Interactive comment on "Reconstructing long-term gully dynamics in Mediterranean agricultural areas" by Antonio Hayas *et al.*

Referee #1

This paper is a great topic and a novel approach to understanding dynamics through extrapolation and quantification of uncertainty through Monte Carlo. The detailed network mapping is impressive. However, the field sample sizes are fairly small, and a bit of a concern for understanding true field variability. More emphasis could be put on field interpretation of geomorphic processes and channel evolution, to match better with observations and extrapolation from air photo interpretation and generalized land use metrics.

Moderate revisions are recommended to correct some misunderstandings and also more importantly bring out the underlying geomorphic processes that could be operating in this catchment.

Moderate Comments There is nothing overly special about Mediterranean climates and gully erosion. Other areas of the globe with highly variable rainfall also have gullies. Perhaps downplay the Mediterranean climate emphasis or compare to other areas of the globe. Soils and geology and topography are also big factors with gully erosion.

We appreciate the encouraging comments of the Reviewer. We agree with the observation about the emphasis on the Mediterranean conditions. Therefore we have changed the text where such references were made:

-abstract, we deleted *especially*, leaving the sentence as *Gully erosion is an important erosive process in Mediterranean basins*
-The sentence of page 1 lines 1-3 has been rewritten as:

Understanding gully erosion dynamics under changing land use and climate conditions is essential for soil and water conservation

-Also on page 2 line 30-33 we have deleted the reference to Mediterranean climates:

Under different climates, especially where rainfall is less uniform and more concentrated in time

Page 2 Line 17 Focus more on field empirical data and methods of other authors, as models you review here do not really help estimate gully volumes at the catchment scale. Are there more references you can cite that use your methods of estimating gully volume from remote measurements of width and length, and field measurements of depth? And then the use of m Carlo simulation to extrapolate from small sample sizes of 35 and 27?

We have performed an additional literature search and have located several additional references on this topic, although most are on local and not regional scales. The study by Nachtergaele and Poesen (1999) uses a similar methodology, where gully cross sections are measured in the field, while gully length is extracted from aerial photos.

With respect to the use of Monte Carlo simulations in gully erosion research, this is relatively novel to our knowledge, so we found no references on this topic, irrespective of sample size. Further on, we have replied in more detail to the issue of sample. However, Istanbulluoglu *et al.* (2002) used a similar Monte Carlo approach to estimate channel incision locations in a catchment of about the same size as ours. For the different input parameters they have between 19-25 field measurements (see their Table 1), which they use to fit a theoretical probability distribution, similar to our method.

We have deleted:

More research will be needed in order to develop full three-dimensional models, capable of accurately predicting gully erosion volumes

We have included more references related to field empirical data in page 2, Line 20:

Different methodologies, apart from traditional field measurements with total station, laser profilemeters and poles (Castillo et al., 2012), have been proposed and successfully applied to estimate gully volumes. For instance, at the individual gully scale, 3D reconstruction from high resolution aerial photography and digital photogrammetry has been widely applied (e.g. Marzolff and Poesen, 2009). Recently, terrestrial imagery modelling and Structure from Motion-Multi View Stereo (SfM-MVS) procedures have also been used to determine *gully volumes (Gómez-Gutiérrez et al. 2014; Frankl et al., 2015 and Castillo et al., 2015). Terrestrial LiDAR has been applied to measure rills or gullies at both laboratory and plot scale (Vinci et al.,2016; Momm et al., 2011, 2012). Nevertheless, at the catchment scale, the number of studies is limited. At this scale, most studies focus on the areal extension of gully networks, using aerial photos or other remotely sensed imagery. Few studies report gully volumes, due to the inherent difficulties of determining depths for the whole gully network. Nachtergaele and Poesen (1999) determined gully length from aerial photos and by additional field measurements they established a mean cross section to calculate volumes of small ephemeral gullies in the Belgian loess belt. Martínez-Casasnovas (2000) mapped and quantified the erosion produced in a gully systems of big dimensions by processing multitemporal orthophotograms and DEMs in a GIS for a 25 km2 catchment located in NE Spain. Frankl et al. (2011) used sequential photographs to link long-term gully and river dynamics to environmental change in Northern Ethiopia. More recently, Peter et al. (2014) used UAVs and photogrammetric analysis to quantify gully erosion, albeit at local scale in the Souss Basin (Morocco).*

Page 2 Line 70 See recent paper by Shellberg *et al.* 2016 in Geomorphology as an example on the rates and dynamics of gully erosion pre- and post European settlement in Australia... and over long time periods. Many other examples around the world too.

