S1 Runoff threshold quantiles

Table S1. Summary of threshold quantiles for runoff [m^3/s] in the Verzasca and Pincascia catchments.

	$q_{0.5}$	$q_{0.6}$	$q_{0.7}$	$q_{0.8}$	$q_{0.9}$	$q_{0.95}$	$q_{0.975}$	$q_{0.99}$
Verzasca	8.4	11.2	13.6	17.2	25.3	46.9	70.8	135.0
Pincascia	1.2	1.6	2.4	3.3	6.3	13.8	23.8	37.4

S2 ROCa

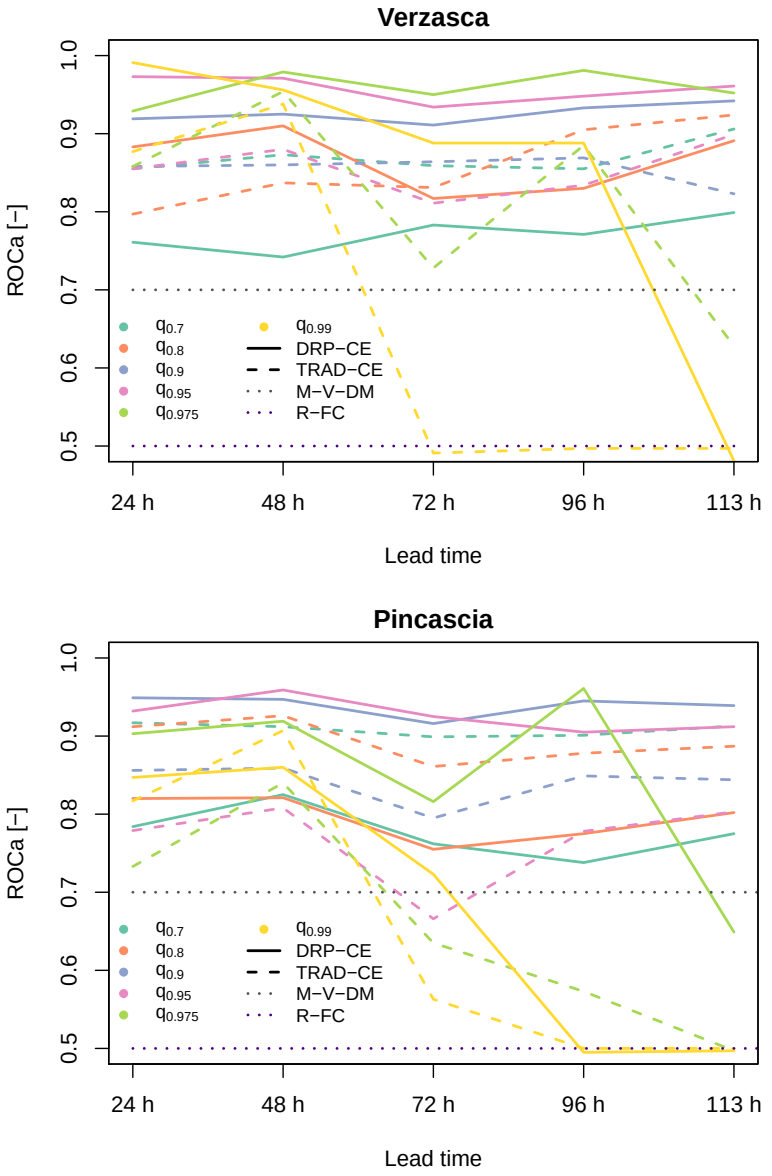


Figure S1. Evolution of ROCa for probabilistic DRP-CE (solid) and TRAD-CE (dashed) as a function of lead time for Verzasca in the upper and for Pincascia catchment in the lower panel for several quantiles. These values served as basis for the ROCa summary in the paper. Grey dotted line (M-V-DM) indicates ROCa of 0.7, which is minimum value that is still useful for decision makers (Buizza et al., 1999). An unskilful forecast would yield a ROCa of 0.5, which is indicated by the purple dotted line (R-FC). A window of 24 hours was taken for the computations, e.g. values from 25 h to 48 h were considered for the 48 h lead time.

