Final Author Comments:

First of all, the Authors would like to thank the Referees for the time spent on reviewing this paper and for the useful comments provided to improve it.

Referee 1

This manuscript deals with flood frequency analysis in the bivariate framework by focusing on flood peak and volume. More precisely, it is oriented to hydrologic dam design. The manuscript is generally well written and presents some interest. However, on the basis of the following comments, I recommend major revision.

1.

a) The main contribution of the manuscript is not clearly highlighted in the introduction and not enough developed in the methodology section.

R.: The Authors thank the Referee for this comment. Both the introduction and the methodology have been rewritten in depth to make these contributions more clear. The main contributions of the study are: i) the introduction of a methodology based on copulas for obtaining the empirical return period linked to the risk of dam overtopping and ii) conducting a comparison between the traditional joint return periods based on the probability of occurrence of a flood event and the empirical return period based on the risk of dam overtopping.

b) In the conclusion, text on top page 575 is not connected to the methodology or the results sections.

R.: The Authors apologise for not explaining the connection clearly. The text has been rewritten as follows: "Thereby, although a previous bivariate analysis is always ~~necessary~~ recommended, there could be ~~are~~ cases in which a univariate return period analysis should ~~could~~ be enough ~~considered~~ depending on the characteristics of the dam."

c) How the methodology (especially the novel part if any) could be applied to other datasets?

R.: The Authors consider that following the steps presented in the proposed methodology, the empirical return period in terms of risk of dam overtopping based on a bivariate analysis on flood peak and volume using a copula model can be applied to other datasets. The comparison to the theoretical joint return period curves obtained through the fitted copula can also be done. Nevertheless, the methodology has been improved and extended, especially Section 2.3, which has been divided into Section 2.3 and 2.4, to improve its understandability.

2. Information in some places are not complete or unclear which prevent a full understanding some parts, such as:

a) page 561, sentence in lines 3-5 is incomplete;

R.: The Authors apologise for not being clear. The text has been rewritten: "[...] ~~According to the maximum water level reached at the dam, floods were classified in different areas from the risk associated with the primary return period~~. The maximum water level reached at the dam by each flood was estimated, to graphically analyse the relation between this level and the value of the primary return period for each event".

b) lines 8-10 should be justified;

R.: The Authors thank the Referee for this comment. This paragraph has been rewritten and extended: "In this paper, a bivariate ~~copula model~~ flood frequency analysis was carried out by a copula model ~~is used for generating for generating a set of synthetic peak-volume pairs.~~ to conduct a comparison between the return periods that estimate the natural probability of occurrence of floods and an empirical return period defined in terms of risk of dam overtopping, to assess the influence of the routing process on them. Traditional return periods based on the joint probability of occurrence were estimated from the fitted copula. In addition, the fitted copula was used to generate a large set of synthetic peak-volume flood pairs in the catchment. Synthetic hydrographs ~~are estimated using observed hydrographs to be ascribed an adequate shape.~~ were generated to ascribe a shape to the synthetic peak-volume pairs. ~~Flood~~ The set of synthetic hydrographs ~~are~~ was routed through ~~a~~ the reservoir to obtain the maximum water level reached at the dam during the routing process, in order to assess the hydrological risk of dam overtopping. ~~Both curves that represent the risk to the dam and joint return period curves that represent the probability of occurrence of floods are compared.~~ The empirical return periods based on the risk of dam overtopping were estimated.".

c) 21-24 does not reflect what is presented at the end of section 2 (there is no “procedure”);

R.: The Authors thank the Referee for this comment. This paragraph has been modified to be adapted to the improvements that have been done in Section 2: "First, the steps followed to select the copula model from observed data are described. Then, the traditional joint return periods are introduced. ~~Finally,~~ Thirdly, the procedure to generate ~~a set of~~ synthetic hydrographs ~~and the purpose of routing the hydrographs through the reservoir are explained~~ is presented. Finally, the procedure to obtain the empirical return period in terms of risk of dam overtopping is offered.". Further changes have been included in Section 2.3 and 2.4 to clarify this misunderstanding.

d) the function K(t) is not defined and what is its “generalized” version (on page 563);

R.: The function K(t) was briefly defined, consequently additional information has been added to improve its understandability: "[...] being $K(t)$ the probability that the copula function is equal or smaller than $t \in [0,1]$, i.e, the cumulative distribution of the copula value.". Moreover, The "generalized K-plot" is just the widespread name of the graphic presented by Genest and Rivest (1993), which is based on the comparison of the parametric and non-parametric estimates of K(t).