We have modified Line 72 to include a new citation:

Also Shellberg et al. (2016) observed an increase in the gully erosion by the changes in the land use produced by post-European settlement in the Mitchell River fluvial megafan (Queensland, Australia). This relationship between pioneers and gully erosion was previously suggested by Leopold (1924) in the US.

Page 3 Line 48 What were the various scales that these air photos were originally collected at? Could differences in scales influence the accuracy of results between years, even with the working scale for measurement was set at 1:5000? Also, the data gas between 1956 and 1980 and 1984 and 1999 are large, compared to the other years. It would be good to highlight the dynamic uncertainty more for these years, as this data limit the ability for "high temporal resolution" analysis. Use the high resolution years for discussion if uncertainty in other years.

The scales of the air photos have been added to Table 1.

We believe scales do not influence the accuracy in our case. Our results on gully network evolution and drainage density for the years 1956 and 1980 (fig. 7) suggest that differences between scales (in this particular case) are not significant for our purpose. If scale differences were causing an impact on the accuracy of our results, a lower drainage density in 1956 than in 1980 could be expected, since the original scale of the air photos, *flight scales*, in 1956 and 1980 are, respectively, 1:33,000 and 1:18,000. The respective restitution scales vary between 1:10,000 and 1:5,000. However, our results show a different trend. This trend could be attributed to the frequent infilling operations favoured by the advent of the tractor in the middle of the decade of 1960-70. In addition, the large majority of the gully sections surpass widely the minimum raster resolution in the orthophotos (1 m), suggesting that the uncertainty is limited.

We agree that these 1956-1980 and 1984-1999 periods are longer and are aware of the limitations. We have noted this already in the Discussion section, on page 6 line 77-82:

However, given the length of this periods and since there are some particular years (i.e. 1961-1962) with extreme rainfalls it is likely that positive gully growth occurred during this period that was later masked by infilling.

We have added to this discussion:

This shows that longer periods, such as 1956-1980 and 1984-1999, are subject to a higher uncertainty with respect to the post-1999 period, when a higher temporal resolution is available.

Page 3 Line 74 Were these 35 stretches the same as the 27 sections visited below? If not, why were not the 35 stretches visited in the field to confirm measurements from air photos? There seems to be a mismatch here/

No, these 35 and 27 sections are not always the same. The sections used in the orthophoto interpretation (n=35) were established first. Afterwards, during the field truthing, some of the sections had to be moved due to practical limitations, but maintaining the essential morphological characteristics since the gully sections were

proximal. Practical limitations include accessibility, but mostly the alteration of gully sections due to farmer operations. Therefore, we always aimed to measure an undisturbed gully cross section.
To clarify this to the reader, we changed page 3, line 79:
Gully top width and depth were measured at 27 representative sections that were located as close as possible to the 35 sections used in the photointerpretation

Page 3 Line 74 Are there more references you can cite that use your methods of estimating gully volume from remote measurements of width and length, and field measurements of depth? And then the use of Carlo simulation to extrapolate from small sample sizes of 35 and 27? Will also need more convincing of the readers that 27 or 35 sample points are enough to extrapolate to a 20 km^2 catchment.

With regard to similar references, see our previous comment, we have added more references such as Frankl *et al.* (2011) which determined gully erosion from oblique photos and Nachtergaele and Poesen (1999) who used field measurements of gully sections in combination with gully length derived from aerial photos.

We added to the text page 3, line 85:

This methodology of combining photointerpretation with field measurements of gully morphology is similar to Nachtergaele and Poesen (1999).

