Answers to Anonymous Referee #1

The manuscript introduces a new and apparently user-friendly software, called r.slopeunits, for the automatic subdivision of the terrain into terrain units (i.e. slopeunits; SU). From a geomorphological perspective, the parameters "a" (minimum size of slope-units) and "c" (circular variance of slope aspect) represent the crucial parameters to control the size and orientation (i.e. aspect) of the SU. An optimal subdivision of the terrain into SU (measured with the introduced segmentation metric F(a,c)) is characterized by a high internal homogeneity (i.e. low local aspect variance within a SU) and a high external heterogeneity (i.e. high variability between SU) of SU. The authors also propose an approach to identify the "best" combination of "a" and "c" to generate SU for statistical landslide susceptibility modelling. This procedure is based on the previously mentioned SU segmentation metric and the fitting performance of the generated landslide susceptibility models (measured with the metric AUROC). The optimal terrain partitioning for landslide susceptibility modelling is then based on a combination of high segmentation performance of SU (F(a,c)) and a high fitting performance (i.e. high AUROC) and measured via the introduced function S(a,c). Software and optimization approach were tested in Central Italy, where landslide susceptibility was modelled using logistic regression and a large set of potential predictors. The quality of the present discussion paper is - for its most part - high and contains very useful graphs. From my perspective, the application of r.slopeunits appears to be very useful for a variety of purposes, but especially for empirically-based landslide modelling (e.g. landslide susceptibility modelling, probabilistic hazard modelling, etc.). The presented model is able to account for multiple important details (e.g. minimum area, removing odd-shaped units) and seems to be sophisticated from a technical perspective. I think that the usability/limitations of SU-based approaches, as well as the presented optimization-approach, should be discussed more thoroughly (see comments below).

I recommend a moderate revision.

We thank the Referee for ranking our manuscript high-quality, and for judging our approach useful and sophisticated. Below, we will try to address general and specific comments and to describe the modifications we have made to the manuscript.

I believe that the paper would improve by addressing the following issues:

(i) Adding some discussion on advantages and limitations of the presented model and optimization-approach. E.g. As a potential user of the software, I am very interested on why I should favour SU (e.g. over more easily applicable gridbased approaches) for the purpose of landslide susceptibility modelling. What are potential advantages and disadvantages of the presented software and SU-based landslide susceptibility models in general? (ii) Shifting some text parts to other sections respectively slightly restructuring the paper (see comments below) (iii) Reducing some redundancies within the text (see comments below) (iv) Modifying some figures (see comments below)

Specific comments:

1. From my experience, the boundaries of SU do not directly correspond to the units used by spatial planners (i.e. spatial planning-units are not equal to geomorphic units which is somehow addressed also in p. 11 line 21f)). For me, grid-based landslide susceptibility models appear more flexible in this respect as they can be (with some limitations) regularly and easily adapted to such boundaries (e.g. aggregating pixel values etc.). Discussion of such issues would shed more light on the usability of the presented software.

In the original version of the manuscript, in the "Introduction", we stated that "*Compared to other terrain subdivisions, including grid-cells or unique-condition units, SU are related to the hydrological and geomorphological conditions and processes that shape natural landscapes. For this reason, SU are well suited for hydrological and geomorphological studies, and for landslide susceptibility modelling and zonation".* There are conceptual, technical and practical reasons to favor slope-units as "best" (i.e., "optimal") mapping units for landslide susceptibility assessment and zonation. As a matter of fact, landslide phenomena are naturally bounded to occur within hillslopes, which makes slope-units the natural choice as a mapping unit of reference for landslide susceptibility. Indeed, there is a causal relationship between hydrologically consistent slopes and (most of the) landslides. The same cannot be said for other types of mapping units including e.g., grid cells or unique conditions units [A detailed description of the advantages and limitations of most of the terrain mapping units commonly used for landslide susceptibility assessment can be found in Guzzetti F. (2006) Ph.D. Thesis, Landslide Hazard and Risk Assessment. Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität University of Bonn, Bonn, Germany]. A suitable definition of such "slopes" and their delineation (i.e., the identification of "slope units") is the aim of our work. From the technical point of view, the use of a regular grid for landslide susceptibility zonation requires the use

