Hydrol. Earth Syst. Sci. Discuss., 10, C4872–C4890, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C4872/2013/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Precipitation bias correction of very high resolution regional climate models" *by* D. Argüeso et al.

D. Argüeso et al.

d.argueso@unsw.edu.au

Received and published: 12 September 2013

We thank all the reviewers for their suggestions, remarks and discussion, which has definitely strengthened and improved this paper. We consider that the manuscript is now easier to follow as well as more rigorous.

All three reviewers are generally positive and highlighted that the manuscript is well written and concise. They also believed that the manuscript has potential influence in the field. Two major concerns are shared by most of the reviewers: a) The fact that the model could still produce fewer days than the station observations and thus the problem is not completely solved but reduced and b) the selection of the parameters. We paid especial attention these two issues both in our replies and the revised version

of the manuscript. We also tried to address all remarks by the reviewers.

Reviewer comments are here repeated in **boldface**. Our replies to the reviewer follow in normal typeface.

The authors present a new method for correcting model biases by using gauge data directly, instead of the common method of using gridded observational data. As point source measurements can have fewer wet days than measurements aggregated over larger areas, this reduces the number of wet days in the reference data, thus allowing the model to be drier and still fulfilling the assumption of being wetter than the reference data. Such an assumption is necessary to correct bias with most methods, without generating artificial events. The paper is well written, concise, and the case is mostly clearly presented.

Main comments:

1. The main point of the paper is to provide a method for bias correction for model simulations where the number of wet days in the model is lower than the observations. It is implicitly assumed that point source data (gauges) have less wet days than area integrated data (gridded). Although I do not doubt the validity of this assumption, it is not sufïňĄcient. The model could anyway have too few wet days, the proposed method has only lowered the threshold for the observations. So the original problem is not solved, as the authors argue, but rather reduced.

We agree with the reviewer that the model could have too few wet days and thus the method did not completely solve the problem, but reduced it. We have moderated our statements accordingly and this issue is now discussed throughout the manuscript. We have also added a figure to the manuscript which illustrates WRF seasonal biases in the number of rain days with respect to GHCN and AWAP (now Figure 5 in the manuscript). Indeed, the westernmost region (region 5) is affected by this problem. The model tends to produce less wet days than even the station observations along the west border of region 5 for most of the seasons, except winter, when the negative bias extends to a larger portion of region 5. Therefore, we take the opportunity to use that region as an example of possible inaccuracies of our method and demonstrate its limitations when the model is excessively dry. It should also be mentioned that the biases in the number of rain days with respect to AWAP are systematically negative, providing further evidence that the use of gridded datasets to correct very high-resolution model outputs is not appropriate and highlights the benefit of using station data in this situation.

2. Furthermore, using gauge data (practically a point source) will inïňĆate the variance of the pdf (Maraun, 2013). The authors should at least discuss this issue, or preferably investigate it further with their data.

We agree that using gauge data to directly correct model outputs is likely to inflate the variance of the pdf because of the spatial scale disparity between the stations and the model grid, at least partly. Although we do not question the conclusions drawn by Maraun (2013), we consider that the situation addressed in our study substantially differs from the problem raised by Maraun (2013). He examines the correction of a single 25km-by-25km grid point with respect to 20 different stations contained in the area represented by such a grid point and analyses whether the corrected model timeseries capture the local scale represented in the station timeseries. Then he aggregates the 20 observed and 20 corrected model timeseries (which originally come from the same non-corrected model grid point timeseries) to calculate the quantiles and generate a quantile-quantile plot. In our study, a large number of 2km-by-2km grid points are corrected toward a set of stations that are not necessarily within the area of the grid cell. The number of grid points makes this assessment statistically more representative, but also the spatial resolution is very likely to make a considerable difference, since the spatial scale represented by the model grid points is much closer (and thus comparable) to the stations' local scale. In order to determine whether our method inflates the variance of the model pdf under the conditions of our study, we have conducted a

quantile-quantile analysis, similar to that carried out by Maraun (2013). For each of the regions, the quantiles were calculated using all grid points (stations) within each region for the corrected and non-corrected model outputs (station observations) and displayed in a quantile-quantile plot (Figure 1 in this comment). Figure 1 (in this comment) shows that the method accurately corrects quantiles in all regions and the corrected outputs match the observed variance much better. The original model variance was actually reduced in regions 1, 2 and 3; slightly increased in region 4, and a larger increase was applied to region 5. However, only the very end of the distribution tail in region 5 (and the last percentile in region 4) is above the observations. Therefore, we consider that the inflation of the pdf variance is not an issue in our study. We have added a sentence in the manuscript to mention the results of this analysis and included Figure 1 (in this comment) as supplementary material.