With regard to the reduced number of sample points commented by the Reviewer, we are aware of this limitation. However, the objective of the Monte Carlo method is exactly to reduce the number of observations and avoid having to measure the cross section of each gully section. We do think that the gathered information is enough to develop a Monte Carlo method. Hammersley and Handscomb (1964, section 1.1) appositely explain that one of the reasons for adopting such a method is the inherent difficulty and high economic cost to get field information. In any case, the Monte Carlo simulates in the area occupied by gullies, not the entire catchment, which is less than a 5% of the total catchment area (20 km^2). As noted above, Istanbulluoglu *et al.* (2002) used a similar Monte Carlo approach to estimate channel incision locations in a catchment of about the same size as ours. For the different input parameters they have a similar amount of field measurements, between 19-25 (see their Table 1), which they use to fit a theoretical probability distribution, similar to our method. As long as the field measurements allow to characterize the probability distribution function (pdf) of the variable in question, we are confident that the Monte Carlo method gives a reliable estimate of the uncertainty. We believe that the results shown in Table 4, where we quote the fit between the observations and the theoretically fitted pdfs, with p-values of 0.64-0.98 demonstrate the good fit of these relations.

We have added to the discussion section some more explanation on this topic at page 7 line 39:

Although more field measurements of gully sections would be advantageous in order to reduce uncertainty, time and money spent on ground truthing will increase accordingly. However, the high p-values of 0.64-0.98 obtained here for the fit between the theoretical probability distribution function and the experimental data suggests satisfactory results can be obtained, even with a limited field sample. Moreover, also Istanbulluoglu et al. (2002) successfully used a Monte Carlo approach to estimate gully incision locations using a similar amount of field data.

Page 5 Line 1 Is this % of total catchment area? or % of all land uses? Or just % of occupied land by humans?

The comment is correct. In all cases we refer to total catchment area. Therefore to reduce the ambiguity of the terminology the sentences of lines 1 to 6 in page 5 have been modified as follows:

In the study period, olive orchards grew from 13% to 63% of the total catchment area. At the same time herbaceous crops decreased from 85% to 35% of the total catchment area. The main land use change occurred between 1984 and 1999, when the olive orchards passed from occupying 25% to 48% of the total catchment area.

Page 5 Line 64 Was there a correlation between farmer infilling and levelling activity during one period with an increase in erosion during the next period? That is, a lag effect of the influence of actual machine disturbance on gully erosion and perhaps water yield. Generally if you disturb a gully, it will erode faster for a while into the future until it settles back down.

We totally agree with this comment. This was also demonstrated by Gordon *et al.* (2008) who obtained much higher erosion rates for infilled gullies compared to gullies left alone. However, we cannot appreciate a clear correlation after a close inspection of our data, shown in the figures 6 and 13. Intense land use change between the photographs of 2005 and 2007 was followed by only a moderate gully erosion phase in the 2007-2009 period. Only in 2009-2011 gully erosion rate increased significantly, due to extreme rainfalls in this period. In other periods, as for instance 1984 - 1999 land use changes were more intense. However in the next period (1999-2001), erosion rates were negative.

We have included this now in the discussion, page 6 line 82:

It could also be expected that infilling phases were followed by phases of higher erosion rates. Gordon et al. (2008) obtained higher erosion rates from periodically infilled gullies compared to gullies that were left undisturbed. However, our results do not show such a trend. For example, land use change and infilling between 2005 and 2007 was followed by only a moderate gully erosion phase in the 2007-2009 period.

Page 5 Line 73 This is a key hypothesis, and needs to be brought out more here and in discussion, as you say that extreme rainfall is the dominant variable, but you do not have any other data on land use besides just % area covered, and thus do not have a metric that links runoff accelerated by land use to gully erosion.

It is true that more research on this issue is needed, we are currently investigating the effect of olive roots versus cereal roots on gully incision rates. At this point, the main conclusion that we draw from our data is the importance of rainfall. However, with the cited paragraph, we wanted to indicate that land use could play a secondary role to lower the resilience of the system. This is explained further in a following comment where we include the reviewers comments on thresholds.