e) the justification of focusing on the upper tail is not convincing (page 565, lines 3-5);

R.: This paragraph has been modified deeply to improve its understandability as follows: "In this work more attention is paid on the upper tail dependence (Serinaldi 2008), not being relevant the analysis of the lower tail dependence due to the focus ~~of this work~~ on the frequency analysis of extreme flood events for dam safety analysis". The added reference Serinaldi (2008) highlighted the need to analyse the upper tail dependence in evaluating the occurrence of extreme events. That paper also provided other useful references in which this issue was studied. Further changes have been introduced in this Section regarding this comment.

- **f)** no interpretation is given for the quantity lambda_U (eq 5); what is the goal of this comparison (page 565, lines 13-15)?

R.: The Authors apologise for not being clear. With the aim of facilitating both, the understandability of the meaning of this quantity and the objective of the comparison, Section 2.1.4 has been rewritten.

- **g)** The first paragraph on page 566 is not clear;

R.: The Authors would refer the Referee to the part of the introduction in which the literature review was shown and the idea presented in this paragraph was introduced: "~~Moreover, bivariate flood frequency analyses require the estimation of~~ Different bivariate return periods estimated by copulas have been developed recently . Salvadori and De Michele (2004) studied the unconditional and conditional return periods of hydrological events using copulas, focussing on the joint return period in which either *x* or *y* are exceeded (primary return period) and on the joint return period in which both *x* and *y* are exceeded.[...]". Moreover, this paragraph has been rewritten as follows: "Different ~~The estimation of~~ joint return periods estimated by the fitted copula have been developed for the case of ~~is required in~~ a bivariate flood frequency analysis. [...]".

- **h)** the secondary return period is not defined and the relation in 11 is not well defined and not clear;

R.: The Authors apologise for not being clear. With the aim of improving both, the identification of the secondary return period (that was defined in the second part of the Eq. 11) and the understandability of the relation showed in Eq. (11), the first part of this equation has been removed and transformed into text: "An additional return period is also studied, the secondary return period $\rho_t^{\vee}$. ~~As can be seen in Eq.(11),~~ The secondary return period (Eq. 11) is associated with the primary return period, as ~~.I~~it can be defined as the mean interarrival time of ~~a critical event for dam when a critical design~~ an event with a primary return period larger than a threshold $\vartheta(t)$ ~~is defined (Klein et al., 2010).~~ That is to say, it is related to the probability of occurrence of an event in the area over the copula level curve of value t (Salvadori and De Michele, 2004).

- **i)** the idea on page 567 lines 4-6 seems important and needs more explanation;

R.: The Authors thank the Referee for this comment. The paragraph has been extended to improve its understandability: "The three joint return periods can be easily obtained using copulas thanks to their formulation. Once the copula selection is completed, the level curves of the fitted copula will be the curves where the events with the same probability of occurrence are located, as the copula value indicates the probability of both *x* and *y* are not exceeded."

j) page 571 line 7, zero of what? The sentence seems incomplete or unclear;

R.: The Authors apologise for the lack of clarity. This paragraph has been modified as follows: "The graphical analysis of the upper tail dependence of the observed data is carried out based on the Chi-plot, only considering the observations located in the upper right corner of the scatter plot (Fig. 7). The analysis indicates that upper tail dependence exists in the data set (what was expected for extreme value data), as the points located in the right edge tend to be far from the zero value of the *y* axis, which is the independence hypothesis ~~show values different from zero (independence)~~." (the initial "Fig. 6" has been renamed as "Fig. 7").

k) lines 14-28 need to be rewritten for more justification and clarity;

R.: These lines have been modified to improve its understandability: "Then, the non-parametric estimator of the upper tail dependence coefficient of the observed data obtained by means of [...].
In summary, the best copula should represent properly both, the dependence structure of the observed pairs of peak and volume and the ~~extreme events~~ behaviour in the upper part of the distribution. Considering the whole tests, the Gumbel copula was selected as the best copula model. It is an extreme value copula, ~~Although the best copula model based on the goodness-of-fit test is the Frank copula, the Gumbel copula is the extreme value copula with the lower value of Sn and a suitable p-value. So,~~ consequently it takes into account the upper tail dependence and ~~represents properly~~ , at the same time, shows a suitable p-value representing properly the dependence structure between both variables. Besides, as the Gumbel copula is also an Archimedean copula, it preserves the useful properties of this family.[...]"

