Supplement of Nat. Hazards Earth Syst. Sci., 24, 1929–1935, 2024
https://doi.org/10.5194/nhess-24-1929-2024-supplement

Natural Hazards and Earth System Sciences
Open Access

Supplement of

Brief communication: SWM – stochastic weather model for precipitation-related hazard assessments using ERA5-Land data

Melody Gwyneth Whitehead and Mark Stephen Bebbington

Correspondence to: Melody Gwyneth Whitehead (m.whitehead@massey.ac.nz)

Overview

This supplement provides the complete set of results from the statistical analyses on both exemplar realisations: *Realisation 1* and *Realisation 2* each of which represent 999 simulated datasets. These are compared to the single real dataset for the Tarawera/Rangitāiki catchment: hourly rainfall data across a 11x14 grid of longitude: {176° E, 176.1° E, …, 177.0° E} and latitude: {37.8° S, 37.9° S, …, 39.1° S}, for 40 years (1981 – 2020) downloaded from ERA5-land.

Additional comparisons are provided here for rainfall quantiles by month (Figure S3), comparison of monthly rainfall totals (Figure S4), and comparisons of mean, variance, skewness, and proportion of wet periods (Figure S5).

S1 Monthly means and variance

The Shapiro-Wilks normality test (Royston, 1982) was run using the *shapiro.test()* function from *base* R. A significance level of 0.05 was applied, i.e., any test result with $p < 0.05$ was deemed to have failed the normality test (Table S1).

The Levene test for equal variance was run using the *leveneTest()* function from *car* (Fox and Weisburg, 2019). A significance level of 0.05 was applied, i.e., any test result with $p < 0.05$ was deemed to have failed the equality of variance test (Table S2).

Student's t-test for equality of means (Student, 1908) was run using the *t.test()* function from *base* R. A significance level of 0.05 was applied, i.e., any test result with $p < 0.05$ was deemed to have failed the normality test (Table S3).

Table S1: Shapiro-Wilks normality test failure rate (%) by month on mean monthly rainfall

Month	**ERA5-land**		**Realisation 1**	**Realisation 2**
January	p = 0.0004647	1/1 = 100 %	754/999 = 75 %	820 / 999 = 82 %
February	p = 0.001508	1/1 = 100 %	681/999 = 68 %	779 / 999 = 78 %
March	p = 0.001861	1/1 = 100 %	466/999 = 47 %	548 / 999 = 55 %
April	p = 0.009444	1/1 = 100 %	613/999 = 61 %	689 / 999 = 69 %
May	p = 0.003474	1/1 = 100 %	486/999 = 49 %	587 / 999 = 59 %
June	p = 0.06495	0/1 = 0 %	317/999 = 32 %	299 / 999 = 30 %
July	p = 0.017	1/1 = 100 %	432/999 = 43 %	515 / 999 = 52 %
August	p = 0.0947	0/1 = 0 %	217/999 = 22 %	341 / 999 = 34 %
September	p = 0.5795	0/1 = 0 %	182/999 = 18 %	261 / 999 = 26 %
October	p = 0.404	0/1 = 0 %	225/999 = 23 %	441 / 999 = 44 %
November	p = 0.0003	1/1 = 100 %	283/999 = 28 %	468 / 999 = 47 %
December	p = 0.434	0/1 = 0 %	253/999 = 25 %	447 / 999 = 45 %
Overall	**7 / 12 = 58.3 %**		**594 / 1140 = 52.1 %**	**6195 / 11988 = 51.7 %**

Table S2: Levene variance equality test failure rate (%) by month on mean monthly rainfall pairs (real:sample)