3. The "drizzle effect" is a model phenomenon where the model to readily produces precipitation. It is not correct to refer to the increase in low intensity precipitation with spatial (or temporal) averaging as a "drizzle effect".

Two occurrences of the term "drizzle effect" were removed. The sentences now read: "Models often display an increase in low intensity precipitation as a function of spatial resolution due to various reasons (i.e., model physics, spatial representativeness), producing more frequent but less intense precipitation than the station measurement as the resolution decreases. Gridded observations are also affected by similar behaviour, mostly due to the gridding methodology that involves some sort of spatial averaging" "A histogram equalisation method (Piani et al. 2010) was adapted to be used with stations, which are not subjected to more frequent low intensity precipitation due to spatial averaging. . ." It is still used once but only referring to Regional Climate Models and it is mentioned along with the spatial averaging effect: "...the model produces more rain days than the observations, since it could be affected by very strong biases that are not compensated by light precipitation events due to spatial averaging or the 'drizzle' effect."

4. A new ïňĄgure should be provided that shows the bias in number of wet days for the WRF simulation at 2km, and for each season. This should be positive at all locations, otherwise the method was not successful in fulïňĄlling the assumption of a "wetter model". Possible failures of the proposed method should be discussed.

A figure showing the biases in the number of wet days for the WRF simulations at 2km is now provided (now Figure 5 in the manuscript). The biases were calculated using the exact same method to weight the correction factor in the bias-correction methodology, that is, for each grid point the numbers of rain days from the nearest 5 stations are averaged using inverse distance weighting and a penalisation in the weighting of 0.5 for the stations outside the grid-point region.

As mentioned above (see first comment of this reviewer), the model shows negative biases in the number of rain days along the westernmost border of region 5 for most of the seasons, except winter, when the negative bias extends to a larger portion of region 5. Although the use of stations makes it more likely that the assumption of a "wetter model" be valid, it does not guarantee that it will be fulfilled in all circumstances. We have acknowledged this limitation in the manuscript and have used results in region 5 as an example of this situation.

We have also added the biases with respect to AWAP, so that the benefit of using the station data is displayed too.

5**. It seems the authors used the same data for both calibration and validation of the method. This is generally not good practice, but given that the topic of this paper isto present a method, not a corrected data set that will be used for further analysis, I think it is ïňĄne if mentioned. Furthermore, given that the corrections were made for the calibration period, the results of the bias correction are rather poor. I assume the remaining bias is partly due to the averaging of corrections over ïňĄve stations, but are there also other effects? Is the gamma distribution**

not a good approximation? Please discuss also such aspects.

We agree with the reviewer that for future use of the method, it must be applied to two separate periods for calibration and validation purposes. As he rightly noted, this paper calls the attention to an issue that is likely to become frequent in the upcoming regional climate simulations and suggests a method to reduce its impacts. It has been mentioned in the text that the application of the method to produce corrected dataset requires its calibration and validation using different periods.

Despite the fact that we used a single period in this study, a perfect match in the comparison we performed cannot be expected. There are various reasons that might explain the differences between the corrected model outputs and the observations, which are actually very small in most of the domain, except region 5 (Figure 6 e-h and m-p, and Figure 7 in the manuscript; Figure 1 in this comment). As pointed in comments 1 and 4, region 5 is particularly dry in the model and does not fulfil the assumption of a wetter model, which results in inaccurate correction of the model rainfall. This has been discussed in the manuscript, as stated in comments 1 and 4. Furthermore, only Figure 5 in the manuscript compares each station with the nearest grid point. In that situation, the averaging over 5 stations definitely has a contribution to the small differences. For that reason, not only the comparison with stations was provided, but also with AWAP to provide a further reference for the methods capabilities. The remaining comparisons were made on a region-by region basis because, even after correction, the model is not always able to fully reflect the spatial scale of the stations and its local characteristics, as also noted by Maraun (2013).