S3 BSS

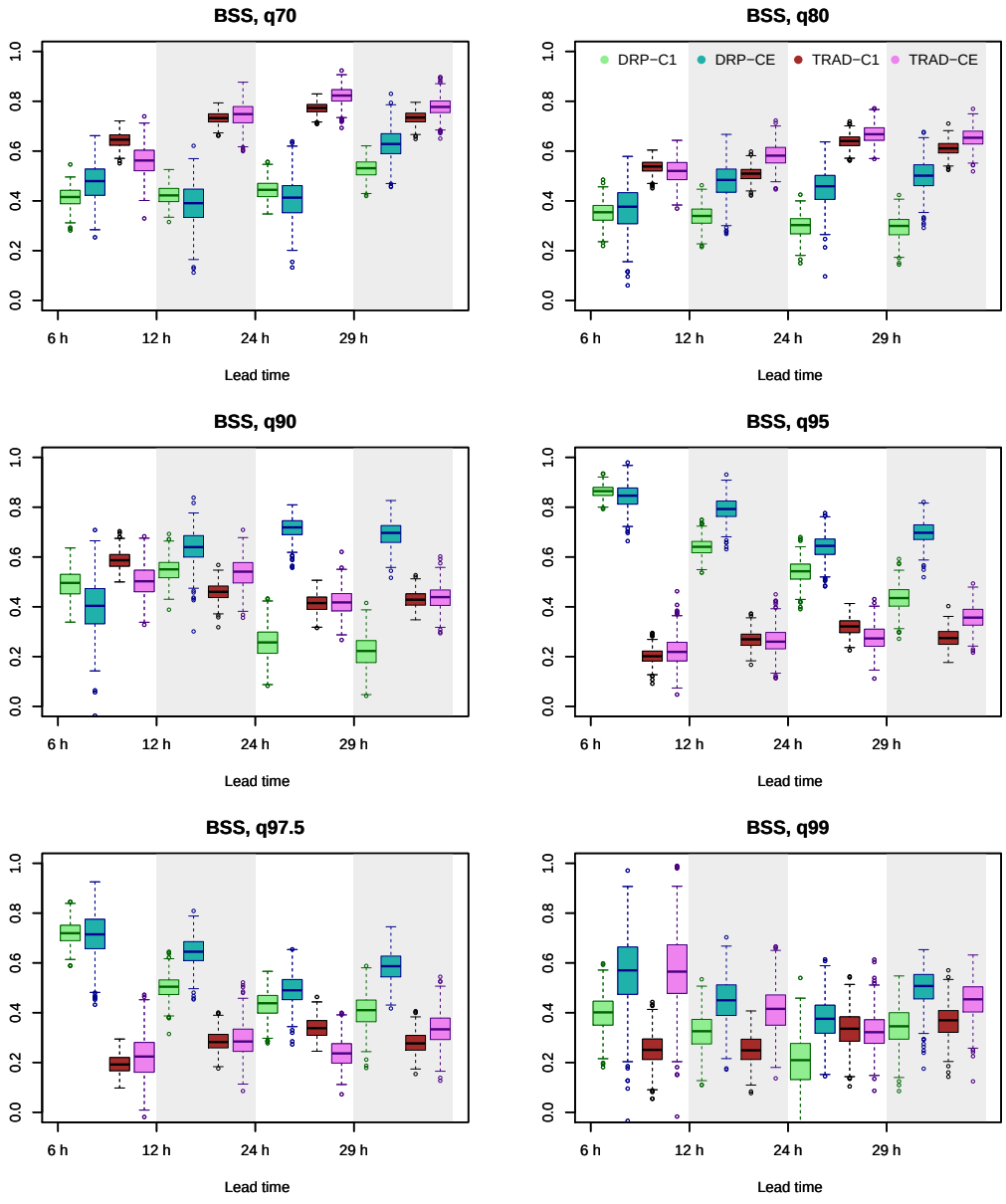


Figure S2. Comparison of BSS in Pincascia catchment for deterministic DRP-C1 and TRAD-C1 and probabilistic DRP-CE and TRAD-CE as a function of lead time for several threshold quantiles. A window of 6 hours was taken for the computations, e.g. values from 19 h to 24 h were considered for the 24 h lead time. The boxplots represent the sampling uncertainties of the score computations obtained with 500 iterations of bootstrapping, which is further explained in the paper.