l) q_T and v_T are not defined;

R.: The Authors apologise for not defining q_T and v_T. For this purpose a new equation has been added (Eq. 13) and the associated paragraph has been extended: "The joint return periods [...] associated to the theoretical events with peak equal to q_T and volume equal to v_T for return periods (T) equal to 10, 100 and 1,000 years are estimated for both Gumbel and Frank copula, being q_T and v_T the quantiles obtained from the Gumbel marginal distributions (Eq. 13).

$$x_T = \mu - \gamma \log[-\log(1 - \frac{1}{T})] \tag{13}$$

where μ is the location parameter and γ is the scale."

m) since the graph is not defined, the interpretation on top page 573 is not clear.

R.: The Authors apologise for this misunderstanding, the "graph" referenced the aforementioned "Fig. 10". This issue has been solved by changing "This graph" to "Fig. 10". It should be noted that the initial "Fig. 9" has been renamed as "Fig. 10".

3. The structure of the manuscript, even though generally good, needs to be improved.

a) For instance, section 3 is very short and could be included in a section like Applications;

R.: Section 3 has been extended widely.

b) the abstract does not fellow usual order (it is like a mix between the context, the aims and the results);

R.: The Authors thank the Referee for this comment. The abstract has been rewritten as follows: "~~Hydrologic frequency analyses are usually focused on flood peaks.~~ A ~~M~~multivariate analys~~i~~es on flood variables is needed to design some hydraulic structures like dams, as the complexity of the routing process in the reservoir requires a ~~have not been so exhaustively studied despite the fact that they are required to~~ representation of the full hydrograph~~, which is essential for designing some structures like dams~~. In this work, a bivariate copula model was used to obtain the bivariate joint distribution of flood peak and volume, in order to know the probability of occurrence of a given inflow hydrograph. However, the risk of dam overtopping is given by the maximum water elevation reached during the routing process, which depends on the hydrograph variables, the reservoir volume and the spillway crest length. Consequently, ~~A~~an additional empirical bivariate return period was defined in terms of ~~acceptable~~ risk of dam overtopping ~~to the dam through the~~ based on this maximum water elevation ~~reached during the routing process, in order to perform a risk assessment of dam overtopping.~~ obtained after routing the inflow hydrographs. ~~A Monte Carlo procedure was developed to compare t~~The probability of occurrence of ~~a~~ floods was compared with the return period linked to the risk of dam overtopping, as in both cases hydrographs with the same probability will draw a curve in the peak-volume space. The procedure was ~~is~~ applied to the case study of the Santillana reservoir in Spain. ~~A set of synthetic peak-volume pairs was generated by the fitted copula and synthetic hydrographs were routed through the reservoir.~~ Different reservoir volumes and spillway lengths were considered to investigate the influence of the dam and reservoir characteristics on the results. The methodology ~~Hydrographs with the same risk were represented by a curve in the peak-volume space. These curves were compared to those linked to the probability of occurrence of a flood event, in order to~~ improves the estimation of the Design Flood Hydrograph and can be applied to assess the risk of dam overtopping.".

c) the introduction is almost a literature review and does not contain important elements such as the motivation of the study, the novelty, the problematic and critic of other similar studies;

R.: The Authors thank the Referee for this comment. All these suggestions have been taken into account and the introduction has been rewritten. In summary, the motivation of the study consists of providing an empirical bivariate return period considering peak and volume in order to ascribe the return period in terms of risk of dam overtopping to the fixed return period for dam design considered by National laws and guidelines. The novelty of the study is based on the use of copulas for obtaining that empirical return period, as well as on the introduction of a methodology to compare that empirical return period to the theoretical ones obtained also by copulas.

d) the methodology section is mainly composed of wellknown techniques and neglected the development of the new part (I guess it is part 2.3).

R.: The methodology has been extended, especially Section 2.3, which has been divided into Section 2.3 and 2.4 to improve its understandability.

4. The results in Table 5 are problematic.

R: As "Table 1" has been removed, "Table 5" has been renamed as "Table 4".

a) First, the SE quantity is defined and obtained?

R.: The Authors thank the Referee for this comment. The SE is just the standard error of the parameter estimation obtained by simulation. For the sake of brevity, a reference has been added: "[...] (Kojadinovic and Yan, 2010)".

b) The estimators are very different and particularly the p-values where for each case (except Frank and Plackett) it is very close to reject with MPL and the opposite with Tau (e.g. with alpha = 10%). These results are important and require a careful checking and more justification.