The fitting of precipitation to a particular theoretical distribution function could also explain the differences and other functions could be chosen instead, but in view of the results, we consider the choice of the gamma distribution be appropriate. This choice is also supported by a number of previous studies that used a gamma fitting, with satisfactory results. See references in the manuscript that evaluated the method proposed by Piani et al. (2010), but also Haerter et al. (2011) -and references therein-, Wilks (2006).

6. The skill score used to evaluate the PDFs is mostly determined by the high probability values, i.e. the high skill scores for the corrected data are basically determined by the correction of the ïňĄrst few bins. Although a claimed limit for wet day is >0mm/day, the actual limit is higher (probably closer to 0.1) due to the gauge precision. A model is not affected by such limits and will therefore deviate in the ïňĄrst bin of the pdf (depending on the bins used for the pdf, I assume they start at >0 as nothing else is stated). This should be mentioned when discussing the results.

The limit for a wet day is set 0.0 mm/day as discussed in detail below (minor comment P8152, L5). However this is equivalent to 0.1 mm/day in the case of observations according to the instruments' sensitivity. The model outputs are not affected by this sensitivity limit the wet-day threshold is effectively 0.0 mm/day. The PDF of wet days (all dry days removed) is described using 1mm bins starting at 0 mm (e.g., [>0 mm,<= 1 mm]; [>1 mm, <=2mm];. . .).

A sentence mentioning the contribution of rainfall events in the range 0-0.1 mm/day has been added.

Minor comments: P8146, L8-9: Please remove "than the gridded observational products" for increase readability

Corrected.

P8146, L11: Please change to "model outputs" or "resolution is compared"

Corrected

P8146, L15: Please remove "selected and". Also, please add a sentence here describing why the use of gauge data alleviates the problem, and indicate the limitations.

"selected and" has been removed.

The paragraph has been rewritten to make clear that the use of gauge data makes it more likely the assumption of wetter model to be valid and to indicate the limitations of the method. It now reads: "...The gridded observations are found to be inadequate to correct the high-resolution model at daily timescales, because they are subjected to too frequent low intensity precipitation due to spatial averaging. A histogram equalisation bias correction method was adapted to the use of station, alleviating the problems associated with relative low-resolution observational grids. The wet-day frequency condition might not be satisfied for extremely dry biases, but the proposed approach substantially increases the applicability of bias correction to high-resolution models. . ."

P8146, L22: "RCMs' "

Corrected

P8148, L1: Please include that the gridded data are more compatible with model data due to the representativity of an area as opposed to a point.

In our opinion, the comparability between gridded data, station observations and model outputs highly depends on the grids resolutions, and this is precisely one of the aspects we attempted to emphasize in our study. However, we have included in the manuscript a sentence to state that the spatial compatibility between gridded observations and model outputs of similar horizontal resolution has also been a reason to use gridded datasets for model bias correction.

P8148, L8-9: "will fail to adequately". That is not obvious. If you correct empirically with histogram equalization, you will expect to have a perfect bias correction if enough wet days are present in the model. If that condition is not fulïňĄlled, the correction will not be "perfect", but might still be "adequate" depending on some possibly subjective requirements. If a distribution-based algorithm is used, the lack of wet days in the model might give a smaller error

than the assumption for the applicability of the ïňĄtted distribution. So, please change to "...migh fail to ..." or similar.

Corrected

P8148, L11-12: Same as above, "might be unrealistically. . ."

Corrected

P8148, L15-16: Please clarify what is meant with the "drizzle effect" for gridded observations. It is different from the drizzle effect in models, see major comment.

The term "drizzle effect" has been removed when referred to observations/ Please see major comment 3.

P8148, L20: I propose, "Such RCMs are likely. . ."

