

## 1 General comments of the first reviewer

The manuscript is very well written and gives sufficient context to understand the relevant developments and issues in hydrological impact analysis. The authors clearly motivate why a stochastic rainfall-evapotranspiration model is useful in this context. Their proposal is based on vine copulas, a modern statistical tool for modeling stochastic dependence between multiple variables. This is a laudable effort, but the way this methodology is applied and its performance is evaluated is problematic in several ways. I fully acknowledge that HESS is not a statistics journal and statistical subtleties may not matter in specific applications. But the extent to which they do in this particular context are unclear and needs to be addressed. Below I identify three major issues and explain why they are problematic from a statistical perspective. I urge the authors to thoroughly evaluate the implications for their hydrological model. Where possible, I try to make suggestions for alternative methods that may improve their model and its assessment. Since the first two issues may be equally relevant for other readers, my comments will be more elaborate than what is common in a closed review.

*We would like to thank the referee for his valuable comments and extensive explanation of the statistical background. We hereby list the comments of the referee and formulate our answers in italics.*

## 2 Specific comments of the first reviewer

### 2.1 Major issues

#### 2.1.1 Seasonal effects

A copula models the dependence between two random variables  $X_1$ ,  $X_2$  with marginal distributions  $F_1$  and  $F_2$ . Its parameters can be estimated from observations of these variables,  $\mathbf{X}_t = (X_{1,t}, X_{2,t})$ ,  $t = 1, \dots, T$ . The usual assumption for the validity of the estimate is that the data  $\mathbf{X}_t$  are independent and identically distributed (iid). In particular, the distribution of  $\mathbf{X}_t$  should not change with  $t$ , which is usually violated by climatic variables. The authors acknowledge that by fitting multiple models, one for each month. I am afraid this may not be enough, because climatic trends also exist within months. For example: in central Europe, the end of April is on average much warmer than the beginning of April. Suppose that additionally, the average precipitation is decreasing during April. Then high

temperatures will likely coincide with low levels of precipitation and vice versa. A copula fitted to this data will show negative dependence, which merely reflects the two deterministic within-month trends working in opposite directions, but not the stochastic dependence between the time series. Whether or not within-month trends exist can be easily checked visually or by formal statistical tests (e.g., Harris and Sollis, 2003, Chapter 3). If they do exist, they should be accounted for on a finer time scale. Splitting the data into weeks or even days could be a solution, but significantly decreases the number of observations available for fitting the copula model. A good alternative is to center and scale the time series by their seasonal mean and standard deviation. More specifically, if  $X_{j,d,y}$  denotes variable  $j$  observed at day  $d$  of year  $y$ , set

$$\tilde{X}_{j,d,y} = (X_{j,d,y} - \mu_{j,d})/\sigma_{j,d}, \quad (1)$$

where  $\mu_{j,d}$  and  $\sigma_{j,d}$  are the mean and standard deviation of  $X_{j,d,y}$ ,  $y = 1, \dots, 72$ . If necessary, trends in the skewness of  $\tilde{X}_{j,d,y}$  can be removed similarly using a Box-Cox transformation. This transformation is usually sufficient to account for deterministic seasonal effects. We can now build a copula for the stochastic dependence in  $\tilde{\mathbf{X}}_t = \tilde{\mathbf{X}}_{d+365(y1)}$ . Simulated data from this model can be transformed to the original scale by inverting (1).

*We thank the referee for this elaborative explanation. On the basis of his comments, we checked for each variable and each month for the existence of within-month trends. For each month and each variable, we therefore first visualised the data using box plots, each containing the 72 observations for each day of the month, cf. we have a 72-yearly time series. Figure 2.1.1 illustrates these box plots for temperature, precipitation and evapotranspiration for the month March.*

*This figure clearly demonstrates the existence of a within-month trend for temperature and evapotranspiration. This was also confirmed by an ANOVA test, employed when distributions were homoscedastic, a Welch ANOVA test, employed when distributions were heteroscedastic or a Kruskal Wallis test, employed when distributions were not-normal and heteroscedastic. For these tests the significance level  $\alpha$  was set to 0.001. For precipitation no trend was observed, which was also confirmed by the results of the statistical tests. Therefore, we decided to standardize the temperature and evapotranspiration data as suggested by the referee. For each month, vine copula models will then be built on the basis of standardized temperature and evapotranspiration data and non-standardized precipitation data.*

