

Supplement of Nat. Hazards Earth Syst. Sci., 17, 1623–1629, 2017
<https://doi.org/10.5194/nhess-17-1623-2017-supplement>
© Author(s) 2017. This work is distributed under
the Creative Commons Attribution 3.0 License.



Supplement of

Effects of sample size on estimation of rainfall extremes at high temperatures

Berry Boessenkool et al.

Correspondence to: Berry Boessenkool (boessenk@uni-potsdam.de)

The copyright of individual parts of the supplement might differ from the CC BY 3.0 License.

Abstract. For the analysis in our corresponding article, we have fitted the General Pareto Distribution (GPD) to the top 10 % of samples via L-moments to obtain sample size independent quantile estimates. This supplement explores other fitting methods, namely the maximum likelihood estimation (MLE) and probability weighted moments (PWM) as implemented in different R packages. We also examine different truncation proportions used for the threshold in the Peak Over Threshold (POT) approach.

We find that different implementations of the same method can yield different results. The estimates from MLE methods are closer to each other than from moment based methods. In very small samples ($n < 100$), however, they also differ significantly in their estimates. As a disadvantage, MLE methods have a stronger sample size bias than moment methods while the latter have a stronger dependency on truncation percentage. Overall, the best quantile estimator for small samples seems to be the L-moments approach from the `lmomco` R package.

1 Introduction

We use hourly observations from 142 weather stations located across Germany, see the map in Fig S1. The code creating the figures in this supplement is available at https://github.com/brry/prectemp/blob/master/Code_supplement.R.

2 Sample size dependency

As shown in the main article, high quantiles are underestimated in small samples if empirical order based statistics are used. Parametric estimators (from a fitted distribution function) show a smaller bias. The convergence with the estimate in large samples also appears already at much smaller sample sizes.

Although using the GPD generally aids in reducing this sample size bias, the effectiveness depends on the method used for fitting. The median results from the 1000 simulations of the main article are presented in Fig. S2 for several GPD methods. MLE estimates generally show a much larger bias (stronger underestimation) at low sample sizes than moment based estimators. This could explain why some of the cited articles using GPD based (instead of empirical) quantiles still show a rainfall intensity drop at very high temperatures.

Implementations of the same method in different R packages (see table S1) can behave quite differently. The differences may stem from different optimization methods for parameter fitting, bugs or implementation errors. A thorough review of all the packages including benchmark testing is unfortunately beyond the scope of this article.

3 Threshold dependency

In a Peak Over Threshold (POT) approach, the distributions are fit closely to high values. This is useful, as the observed extremes are especially important for high quantile estimation. The truncation threshold is usually determined as a certain quantile of a sample. Higher values lead to better fits but also increase sample size demand. For the main simulations we chose a truncation proportion of 90 % as trade off, thus using the top 10 % of a sample. Berg et al. (2009) used the top 20 % of precipitation intensities to fit a GPD while Lenderink and Meijgaard (2008) used the top 5 % and Lenderink et al. (2011) the top 4 %, see table 1 in the main article. Haerter et al. (2010) found that GPD quantiles do not depend on truncation percentage. We observe the same result for MLE methods, but for moment methods, the estimates rise with truncation percentage, see Fig. S3. With higher thresholds, the different estimates for `lmomco` appear to converge, but necessary sample size increases.

4 Application to rainfall temperature relationship

The different methods described above are applied to examine the Clausius Clapeyron relationship as explained in the main article. They show different estimate locations but a similar general pattern in Fig. S4. Consistent with Fig. S2, the MLE methods for the high temperature bins (with few observations) show slightly smaller values than the moment methods. All methods consistently differ markedly from the behavior of the empirical quantile estimate.

For the POT method to be valid in theory, there may be no dependency among the occurrences of extremes, which can for example be achieved by using event maxima. As event separation is subjective, see for example Ribatet et al. (2009), we compare only 3 different event definitions. Distinct events are defined after intermissions of 1, 5 and 12 hours with no rain (≤ 0.5 mm). Fig. S5 shows quantile estimates for the maximum rainfall per event (with the temporally corresponding temperature value). By looking only at event maxima, the estimates are naturally higher than the estimates for all observations, but the qualitative behaviour remains the same for all these definitions. For the main analysis we hence decided to use all the observations to avoid an arbitrary choice of event definition.

