Recommendations

We thank very much the reviewer for his constructive comments and critics, which we will follow in our revision of manuscript. The reviewer is concerned with physical explanations of the narrow-valley effect concluded as a hypothesis upon the statistical analysis. We agree that relating the alignment of valley with CP to moisture flow is rather an assumption than a conclusion, and would like to emphasize this point in our revision to highlight statistical results and to avoid misinterpretation. In the following, we will address each general and specific comments mentioned by the reviewer.

General comments

1. We have used the Komogrov-Smirnov test to check the null hypothesis whether the CDF of the two stations are the same, and the One-way ANOVA to test whether the mean of the two stations are the same. The reviewer is definitely correct that we can only get a weak conclusion, i.e. we cannot reject the null hypothesis, which doesn't mean the null hypothesis (the CDF or the mean of the two are the same) is true. But this is a general limitation of any hypothesis test, because statistically, there is no way to prove that null hypothesis is true.

   The reviewer also pointed out that in some cases the p-value is relatively low, which is strongly related to the length of available time series. The power of the statistical test (1-β, β is the probability of making type II error) is depending on the data length, the longer the data, the stronger the power of the statistical test. Unfortunately, there are no general rules to determine the power, because β is related to some kind of true value of the underlying data which we don't really know. Theoretically, the null hypothesis will always be rejected at any significance level, given that the data is long enough. In our case, we have more than 10 years data, which gives our tests a very strong power.

   As an example, One-way ANOVA for two sample tests is equivalent to the independent t-test. Take the pair of Zastler and Hofsgrund at CP14 as an example: the standard deviations of the two stations are 11.7 and 12.3 respectively. For CP14, there are 343 days with valid data in total. If we set α=0.05 and β=0.1, i.e. the power is 0.9, the difference we can detect is

   Sqrt$((z_\alpha+z_\beta)^2(11.6^2+12.3^2)/343)$ = sqrt$((z_\alpha+z_\beta)^2(11.6^2+12.3^2)/343)$ = 2.67 mm

   Which means we will correctly reject the null hypothesis if there is a difference of 2.67mm between the two stations with 90%.

2. We thank the reviewer for the constructive suggestions of mapping similarity to reflect the precipitation patterns. Actually we tried several means to demonstrate the precipitation patterns: (1) to plot the empirical cumulative distribution functions of several stations in one neighborhood, which shows the difference among stations without spatial information. The reader still needs to refer to the map to relate the location of the station to the curves; (2) to plot the spatial mean precipitation of difference CPs and general days. This is a straightforward message, which needs interpretation. Therefore eventually the scatter plot which contains more information is chosen.

We do agree, as we have mentioned in the manuscript, the rainfall pattern we have discovered, can only be found with a good data set. The pattern itself is general, but we need proper stations to capture it. In the manuscript, we didn't intentionally select any preferred stations, but stations with a clear valley-mountain-open area geography.

3. This manuscript is to investigate the rainfall pattern with a statistical approach. Given the patterns we have found through the data, we just hypothesized some "physical principle" behind it. Without detailed physical models, we cannot prove it is correct or incorrect, therefore we didn't claim it is correct, but just intend to postulate a hypothesis for further investigation of atmospheric scientist. So, in the revision, we will make this point more clearly.

4. The reviewer has mentioned a very important issue we have forgotten to state in the manuscript. The rainfall uncertainty due to the perturbation at small-scale by wind. As we have found out that in some cases the valley station receives more precipitation than mountain stations, which is definitely due to the rainfall uncertainty. However, as the reviewer has pointed out, it is hard to quantify such uncertainty without detailed information, therefore we will leave it.

   The seasonality issue mentioned by the reviewer is a brilliant comment, which we should consider. In our case, it is already considered through the CP classification. CP is more detailed than the seasonality.

5. The reviewer has mentioned a very important issue we have forgotten to state in the manuscript. The rainfall uncertainty due to the perturbation at small-scale by wind. As we have found out that in some cases the valley station receives more precipitation than mountain stations, which is definitely due to the rainfall uncertainty. However, as the reviewer has pointed out, it is hard to quantify such uncertainty without detailed information, therefore we will leave it.

   The seasonality issue mentioned by the reviewer is a brilliant comment, which we will check in the revision.

6. The reviewer has pointed out an important message we have neglected. We used the bias instead of absolute value of the bias. Therefore we end up with a very small value of 0.4mm and 0.1mm. However, the smaller average bias with surrogate elevation does show the bias is more symmetric. To better demonstration the improvement of surrogate elevation, we will try to quantify the normality with some test statitics.

Specific Comments

1. The range listed here is a station-wise annual (not areal average). In the extreme drought year of 2003, the station Tuebingen get only around 500mm rainfall. To avoid the misunderstanding, we will use the long-term station-wise average precipitation in the revision, which is 700mm ~ 1780 mm

2. There are several different types of rain gauges used by German Weather Service (DWD). However, they are all calibrated and corrected internally by DWD to provide a consistent measurement. We are inquiring DWD for the detailed information of the weather stations.
3. We have added one of the classical paper authored by Bardossy et al. (1995) concerning fuzzy-rule based CP classification, which contains a well-stated definition of CP.
4. The two-sample one-way ANOVA is equivalent to student t-test, which compare the means of the samples. According to the Central Limit Theorem, the mean is always approaching normal, even though the population may be highly skewed. In general, if the sample size is over 30, it is reasonable to consider the distribution of sample mean as normal. Therefore, we think the violation of normality of the population here doesn't affect the test statistic of the means that much.
5. Standard error is the standard deviation of the sample mean, which can be expressed as s/sqrt(n), where s is the sample standard deviation and n is the size of the sample. This statistic shows the variation of the sample mean.
6. For different stations and CPs, the sample size ranges from several ten days to more than thousand days. If CP type is not considered, it can be more than 10 thousand days. Therefore, the sample size does affect the p value, and we have to distinguish between small and large sample size. But the reviewer is right, we should assign a range for the sample size to make it more viable in the revision.
7. In this manuscript, we have just taken an experiment to consider the average height of points in the 2.5km upstream and 2.5 km downstream in the CP direction as the surrogate elevation. This is just an arbitrarily chosen range, which we think to be reasonable. In a following-up research, we are considering this issue more in details. Here, we just want to show there IS an improvement.
8. Yes, we will follow the reviewer's comment to provide a more quantitative comparison of the bias.
9. Yes, we will provide the sample size as well as the significance level in the table.
10. We will revise the figures missing the legends.
11. We are not sure which y axis the author is talking about. In this figure, we have already two y axis, one for the fraction of gauges (number of gauges in certain elevation band divided by total number of gauges) and the second one the mean precipitation of the gauges in one elevation band.
12. Sorry for forgetting the scale bar, we will add it.