S4 POD, FAR, FB

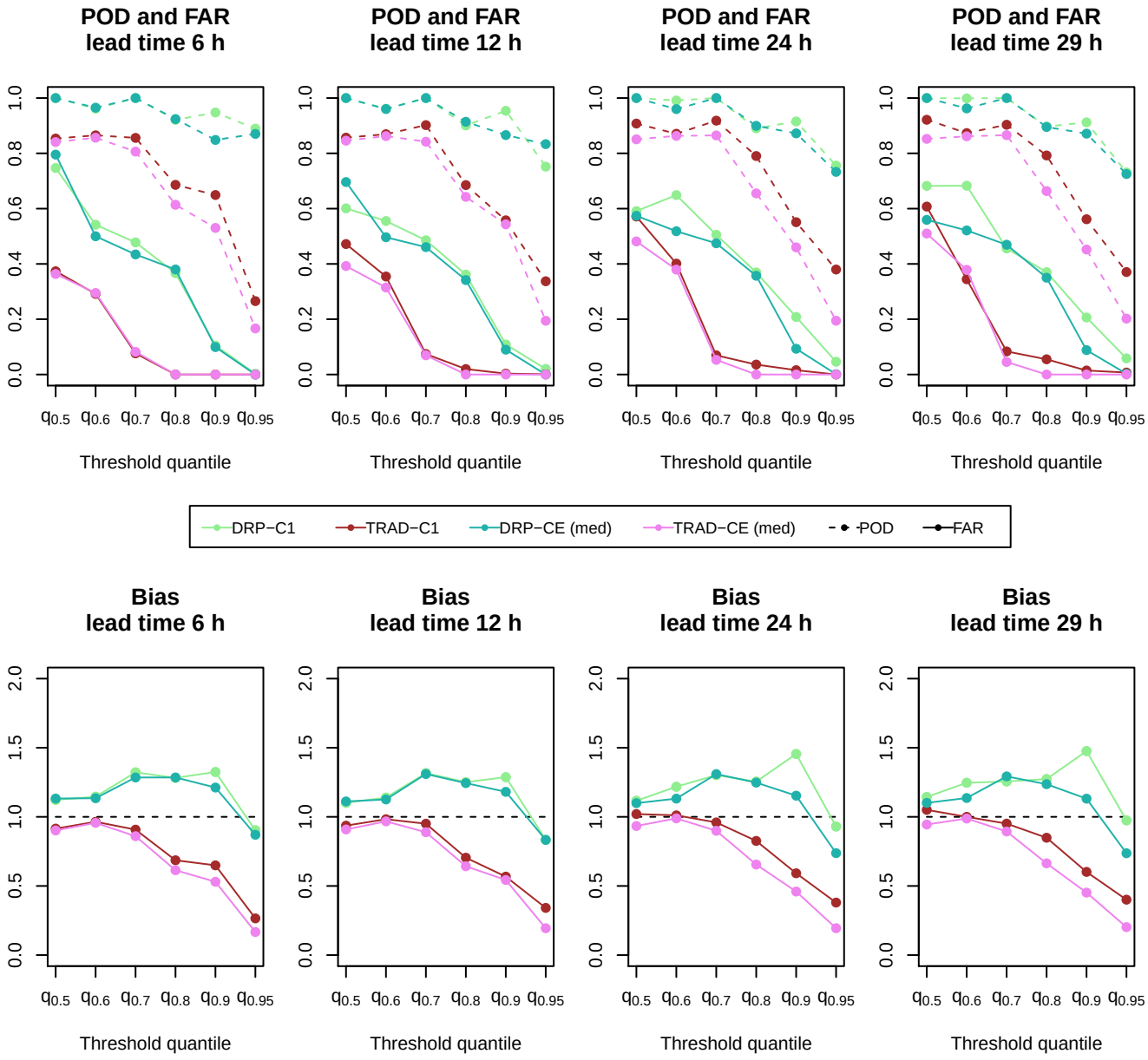


Figure S3. POD, FAR (upper panel) and FB (lower panel) for Pincascia catchment as a function of threshold quantile and for several lead times for DRP-C1, TRAD-C1, DRP-CE (med) and TRAD-CE (med). A window of 6 hours was taken for the computations, e.g. values from 19 h to 24 h were considered for the 24 h lead time.

S5 NSE, KGE (including reference runs)