The difference between empirical and parametric quantile estimates gets smaller for smaller quantiles. Below 99 %, the difference is negligible, see Fig. S6. In earlier versions of the manuscript draft, different distribution functions were aggregated to a weighted average. That showed a difference also in smaller quantiles (albeit less pronounced), which can be found on page 2 and 6 of <https://github.com/brry/prectemp/raw/519fc0a0757e5f7386b86d4906d0d0511bf63517/fig/PTQ.pdf>. This may be related to the fact that the GPD (as compared to other distributions) specifically fits well to extreme distribution tails.

To show the relation of precipitation quantiles with sample size, the number of observations per temperature bin are presented in Fig. S7.

15 5 Conclusions

We compared several methods to fit the GPD (and other distributions) to a sample to obtain robust quantile estimates. Fitting through maximum likelihood estimation (MLE) shows a stronger underestimation in small samples than L-moments or probability weighted moments. The moment methods, on the other hand, show a stronger dependency on threshold choice. The quantile estimates are hence sensitive to the distribution fitting procedure, especially the parameter estimation. Different software implementations also yield different quantile estimates, especially with moment based fitting. An ideal estimator quickly converges with increasing sample size and also shows a stable behavior over different truncation thresholds. There is no single combination that is superior over both criteria, but the L-moments based implementation in the `lmomco` R package seems to be the best option. They have almost no sample size bias and are stable with regard to the threshold if it is larger than 0.8 (in this particular case).

References

- Berg, P., Haerter, J. O., Thejll, P., Piani, C., Hagemann, S., and Christensen, J. H.: Seasonal characteristics of the relationship between daily precipitation intensity and surface temperature, *Journal of Geophysical Research: Atmospheres*, 114, D18 102, doi:10.1029/2009JD012008, 2009.
- 5 Haerter, J. O., Berg, P., and Hagemann, S.: Heavy rain intensity distributions on varying time scales and at different temperatures, *Journal of Geophysical Research: Atmospheres*, 115, D17 102, doi:10.1029/2009JD013384, 2010.
- Lenderink, G. and Meijgaard, E. v.: Increase in hourly precipitation extremes beyond expectations from temperature, *Nature Geoscience, Letter*, 1, 511–514, doi:10.1038/ngeo262, 2008.
- Lenderink, G., Mok, H. Y., Lee, T. C., and van Oldenborgh, G. J.: Scaling and trends of hourly precipitation extremes in two different climate zones - Hong Kong and the Netherlands, *Hydrology and Earth System Sciences*, 15, 3033–3041, doi:10.5194/hess-15-3033-2011, 2011.
- 10 Ribatet, M., Ouarda, T. B. M. J., Sauquet, E., and Gresillon, J.-M.: Modeling all exceedances above a threshold using an extremal dependence structure: Inferences on several flood characteristics, *Water Resources Research*, 45, doi:10.1029/2007WR006322, w03407, 2009.

Table S1. R packages implementing GPD fitting. The Bayesian method is not used because of the high computational demand.

R package	Documentation URL	Methods, notes
evd	https://www.rdocumentation.org/packages/evd/topics/fpot?	MLE
evir	https://www.rdocumentation.org/packages/evir/topics/gpd?	MLE, PWM
extRemes	https://www.rdocumentation.org/packages/extRemes/topics/fevd?	MLE, LMO, GMLE, Bayesian
fExtremes	https://www.rdocumentation.org/packages/fExtremes/topics/gpdFit?	MLE, PWM
ismev	https://www.rdocumentation.org/packages/ismev/topics/gpd.fit?	MLE
lmomco	https://www.rdocumentation.org/packages/lmomco/topics/pargpa?	LMO
Renext	https://www.rdocumentation.org/packages/Renext/topics/Renouv?	Uses evd::fpot
extremeStat	https://www.rdocumentation.org/packages/extremeStat/topics/q_gpd?	Compares all above

