

***Interactive comment on “Copula-based
assimilation of radar and gauge information to
derive bias corrected precipitation fields” by
S. Vogl et al.***

S. Vogl et al.

stefanie.vogl@imk.fzk.de

Received and published: 11 April 2012

We greatly appreciate the thoughtful comments of Referee 2 which helped to clarify some important aspects within the manuscript. All comments will be individually addressed in the sequel. Please note that the Referee's comments are bolded and our responses are in regular font format.

The paper introduces and describes a procedure to assimilate rainfall radar and gauge information using copula function. The topic is interesting and the proposed approach potentially useful, however there are several issues to clarify. In the following general and detailed comments are listed.

C776

I. General Comments:

1) The manuscript is too elaborate. Different methodologies are included and explained in details making difficult an easy comprehension. So, the first suggestion would be to exclude from the text some useless analytical developments and include it in an appendix.

The methodology section (section 3) was shortened and some details about empirical and theoretical Copulas are now given in the Appendix.

2) The manuscript introduces two methods (Multiple theta and Maximum theta). I do not think that there is much difference between them and, at the same time, keeping both the paper is heavy to read. So, the suggestion is to remove one of them or including it in the discussion section just to compare the results.

We would like to keep both methods included in the article due to the following reasons: both methods, namely the Maximum Theta and Multiple Theta are similar, but the two methods have very fundamental differences in terms of applicability. The Maximum Theta approach is only applicable in cases where one unique and one-parametric Copula family is appropriate for the used data. Even if that is a potential limitation, the Maximum Theta method could be of interest for many applications where Copula GoF-tests allow to restrict the analysis to one unique Copula. For these cases, the Maximum Theta method is preferable because it is computationally less demanding and performs better (at least in our study). However, if different Copulas with multidimensional parameter space have to be considered, the Multiple Theta approach must be used, because it allows for full flexibility of both the Marginals and Copula family. These differences are also discussed now in more detail in section 3.3.1 and section 3.3.2 in the manuscript.

3) The manuscript would like to propose a general methodology but, actually, the adopted simplifications reduce drastically its generality. I am referring to the three main hypotheses proposed by the authors:

C777

3.1) The rainfall time series are affected by autocorrelation. In this case authors suggest to apply the ARMA-GARCH approach, but since the case study does not show this behavior they skip this step in the paper. What happen when an analyst finds the autocorrelation in his time series? Probably an autocorrelated case study would have been more general.

We want to thank the referee for his helpful comment which gave the impulse for a careful check of all codes and routines that were used to test for autocorrelation and heteroskedasticity. During this check we found, that our previous conclusion was based on a misinterpretation of the p-value in the hypothesis testing. Our corrected analysis now indeed shows that the time-series are not iid. In fact, the data are affected by relatively weak autocorrelation and heteroskedasticity. The new test results have been confirmed by visual inspection of the autocorrelation functions. The Ljung-Box Q-tests have been repeated and the new results are shown in the revised manuscript. The recalculation of all results and the consideration of non-iid time series was leading to a complete revision and significant extension of the manuscript. Now the application of the suggested methods is even more general as it also includes the iid-transformation: The data have been transformed in a first step by using an ARMA-GARCH algorithm. It is also shown that autocorrelation and heteroskedasticity can be removed sucessfully by applying an ARMA(1,1)-GARCH(1,1) time series model. Based on the generated iid-residuals for both gauge and radar data, all the subsequent calculations have been repeated including the estimation of the Copula models and the simulation of the rainfall fields.

3.1) The marginal distribution was chosen equal for all time series.

The marginal distributions for both radar and gauge data are chosen according to the results of the applied GoF-tests. These tests (Kolmogorov-Smirnov, Chi-Squared, AIC, BIC) show the Weibull distribution to be the best choice in about 99.7 % of the radar grids and for all 31 gauge stations. Therefore the choice of only one marginal distribution is not an arbitrary restriction but a consequence of the analysis of the data.

C778

Both methods are flexible to the choice of the marginal distributions and therefore may also be applied without restriction in cases where more than one marginal distribution is identified. The details about the choice of the marginal distributions are given in section 4.1 in the revised manuscript.

