. Large landslides also appear to be under-represented relative to model predictions. There are several possible reasons for this, including an incomplete dataset, an incomplete landscape (i.e. relatively steep slopes are under-represented), and/or a reflection of a transient landscape response as the UK emerged from the last glacial maximum through a highly variable climate and toward a generally more stable late Holocene state. The proposed process of landscape annealing and a transient response of the landscape has the consequence that it is not possible to use the statistical properties of the current SLI database to rigorously constrain probabilities of future landslides in the UK.