Corrected

P8149, L12: Does the 2km domain have its western boundary in the mountain range? Might this be the reason for the bias over the mountains, and perhaps as a consequence east thereof?

The western border of the boundary is located over an area of about 600m altitude. Although it is generally desirable to locate the borders away from complex terrain, it is often not possible to completely avoid mountain ranges and a compromise must be made. We have located the border in this experiment to the west of the ridge, over a relatively homogeneous area, where the topography is likely to be described similarly in both the 10km and the 2km and thus the impact due to inconsistencies should be small.

Nonetheless, there is still the possibility that a different location of the western border will yield better values. We are very much interested in the biases in the western region and over the mountains, and we are currently investigating possible causes, which include not only the domain design, but also other parameters such as the vertical

resolution or the use of cumulus scheme.

P8149, L24: Please provide the lower measurement limit of the gauges.

The lower limit of the rain gauges is 0.1 mm. This has been added to the text.

P8150, L5-6: First, please change "number of days..." to "number of wet days...". Second, there are many reasons for the number of wet days to decrease with increasing resolution of the model, e.g. model physics and paramterizations and spatial averaging. The two are easily distinguished by averaging the higher resolution data to the coarser grid. Only then are the different data compared on equal terms. Just a comment.

Changed "number of days" to "number of wet days".

We have introduced a sentence (please see major comment 3) to acknowledge that various reasons might be responsible for the decrease in number of wet days with increasing resolution of the model.

P8151, L12: Please provide some comments on why 5 stations were used, and a sensitivity analysis thereof.

As also suggested by Reviewers 1 and 3, we have conducted a sensitivity analysis to determine the effect of choosing different number of stations and different penalisation. Here we quote the reply to reviewer 1:

We carried out a sensitivity analysis to determine the impact of the choice of both the penalization parameter and the number of nearest stations. Ten possible combinations of these two parameters were selected and their performance was investigated. In particular, three values of the penalisation parameter (0.1, 0.5 and 0.9) and three values of the number of stations (3, 5 and 7) were examined, which yield 9 possible combinations. An additional configuration using a single nearest station is also included (penalisation does not apply in this case).

The bias correction was applied 50 times for each of the parameter configurations removing 10% of the stations each time as a means of perturbing the corrected solutions and thus studying the method sensitivity to these choices. Figure 2 (in this comment) illustrates the results of this sensitivity analysis through two different statistical measurements. Corrected WRF outputs were compared with GHCN (each station with the nearest grid point) on seasonal timescales and the mean absolute error (MAE) is calculated averaging over all stations and all seasons to generate a boxplot including the 50 realisations of the bias correction for each of the parameter configurations. The pattern correlation is also calculated to measure the spatial pattern similarity between corrected WRF outputs and AWAP, and also represented using a boxplot.

The seasonal MAE shows that the choice of the parameter has a minor impact on the results and all possible configurations give a similar value (8 mm/month), including the use of a single station, for which penalisation does not apply. However, the analysis of the pattern correlation reveals that in this particular case, the spatial structure is better represented using more than a single station, showing a very slight decay with larger penalisation values. The stability of the method with respect to the parameters is probably enhanced by the network density and the inverse distance weighting, which tends to give much larger importance to closer stations. In this particular case, we have chosen a moderate number of stations and penalisation in order to provide a test of the method. The sensitivity test suggest that different choices should not significantly affect the results, although this is likely dependant on the characteristics of the region and the network and thus different values might be more suitable under different conditions.

We have included a sentence on the manuscript to mention that the method was found to be slightly sensitive to the choice of the parameters, except that the use of several stations is generally preferred to improve the spatial structure of precipitation climatology. Figure 2 in this comment was also added as supplementary material.

P8151, L15-16: What about corrections for orography etc in gridded observations. This is completely discarded in the proposed method. Perhaps note in the

paper that this would be a possible extension of the method.

Orography correction is a usual approach in temperature bias correction methods, although it sometimes introduces some uncertainty (i.e. which lapse rate is used). In the case of precipitation, the necessary degrees of freedom that must be included in an orography correction are much more (e.g., elevation, slope, orientation, wind direction, location of water masses…) and thus the associated uncertainty would be much larger.