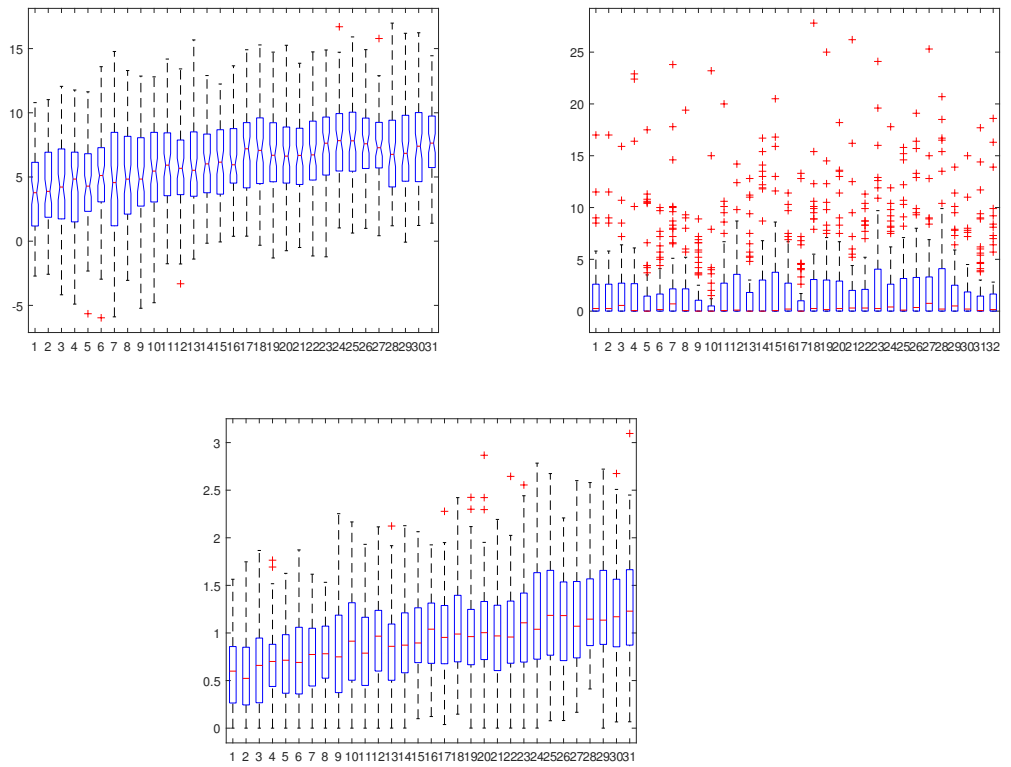


Figure 1: Boxplots per day of the month for temperature (top left), precipitation (top right) and evapotranspiration (bottom) in March

### 2.1.2 Inter-serial dependence

Even when the distribution of  $\tilde{X}_t$  is the same for each  $t$ , subsequent observations of the time series may not be independent. Such data is called stationary which is less restrictive than *iid*. Typically, stationarity is sufficient to allow for valid estimation of the marginal distributions and copula of  $\tilde{X}_t$ . But inference tools (like confidence intervals and goodness-of-fit tests) derived under the *iid* assumption are no longer valid. Another potential issue is that inter-serial dynamics can play an important role in applications. If so, these dynamics should be modeled explicitly. In the context of hydrological discharges, this is likely the case. Large discharges often occur when extreme weather conditions have been persistent for several days, and persistence is a sign of inter-serial dependence. A simple way to check whether such dependence is present is to look at the autocorrelation of the time-series, i.e., the correlation between  $\tilde{X}_t$  and  $\tilde{X}_{t1}$  (and their squares). If the correlation is small, one can test statistically whether it is zero. If there is dependence, there are two popular ways to capture it:

1. **Copula models:** This route is taken by the authors in 2.3.3, but only for the temperature variable. Similar models for the inter-serial dependence in evapotranspiration and precipitation should be employed in addition. If  $F_{j,t,t1}$  is the joint distribution of  $\tilde{X}_{j,t}$  and  $\tilde{X}_{j,t1}$ , the between-variables dependence can be modeled by a copula for the variables

$$U_{1,t} = F_{1,t|t1}(\tilde{X}_{1,t}|\tilde{X}_{1,t1}), \quad U_{2,t} = F_{2,t|t1}(\tilde{X}_{2,t}|\tilde{X}_{2,t1}), \quad (2)$$

where  $F_{j,t|t1}$  is the conditional distribution of  $\tilde{X}_{j,t}$  given  $\tilde{X}_{j,t1}$ .