of parameters which are not related to the "real world" including e.g., the resolution of the grid, or the distance from landslide features (pixels or polygons) to be considered relevant for landslide zonation. These problems are overcome by using geomorphologically meaningful mapping units. From the practical point of view, we maintain that the fact that planners are not currently using slope-units is also (or even largely) due to absence of susceptibility maps prepared on a slope-units basis. We also maintain that a grid-based susceptibility zonation, and the associated map, is not more practical than a slope-unit based zonation, as it is virtually impossible to recognize grid boundaries in the field, whereas slope units are "*easily recognizable in the field*". On the other hand, it is equally possible to overlay the slope-units and administrative planning-units as it is to aggregate pixels within the same planning-units.

2. SU are related to hydrological and geomorphological conditions and therefore well suited for landslide susceptibility modelling (cf. p.1, line 20f). The delineation of SU does not consider any quantity other than terrain aspect. Thus, I assume that a single SU, which is represented by similar slope aspects, may as well exhibit a high intervariability of other landslide-influencing topographical properties (e.g. high slope angles in the upper part and low slope angels in the lower part). Thus, there might be as well a high intervariability of landslide susceptibility within a single SU, because the inclination of a slope is directly related to shear stresses. This might be one drawback of SU within the context of statistical landslide susceptibility modelling and should be discussed within the paper. Furthermore, I wonder if a consideration of the variability of slope angles (or a combination of slope angels and terrain aspect) would be possible or reasonable when dividing the terrain into homogeneous areas for the purpose of landslide susceptibility modelling?

It is true that variability of aspect direction is not the only possible choice for driving a terrain partition into distinct units. In the first place, it should be noted that our algorithm first operates a hydrological subdivision of the terrain, and variability is determined within the obtained half-basins at each iteration. An aspect-based measurements of variability determines the degree to which half-basins are facing different *directions*, while a slope-based variability would determine - in many places and morphological settings - a completely different situation. The example presented by the Referee is a good one. For our purposes, slopes containing two or more modes of the slope distribution (e.g., along the downslope direction) pertain to the same slope-unit, since a landslide can actually travel from one region of the slope-unit to another even if, and sometimes even because, the two regions differ in slope angle. However, landslides cannot travel horizontally from one region to another, if they have substantially different average aspect direction. We have coded in our algorithm a desired feature, by considering aspect direction variability instead of slope variability. A second reason for not considering slope variability is that slope is known to be the main explanatory variable in many landslide susceptibility models. We want our slope-units delineation algorithm not to be directly dependent on such variable, in order not to introduce a bias on the subsequent susceptibility zonation. From a technical point of view, any raster map could be used as a variability measure (and as a measure of segmentation quality) and a more general (unpublished) version of our code actually contains such feature, useful to use the algorithm for different purposes other than landslide-related studies.

3. If I am right, SU affected by only one landslide are treated equally as those units affected by e.g. 10 landslides (= landslide-presence unit). In other words, SU-based approaches do not differentiate between slope units slightly (e.g. 3% of the area) or highly affected (e.g. 80% of the area) by landsliding. From my perspective, such tendencies might also affect the final susceptibility modelling results. This argument further highlights that a detailed discussion of advantages and disadvantages of SU (in the context of landslide susceptibility modelling) may be highly beneficial for potential users of r.slopeunits.

The way we use landslide presence is stated in Section 5, where we write "Adopting a consolidated approach in our study area, SU with 2% or more of their area occupied by landslides are considered unstable (having landslides), and SU with less than 2% of the area occupied by landslides are considered stable (free of landslides). The 2% threshold value depends on the accuracy of a typical landslide inventory map. We acknowledge that the selection of the 2% threshold may influence the selection of the appropriate SU subdivision, and may affect the results of the LS zonation. Examination of different thresholds is not investigated in the present work, because is not an input parameter of the r.slopeunits software, and does not change the logic of the approach or the rationale behind our optimization procedure". The threshold value used in this work was established previously in the same study area, and here we assume it is a good estimate. We acknowledge that the threshold may be different in other study areas, or for different thresholds or even a different approach can be selected by the user to subdivide stable/unstable slope units. In addition, we comment on the usefulness of slope-units on this point. A slope-units is consistent with the fact that each slope-unit can be considered as the result of a given set of geomorphological processes and hence