P8151, L28-29: I believe only stations outside the precipitation region are penalized, those inside do not get additional larger weight. Please clarify sentence accordingly.

We have clarified the weight applied to each of the stations depending on whether they are in the same region as the grid-point. In our sentence we intend to say that the stations belonging to the same region as the grid-point have a larger weight than the stations that do not. The sentence now reads: "Using the regionalisation, we are able to give larger weight to stations that belong to the same region as the model grid-point and penalise than those that are likely to have different precipitation regimes. A penalisation factor (Ps) of 0.5 is applied to the stations weights when they are located in a region different to that of the grid point. Otherwise the factor is 1."

P8152, L5: Please verify the actual lower limit as determined by the measurement threshold of the gauges. Also, please provide a sensitivity analysis due to the choice of this low wet day limit, given that others have often used a higher limit.

As mentioned above, the lower measurement limit is 0.1 mm in all rain gauges.

We defined the days with any precipitation (>0.0 mm) as wet days for both model and observational datasets. However, this condition is equivalent to 0.1 mm in the observations due to the sensitivity of the instruments. This has been stated in the text. We chose the lowest possible threshold in order to make the assumption of fewer rain days in the observations more likely. Bearing in mind Figure 6, this requirement is less likely to be met with higher threhsolds. In addition, the 0.0 mm threshold has been previously used in the context of bias correction (Teutschbein and Seibert, 2012). Finally, Berg et al. (2012) found that a similar histogram equalisation method was not sensitive to the choice of the threshold in the range 0-1 mm.

P8152, L9: This sentence makes it clear the an excess of wet days in the model is assumed. But no analysis is provided to validate this assumption. This is serious, since the proposed method is dealing with that very assumption. I propose that a new ïňĄgure with the bias in the number of wet days for the WRF-2km simulation is provided, with data for each season. This should be positive at all locations. If not, the this should be discussed.

The sentence was modified to clarify that the excess of wet days is not assumed. We agree that the occurrence of wet days in the observations and the assumption of a wetter model in terms of rain-day frequency must be discussed. We have included a figure as suggested by the reviewer and have discussed it in the text. Please see major comment 4 for further details.

P8152, L14: "each of the seasons"

Corrected

P8152, L16: "outputs"

Corrected

P8153, L24 and below: The expression "events" is used here, but I think the more precise term would be wet days.

We have followed the reviewer's suggestion in some of the occurrences of the term "events" and have used "wet days" instead of "events". (e.g., in the introdcution: "...do not strictly require an equal or larger number of wet days. . .")

However, in this particular sentence we meant the contribution from rainfall events of different intensity, as mentioned in the paragraph above. Therefore, we consider that replacing "events" with "wet days" will not make the sentence clearer. We have decided to use the term "rainfall events of different intensity" to enhance the clarity". In other occurrences of the term "events", we have generally opted either for "precipitation intensity", "precipitation events", "rainfall events" or "rain events" to avoid any possible misunderstanding.

P8154, L13: "method's skill"

Corrected

P8154, L15: Which reference data is used for the skill score calculations? Also, see major comment on the interpretation of the skill scores.

The original sentence says that the "in-situ observations" were used to calculate the skill scores. "...the non-corrected model outputs and the in-situ observations share. . ." For further comments on the skill score interpretation, please see major comment 6.

P8155, L7: Again, the drizzle effect is a model phenomenon.

Corrected. The term "drizzle effect" was removed when referring to observations. It now reads: "...which are not subjected to more frequent low intensity precipitation due to spatial averaging and. . ."

P8155, L7: "and thus might make this assumption valid again". It might not be sufïňĄcient depending on the model bias.

Corrected. It now says: "...and thus are more likely to make this assumption valid again"

P8155, 20: Please add discussion on limitations of the method (when it can fail etc) and possible extensions of the method.

We have acknowledge that limitations in the applicability of the method still exist. We have also pointed to possible extensions of the method.

Fig.1c, Fig.4legend, and Fig.6,legend: It is difiňĄcult to see the black dots and text with the dark colors. Please change color scheme for increase readability

Corrected. A brighter color scheme was used.