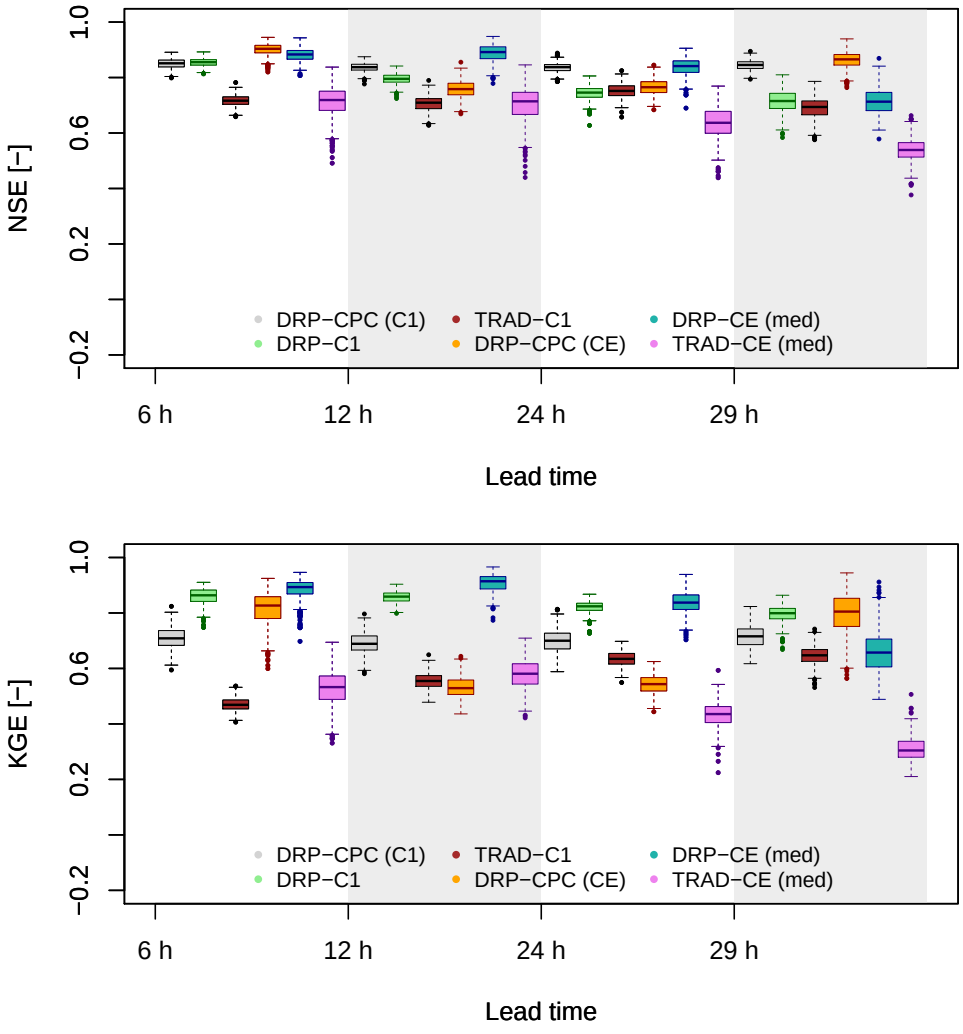
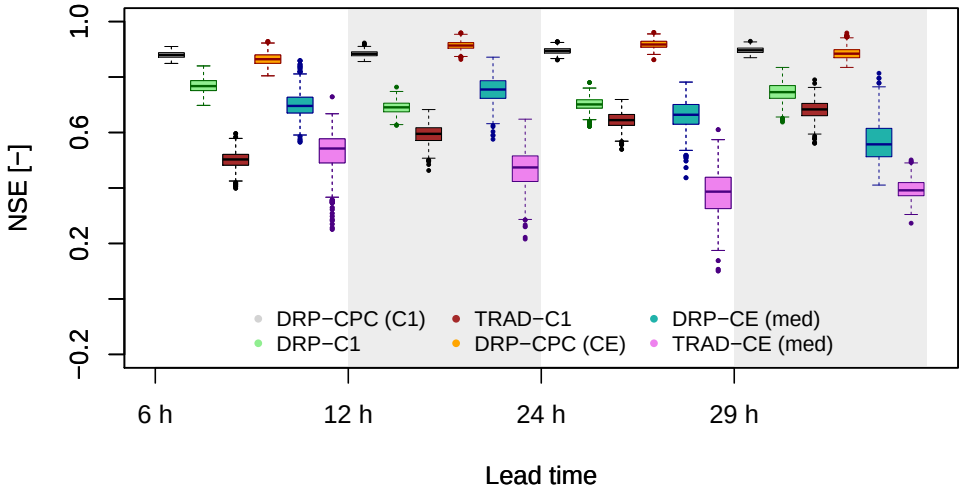


Figure S4. NSE and KGE for Verzasca catchment as a function of lead time for DRP-C1, TRAD-C1, DRP-CE (med) and TRAD-CE (med). In addition, values for the reference runs DRP-CPC (C1) and DRP-CPC (CE) are shown. The reference runs are completely forced with CombiPrecip data. A window of 6 hours was taken for the computations, e.g. from 19 h to 24 h for the 24 h lead time. The boxplots represent the sampling uncertainties of the score computations obtained with 500 iterations of bootstrapping, which is further explained in the paper.



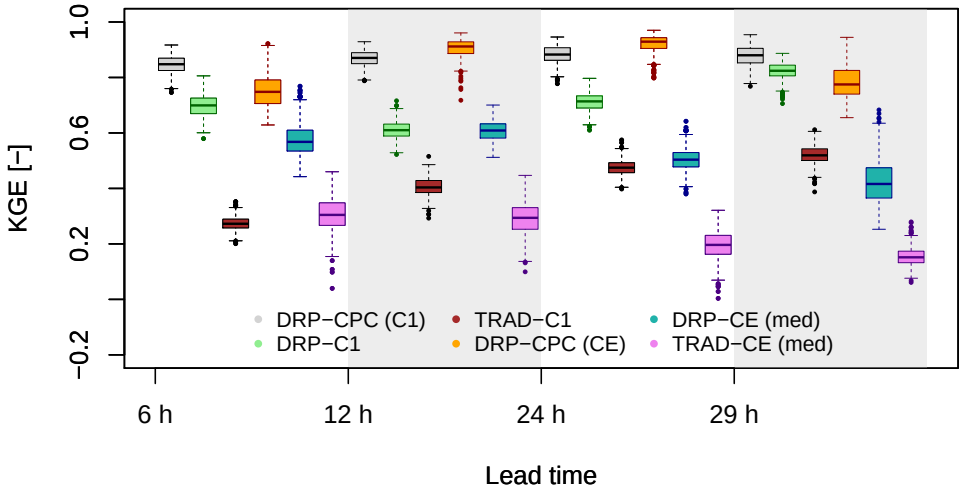


Figure S5. NSE and KGE for Pincascia catchment as a function of lead time for DRP-C1, TRAD-C1, DRP-CE (med) and TRAD-CE (med). In addition, values for the reference runs DRP-CPC (C1) and DRP-CPC (CE) are shown. The reference runs are completely forced with CombiPrecip data. A window of 6 hours was taken for the computations, e.g. from 19 h to 24 h for the 24 h lead time. The boxplots represent the sampling uncertainties of the score computations obtained with 500 iterations of bootstrapping, which is further explained in the paper.