R.: The Authors thank the Referee for this comment. The paragraph has been modified as follows: "The parameter of the studied copulas has been estimated using two different methods, the Inversion of Kendall's Tau and the MPL method. Consequently, not only the comparison among all the S_n values is done, but also the comparison among the S_n values provided by each method, as the estimations provided by different methods can lead to significant differences in results. It can be seen that ~~The results show that the~~ S_n leads to better results by the inversion of Kendall's tau method than by the MPL method for all copula models. Hence, in this case, the Inversion of Kendall's Tau method behaves better than the MPL method.".

5.

a) The quantile notion is not considered here even though it is closely related to the return period.

R.: The Authors thank the Referee for this comment. In the bivariate case the quantile notion is not as simple as in the univariate case. The relation between peak-volume and a given return period is defined by a curve, as several peak-volume pairs can have the same return period. So, because of the aim of this study, even though the quantile and the return period notions are very close, the Authors consider more useful to work with return periods. Further details about this comment can be found in Chebana and Ouarda (2011).

b) In addition, the curves in Figs 9 and 10 are very similar to those of the bivariate quantile in Chebana and Ouarda (2011, Environmetrics).

R.: The Authors understand that this fact can lead to misunderstanding. Firstly, it should be noted that the initial "Fig. 9" and "Fig. 10" have been renamed as "Fig. 10" and "Fig.

11", respectively. The dashed lines in Fig. 10 represent the theoretical return period curves estimated from copulas. These return periods curves are well-known curves that are shown in several studies (Chebana and Ouarda, 2011; Klein et al., 2010; Salvadori and De Michele, 2004). On the other hand, the solid lines represent the empirical return period curves obtained in this particular study, by routing the hydrographs through the reservoir. Hence, although the empirical curves can seem equal to the theoretical ones, they have been generated in a different way. In order to facilitate the understandability of this point, an illustration has been added to better explain how to obtain this empirical return period different from the widely found in the existing literature (Fig. 1). In addition, Section 4.3 has been renamed and the following changes have been done:

- "Fig. 10 shows the comparison among the empirical curves that represent the risk to the dam (T_{dam}) and the theoretical curves associated with the joint return periods $T^{\vee}_{X,Y}$, $T^{\wedge}_{X,Y}$ and $\rho^{\vee}_{t}$ estimated from the fitted copula."

- "Figure 11 displays the return period curves related to the risk of dam overtopping for different reservoir volumes given by [...]."

c) The bivariate quantile or return period are curves whereas in Tab 7 and 9 they are presented as numbers. In these tables, what is t? Why K(t)?

R.: First of all, it should be noted that as "Table 1" has been removed, "Table 7" and "Table 9" has been renamed as "Table 6" and "Table 8", respectively. In these tables specific events of the return periods curves are just presented for didactical purposes. For example, in Table 8 only two events (peak-volume) were selected. The aim is to compare the values of the different kind of return period curves associated to each event. The values of the return period curves showed in the table can also be seen graphically in Fig. 10 (a, b, c) by locating the event in the figure. Moreover, as the variables t and K(t) (that is needed to obtain the secondary return period value associated to each specific event) were introduced previously in Section 2.1.3, they have not been presented again.

6. The use of the p-value as a selection criterion is not correct (page 564, lines 22-23). The correct use of the p-value is only to tell “accept” or “reject” but not to “make an order” among those accepted. However, for this purpose, one can use criteria like AIC. This has an impact on the results on page 570.

R.: The Authors think that there could be a misunderstanding about this point. We agree with the Referee about the p-value is only to "accept" or "reject", assumption that was considered in this study. However, the goodness-of-fit test statistic (S_n) is the one that can be used to make an order. Therefore, in this point we consider that the selected copula should be the copula that presents the lower S_n value (the lower error between the observations and the values provides by the copula) and has an appropriate p-value (larger than 0.05 in order to can be "accepted"). The following lines have been added to Section 2.1.3 to make it more clear: "The S_n statistic based on the empirical copula was the goodness-of-fit test utilised in the present paper. The statistic value is used to classify the copula models as the p-value can only be utilised to accept or reject each copula model (Salvadori and De Michele, 2011). Consequently, Tthe selected copula

should have the lower value of the statistic with an admissible p-value (i.e, larger than 0.05)."
Moreover, the AIC criteria is used overall when the distributions that are compared have different number of parameters, but in this case all the copula models have only one parameter.