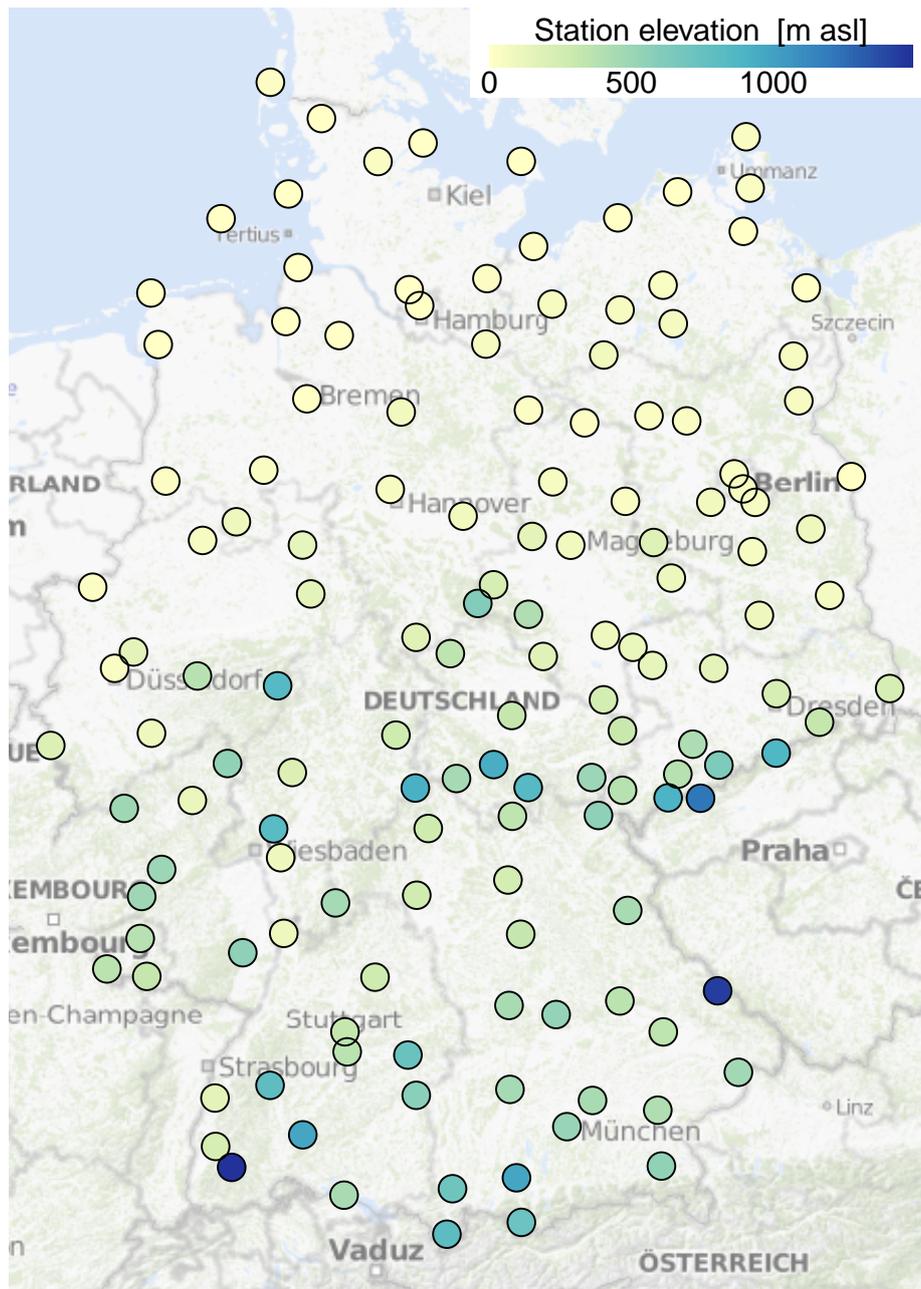


Figure S1. Map of DWD stations used in main analysis

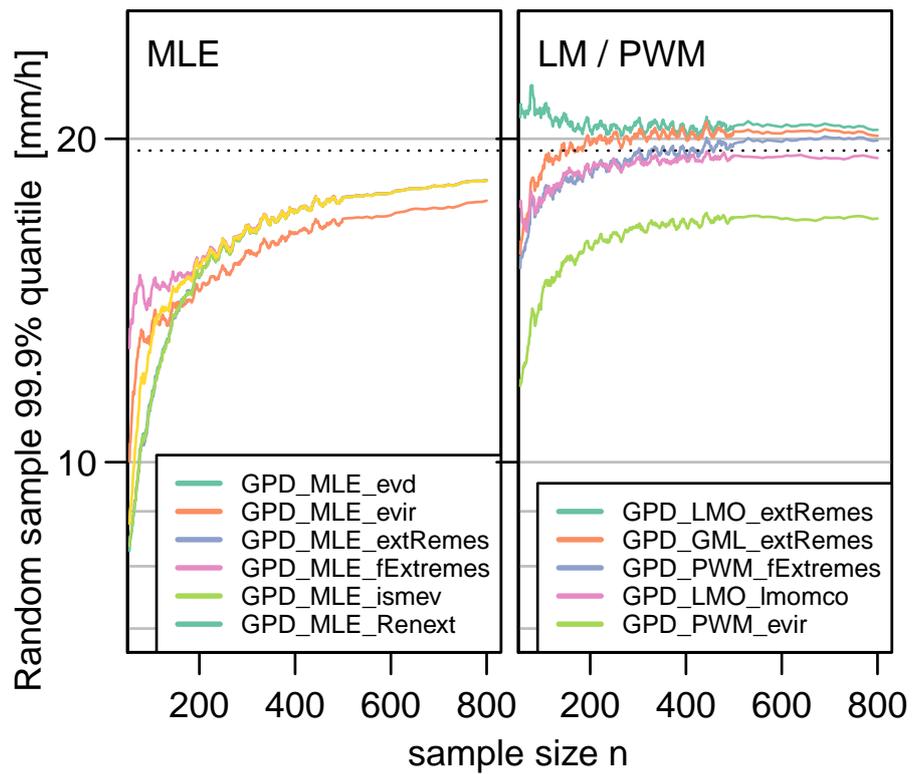


Figure S2. Sample size dependency of different GPD fitting methods