2. **Classical time series models:** Classical time series models (see, e.g., Shumway et al., 2000, Chapter 3) assume that the variable  $\tilde{X}_{j,t}$  is a linear combination of the preceding values ( $t' < t$ ) and *iid* noise. For example, the autoregressive model of order  $p$  is

$$\tilde{X}_{j,t} = \sum_{k=1}^p \phi_{j,k} \tilde{X}_{j,t-k} + \epsilon_{j,t}, \quad (3)$$

where  $\phi_{j,k}$  are model parameters and  $\epsilon_{j,t}$  is *iid* noise with mean zero. The sequence  $\epsilon_{j,t}$  is commonly called innovation or residual series. The stochastic between-variable dependence can then be captured by a copula model for  $(\epsilon_{1,t}, \epsilon_{2,t})$ . More complex models are required when  $\tilde{X}_{j,t}$  is autocorrelated (Harris and Sollis, 2003, Chapter 8).

Again, we thank the referee for this elaborative explanation. We statistically checked for the existence of autocorrelation on the monthly time series by employing a Ljung-Box  $Q$ -test with a significance level  $\alpha$  of 0.001. As the total time series covers 72 years of data, this test was performed for each month and repeated 72 times. On this basis we found that autocorrelation exists for temperature and evapotranspiration. We will first take this autocorrelation into account by extending the vine  $V_{TPE}$  with  $E_{t-1}$ . In a further stage, we will couple the vines  $V_{T_pPT}$  and  $V_{TPE_pE}$ , where  $T_p$  and  $E_p$  denote temperature and evapotranspiration of the previous day, by fitting a copula between the conditioned values  $F_{T|T_pP}$  and  $F_{E|TPE_p}$ . This coupling will then be taken into account in the sampling procedure as to generate coupled time series of temperature and evapotranspiration.

### 2.1.3 Assessing the quality of the vine copula model

There are multiple issues with how the quality of the model is evaluated:

1. To check the models validity, the authors merely look at the density/cdf of the observed and simulated values of a single time series. This is only weakly related to the vine copula model and not a good indicator for its fit. Under this measure, just simulating from the distribution  $F_E$  (thereby assuming that  $E$  is independent of  $T$  and  $P$ ) would lead to results that are at least as good as the ones from the vine copula. To adequately assess the quality of the dependence model, pair-wise comparisons should be made. For example, one can look at the scatter plots of observed and simulated pairs  $(X_{1,t}, X_{2,t})$ . Another alternative are contour plots of the estimated joint density of observed vs. simulated pairs. Such comparisons should be made for all variable combinations. Similarly, multivariate return periods should be considered instead of single-variable return periods (see, Salvadori et al., 2011).

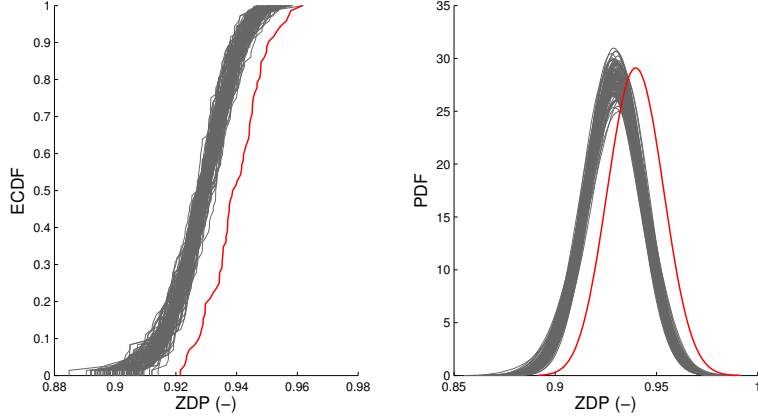
*The vine copula models are employed to generate time series of temperature or evapotranspiration that reflect the properties (statistics) of the original observed time series. To this end, different time series were generated by the vine copula models and their distributions were compared to the distribution of the observed time series. Furthermore, in order to check whether or not the dependence between the generated variables is maintained, the values of Kendall's tau were examined (see Figure 5 of the original manuscript). By examining scatter plots be-*