Month	ERA5-land : Realisation 1	ERA5-land : Realisation 2
January	10 / 999 = 1 %	10 / 999 = 1 %
February	66 / 999 = 7 %	63 / 999 = 6.3 %
March	3 / 999 = 0.3 %	5 / 999 = 0.5 %
April	9 / 999 = 0.9 %	31 / 999 = 3.1 %
May	410 / 999 = 41 %	52 / 999 = 5.2 %
June	103 / 999 = 10 %	189 / 999 = 18.9 %
July	5 / 999 = 0.5 %	5 / 999 = 0.5 %
August	9 / 999 = 0.9 %	8 / 999 = 0.8 %
September	23 / 999 = 2.3 %	36 / 999 = 3.6 %
October	4 / 999 = 0.4 %	36 / 999 = 3.6 %
November	26 / 999 = 2.6 %	2 / 999 = 0.2 %
December	11 / 999 = 1.1 %	8 / 999 = 0.8 %
Overall	**679 / 11988 = 5.7 %**	**445 / 11988 = 3.7 %**

Table S3: Student's t test failure rate (%) by month on mean monthly rainfall pairs (real:sample)

Month	ERA5-land : Realisation 1	EAR5-land : Realisation 2
January	9 / 999 = 0.9 %	1 / 999 = 0.1 %
February	1 / 999 = 0.1 %	1 / 999 = 0.1 %
March	2 / 999 = 0.2 %	4 / 999 = 0.4 %
April	7 / 999 = 0.7 %	4 / 999 = 0.4 %
May	13 / 999 = 1.3 %	15 / 999 = 1.5 %
June	21 / 999 = 2.1 %	21 / 999 = 2.1 %
July	16 / 999 = 1.6 %	3 / 999 = 0.3 %
August	6 / 999 = 0.6 %	3 / 999 = 0.3 %
September	6 / 999 = 0.6 %	11 / 999 = 1.1 %
October	6 / 999 = 0.6 %	0 / 999 = 0 %
November	12 / 999 = 1.2 %	10 / 999 = 1 %
December	6 / 999 = 0.6 %	3 / 999 = 0.3 %
Overall	**105 / 11988 = 0.9 %**	**76 / 11988 = 0.6 %**

S2 Significance of month and source for rainfall prediction

Linear models were built for each real : realisation set using the *lm()* function in base R with both month and source as a factor, with m1 allowing for an interaction term (between month and source), and m2 not. Accompanying files: Tukey_HSD_Realisation1_results.txt and Tukey_HSD_Realisation2_results.txt provide all outputs from check 2 for each realisation. While the *summary()* function for each model does provide an estimate of whether a model coefficient is statistically significant, these p-values are unreliable for pair-wise comparisons because the probability of false detection is over inflated (the family-wise error rate). Thus, Tukey's Honest Significant Difference function is used instead, this was applied using the *TukeyHSD()* function from *base* R. In all cases, the p-value was high (> 0.05), indicating that source is not a statistically significant factor in the prediction of rainfall data.

Realisation 1, model 1: *TukeyHSD()* for whether source is a statistically significant factor, $p = 0.9871253$
Realisation 1, model 2: *TukeyHSD()* for whether source is a statistically significant factor, $p = 0.9871252$
Realisation 2, model 1: *TukeyHSD()* for whether source is a statistically significant factor, $p = 0.9891186$
Realisation 2, model 2: *TukeyHSD()* for whether source is a statistically significant factor, $p = 0.9891186$

S3 Distribution of monthly rainfall totals

For each of the 999 simulated data in each sample, empirical Cumulative Distribution functions (eCDFs) were built using the *ecdf()* function in *base* R. These were then plotted and overlain by the real eCDF to look for departures, i.e., any locations where the real (ERA5-land) data fell outside of the envelope drawn by the simulated data. Results for *Realisation 1* and *Realisation 2* are provided in parallel (by month) as Figure S1.

S4 Temporal trends on daily and monthly timescales

For each of the 999 simulated data in each sample, autocorrelation functions (Venables and Ripley, 2002) were built using the ac*f()* function in *base* R. These were then plotted and overlain by the real autocorrelation function to look for departures, i.e., any locations where the real (ERA5-land) data fell outside of the envelope drawn by the simulated data. Results for *Realisation 1* and *Realisation 2* are provided in parallel (by month) as Figure S2.