homogeneous for what concerns landslide predisposing factors. It follows that it does not really matter *where* and/or *how many* landslides have occurred in a unit, but only if the unit is large enough (larger than a meaningful threshold). This is particularly true when landslide inventories are incomplete (a common case in geomorphology). The abundance of recognized landslides in a mapping unit does not mean that the unit has a higher instability than a nearby unit where the number of observed landslides is lower. The concept is particularly relevant if we consider a landslide inventory as a "sample" of the real landslide history of a slope unit. In this sense, just one single landslide should be enough to define a slope unit as unstable.

4. The authors mention that "a detailed terrain partitioning, with many small SU, is required to capture the complex morphology of badlands, or to model the susceptibility to small and very small landslides (i.e. soil slips)" (p. 15, line 2f). Please provide clarification if or how the proposed optimization approach accounts for terrain complexity (i.e. are there other parameters than terrain aspect?). For instance, does the optimization approach favour small SU in the case of complex terrain morphology or very small landslides or should the users decide by themselves on those parameters (which would be in contradiction to the proposed optimization approach)? Is the portion of SU considered as presences (= affected by landsliding) higher in the case the respective units are larger? If yes, does such a change in the ratio between landslide-presences to absence systematically change subsequent modelling results? E.g. Figure 10 suggests that the total areal extent of susceptible areas increases with an increasing SU-size. Could this also influence the apparent fitting performance of the model? Please discuss.

The r.slopeunits algorithm is designed to take into account the different degree of complexity in different regions within the same study area. From a technical point of view, this feature is provided by the iterative procedure. This is explained in detail in Section 3.1, "*Slope-unit delineation algorithm*", and illustrated in Figure 2. The idea behind the algorithm is stated in Section 3, where we write "*The [second] strategy defines an initial small number of very large areas, and reduces progressively their size until a satisfactory result is obtained*". Then, Section 3.1 explains in detail that the progressive reduction of the size of the slope units is stopped at different levels of granularity in different regions of a study area, depending on whether the constraints posed by the *a* and *c* parameters are fulfilled. The value of *c*, in particular, causes the iterations to stop earlier in the regions where variability is lower, resulting in larger slope units, whereas in more detailed (and more geomorphologically complex) regions the iteration is continued, resulting in smaller slope units. The desired "adaptive" feature of the algorithm is thus achieved.

As far as the optimization procedure is concerned, which is not contained in the r.slopeunits software itself, but it is a further action described in this work, it does not explicitly take into account the degree of complexity captured by the slope-unit delineation, but it considers (i) the aspect segmentation quality and (ii) the susceptibility model performance. In other words, (i) capturing the complexity of the terrain is a task accomplished by r.slopeunits; (ii) the optimization procedure is designed in such a way that a combined aspect segmentation quality and susceptibility model performance is maximized simultaneously. In the specific case of landslide susceptibility modelling, we can only expect that the LRM performs better when slope-units are better suited (because of their shapes and sizes) to a particular landslide inventory, since the model uses the explanatory variables more efficiently.

5. The findings demonstrate that the number of significant variables generally increased with a decreasing size of slope-units. Predictors related to local terrain settings (e.g. slope angle) were frequently neglected by the models while especially predictors related to lithology controlled the final landslide susceptibility modelling results. From the text, I deduce that the dominance of lithological parameters is related to an increasing SU-size. The authors finally infer that a logistic regression model generated for an area represented by very large SU "can be replaced by a simple heuristic analysis of the lithological map" even though high AUROC values might be achieved (p .13, line 20f). I think that this observation further indicates that a purely quantitative optimization of a SU-partitioning might (in some cases) not be sufficient to produce high qualitative SU-based landslide susceptibility maps. Please discuss.