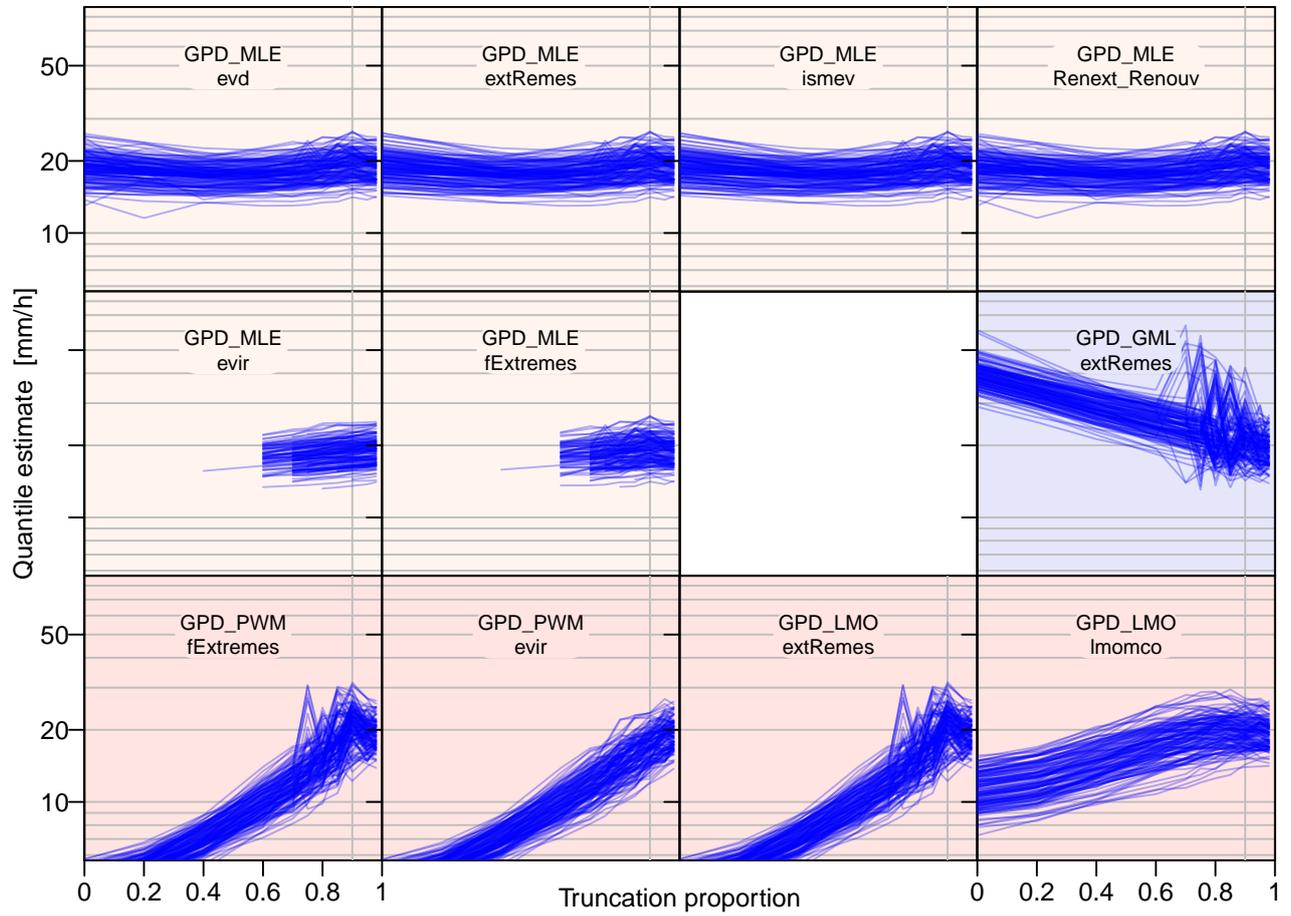


Figure S3. Threshold dependency of different GPD fitting methods

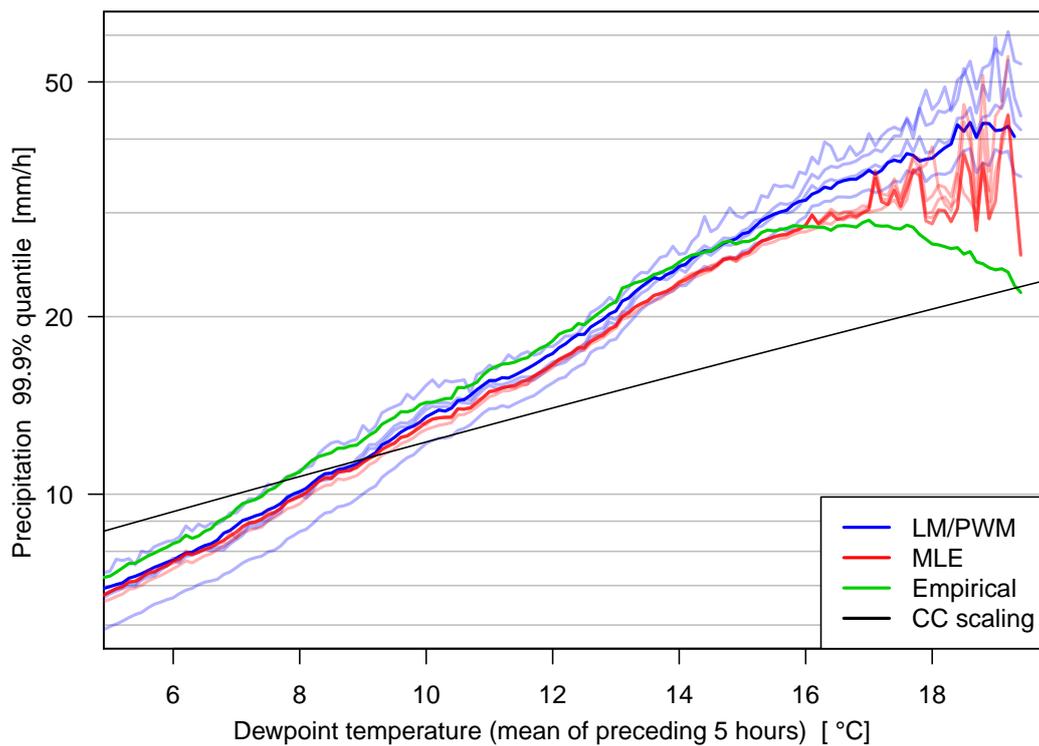


Figure S4. Dependency of rainfall intensity quantile estimations on temperature. The lines mark the per temperature bin average for the 142 stations. The thick blue line marks the Imomco implementation used in the main article.