We have added some more explanation on this in the Discussion section page 7 line 15:

Gioia et al. (2008) stressed the importance of different runoff thresholds to explain flood occurrence in the Mediterranean areas. Ordinary flows are produced when rainfall rate exceeds the infiltration rate of the soil in a small area, a typical case of Hortonian runoff generation, or Hortonian threshold, while what Gioia et al. (2008) denominated outlier events, occurred when the water of almost continuous rain spells surpassed the storage capacity of the soil in a large area of the catchments, extreme rainfalls for explaining the runoff behaviour of catchment, or Dunnean threshold. This behavior is similar to the complex response to the geomorphic thresholds discussed by Patton and Schumm (1975)

Our data seems to indicate that land use did not play a dominant role, although we cannot exclude that land use changes to olives and soil management have lowered the resilience towards gully incision.

Page 6 Line 27 Could this be partially due to the higher quality and resolution of air photo data late in the period? Mismatch of data sources here for these different periods.

We do not believe there is a quality or resolution effect at play here. The air photos of 2005, 2007 and 2009 present a similar quality and resolution. If this were a quality effect, we would expect a gully volume increase already in 2005. However, we do not observe such increase. Nonetheless if we look at the standardized annual rainfall (fig. 4) we see that rainfall in those periods have been below the average, which seems to corroborate our hypothesis that gully growth is conditioned by rainfall.

Page 6 Line 37 ha of total catchment area? Or ha of total gully area? ton or tonnes? which units? Metric or English? This is an issue through the paper.

We understand that the terminology used could be ambiguous. Units used through the paper are referred to the metric system. "ton" have been replaced by "t" all through the paper. We always refer to ha of total catchment area, never calculated with respect to gully area.

Line 36 in page 6 has been modified as follows:

Dynamics of gully erosion rate referred to the total catchment area are shown in Figure 13.

Page 6 Line 53 again, are these yields based on total catchment area or gully area? That is, are the gully rates averaged across the whole catchment area?

No, gully erosion rates cited here are reported with respect to the total catchment area.
Line 46 to 47 has been modified to clarify this point:

The average gully erosion rate of 39.7 t $ha^{-1} yr^{-1}$ *for the total catchment area obtained in this study, ...*

Correction in Line 54 to refer it to the total catchment area:

The highest gully erosion rate of 331 t ha-1 yr-1 referred to its catchment was found ...

Page 6 Line 84 Many of us in the world would consider Mediterranean Climate a type of Temperate Climate. Under the Koppen climate classification system, a Mediterranean climate is a type of the temperate climate group ("C" climates). Please expand more or change terms and classification. Maybe deemphasize the climate variable and talk more about soils and geology, and the build-up of colluvium/alluvial soils and then their release during major events and gully erosion.

The Reviewer is right, again. Both climates are temperate, what makes very unfortunate the sentence. Using the revised version of the Köppen-Geiger world climate types of Peel et al. 2007, we are rewritten the sentences in the text. We agree and have modified the original version, Line 83 now says:

Moreover, the data presented here clearly show than in Mediterranean-climate areas (Csa type in the Köppen classification) the gully growth dynamics are different than in Temperate Oceanic west-European areas (Cfb type) for instance.

And now Line 89 says:

Data for this study was from the Temperate Oceanic (Cfbn type) loess belt...

Page 6 Line 91 See recent paper by Shellberg et al. 2016 in Australia on continued near linear increases of gully erosion over time in highly erodible soils.

We appreciate the suggestion and consequently the reference has been included in the text:

This observation is not unique since in other environment Shellberg et al. (2016) have detected an almost continuous increasing trend in the gullies of the Mitchell River in Queensland.

Page 6 Line 103 Dams often decrease sediment yields and can create incision in main channels, and thus rejuvenate tributary systems. You cannot discount this without some references or data for overall channel stability in the catchment. Also over long time periods, most catchments are in a degradation state, and thus maybe you are in some part of a bigger cycle? No doubt human land use has also increased erosion. Plenty of references there. But is it from upstream or downstream impacts and controls, or both?

In a broader geomorphological context, the main Guadalquivir channel is in an incision stage. However, since the 1956 many new dams were constructed that control the stream dynamics. Although it is true that dams induce complex upstream and downstream impacts, in this particular case there are no dams upstream which could create incision in the main channels due to sediment deprivation. There is only a downstream base level control due to the dam, limiting any incision effect in the main channel network of our catchment.