The understanding of the Referee about the dependence of the significance of explanatory variables as a function of the slope-unit size is correct. As a matter of fact, we investigated wide ranges for the *a* and *c* parameters, a few combinations of them resulting in unrealistically large/small slope units. The discussion quoted by the Referee is meant to illustrate the inadequacy of such slope-unit delineations, and the limitations posed in those regions of the (*a*, *c*) parameters space, by an uncritical use of AUC_{ROC}. In the region of very large slope-units AUC_{ROC} gives a result which is not acceptable: large values of the metric correspond to a poor statistical significance of the explanatory variables and to results which is not of any practical use, since susceptibility maps prepared with large units have no meaning. The fact that slope angle is not significant indicates that slope units are probably too large, and each of them containing any value of the terrain slope angle, so that the LRM cannot discriminate between values of slope causing or not causing landslides. The behavior of the AUC_{ROC} is known in the literature, and it is simply due to the metric

being a synthetic index that summarizes in one score the results pertaining to an arbitrarily large study area, with no further dependence on different geographical locations. To our knowledge, a metric that can take into account the complexity of the problems like the one we discuss in our paper does not exist, and we can only rely on existing tools. In our discussion we also clarify that we proposed one possible metric. Users can select other metrics, and replace AUC_{ROC} with a different and possibly more suitable metric, according to a specific problem and their personal judgment and understanding of the problem at hand.

We stress that the merit of our optimization procedure is that we provide a *quantitative* way of rejecting situations like the one described above, rather than relying on common sense (*i.e.*, using arguments like "slope units are too large or too small"), pre-existing results or maps (*i.e.*, a different method produced a susceptibility map from which our result cannot be too different), or other forms of personal or heuristic judgement. We proposed an overall performance metric, S(a,c) = F(a,c)R(a,c). The function is the sole metric we use to determine the optimal values of the parameters, and the maximum of S(a,c) in our test case is indeed far from the limiting-case situations of slope-unit with unrealistic sizes. In conclusion, we proposed a working solution that can be criticized or improved, but it certainly is a rigorous solution.

6. The proposed combination of the aspect segmentation metric and the AUROC metric appears logical from a purely quantitative perspective. The subsequent final metric S(a,c) is then based on a multiplication of both previously mentioned and normalized metrics. It would be interesting to know why the authors chose to multiply the respective metrics (instead of using an average) to get S(a,c). Is it because a very low value of one metric (e.g. 0.1) prevents an "acceptable" S(a,c)-value in the case the other metric is relatively high (e.g. 0.8) (e.g. multiplication: 0.08 vs. average: 0.45)?

The Referee is correct. Multiplying the two quantities, F(a,c) and R(a,c), after normalization to the [0,1] interval, causes the final metric S(a,c) to be negligible where any of the two factors is negligible, so that no unacceptable situation arises. The procedure seems to be not only logical, as the Referee observed, but it proved to be effective in rejecting unrealistic results.

7. The usage of S(a,c) appears to be based also on the assumption that a high AUROC value is directly related to a higher quality/usability of the underlying logistic regression based model. In this context, I do not see a problem to use the calibration set to measure this metric since logistic regression models are relatively robust and do not tend to (strongly) overfit on training data. However, several studies outline potential limitations of AUROC-based measures for spatial distribution models (Frattini et al., 2010; Lobo et al., 2008; Steger et al., 2016). Therefore, I assume that there might be a risk that the conducted direct deductions (i.e. by interpreting solely the metrics) can lead to misleading conclusions. I think that a discussion of potential drawbacks of the proposed metrics (S(a,c)) would also be valuable. Frattini P, Crosta G, Carrara A (2010) Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology 111:62–72. Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17:145–151. Steger S, Brenning A, Bell R, Petschko H, Glade T (2016) Exploring discrepancies between quantitative validation results and the geomorphic plausibility of statistical landslide susceptibility maps. Geomorphology 262:8–23.

We agree with the Referee. However, we strongly stress that we did not "conducted direct deductions", as we did not make a straightforward use of the AUC_{ROC} metric, as we discussed already, but we rather produced a very different overall metric. In our opinion, this is now clear in the Introduction, throughout the manuscript, and in the Conclusions.

8. I also suggest a minor restructuring of some text parts to enhance readability and to reduce redundancies:

From my perspective, several "non-discussion-parts" of the manuscript already contain useful discussions (e.g. p8 line 28-30; p. 9 line 28f; p. 10 line 11f etc.). As a reader, I would prefer to read those text segments within a well-structured discussion, which summarizes those issues (e.g. one part may relate to the SU-segmentation, another to advances/limitations of the metrics, another to the conducted susceptibility modelling etc.). I propose to restructure and expand the discussion part (e.g. including subsections, separating discussion and conclusion).