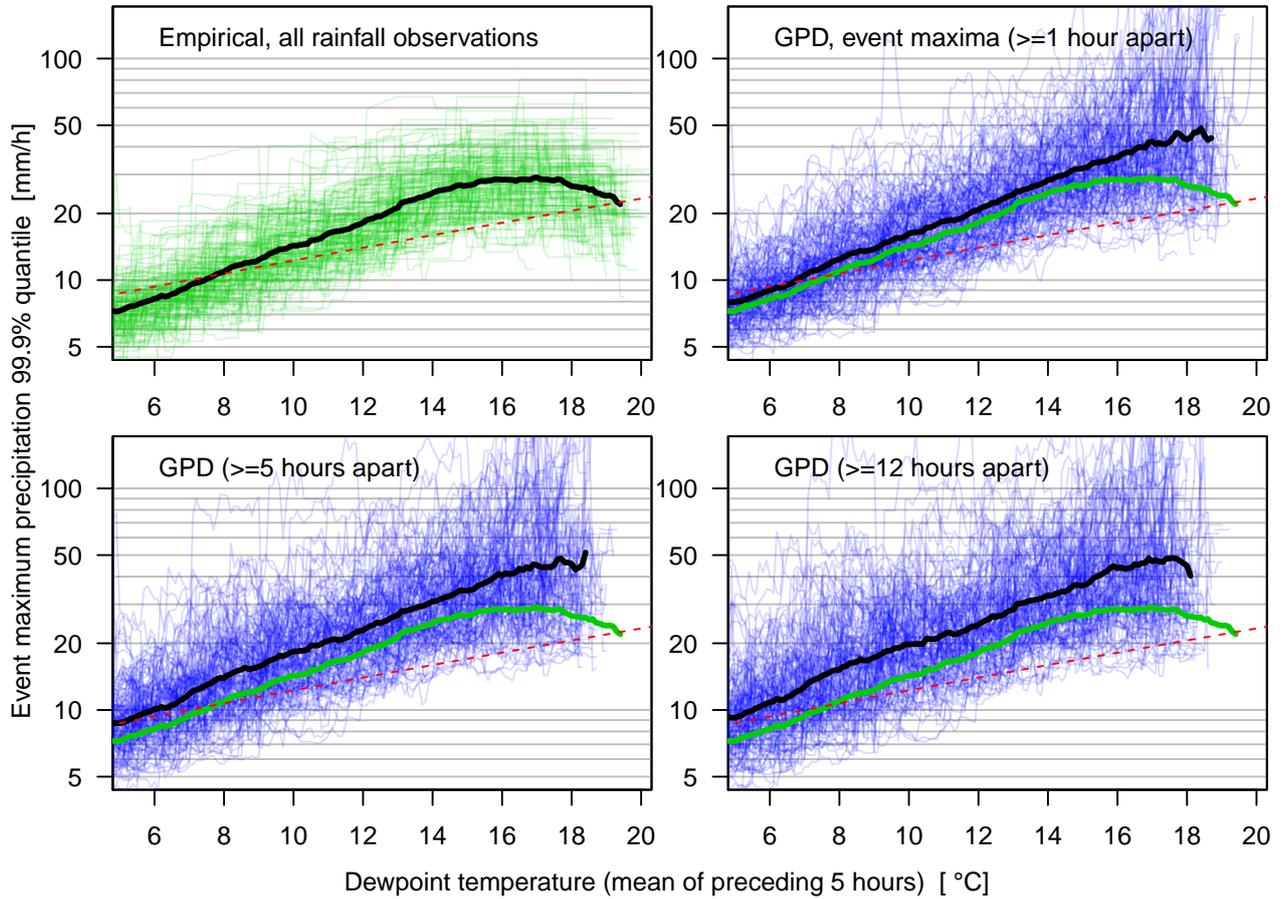


Figure S5. P-T relationships at the 142 stations (as in Figure 3 of the main article) using event maxima instead of all rainfall observations. Events are considered distinct when more than 1, 5 or 12 hours with no rainfall (<0.5mm) separate them.