We have added some more discussion on this on page 6 line 100:

progressive increase on the erosion rate. During the Quaternary, the main Guadalquivir River is in an incision stage due to base level fall. However, this incision has been slow, as demonstrated by Uribelarrea and Benito (2008), who found evidence of a 1.2 m incision over the last 500 years. In any case, since the 1950-60s, when many dams were constructed, the Guadalquivir is a highly regulated river. Such dams are known to have both a downstream incision effect due to removal of sediment load, and an upstream aggradation effect. With respect to our study area, there are no upstream but only downstream dams. Therefore, it is expected that the influence of the incision stage has been artificially limited in this catchment since the 1950s and that the observed changes in the gully network can be fully attributed to upstream changes in rainfall or land use regime.

Page 7 Line 37 How would this uncertainty be reduced if you had much more field data. 40 data points is a small sample size to be sure you have calculated uncertainty correctly.

See previous comments. This is difficult to know without measuring more profiles. We considered that in our study zone this data is enough since it comprises the whole range of widths present in the area and their frequency could be considered representative from that in the study zone. Although, we understand that more field data could be desirable, it is difficult to obtain more field data (due limit access to parcels, time and expenses limitations) which motivates the Monte Carlo simulation.

Page 7 Line 54 Please do not recommend more modelling. Field measurement is essential to understand gully dynamics. Or Repeat airborne LiDAR. or small scale terrestrial lidar.

We agree that field measurement is essential. We have modified the final version to not emphasise modelling in Line 54.
...further research of runoff, gully headcut retreat rates, and sidewall dynamics should be made at this last point

Page 7 Line 67. again, need field measurements not models to understand these dynamics.

As expressed above, we also acknowledge that field measurements are essential to improve our understanding on gully erosion, but this information needs to be inserted in a structure which help us to interpret how gullies start and evolve. Line 67 has been modified:

Further studies with more field data are needed to improve ...

Line 68 has been modified as well:

Implementation of physically-based models of gully retreat rates and sidewall collapse as well as more field measurements and interviews of local farmers on soil management practice could contribute to a better understanding...

Page 9 Figure 1 Figure 1 needs to be improved. The stream map shows not much. Perhaps this would be a great place to show the high resolution air photos so we can see what the landscape and land use looks like? Some field photos of gullies and land use would be really useful for the geomorphic reader. Page 10 Table 1 What about scale? 1:50000 etc...

Figure 1 has been reformed to inform better about the study site. The original scale of the aerial photos has been included in the new version of Table1.

Page 13 Figure 10 Were these from the field measurements at the 40 sites or from extrapolated estimates?

Figure 10 shows width dynamics obtained from the photointerpretation-measured widths. Caption has been modified to make it clear:

Gully top widths dynamics in the period 1956 - 2013 derived by measuring by photointerpretation.

Page 14 Figure 12 This is a major increase in gully volume. If real it shows a major permanent shift and increase in erosion.. Perhaps a threshold was reached due to intrinsic variables like sediment storage in valleys, or land use or rainfall, or combination. See threshold complex response by Schumm. Shellberg 2016 also has a decent literature review on the subject, as well as Tucker in USA.

This is an apposite observation. Certainly the threshold complex response is a common characteristic of natural phenomena. See the answer to the comments of Page 5 Line 73.

Page 15 Table 2 Should not crops just be used for crops, pasture for grazing, and dense scrub-land for not grazed or cropped areas? Are there only two simplified classes or 3?

This was an error carried on in the version of the manuscript used by the reviewer, not in the version appearing in the Journal web page (http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-239/hess-2016-239.pdf). The error was corrected in the final version. Pasture, dense scrubland, streams and natural watercourses and agricultural buildings and farms are grouped in a third class called "Other land use" since they comprise less than 5% of the total catchment area.

Minor Comments There are many grammatical English errors in the document, which is understandable for multiple-language speakers. There are too many errors to list here one-by-one, but a PDF of the original document is included/attached with "sticky-note" locations with grammar issues or suggestions.

We thanks the corrections of the Reviewer, completing a new grammatical revision of the manuscript.

Many paragraphs spaces are need to help flow of text.

We revised this point as well.