Fig.5: Please use the same color scheme for points and grids, i.e. remove the legend to the right and use only the legend at the bottom. Furthermore, it is difiňĄcult to separate the colors, so please change to fewer steps and increase contrast between them. It is also useful to have e.g. gray color for the steps just above and below zero.

Corrected. The same color scheme was used for both points and grids, and the legend to the right was removed. The number of steps in the colorbar has been reduced to half and the contrast between them was increased. White has been replaced by gray in the steps above and below zero.

Fig.6: It is difiňĄcult to separate the colors for WRFBC and GHCN, please change one of them to a lighter shade or different color. Also, a lin-log or log-log scale might be more revealing.

Corrected. A brighter color has been used for WRFBC so it is easy to distinguish from GHCN black.

The use of a log scale in the representation of the probability distribution function was also considered (or the contribution from different intensities, please see minor comment 2 from reviewer 1 for further discussion on the issue). The log scale tends to emphasize the small contribution from extreme events and masks the large contribution from light-to-moderate events, as shown in Figure 3 (in this comment), which is the same set of plots as in Figure 7 in the manuscript but using log scale. We consider that this study is strongly focused on the events below 30 mm/day or so, and thus the linear scale represents much better the magnitude of the deviations we tried to deal with in

our study.

Fig.6,caption: “events”, see comment above. check spelling of precipitation.

Corrected according to response to comment above. The term "events" has been replace with "rainfall events". Spelling has been corrected.

References: Maraun, D. (2013) Bias correction, quantile mapping, and downscaling: Revisiting the inïňĆation issue, J. Clim., 26, doi:10.1175/JCLI-D-12-00821.1.

REFERENCES Berg P, Feldmann H and Panitz HJ (2012) Bias correction of high resolution RCM data. Journal of Hydrology 448-449: 80–92: doi:10.1016/j.jhydrol.2012.04.026. Haerter JO, Hagemann S, Moseley C and Piani C (2011) Climate model bias correction and the role of timescales. Hydrology and Earth System Sciences 15(3): 1065–1079: doi:10.5194/hess-15-1065-2011. Teutschbein C and Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. Journal of Hydrology 456-457: 12–29: doi:10.1016/j.jhydrol.2012.05.052. Wilks DS (2006) Statistical Methods in the Atmospheric Sciences. Academic Press, Elsevier.

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/10/C4872/2013/hessd-10-C4872-2013-supplement.pdf

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 8145, 2013.