S6 Reliability diagrams

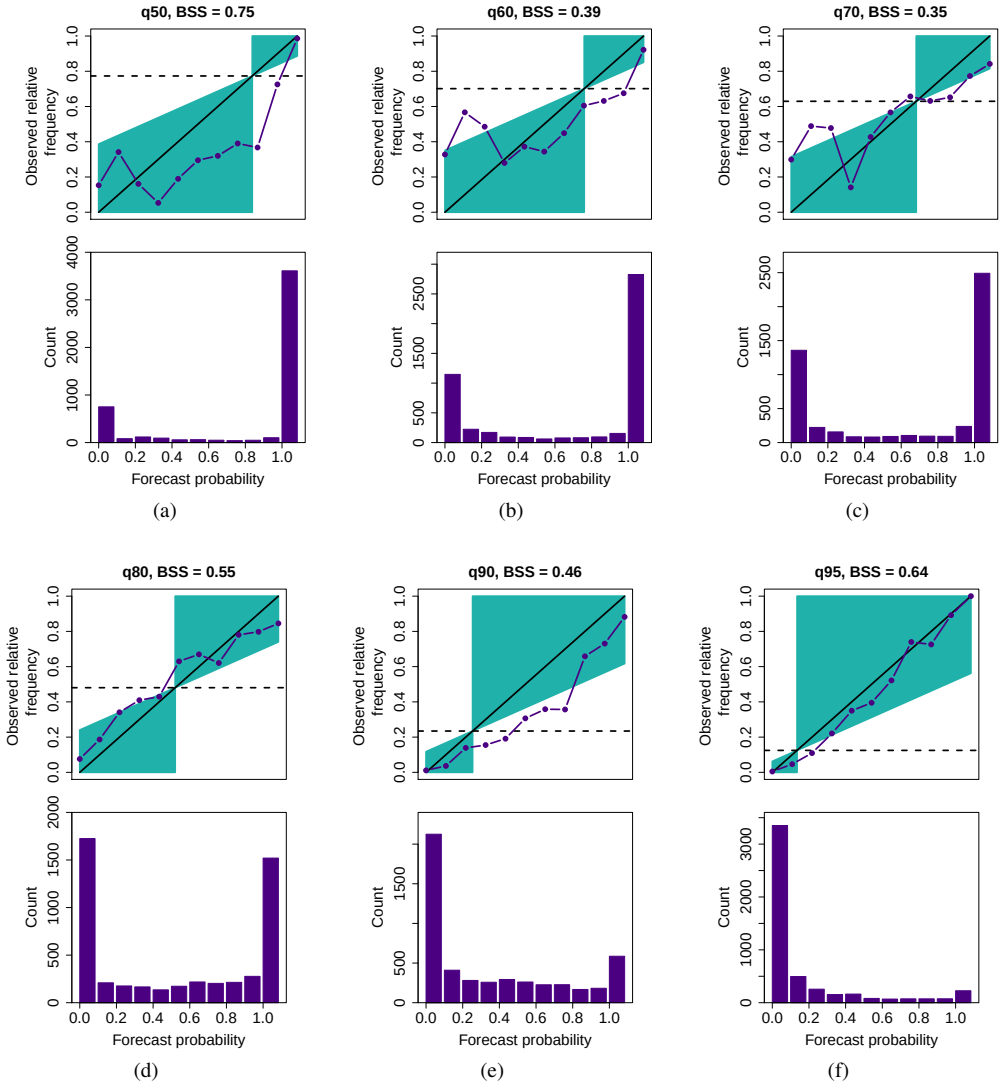


Figure S6. Reliability diagrams for DRP-CE in Verzasca catchment for various quantiles from (a) to (f). The upper panel shows the calibration function, which is observed relative frequency as a function of forecast probability. Horizontal dashed line indicates the no-resolution line. Points within blue shaded region contribute to a positive BSS. The lower panel shows the refinement distribution, a histogram indicating how often each probability forecast was issued. All values up to a lead time of 113 hours were considered for the computations.