*tween observed and simulated pairs, as the referee suggests, one would rather investigate whether the simulated values approximate the observed values, as in a prediction model. In this study, however, it is not the aim to match the observed values for each time step  $t$  as closely as possible, but rather mimic the behaviour of the observed time series w.r.t. statistics and extremes.*

2. Figures 6 and 9 use empirical cumulative distribution functions (ECDF) instead of densities for no obvious reason. I advise against using ECDFs because they suggest a misleading sense of closeness between distributions. Since ECDFs are necessarily monotone functions with boundary values 0 and 1, their shape is quite restricted. For example, the left panels of Figure 9(d) show that the distributions are different, but the ECDFs still look somewhat similar. But the corresponding densities would show almost no overlap and more clearly communicate the dissimilarity.

*It is not clear to us what the referee means with his reference to the left panels of Figure 9(d). These panels clearly show that the distributions (the ECDFs) of the observation and the simulation differ. The left panel shows that the ZDP values of the simulations underestimate the ZDP values of the original time series, the second panel from the left shows the opposite. By showing the probability density distributions instead of the cumulative distributions, one would also see this under- or overestimation. (See Figure 2). Yet, we will change the figures and use densities instead of ECDFs.*

3. In Section 4, the uncertainty in the simulation model is assessed for various degrees of data availability. From the spread of estimated densities in Figures 11-15, the authors conclude that uncertainty increases when a variable is not observed and needs to be simulated. This is likely true, but can not be inferred from these figures. The density plots for cases 1-3 are based on a different number of observations. The spread seen on 125000 simulations will naturally be larger than the spread on 50 observations even when the actual distribution is the same. Hence, the spreads should only be compared when they are based on the same number of simulations.



We are aware of the fact that the density plots for cases 1-3 are based on a different number of observations. This is due to the fact that in case 1 one stochastic model is employed, and hence “only” 50 time series have to be generated, whereas two or three stochastic models are employed in respectively case 2 and 3. In the build-up of this manuscript, at first we “only” generated 50 time series of evapotranspiration in all cases. However, this means that in cases 2 and 3, only one time series of temperature respectively temperature and precipitation was generated. In such a setting, the stochasticity of  $V_{T_{pPT}}$  and the MBL model would be ignored. For this reason, we decided to also generate 50 time series of temperature in case 2 and for each of these time series 50 time series of evapotranspiration. In case 3, we hence generated 50 time series of precipitation and for each of these time series, we generated 50 time series of temperature. For each of the generated temperature time series, we also generated 50 time series of evapotranspiration conform case 1. We are aware that this increases the spread, which is also indicated in the paper, but we prefer not to ignore the stochasticity of the other employed models in cases 2 and 3.

In order to compare the spreads on the basis of the same number of simulations, we will carry out the following exercise. We will also generate for one time series of  $P$  and  $T$ , 50 time series of  $E$  and compare these separately w.r.t. the spread. By repeating this several times, we can hence compare the spread on the basis of the same number of observations and still take into account the stochasticity.

## 2.2 Minor issues

1. Since vine copulas are the essential ingredient in your model, I suggest to indicate this in the abstract.

*We will change the abstract to better stress this.*

2. p. 4, p. 165: If unconditional bivariate copulas are used (as is common), a vine copula is not a decomposition, but a construction. A decomposition is called non-simplified vine copula and involves conditional bivariate copulas (see, e.g., Stöber et al., 2013). I suggest to rephrase this sentence.

*We will rephrase this sentence.*

3. p. 4, l. 167: I suggest to change all types of dependence to a wide range of dependence structures. All types can only be modeled by a non-simplified vine copula.