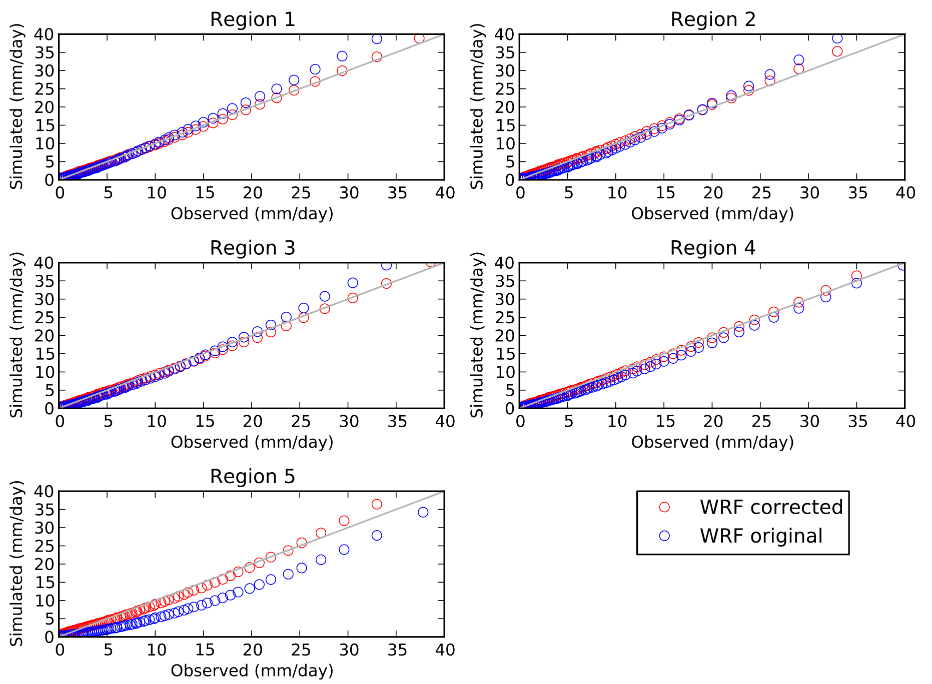

Fig. 1. Quantile-quantile plots for each of the regions from the original (blue) and the corrected WRF outputs, compared to GHCN stations

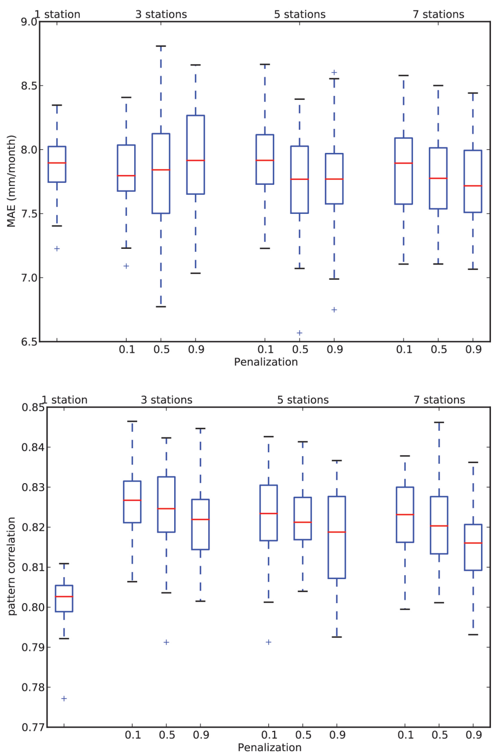

Fig. 2. Top: Seasonal MAE between the 50 perturbed realisations of each bias correction con- ïňĄguration and GHCN. Bottom: as top but for pattern correlation and using AWAP as reference

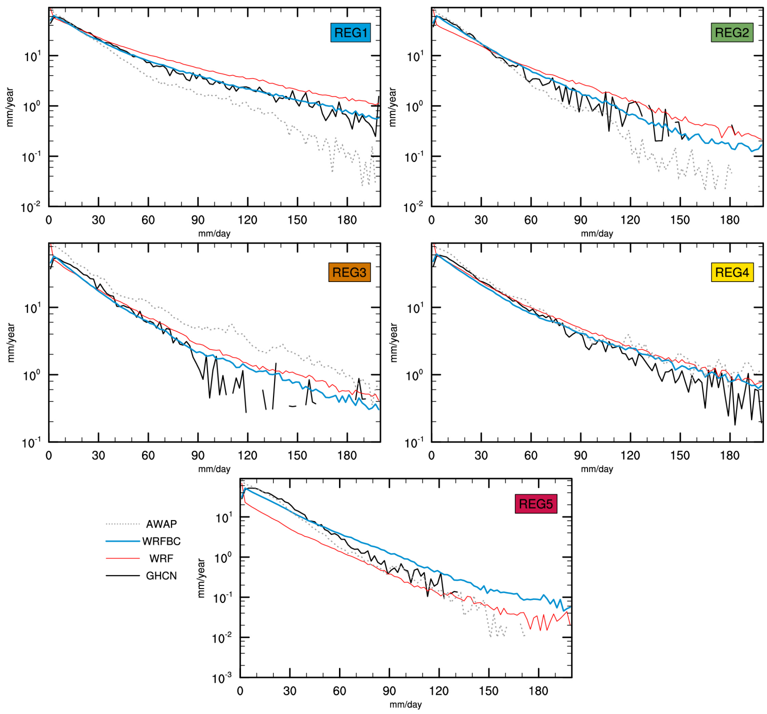

Fig. 3. Contribution to annual precipitation from different-intensity rainfall events in the 5 regions for AWAP, WRF (corrected and non-corrected) and GHCN, using logarithmic scale