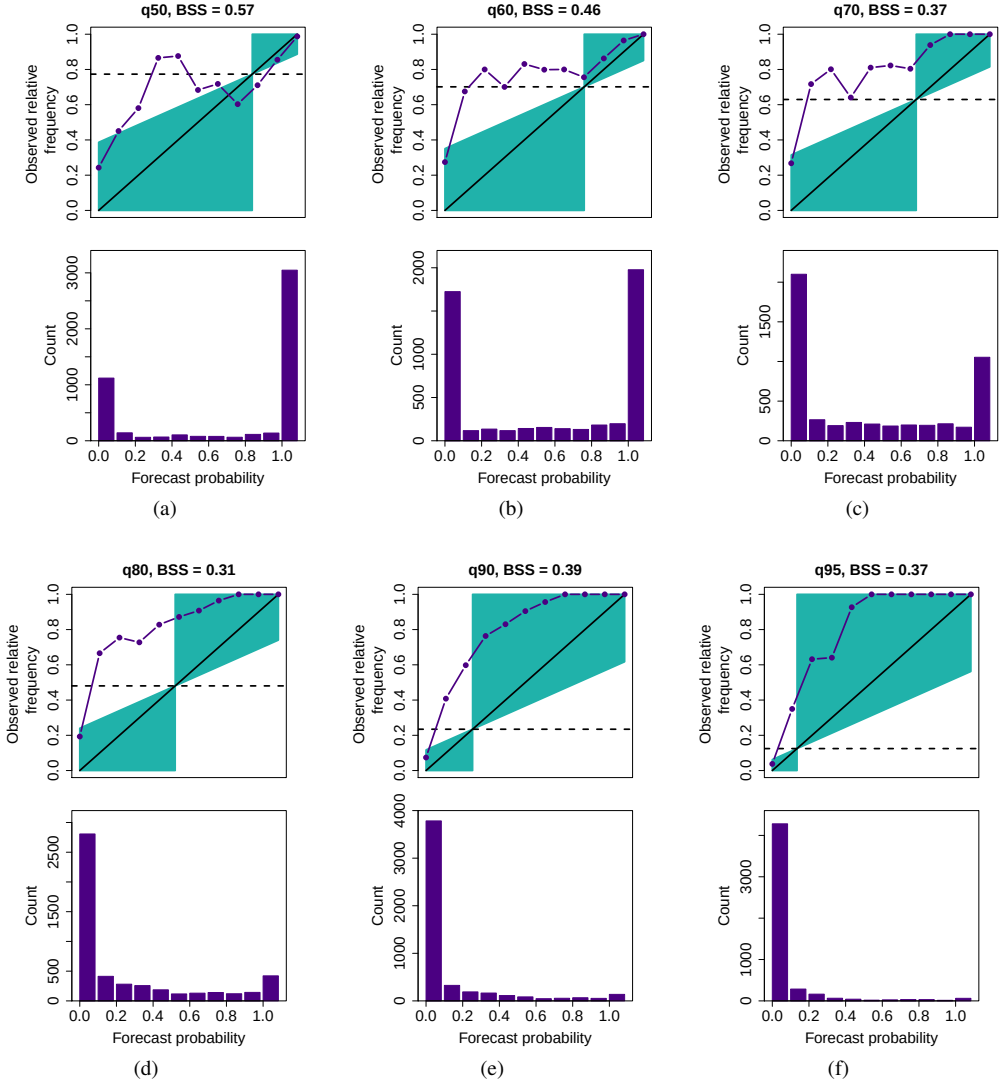


Figure S7. Reliability diagrams for TRAD-CE in Verzasca catchment for various quantiles from (a) to (f). The upper panel shows the calibration function, which is observed relative frequency as a function of forecast probability. Horizontal dashed line indicates the no-resolution line. Points within blue shaded region contribute to a positive BSS. The lower panel shows the refinement distribution, a histogram indicating how often each probability forecast was issued. All values up to a lead time of 113 hours were considered for the computations.