7. A general procedure to be applied to other data sets is missing. For instance, section 2.3 needs more clarity and developments (how related to other steps, to the literature, an illustration could be helpful, justify the choice of 100 000, explicit formulations, etc).

R.: The methodology has been extended, especially Section 2.3, which has been divided into Section 2.3 and 2.4 to improve its understandability. An illustration has also been added to clarify the procedure (Fig. 1).

8. In section 3, the choice of Gumbel for the margins needs justification (at least with which method).

R.: The Authors agree with the Referee. The following justification has been added to Section 3: " The marginal distributions for both variables were fitted to a Gumbel distribution, estimating parameters by the L-moments estimation method (Table 1). A prior study carried out in Spain showed that in this Spanish region, the Gumbel and the Generalised Extreme Value marginal distributions are appropriate for fitting the annual maximum flood peaks (Jiménez-Álvarez et al., 2012). Consequently, for the sake of simplicity, the Gumbel distribution was considered to fit the data.".

9. The generation of hydrograph seems to generate pairs of peak and volume but not hydrograph. May be I am missing something.

R.: The Authors would refer the Referee to the first part of Section 2.3. The generation of hydrographs from the peak-volume pairs obtained through the fitted copula is done by the procedure based on the use of a set of observed hydrographs proposed in Mediero et al. (2010).

10. In section 4.1,

a) why not present first the scatter plot (Q,V)?

R.: The scatter plot (Q,V) has been presented as Fig. 3a. The following text has been added to Section 4.1: "[...]. The scatter plot of the observed data is displayed in Fig. 3a."
At first, the scatter plot (Q,V) was not presented just because the Authors were focused on the scatter plot of the ranks derived from the data set (Q,V) (Fig. 3b, previously named "Fig. 2").

b) It is not clear how the conclusion is drown from Figure 3a? And the quantities in the axes of this figure are not defined. The K-plot (in fig 3b) is not clear how obtained? If it the one mentioned in section 2, then it is not aimed for this purpose.

R.: The Authors apologise for not being clear. Firstly, it should be noted that the initial "Fig. 3a" and "Fig. 3b" have been renamed as "Fig. 4a" and "Fig. 4b", respectively. Following also the suggestion of the Referee2, the meaning of these two figures has been explained in Section 2.1.1: "A graphical analysis of dependence can be displayed by the scatter plot of the pairs of ranks [...] and by other two rank-based scatter plots: the Chi-plot ~~(Fisher and Switzer, 1985, 2001),~~ and the K-plot ~~(Genest and Boies, 2003).~~ The Chi-plot displays a measure of location of an observation regarding the whole of the observations (λ_i) against a measure of the well-known Chi-square test statistic for independence (χ_i). Consequently, the larger the distance between the points and the zero value in the *y* axis, the larger is the dependence. The dependence is positive if the points are above an upper control limit and negative if they are located below the lower control limit (Fisher and Switzer, 1985 and 2001).
The K-plot relates the order statistics estimated from the observed data (H_i) to the expected value of these statistics generated under the null hypothesis of independence between U and V ($W_{i:n}$). Therefore, the larger the distance between the points and the diagonal line, the larger the dependence. Hence, if the points are located above this line the dependence is positive, while it is negative otherwise (Genest and Boies, 2003).".

11. The choice of 100 000 simulations is based on what? And why it is the same for the different situations (coincidence!)?

R.: Different authors have considered different lengths, 1,000 pairs of peak-volume were simulated in De Michele et al. (2005), 10,000 pairs were generated in Chebana and Ouarda (2011), while 1,000,000 pairs were estimated in Klein et al. (2010). No sensitivity analyses were carried out in these previous studies. Nevertheless, the following line has been added to Section 2.3 to clarify this aspect: "A set of 100,000 synthetic hydrographs was ~~were~~ generated by this procedure, to have a large and representative sample to study high return periods".

12. Some tables could be gathered or converted to text. Tables 4 and 5 require a reference source to be indicated.

R.: The Authors thank the Referee for the comment. "Table 1" has been removed and added to Section 3 as text, so "Table 4" and "Table 5" has been renamed as "Table 3" and "Table 4", respectively. Tables 3 and 4 have been generated by consulting the information provide by Joe (1997), Nelsen (1999) and Salvadori et al. (2007). As these references were initially shown in lines 22-23 of page 559 and they are well-known references, they have not been shown again.