We followed the Referee's suggestion to expand the Discussion. We moved the conclusions to a separate section. The new Sections (Discussion, Conclusions) are listed at the end of this comment. Section 8.4 was then removed from the Manuscript, and Figure 1 was changed according to the new sections numbering. The modified Figure is attached to this comment as Figure 1.

p. 2, line 20-23: I suggest to shift these lines to section 5 ("Landslide susceptibility modelling")

We have accepted the suggestion.

p. 13f (section 8.4): I propose to address the calculation of the combined metric S(a,c) within the methods section.

We have accepted the suggestion.

p. 14 line 15 to 20: Those sentences are similar to the text passages in p. 1 line 16-19. I recommend explaining the concept of SU solely within the introduction (p.1)

The suggested text was removed from the Discussion.

9. Figure 5: I would prefer a more intuitive colour selection for the land cover map (e.g. dark green for forests, light green for pastures, brown for arable land, blue for water etc.). Figure 10: I think that inserts of corresponding metrics (F(a,c), R(a,c), S(a,c), fraction of significant predictors) within each of those nine maps would further enhance traceability (i.e. interrelations) of the results. Maybe you can also present the susceptibility map produced by the optimal parameter combination (i.e. "a" and "c").

We have changed the colors in the Figure according to the suggestion. The new figure is attached to this comment as Figure 2.

10. The title clearly reflects the content of the paper. The guidelines of the journal (GMD) state that "the model name and number should be included in papers that deal with only one model." Maybe the authors can add the name of the software "r.slopeunits" to the title?

We have added the name of the software to the title, which now reads "Automatic delineation of geomorphological slope-units with r.slopeunits v1.0 and their optimization for landslide susceptibility modelling".

The new Discussion section reads as follows:

We have run the r.slopeunits software with a significant number of combinations (99) of the (a, c) input parameters, and a corresponding number of realizations of the LS model. Results showed that new r.slopeunits software was capable of capturing the morphological variability of the landscape, and to partition the study area into SU subdivisions of different shapes and sizes well suited for LS modelling and zonation. As a matter of fact, depending on the type of landslides, the scale of theavailable DEM, the morphological variability of the landscape, and the purpose of the zonation, the detail of the terrain subdivision may vary. A detailed terrain partitioning, with many small SU, is required to capture the complex morphology of badlands, or to model the susceptibility to small and very small landslides (i.e. soil slips). A coarse terrain subdivision is best suited for modelling the susceptibility of very old and very large, deep-seated, complex and compound landslides. Coarse subdivisions can also be used to model the susceptibility to channeled debris flows that travel long distances from the source areas to the depositional areas. Subdivisions of intermediate size may be required for medium to large slides and earth flows (Carrara et al., 1995). By tuning the set of user defined model parameters, r.slopeunits can prepare SU terrain subdivisions for LS modelling in different geomorphological settings.

Concerning the LS model, we acknowledge that our selection of the 2% presence/absence threshold may influence the production of the appropriate SU subdivision, and may affect the results of the LS zonation. Examination of different thresholds is not investigated in the present work, because it is not an input parameter of the r.slopeunits software, and does not change the logic of the approach or the rationale behind our optimization procedure.