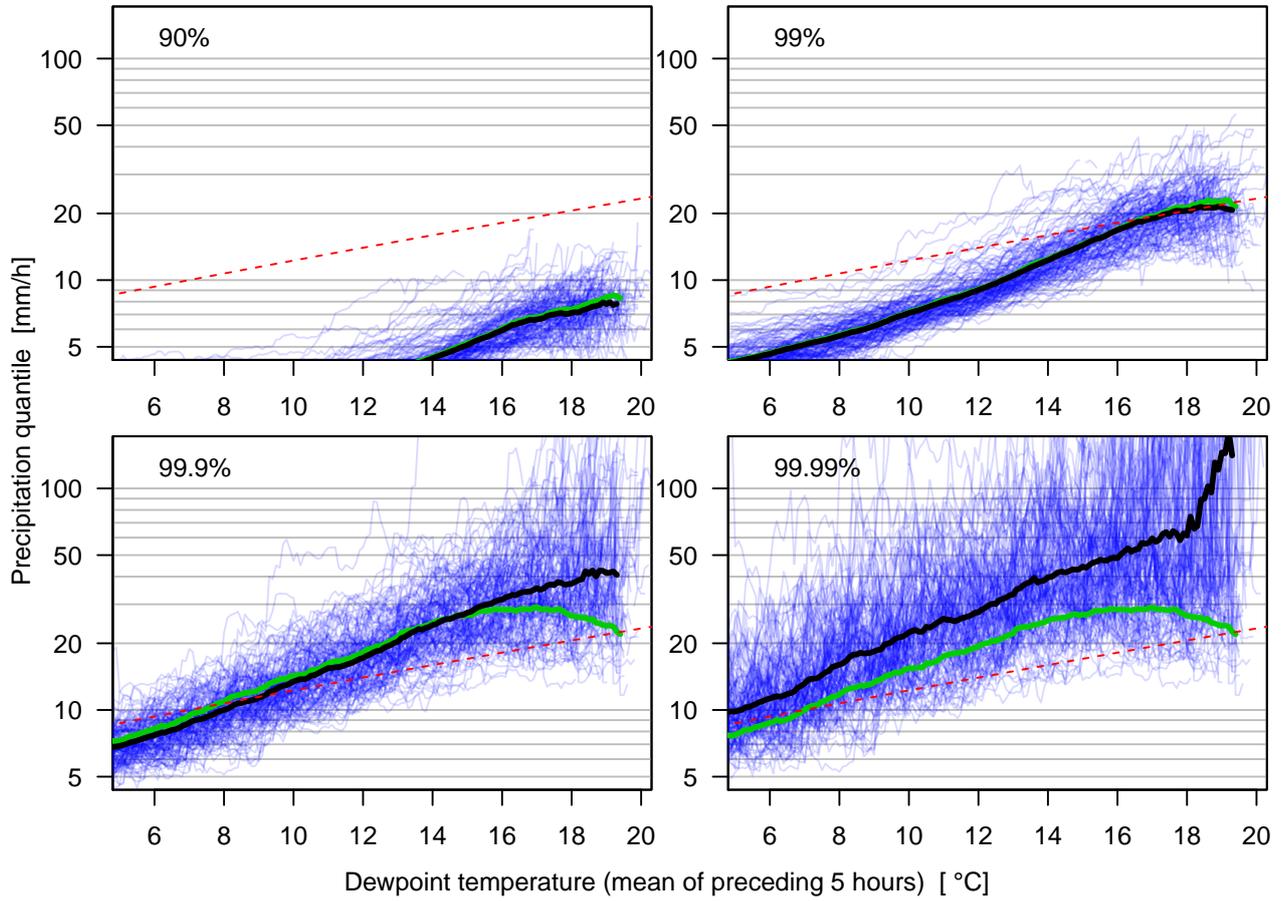


Figure S6. P-T relationships at 142 stations (blue) and their average per temperature bin (black) for different quantiles. The green lines denote the average of empirical estimates across stations.