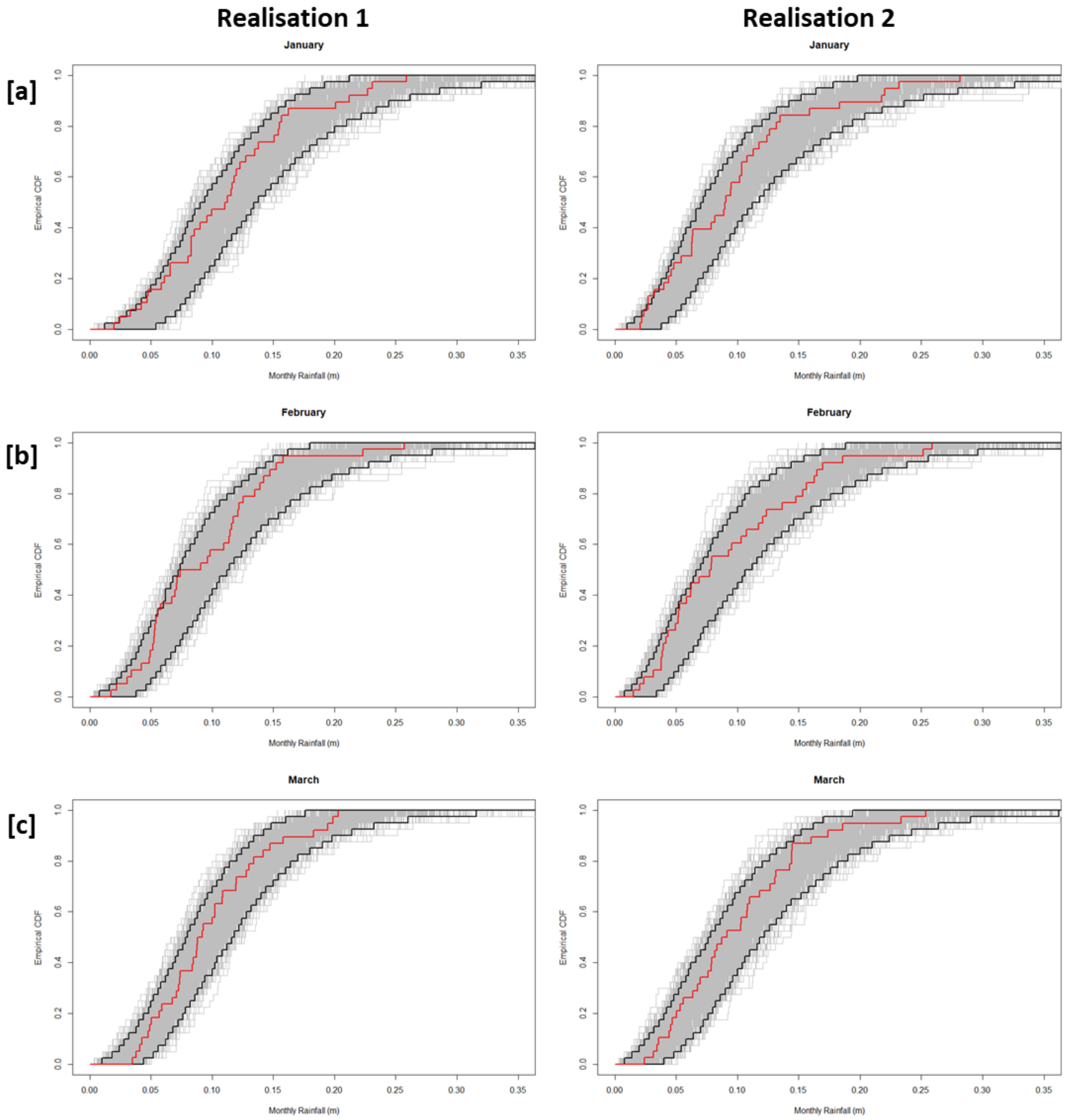


Figure S1: Empirical Cumulative Distribution Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (a) January, (b) February, and (c) March.