3.2) The copula function was chosen equal for all time series.

Similarly to the marginal distributions, the Copula functions are chosen according to the results of the GoF-tests. These tests identify the Gumbel-Hougaard Copula in the vast majority (>95 %) of the possible radar/gauge pairs to provide the best fit. Therefore it is not too restrictive to consider one Copula family solely. To model the dependence structure of all radar/gauge pairs by one single and one-parametric Copula function is the prerequisite for the Maximum Theta approach. However, the Multiple Theta method may be applied without any restriction: it is flexible with respect to the choice of the marginals and the Copula models. The details about the fitting of the Copula functions are given in section 4.2 in the manuscript.

4) The ARMA-GARCH model useful to remove the time dependence is just mentioned. This is an important point. If the analyst has to filter the series to remove the correlation, he should apply the inverse transformation at the end of the analysis to have actual rainfall data. This double steps could affect the data distribution and include additional uncertainty in to the analysis. So it is important to better clarify the ARMA-GARCH role and effect in the procedure.

As mentioned before, we realized that our data is affected by autocorrelation and heteroskedasticity. The results shown in the Discussion Paper are produced without applying an ARMA-GARCH transformation first. The ARMA-GARCH transformation filters out parts of the artifical dependence due to serial correlation. Therefore, the ARMA-GARCH corrected theta parameters are slightly lower. As a consequence, the revised results reveal a decreased performance for both methods compared to the non-transformed data. Nevertheless, the iid assumption is a necessary prerequisite for

C779

the Copula analysis and cannot be neglected. Details about the performance of the ARMA-GARCH transformation are found in table 4 and the results shown section 4 are completely revised.

5) Considering the same marginals and the same copula function for the entire data set is in contradiction with using the copula approach. The added value of copula is to provide a flexible tool that allows to vary marginals and dependence structure. If the analyst does not use this potentiality probably it would not be worth to use copula function. My feeling is that using a simple linear regression tool, same results would be obtained.

The choice for the marginal distributions and the Copula models is derived from the applied GoF-tests and is not arbitrary here. The Maximum Theta method is restricted to the special case of one single theoretical Copula model, but not to one single marginal distribution. The Multiple Theta method is fully flexible in terms of Copula and marginal distribution functions. A simple linear regression can only correct for linear bias, while the Copula-based correction allows different correction factors depending on the rank of the data (conditioning variable). It delivers, in contrary to one correction factor as obtained by the linear regression, a PDF for each conditioning value (rank space). Non-linear biases as e.g. expected for precipitation data with asymmetric tail dependences, cannot be successfully corrected using linear approaches. The limited applicability of the Maximum Theta approach is now explicitly stated in section 3.3.2.

II. Specific comments:

Introduction

page 939 line 2. the reference Colins and Bolstadt, 1996 is not consisted with the same one shown in the reference list;

The reference was checked and an error in the spelling was corrected.

page 940 line 9 page 940 line 13-17 Dupuis instead of Dupois;

C780

The typo was corrected.

page 940 lines 11- 20. It is my personal opinion that the cited references are not fully representative of the copula literature, there are many other important missed. I understand that it is almost impossible to provide an exhaustive literature review on copula, so I would suggest to mention the website of the ICSH-IAHS (www.stahy.org) where a “complete” list is present and continuously updated.

The webpage of the ICSH-IAHS (www.stahy.org) is now mentioned, so that the interested reader can easily find additional literature concerning Copulas in Hydrology.

page 941 line 11. The acronym CDF is not specified.

The acronym CDF for the cumulative distribution function is now explicitly specified

Section 2

page 942 formula 1. I am not expert in radar analysis and consequently I was wondering on which data set is calibrated the formula 1- I suppose that this is not a local formula, I mean, that the rainfall data used in the proposed analysis have a limited role in the formula 1 parameter calibration.