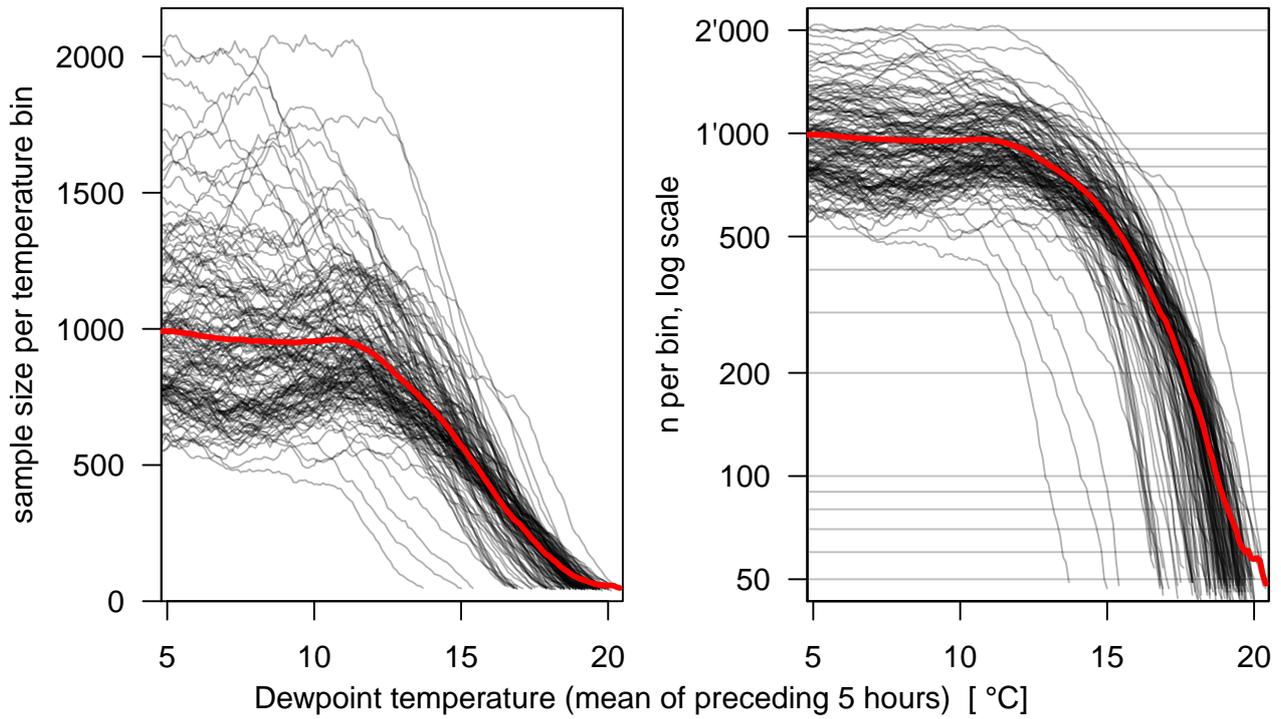


Figure S7. Number of observations per temperature bin for each station (black) and average across stations (red) on linear and logarithmic axis scale.