*This will be changed in the manuscript.*

4. p. 5, l. 179: What do you mean by C-vine copulas are easier to construct than D-vine copulas? In fact, any three-dimensional vine is both a C- and D-vine, which can be easily verified by re-arranging the vertices of the vine graph

*This sentence will be removed from the manuscript.*

5. p. 6, l. 203 ff.: I am afraid a reader without prior knowledge of vine copulas will not understand your paragraph on how the model is estimated. Instead of your explanation, it should suffice to refer the reader to Aas et al. (2009).

*As we think it's important to explain to the reader how we handled the estimation of the vine, we will revise this paragraph to make it more clear to the reader and add a reference to Aas et al. (2009).*

6. How are marginal distributions modeled/estimated?

*Empirical distributions are employed as marginal distributions. This will be added in the manuscript.*

7. Figures 4, 1217 should use a larger smoothing parameter to decrease variability of the density estimates. A large proportion of the observed variability is due to the density estimation technique. This is not the kind of variability you want to assess.



*We will investigate the impact of the smoothing parameter to decrease the variability.*

8. Figure 22: What is  $i$ ?

*The values of the ECDF, indicated in the abscissa, reflect the values of  $i$ .*

### 3 Comments of the second reviewer

This is a nice study that uses copula-based approaches to stochastically generate mutually dependent rainfall and evapotranspiration forcing for rainfall runoff models. This could be used for design purposes where observation is sparse or missing. However, this approach does not seem to be working properly for the extreme events (which are needed for design purposes). Acknowledging this fact, the study is valuable for the areas with no observation. Overall, paper is well written and well structured. I have some comments (most of them major) that could potentially improve the quality of this paper.

*We thank the reviewer for his appreciation for our work and the valuable comments made. Yet, the suggestion made by the reviewer that this study would be applicable for areas without observations is not correct, as the model is fully constructed on simultaneously observed time series of precipitation, temperature and evapotranspiration.*

1. Line 46: Authors use stochastic process models to generate precipitation series. My question is:

How do stochastic process models handle changing characteristics of precipitation? Several studies have shown, for parts of the world, that rainfall events are shrinking in time and expanding in amplitude. Also there is a temporal shift in rainfall events in some parts of the world, let alone the changes in the distribution of rainfall/snow. Addressing these issues could be helpful.

*Two possible (and jointly occurring) phenomena may cause temporal changes in precipitation characteristics (but also of temperature and evapotranspiration). The first is the annual cycle, the second a long-term trend due to e.g. climate change. In our paper, we only account for the annual cycle by building models at a monthly basis. The second process (i.e. long-term changes) is not accounted for and is not the*

*aim of this study. We do not aim at constructing very long time series that reflect both phenomena. We mainly aim at building a model that allows for constructing alternative time series having the same characteristics as those of the current observations (for which it was shown that no long-term temporal trend is present). This model allows for constructing much longer time series than the observations, allowing, to a certain extent, to assess extreme events with a return period larger than the length of the original time series. This model also allows, as is shown in the paper, to generate alternative time series of the same length of the observed time series (i.e. the model predictions and the observed time series are likely outcomes of the meteorological process the observations are just one realisation, the model mimicks alternative realisations through a fully data-driven/statistical framework). When these time series are used as input to a hydrologic model, their outputs will provide alternative realisations resulting from the same meteorological process.*

*Given the fact that the precipitation-evaporation (P-ET) model that is developed in this paper is fully based on observed time series, we believe that the issue of temporal shifts in rainfall events in some parts of the world is not relevant. Also the changes in distributions of rainfall/snow is not accounted for. The model is only considering rainfall, as in the observed time series in Uccle, the fraction of snow events is very small. If discerning between snow and rainfall would be needed, the model should be extended.*

*We will improve the text to make sure that the reader is aware that no long-term temporal changes are considered, nor that the model can be applied in areas without observations.*

2. Lines 91-94: I don't understand how the number of stochastically generated forcing data could influence the uncertainty of the rainfall-runoff model's response. Uncertainty is a characteristic of the forcing data (let's neglect the modeling uncertainties for now), not the number of generated time series. So if you find a time series that fit your runoff extremes well, this is just a random phenomenon. This cannot be the basis for prediction, as we can't determine the best forcing for future, and need to rely on the ensemble of forcing data.