13. Figure 8 is not clear how obtained? Fig 9a and c seem to be identical, is that correct? Why?

R.: The Authors apologise for not being clear. Firstly, it should be noted that the initial "Fig. 8", "Fig. 9a" and "Fig. 9c" have been renamed as "Fig. 9", "Fig. 10a" and "Fig. 10c" respectively. The explanation of how Fig. 9 has been obtained has been improved in the new Section 2.4. Figures 10a and 10c seem very similar but they are not identical, Fig. 10a is generated by Eq. (8) while Fig. 10c is generated by Eq. (11). They are similar because the former expression is affected by the copula value, t, while the latter is affected by the cumulative distribution of the copula value, K(t).

14. The English is generally of good quality but needs to be checked in some few places.

R.: The English has been reviewed. Anyway, in case of final publication, the English language will be copy-edited by the HESS journal before publishing, according to the general conditions established by the Article Processing Charges.

Referee 2

General comments: This might potentially be an interesting paper, but in my opinion it needs to be improved in several ways before it can be considered for publication. The current version is difficult to read and it is not always clear what the authors are trying to communicate or how they have conducted their analysis. The authors have clearly studied couplas in great details, but forget that not all interested readers will automatically possess the same degree of knowledge on the topic. In places I think the manuscript could be made more helpful to the reader by more clearly defining the terminology.

1.

a) I would have liked to see a better case for the move from single to multivariate FFA. For example, it is not the multivariate nature of flood events themselves that is a problem (several hydrological design problems can be successfully completed using only peak flow). It is the need for flood volume and the inflow design hydrograph that requires more complex methods.

R.: The Authors thank the Referee for this comment. The introduction has been modified as follows: "Univariate flood frequency analyses have been carried out widely, focusing on the study of flood peaks, which are used for designing most of hydraulic structures. However, ~~when a hydrological event is characterised by a set of correlated random variables, the~~ univariate frequency analyses do not procure a full evaluation of the probability of occurrence of the hydrological event (Chebana and Ouarda, 2011). Moreover, the full hydrograph is of interest in the case of dam design, as the inflow peak is transformed into a different outflow peak during the routing process in the reservoir. Therefore, due to the multivariate nature of flood events, a multivariate frequency analysis of random variables such as flood peak, volume and duration is required to design some structures like dams." . "[...] the flood variables control the return period depending on the dam and reservoir characteristics." (as can be seen in Fig. 11).

b) The authors could also acknowledge that design flood events and continuous simulation models have been proposed (and used) in engineering hydrology.

R.: This suggestion has been taken into account. The following lines have been added to the introduction: "This analysis has been traditionally undertaken through the use of a stochastic weather generator and continuous rainfall-runoff models (Calver and Lamb, 1995; Cameron et al., 1999; Blazkova and Beven, 2004). Although this approach has proven very successful, it is computationally very demanding, especially if extreme events are the focus of the analysis and an estimation of uncertainty is required. Copula models are a valid alternative, because they allow generating arbitrarily long series to extend the observed hydrological data with less computational effort than continuous rainfall-runoff models.".

2.

c) The paper lacks identification of a knowledge gap in need of filling followed by a clear scientific hypothesis.

R.: The Authors apologise for the lack of clarity and thank the Referee for this comment. The introduction has been rewritten to better express the motivation of the study and the knowledge gap that leads to the considered hypothesis. The conclusions have also been extended. Previous studies have developed different joint return periods that represent the natural probability of occurrence of floods in a bivariate space given by the peak and volume variables (among others, Salvadori and De Michele (2004)). On the other hand, the empirical return period based on the risk of dam overtopping was developed in Mediero et al. (2010). However, no comparison between them has been carried out yet, in order to know the influence of the dam and reservoir characteristics on them and, consequently, to know if there could be some cases in which a univariate analysis on one flood variable would be enough for dam design. Therefore, the paper provides a methodology to obtain an empirical bivariate return period considering peak and volume by a copula model, in order to ascribe the return period in terms of risk of dam overtopping to the fixed return period for dam design considered by National laws and guidelines. The comparison between empirical and theoretical return periods is also developed in this work. All the methodology is based on the theory of copulas.

d) The introduction points more towards a more routine application of a set of reasonably well-known models to a case study.