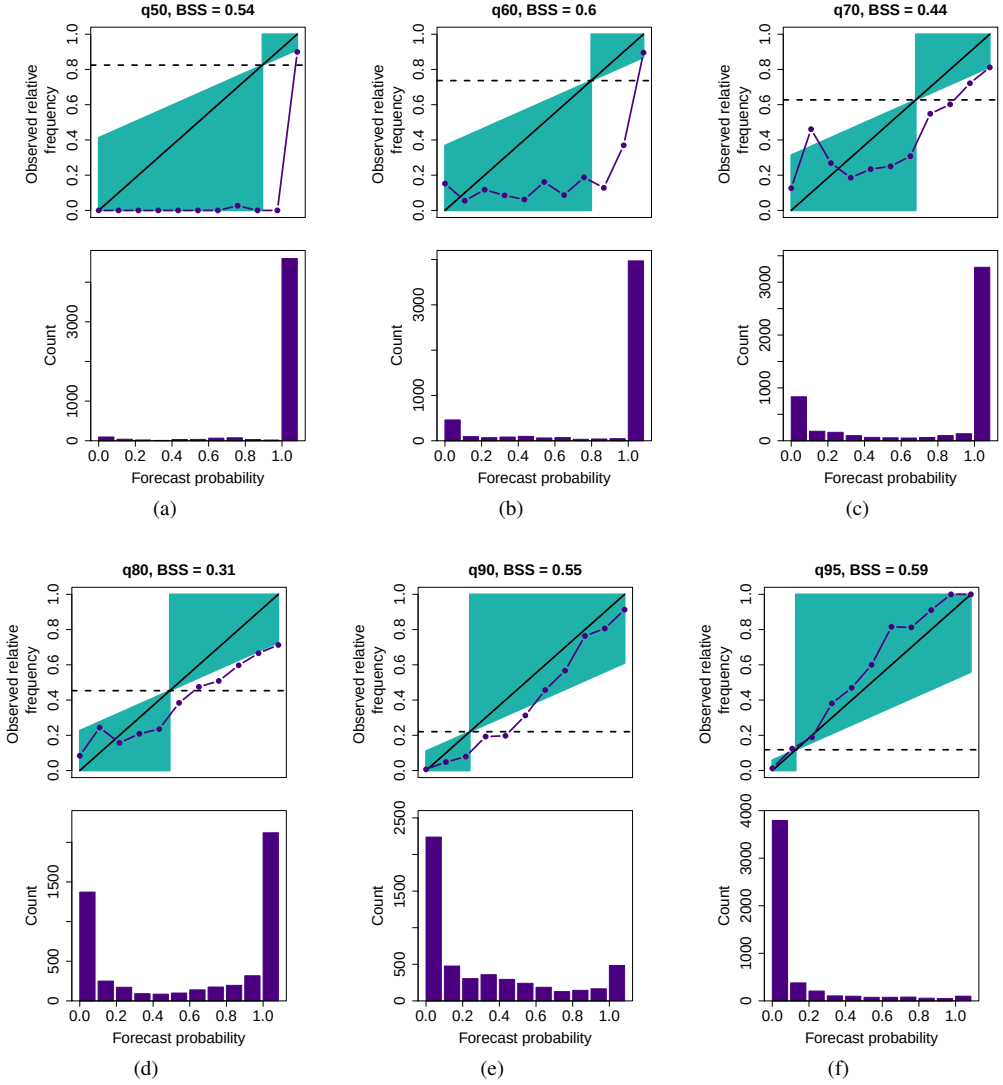


Figure S8. Reliability diagrams for DRP-CE in Pincascia subcatchment for various quantiles from (a) to (f). The upper panel shows the calibration function, which is observed relative frequency as a function of forecast probability. Horizontal dashed line indicates the no-resolution line. Points within blue shaded region contribute to a positive BSS. The lower panel shows the refinement distribution, a histogram indicating how often each probability forecast was issued. All values up to a lead time of 113 hours were considered for the computations.