We clarify that the subdivisions produced by r.slopeunits using different (a, c) parameters are nested i.e., the boundaries of a coarse resolution subdivision encompass the boundaries of intermediate and finer subdivisions (see C, D, E in Fig. 7). This is a significant operational advantage where landslides of different sizes and types coexist, posing different threats and requiring multiple and combined susceptibility assessments, each characterized by a different terrain subdivision (Carrara et al., 1995). Optimal values of the (a, c) parameters have to be determined to obtain the best SU subdivision for a particular goal, in our case, LS modelling. We defined a custom objective function S(a,c) to determine such optimal values. S(a,c) is the product of a segmentation quality measure, F(a,c), and LS model performance in calibration, R(a,c). If only the F(a,c) metric is used to select a particular set of modelling parameters, the resulting ``optimal'' (best) set of SU has the only meaning of ``best partition of the territory in terms of aspect segmentation''. Similarly, the R(a,c) metric considers solely the classification results of the LS model, and not the geometry of the single SU, some of which may be inadequate (e.g., too large, too irregular, too small) for the scope of the terrain zonation. Values of F(a,c) indicate that there are combinations of the cc and ac parameters that result in SU subdivisions that do not satisfy the user requirements in terms of SU internal homogeneity and external heterogeneity (Fig. 8) (Section 8.2). On the other hand, the AUC_{ROC} metric increases with the average size of the SU (Fig. 11) (Section 8.3). To select the optimal terrain partitioning for LS zonation in our study area, we exploit the objective function S(a,c), which simultaneously quantifies (Section 6): (i) the SU internal homogeneity and external heterogeneity (Fig. 13A), and (ii) the (fitting) performance of the LS model (Fig. 13A). Maximization of S(a,c) (Fig. 13B) provides the best combination of the (a, c) modelling parameters for a terrain subdivision optimal for LS modelling, in our study area.

In addition to the AUC_{ROC}, we have analysed the performance of the LRM model by studying the fraction of significant variables used by the LRM, to qualitatively understand the behavior of the classification model as a function of the r.slopeunits software input parameters or, in turn, as a function of the average size of the SU. The LRM is expected to use the input data less efficiently when the average SU size grows, resulting in a smaller number of significant input variables, which is indeed what we observed. This is due to the LRM inability to discriminate between input variables when the SU are too large, since each unit usually contains all the possible values of the variables: using less data makes it easier for the LRM model to produce a high-AUC_{ROC} result, which does not necessarily correspond to the optimal SU set. The complication is removed using the $S(a,c) = F_o(a,c) R_o(a,c)$ function, whose F(a,c) component prevents unrealistic SU sets (both large and small) to have a high overall score.

The function S(a,c) (Fig. 13B), calculated in our test case for different combinations of the a and c modelling parameters, has a maximum value at $a = 150,000 \text{ m}^2$ and c = 0.35. The set of SU that corresponds to the optimal combination of the modelling parameters can be singled out as our ``optimal'' (best) result. The LS results corresponding to the optimal combination was already shown in the central box of Fig. 10, and the optimal SU map is presented in Fig. 14.

The new Conclusions section reads as follows:

Despite the clear advantages of SU over competing mapping units for LS modelling (Guzzetti, 2006), inspection of the literature reveals that only a small proportion (8%) of the LS zonations prepared in the last three decades worldwide was performed using SU (Malamud et al. 2014). The limited use of SU for LS modelling and zonation is due -- among other factors -- to the unavailability of readily available, easy to use software for the accurate and automatic delineation of SU, and to the intrinsic difficulty in selecting a priori the appropriate size of the SU, for proper terrain partitioning in a given area.

To contribute to fill this gap, we developed new software for the automatic delineation of SU in large and complex geographical areas based on terrain elevation data (i.e., a DEM) and a small number of user defined parameters. We further proposed and tested a procedure for the optimal selection of the user parameters, in a 2,000 km\$^2\$ area in Umbria, Central Italy.

We expect that the r.slopeunits software will be used to prepare terrain subdivisions in different morphological settings, contributing to the preparation of reliable and robust LS models and associated zonations. We acknowledge that further work is required to investigate the optimization of SU partitions for different statistically-based tools used in the literature for LS modelling and zonation (e.g., discriminant analysis, neural network). (Guzzetti et al., 2012) have argued that lack of standards hampers landslide studies. This is also the case for the production of landslide susceptibility models and associated maps. We expect that systematic use of the modelling framework proposed in this work (Fig. 1, Section 2) and of the r.slopeunits software for the objective selection of the user defined modelling parameters, will contribute to the production of more reliable landslide susceptibility models. It will also facilitate the meaningful comparison of landslide susceptibility models produced e.g., in the same area using different modelling tools, or in different and distant areas using the same or different modelling tools.

Finally, we argue that the proposed modelling framework and the r.slopeunits software are general and not site or process specific, and can be used to prepare terrain subdivisions for scopes different from landslide susceptibility mapping, including e.g., definition of rainfall thresholds for possible landslide initiation, distributed hydrological modelling, statistically-based inundation mapping, and the detection and mapping of landslides and other instability processes from satellite imagery.