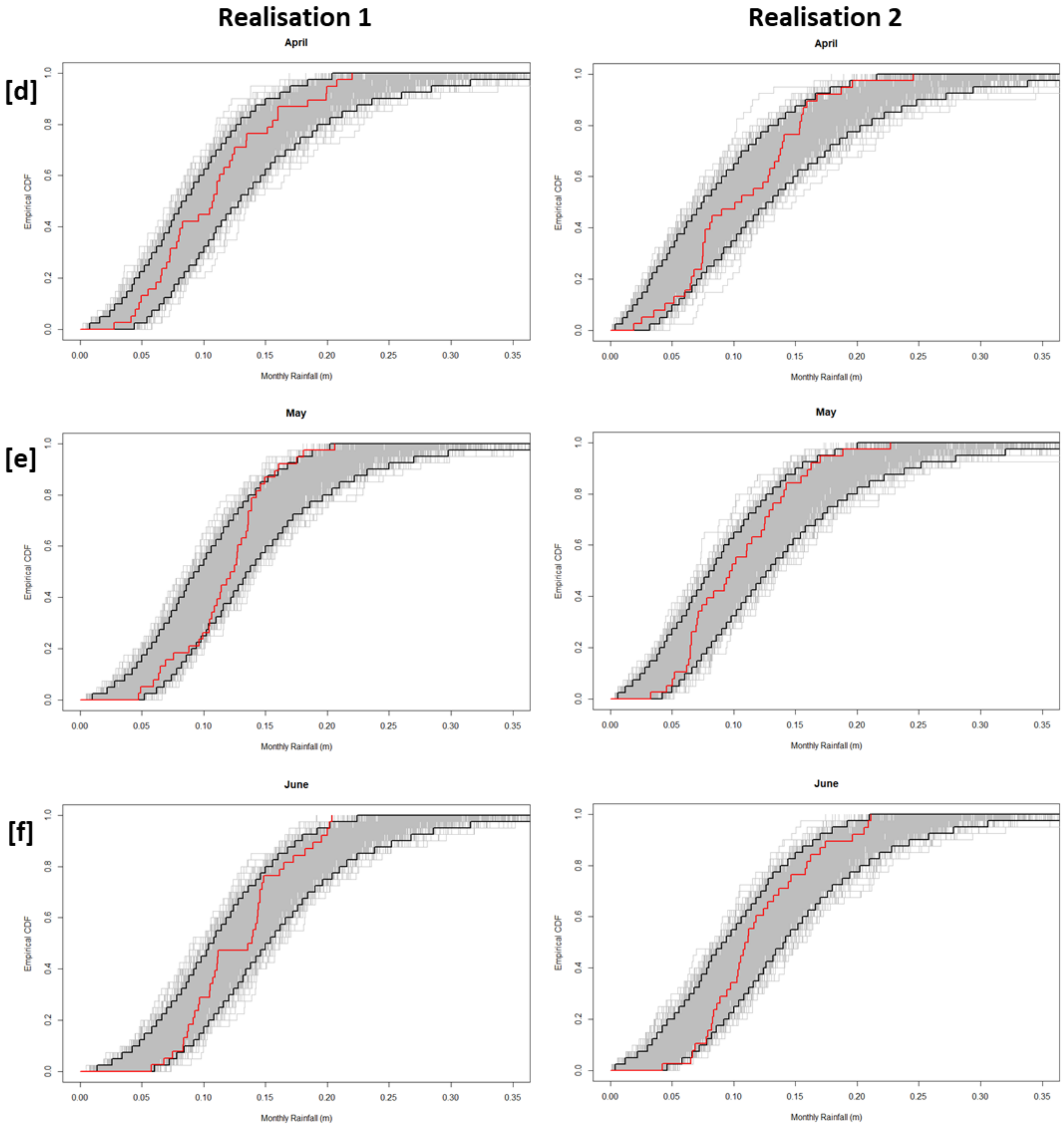


Figure S1: Empirical Cumulative Distribution Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (d) April, (e) May, and (f) June.