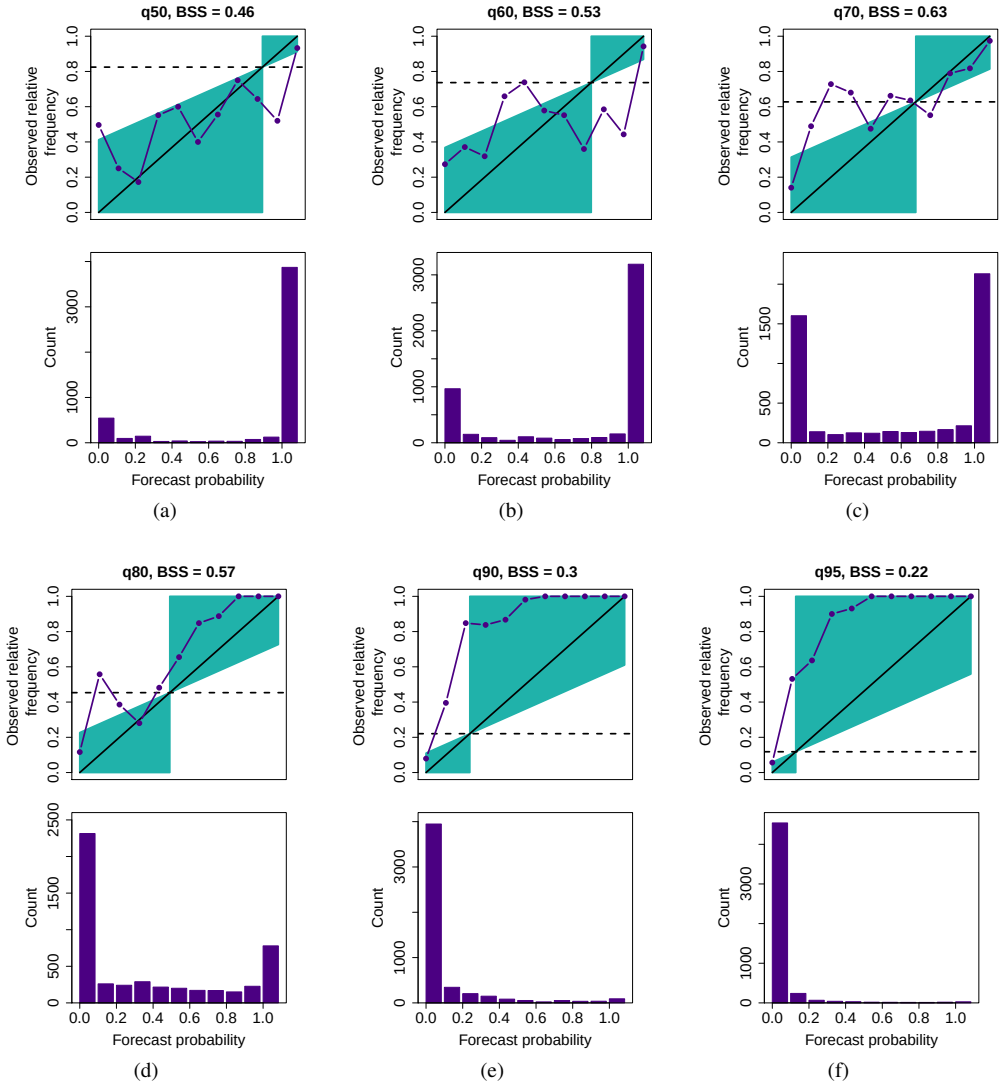


Figure S9. Reliability diagrams for TRAD-CE in Pincascia subcatchment for various quantiles from (a) to (f). The upper panel shows the calibration function, which is observed relative frequency as a function of forecast probability. Horizontal dashed line indicates the no-resolution line. Points within blue shaded region contribute to a positive BSS. The lower panel shows the refinement distribution, a histogram indicating how often each probability forecast was issued. All values up to a lead time of 113 hours were considered for the computations.

S7 Verification rank histograms

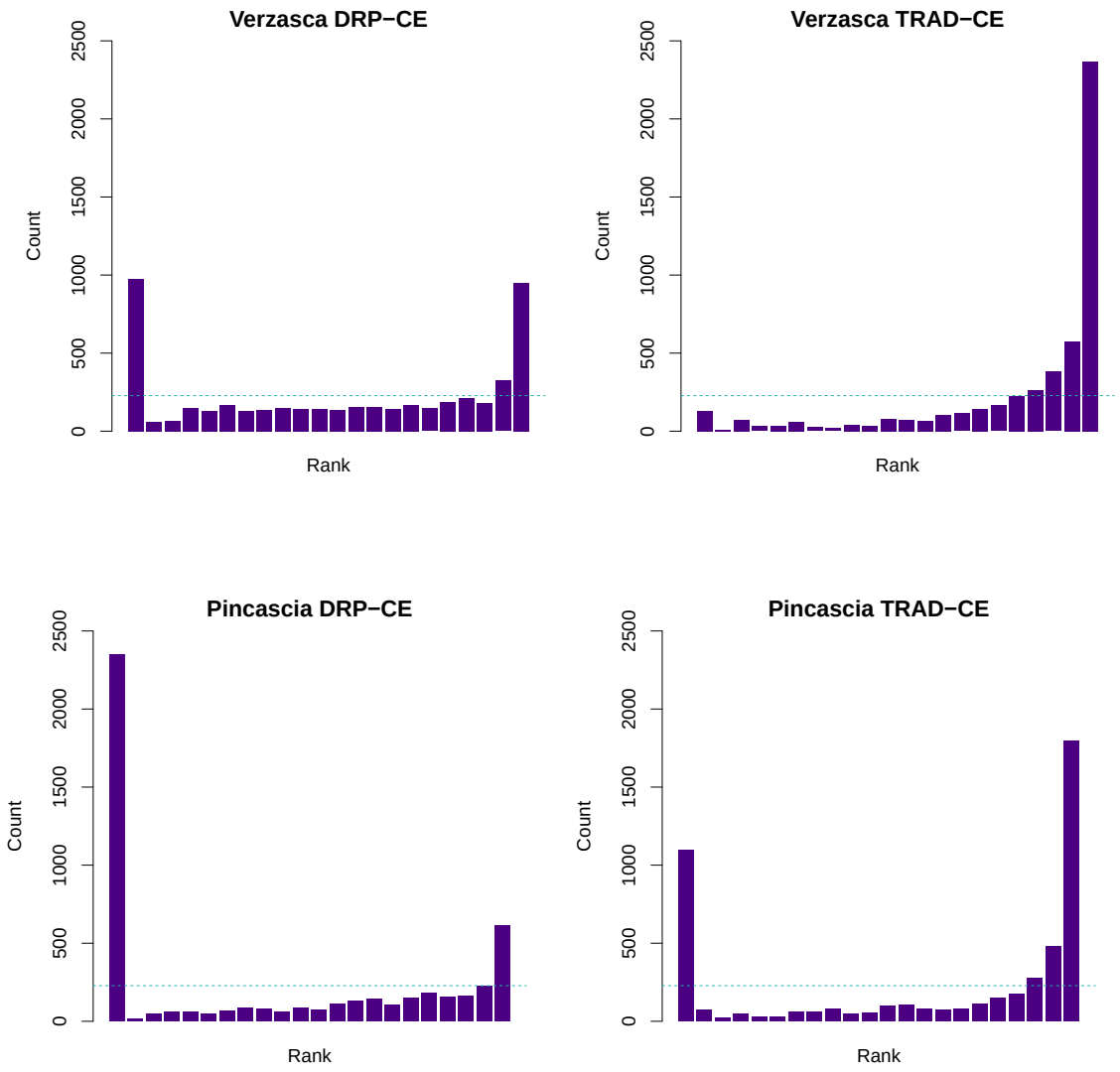


Figure S10. Verification rank histograms for DRP-CE and TRAD-CE in Verzasca and Pincascia catchments. Blue dashed line indicates uniform distribution. All values up to a lead time of 113 hours were considered for the computations.