R.: The Authors apologise for not explaining clearly that this paper is not the application of well-known models to a case study. The introduction and methodology sections have been modified in depth to better explain this point. A methodology based on copulas for obtaining the empirical return period linked to the risk of dam overtopping is developed. This copula model is the base for achieving the aim of the paper, the comparison between the probability of occurrence of a flood event and the empirical return period in terms of the risk of dam overtopping. Furthermore, the influence of the dam and reservoir characteristics on these return periods was studied in the paper. The case study is used only as an illustration of how the proposed methodology can be applied.

e) Reading further it becomes clear that the authors find it not straight forward to select between candidate models based on the available dataset. They try different goodness-of-fit tests which show little difference. In particular, I like Figure 4 which shows that very different results can be obtained using different model structures that all could have been chosen. Of course, the authors give away their true intentions on page 570 line 13-14, where the spread of simulated values in the upper tail is characterised as ’undesirable’. I would say that this is not a very good argument as, by nature, you will know more about the central body of the distribution (that is where you have the most data) and relatively little about the variability of the floods in the upper tail. Thus, to define the spread as undesirable might be correct from the perspective of application, but from a more scientific viewpoint it points towards the need for new techniques to better inform and constrain the upper part of the distribution. In traditional single variable FFA this would typically be done using regional methods, i.e. attempting to bring in more information from nearby locations with flood data from a distribution reasonably similar to the site of interest. Another possibility would be to use a continuous simulation approach and generate multiple events from a coupled stochastic rainfall – rainfall runoff modelling system and see how these events behave. Of course, this would also have limitations based on the properties of the modelling system when extrapolating.

R.: The Authors thank the Referee for this comment and apologise for the use of the word "undesirable" in this paragraph. Some confusion was generated because this word should not have been introduced here as it advances information that will be obtained later, after the upper tail dependence analysis. Therefore, the following changes have been done: "Frank and Plackett copulas include this observation in the set of the generated sample~~, at the expense of an undesirable wider spread in the upper tail~~ as their dependence structure in the upper tail is more spread".
The Authors also are grateful for the comments about the other techniques that can be used to study the upper part of the distribution, but consider that these analyses are out of the scope of this paper. Of course, they will be considered in future studies.

3. Specific comments

a) Section 2.1.1: Please explain what a Chi and a K-plot is. Yes there is a reference, but since you use it in the paper I think this should be explained more clearly here.

R.: The Authors apologise for the lack of information. The explication of these graphs has been extended in Section 2.1.1 in order to the reader can understand them without looking at the references: "A graphical analysis of dependence can be displayed by the scatter plot of the pairs of ranks [...] and by other two rank-based scatter plots: the Chi-plot ~~(Fisher and Switzer, 1985, 2001),~~ and the K-plot ~~(Genest and Boies, 2003).~~
The Chi-plot displays a measure of location of an observation regarding the whole of the observations (λ_i) against a measure of the well-known Chi-square test statistic for independence (χ_i). Consequently, the larger the distance between the points and the zero value in the *y* axis, the larger is the dependence. The dependence is positive if the points are above an upper control limit and negative if they are located below the lower control limit (Fisher and Switzer, 1985 and 2001).
The K-plot relates the order statistics estimated from the observed data (H_i) to the expected value of these statistics generated under the null hypothesis of independence between U and V ($W_{i:n}$). Therefore, the larger the distance between the points and the diagonal line, the larger the dependence. Hence, if the points are located above this line the dependence is positive, while it is negative otherwise (Genest and Boies, 2003).".

b) Section 2.1.2: Consider introducing the theta parameter in connection with Eq. (1)

R.: The Authors thank the Referee for this comment. The text on Section 2.1.2 has been modified as follows: "The estimation of the copula parameter (θ) can be performed through different methods, in order to find the copula with the parameter θ $(C_\theta(u,v))$ of a copula family C that best fits the data." The Eq. (1) has not been modified because is the general way in what the Sklar's theorem is formulated (De Michele et al., 2005; Genest and Favre, 2007; Klein et al., 2010).

c) Section 2.1.2: What is a maximum pseudo-likelihood method?

R.: The Authors apologise for not defining this concept. The following paragraph has been added to Section 2.1.2: "[...]. The former is related to the method of moments, while the MPL is a modification of the traditional maximum likelihood method, in which the empirical marginal distributions are used instead of the parametric marginal distributions."

d) Section 2.1.3: What is a ‘quasi-inverse’ and how is that different from an ‘inverse’?