*The number of stochastically generated forcing data does not influence the uncertainty of the rainfall-runoff (RR) models response. Yet, the idea is that the actual observation time series is only one realisation*

*of the meteorological process, and therefore, the output, i.e. discharge, is also but one realisation. Yet, suppose, due to chaos apparent in the climatological system, an alternative observation time series would have occurred, then an alternative discharge record would have been obtained, different from the actual observed one. The latter will provide other design values (i.e. discharge values corresponding to a given return period) than the actual observed one. Yet, both are realisations of the same process, causing that the discharge values for a given return period actually follow a certain distribution, and thus results in uncertainty on e.g. extreme values, average values. This uncertainty cannot be assessed without evaluating an ensemble of possible outcomes of the meteorological process.*

*We will add some text to the manuscript to better describe the source of the uncertainty.*

3. Lines 95-96: Section 2 should precede section 3!

*This will be corrected.*

4. I am confused about how sections 2.1 and 2.2 are connected. Historical record of climate forcing are obtained for Brussels, and RR model is calibrated for the Grote Nete catchment. How do you use a model calibrated against one watershed, to predict runoff at another watershed?

*The climate forcing data obtained at Uccle are very representative for the climate observed in the Grote Nete catchment (the distance between Uccle (which is a town near Brussels, not a catchment) and the Grote Nete catchment is between 50 and 100 km). The RR model of the Grote Nete catchment was actually calibrated using data (however much shorter in length) observed near the catchment. Yet, given that the meteorological conditions are nearly the same, one can assume that the statistics of the modelled discharge obtained with the forcing data observed near the catchment and those observed at Uccle are negligible (actually, this was shown in studies performed for the Flemish Environmental Agency). Still, in this study, we compare the modelled discharge, obtained from the observations in Uccle, with forcings resulting from the P-ET model. Given the fact that the model is parameterised using the data of Uccle, the results using observed and modelled forcings can be compared. The idea of the paper is to show the impact of different alternative realisations of the P-ET model on discharge predictions, and to show how these alternative realisations*

*differ from the one realisation that correspond to true observations. The RR model used should be considered as a tool to demonstrate this, it is not intended to make predictions (although, studies for the Flemish Environmental Agency have shown that the statistics of the discharge series derived using the Uccle time series are similar to those of forcings measured near the catchment, and to those calculated on the actual discharge time series).*

Moreover, 1 year of data for evaluation is not enough. You will need a couple of years to ensure calibrated model can capture different aspects of a catchment.

*We understand the comment, though we believe it is not relevant. The PDM is in this study used as a tool to convert forcings into discharge, and the same model is used for both observed forcings (in Uccle) as modelled forcings. Yet, to make it realistic, we made use of a calibrated model, which indeed is not fully assessed for all possible hydrological circumstances. However, this model has shown to be able to well mimic the discharge behaviour of the Grote Nete catchment (as well as many other catchments in Flanders) and is used in operational water management. We believe that the model sufficiently mimicks the different hydrologic processes to show the potential of the developed generator for water management purposes. We will add some text explaining that the model used is sufficient for answering the objectives of the paper.*

5. Section 2.3: Copulas characterize dependence structure of different variables. This means there should be a dependence structure. Did you quantify the correlation between evaporation, temperature, and precip? If so, is it significant? At what temporal scale? My understanding is that you perform your analysis at daily scale, and I fear the correlation might not be significant at the daily scale.

*We do not fully grasp the fear of the reviewer, but the correlation between the variables has been checked. Yet, the dependence structures, as is apparent in the input time series, are used to build the copula models.*

6. Line 150: bivariate → It could be multivariate

*True, equation (1) can be written with more than two dimensions. However, as equation (1) only concerns 2 variables, we maintain the used terminology (i.e. bivariate copula).*

7. Line 152: I would reference to Joe 1997 too. Joe and Nelsen both played an important role in introducing copula to the scientific community.

*We will add a reference to Joe (1997).*

8. Lines 171-173: I agree that vine copulas are very flexible, but it comes at a price! A model with 4 degrees of freedom is more flexible than a competitor with 2! However, usually there is not enough information to constrain all parameters. The copula literature usually does not address the parameter uncertainties, and so they neglect the identifiability of parameters. I would address this predicament here. For more info, refer to Figure 6 of: Sadegh, M., E. Ragno, and A. AghaKouchak (2017), Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, *Water Resour. Res.*, 53, doi:10.1002/2016WR020242.

