

## NHESSD Interactive comments

### Response to Review 1

We thank the Reviewer for her detailed review of our work. We take note of your suggestions to grammatically improve the body text. However, in your specific notes there are some questions and comments that demand a direct response:

**Page 3, line 29: which information about landslides location did you collect? Did you digitalize the entire perimeter? Or location refers to the four GPS points? You can explain this better.**

We used the four GPS points collected on the field as reference to draw landslide polygons using QGIS software and Google Earth satellite imagery.

To clarify this point, we modified the text as follows:

“We collected information about the location of each observed landslide, four GPS points (crown, toe, and two flanks), photographs, surrounding area features and information about the landslide type, according to the Varnes, (1996) classification. Each documented landslide was drawn and digitized using its four GPS waypoints recorded and photographs as a reference. QGIS and Google Earth satellite imagery were used for the purpose.”

**Page 3, line 5: why did you consider only shallow landslides?**

We considered that different typologies of slope instabilities are triggered by different mechanisms, which means that the predisposing factors and the way they affect to the occurrence of a given type of landslide can be different. Before the field campaign, we reviewed bibliographical sources, finding that within our study area 753 landslides were inventoried in different studies (INGEMISA, 1995; Gipuzkoako Foru Aldundia, 2013; IDE de Euskadi, 2014). Among all of them, 75% were considered as shallow slide type of movement, 10% were rock falls or rock mass deposits and 3% were flows or complex movements, adding to the 12% of the subset that was labelled as landslide scarp but without specifying the type of landslide. Although those bibliographical sources were not considered appropriate for our study – because they were heterogeneous in type and quality – they showed the most frequent type of landslide, so we focused our research on shallow landslides.

We included the information in the text as follows (section 3.1, page 4):

“During several field trips, 793 individual landslides were collected, and 746 of them were classified as shallow movements. Our observations together with the revised bibliographical sources (INGEMISA, 1995; Gipuzkoako Foru Aldundia, 2013; IDE de Euskadi, 2014) confirm that shallow slides are the most frequent type of landslide in the study area. Consequently, in order to consider only landslides triggered by the same mechanisms, only shallow movements were used as landslide presence when defining the dependent variable in the susceptibility assessment.”

Gipuzkoako Foru Aldundia (2013). Evaluación y gestión integrada de riesgos geotécnicos en la red de carreteras de la Diputación Foral de Gipuzkoa. Technical report, Mugikortasun eta Bide Azpiegituren saila, Unpublished report.

IDE de Euskadi (2014). Infraestructura de datos espaciales de euskadi.

INGEMISA (1995). Inventario y Análisis de las Áreas sometidas a Riesgo de Inestabilidades del Terreno de la C.A.P.V. Technical report, Eusko Jaurlaritza.

**Page 6, line 6: here it's not important that LR is implemented in LAND-SE software. Moreover, you already mentioned that you used this software to perform the analysis. What is important is that you applied the multivariate LR. Is it the most used statistical method for susceptibility in general or for landslide susceptibility? Please, specify.**

It is actually relevant to mention that we used LR as implemented in the LAND-SE software, since the latter is actually a comprehensive package for data preparation, model training and validation, and visualization of the results. Concerning LR, what the review paper discuss is only its application to landslide susceptibility studies. In the text, we modified "susceptibility" to "landslide susceptibility", to avoid ambiguity.

**Page 8, line 9: Why only three random set? What is the implication of using three or more?**

The test the Reviewer refers to was carried out just to confirm that the random selection of the landslide inventory would not affect the model results in a relevant way. Indeed, before starting with the main analysis, three preliminary LR runs were performed only changing the training and validation data sets. In all the cases the model classification performances were very similar. So, in order to choose only one data set for further comparative analysis, we decided to select the one with the best classification result, although we believe that conclusions would not be affected if any other data set would have been used.

We included the information in the text as follows (section 4.3, page 8):

"This exercise allowed us to confirm that the random selection of the landslide inventory would not affect the model results in a relevant way, because in all the cases the model classification performances were very similar."

**Page 8, line 21: Are you sure you mean >0.15% of unstable pixels?**

The total inventoried landslides actually covered 0.15% of the surface of the whole study area. Nevertheless, the presence of one single landslide pixel within a slope unit was not considered enough to label this SU as unstable. Therefore, instead of arbitrarily defining a given threshold value in order to consider a SU as unstable, we decided to use the overall landslide density in the WA.

We included the information in the text as follows (section 4.3, page 9):

"The presence of one single landslide pixel within a slope unit was not considered enough to label this SU as unstable. Therefore, instead of arbitrarily defining a given threshold value in order to consider a SU as unstable, we decided to use the overall landslide density in the WA. For this reason, we considered as unstable those SUs containing equal or more 0.15 % of unstable pixels, and stable otherwise."

**Page 8, line 25: do you mean that from your computation it results 304 unstable SUs? If so, please specify.**

Yes, in 304 cases the SU contained 0.15% or more unstable pixels.

**Page 9, line 32 and Page 10 line 4: why 152 validation SU? They are not 76?**

As it was explained in section 4.3 (Page 8, lines 25-30), 76 SU labelled as unstable were used for validation. Then, the validation sample was obtained by selecting the same number of SU labelled as stable. Thus, the validation sample contained 152 SU (76 unstable + 76 stable).

We included the information in the text as follows (section 4.3, page 9):

“In 304 cases the SU contained 0.15% or more unstable pixels, so we selected at random 228 of them (75%) for training, and the remaining 76 (25%) were used for validation. Like in grid cell approaches, we created two different training samples where unstable SUs were exactly the same, and only the stable SUs were different in each case. The first training sample includes 228 stable SUs selected at random along the WA. The second training sample includes an equal number of stable SUs units selected at random among those that at least partially overlap the ESA.”

“Additionally, 76 SUs labelled as unstable were used for validation. Then, the validation sample was completed by adding a random selection of the same number of SUs labelled as stable and which at least partially overlap the ESA. Thus, the validation sample contained 152 SUs (76 unstable + 76 stable).”