R.: The Authors apologise for the confusion. The "quasi-inverse" is just the way to call the "inverse" when it is "the unique inverse". The text "quasi-inverses" has been changed to "inverses" to avoid misunderstandings.

e) Section 2.1.3, line 13: ‘assess’ rather than ‘prove’

R.: The word "prove" has been changed to "assess".

f) Eq (4): I am not sure how to evaluate the expression in this equation. What is the purpose of the ‘1’ in the summation (I think this must be a typo)?

R.: The Authors apologise for not defining correctly all terms in Eq. (4). 1(A) is the indicator function of set A. This means that it takes a value of 0 when the criterion between brackets is not fulfilled and of 1 when it is. This definition has been added to the text: "[...] and 1(A) the indicator function of the set A."

g) Section 2.1.3, line 19-20: this makes absolutely no sense to me. Where do you get the estimated coupla from?

R.: The Authors apologise for the lack of clarity. These lines have been rewritten in order to improve its understandability: "Being C_n the empirical copula (a non-parametric rank based estimator of the unknown copula), C_{θ_n} the ~~estimated~~ parametric copula with the parameter previously estimated from the observed data [...]".

h) Eq. (5): Is this equation correct? It looks like the probability of a probability?

R.: The Authors thank the Referee for this comment and understand that this equation may seem unusual, but it is correct. The same equation can be found in Serinaldi (2008) and Poulin et al. (2007). Anyway, following also the suggestion of the Referee1, this section has been rewritten and this equation has been better explained to improve its understandability.

- **i)** Section 3, line 18-22: I am a little confused. Are the peaks and the volumes extracted from the same events, and selected based on annual maximum peak flow values?

R.: The comment of the Referee is correct. These lines have been rewritten to avoid confusion: "[...] Observed data are composed of pairs of maximum annual flood peak (Q) and ~~its associated~~ flood volume (V), being the latter the volume of the hydrograph associated to the event with the annual maximum flood peak."

- **j)** Section 3: Is the Gumbel distribution a reasonable choice? Given the sophistication of the couplas method, this choice appears a bit unjustified.

R.: The Authors agree with the Referee. The following justification has been added to Section 3: " The marginal distributions for both variables were fitted to a Gumbel distribution, estimating parameters by the L-moments estimation method (Table 1). A prior study carried out in Spain showed that in this Spanish region, the Gumbel and the Generalised Extreme Value marginal distributions are appropriate for fitting the annual maximum flood peaks (Jiménez-Álvarez et al., 2012). Consequently, for the sake of simplicity, the Gumbel distribution was considered to fit the data.".

- **k)** Section 4.1: The Chi and K-plots have not been sufficiently described for me to find the graphs in the Figures useful or informative.

R.: The Authors thank the Referee for this comment. As a better description of these figures has been included in Section 2.1.1., the Authors consider that the information provided by these figures has been clarified. It should be noted that the initial "Fig. 3" in which these graphs were presented has been renamed as "Fig. 4".

- **l)** Section 4.1: The 'undesirable wider spread in the upper tail' is presumably a function of the structure of the coupla when fitted to your data, but the question is not if it is undesirable but rather if it is true?

R.: The Authors apologise again for the use of the word "undesirable" in the text what has generated misunderstandings. The comment of the Referee is correct. The Clayton copula is not really a good candidate to characterised the data even overcoming the goodness-of-fit test, because its upper tail dependence is zero while the upper dependence of the data is different from zero. Following also the suggestion of the Referee1, the final part of Section 4.1 has been modified to improve its understandability.

- **m)** Section 4.1, line 25-27: I think there is something not right with this sentence? The comparison is between MPL and IKtau, not Sn?

R.: The Authors apologise for this misunderstanding. The comparison is done among S_n values. The S_n statistic is a measure of the fitting error between the studied copula and the dada. The following paragraph has been added as explanation: "The parameter of the studied copulas has been estimated using two different methods, the Inversion of Kendall's Tau and the MPL method. Consequently, not only the comparison among all the S_n values is done, but also the comparison among the S_n values provided by each method, as the estimations provided by different methods can lead to significant differences in results. It can be seen that ~~The results show that the~~ S_n leads to better results by the inversion of Kendall's tau method than by the MPL method for all copula models. Hence, in this case, the Inversion of Kendall's Tau method behaves better than the MPL method.".