Link: <http://onlinelibrary.wiley.com/doi/10.1002/2016WR020242/full>

*We include this remark in the paper and make a reference to the paper suggested.*

9. Line 191: As a minor issue, when someone talks about a 3-dimensional model, I expect the model to have three parameters. When someone talk about trivariate model, I expect a multivariate model that associates three variables.

*In literature, the terminology  $n$ -dimensional copulas is commonly used for a copula that involves  $n$  variates. One wouldn't call the Frank copula a one-dimensional copula as it only has one parameter. We believe this is a misconception of the reviewer. We therefore prefer not to change the terminology.*

10. Line 203: How did you construct the marginal distribution? Empirical? Fitted distribution?

*We used the empirical distributions to construct the marginal distributions. We will add this information in the manuscript.*

11. Eq. 3: how did you calculate inverse of the vine copula? Analytical or numerical?

*The inversions, necessary for sampling a value out of the vine copula, were performed numerically.*

12. Line 253: pvalue larger than 0.05 or smaller?!

*The obtained p-values were larger than 0.05. More details about the theory and the p-values of the White test are given in Shepsmeier, 2015.*

13. Section 2.4: How did you calibrate the modified BartlettLewis (MBL) model, given the stochastic nature of precipitation prediction models? With stochastic models, usually summary statistics of data and simulation are compared, rather than original time series. For this purpose, approximate Bayesian computation is a great framework.

*We will briefly mention in the paper how the model was calibrated. Actually, the model was taken from a previous study (Pham et al., 2013).*

14. Lines 338-344: I cannot disagree more! Forcing and model uncertainties are intertwined, and interact in a nonlinear manner. It is not as simple as you explained. You cannot simply use a RR model calibrated for one watershed to simulate runoff at another watershed! Tens (Hundreds) of papers are available on the regionalization topic, not many of them really provided a sound ground for transferring model parameters from one watershed to another! Worse is that authors assume this modeling uncertainty does not interact with the forcing uncertainty.

*The comment of the reviewer is based on the misconception that the RR model is calibrated for one catchment and applied to another one. As stated before, Uccle is a city (not a catchment) where the headquarter of the Royal Meteorological Office of Belgium is located. At this place, the meteorological data are obtained. These data, which are statistically similar to those observed in the catchment of the Grote Nete (situated less than 100 km from Uccle), are then used in the model of the Grote Nete catchment to model its discharge. The advantage of the data at Uccle is their length (72 years). Such long time series near the Grote Nete are not available. Furthermore, the simulations are made for one single catchment (i.e. that of the Grote Nete), making use of a model calibrated for it. In this sense, the exercise done in the paper allows for partly reducing the uncertainty due to the use of PDM.*

15. Figure 22: 22 figures? Is that many figures really necessary when most of them don't provide any new info?

*We are aware of the large number of figures, however, we believe that they all are necessary to fully get the point we wish to make. Yet, by showing only the results for some months instead of for all months of*

*the year, we can reduce the number of subplots per figure while keeping the same message. By doing this exercise, it may become possible to merge figures attributed each time to one case into one figure showing results for the three cases. ”*

*We are aware of the large number of figures, however, we believe that they all are necessary to fully get the point we wish to make (unless we refer to results that are not shown, but this is quite annoying for readers trying to dig into the results). Of course, instead of showing all months of the year, we can reduce this to a couple of months.*

16. Lines 476-478: I have a hard time accepting this claim. If you generate a much longer synthetic (stochastic) forcing, then lets say predictions at a 100 years return period level improves. I accept this. But I cannot accept the general comment that longer forcing data reduces overall uncertainties. What if I had to estimate a 500 years return period flow?

*The uncertainties that are reduced are the result of working with time series of a given length. This should be better framed in the conclusion: not all uncertainties reduce, only the ones that are due the limited length of the time series reduce, while those caused by the stochastic nature of the climatological variables remain. To answer to the question on what to do when a 500-year return period of discharge would be needed, then the advice should be to model discharge with forcing time series that are a multitude of the return period one aims for. In this case, one should model series of 5000 years or more.*