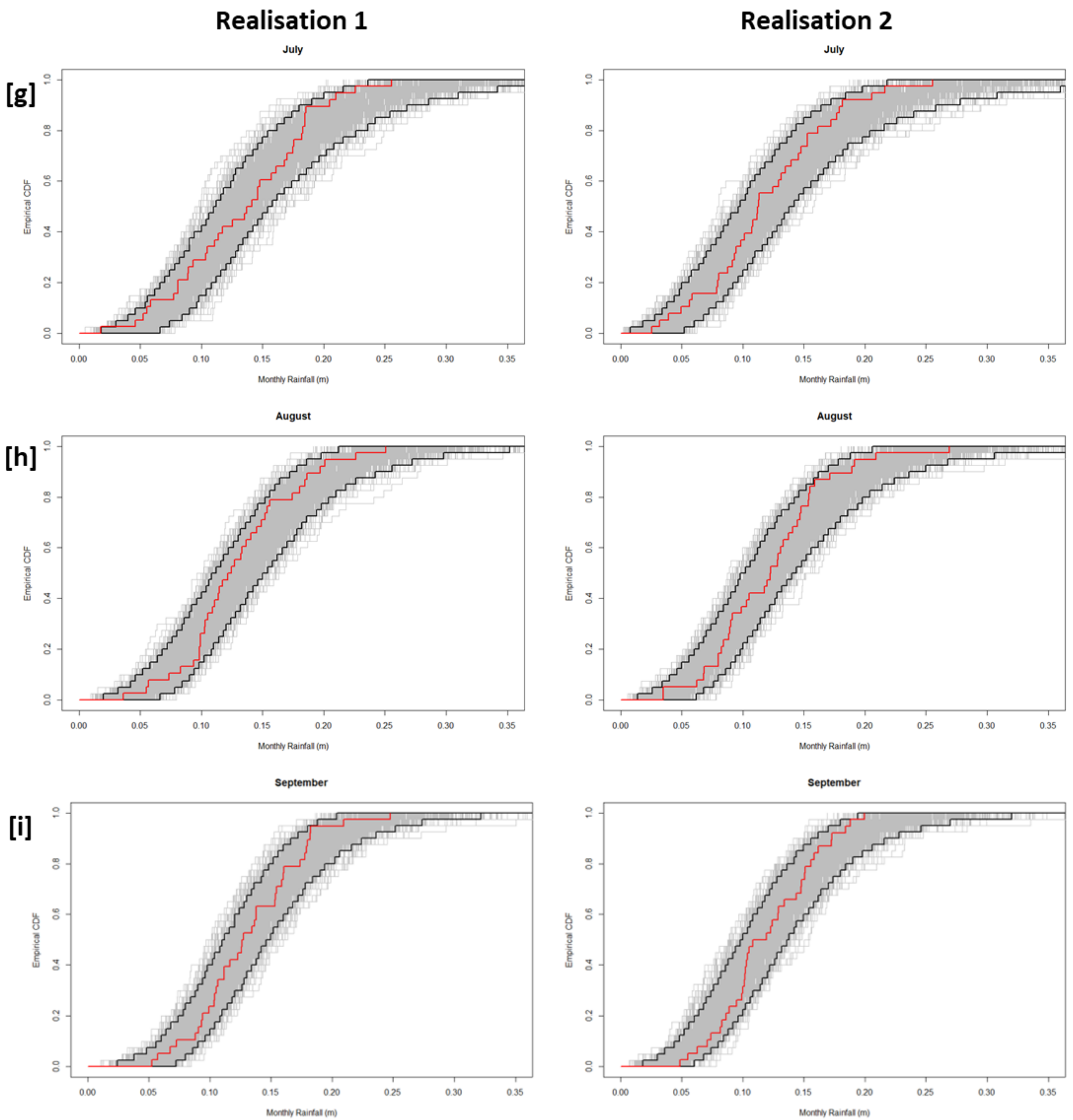


Figure S1: Empirical Cumulative Distribution Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (g) July, (h) August, and (i) September.