The Z/R formula used in this study is based on the theoretical relationship $Z = aR^b$ where the two coefficients a and b have to be derived empirically. Our formula is based on calculations of DWD who estimated the two parameters using precipitation data from gauges, ombrometers and distrometers. It is generally used to calculate precipitation amounts from any radar reflectivities for the respective study area and is not especially developed by investigating the radar data that is used in this study. More details including some references are now given in section 2.1.

Section 2.2

It is not clear which is the rainfall data resolution. At page 945 line 11 it is men-

C781

tioned daily, while at page 952 line 22 it is mentioned hourly. Probably in this section it should be clarified.in this study hourly rainfall data was used.

The results presented in this manuscript are derived by using hourly precipitation time series. This fact is now clarified in the text in section 2.1.

In this section I would also include a picture with the raingauge and grid cells (it could be adapted from/or in figure 1). It could help to clarify the other sections.

For the complete study area we have 10000 radar grid cells (spatial resolution is 1 km x 1 km as detailed in section 2) and 31 gauge stations. This means that unfortunately the network of the radar grids is too dense to show it in Fig.1.

Section 3.1

Formulas 2, 3, 4, and 5 could be included in a Table.

The marginal distributions are now listed in Tab.2.

Section 3.2

Formulas 9, 10, 11, and 12 could be removed.

We removed formulas 10, 11 and 12 to shorten the theoretical introduction.

From page 946 to page 948 line 10 . All this information could organized in appendix and/or Tables cited in the text.

We followed the suggestion of the referee and rearranged the section 3.2. Supplementary material is now provided in the Appendix.

Estimation procedure is missing

It is now clarified how the Copula parameters are estimated. Details are given in section 3.2.1 and section 4.2.

Section 3.3

C782

Although I appreciate the author effort in including a picture and listing the procedure steps, the two methods are still not fully clear to me, but it could be my fault.

We revised the respective section in the text. Hopefully the differences between the two methods are now explained more clearly in section 3.3.

Section 4

page 952 line 25. I am really surprised to see that so long hourly rainfall time series (6months: around 4000 data) do not show at least significant ACF rho1. The autocorrelation is the statistical trace of the rainfall storm time aggregation, so I am surprised.

As mentioned before, it was found during the revision period that the data in fact is showing weak features of autocorrelation and heteroskedasticity. To check this we inspected the ACF of both original and squared time series. These autocorrelation functions show that both radar and gauge data are autocorrelated (see Fig. 4 for an example). This was also confirmed by the Ljung-Box Q-test, which also showed that there is weak heteroskedasticity for the gauge and significant heteroskedasticity for the radar data (see Tab. 5). This means that the data has to be transformed by an ARMA-GARCH algorithm before the Copula based methods can be applied. The details are now given in section 4.

Section 4.1

It is not clear which is the approximation that authors are doing choosing the Weibull distribution. Table 3 and Figure 4 can help the reader to understand what is happening to Garmisch station not for the whole data set.

It is not an approximation to chose the Weibull distribution for both radar and gauge data. This is the result of the GoF-tests that were applied (see exemplary results in Tab. 5). We clarified this statement in the revised manuscript (see section 4.1) and

C783

also discussed the fact that the methods would principally allow for arbitrarily many different marginal distributions.

Section 4.2

The same comment is valid for the copula function choice.

The Gumbel-Hougaard Copula was chosen due to the results of the Copula GoF-tests (see section 4.2). In our case this GoF test allowed us to restrict the analysis to only one unique one-parametric Copula, namely the Gumbel-Hougaard Copula (see Tab. 6). Of course this may differ for other case studies. Note, that only the Maximum Theta approach needs one single and one-parametric Copula as a prerequisite. The Multiple Theta approach allows for full flexibility with respect to marginals and Copula functions (different families, multidimensional Copula parameter space).

Section 5

page 959 lines 1-11. I agree to recall the aim of the paper to introduce the discussion, but probably this period could be shortened.

We have shortened the introduction of the discussion.

Section 6

page 962 line “It is found that the data is intrinsically iid and no transformation is necessary” I do not think that it is a conclusion....it is something that is never shown in the paper but just mentioned.

We have changed the conclusions due to the revised results.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 937, 2012.