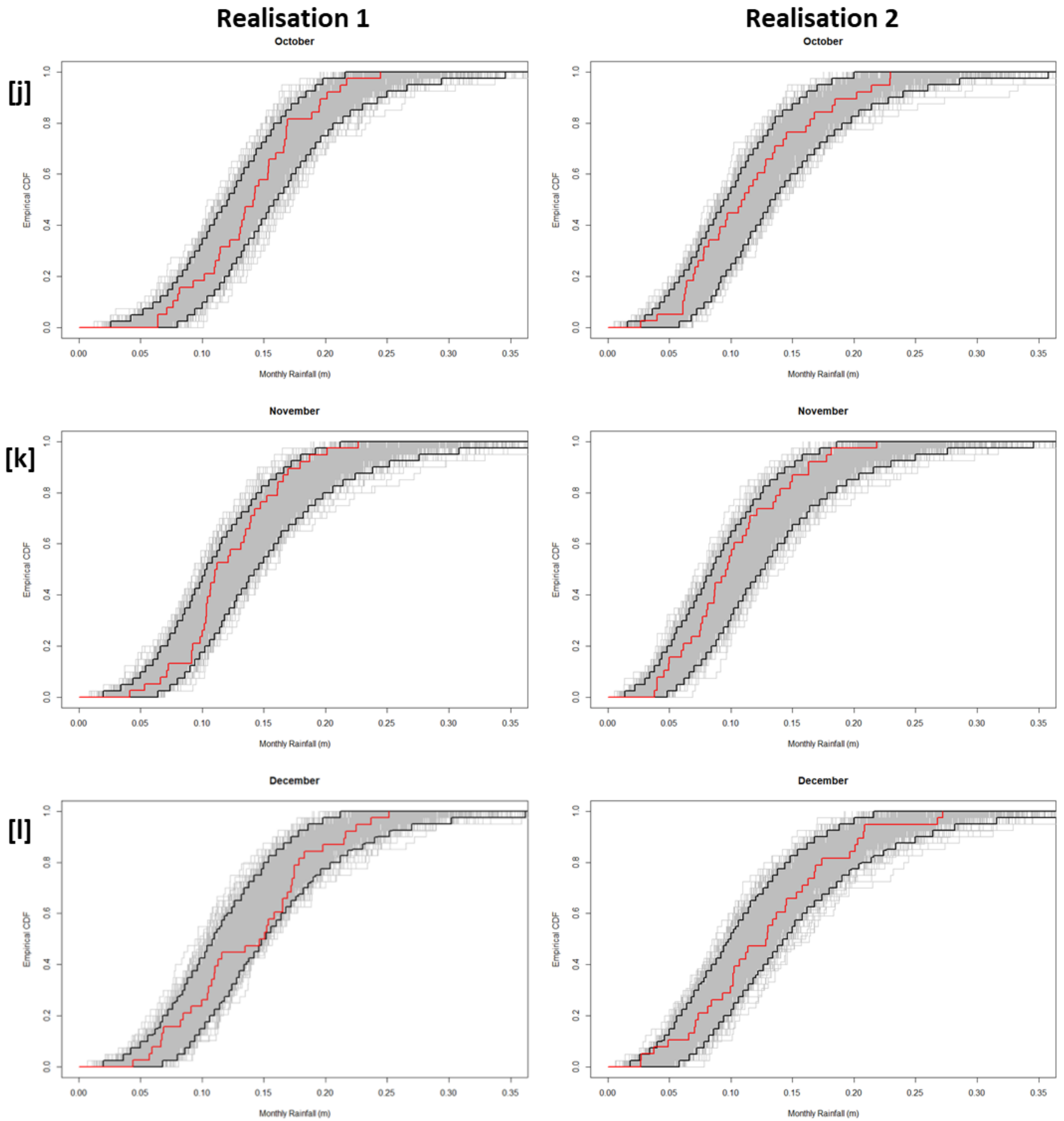


Figure S1: Empirical Cumulative Distribution Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (j) October, (k) November, and (l) December.

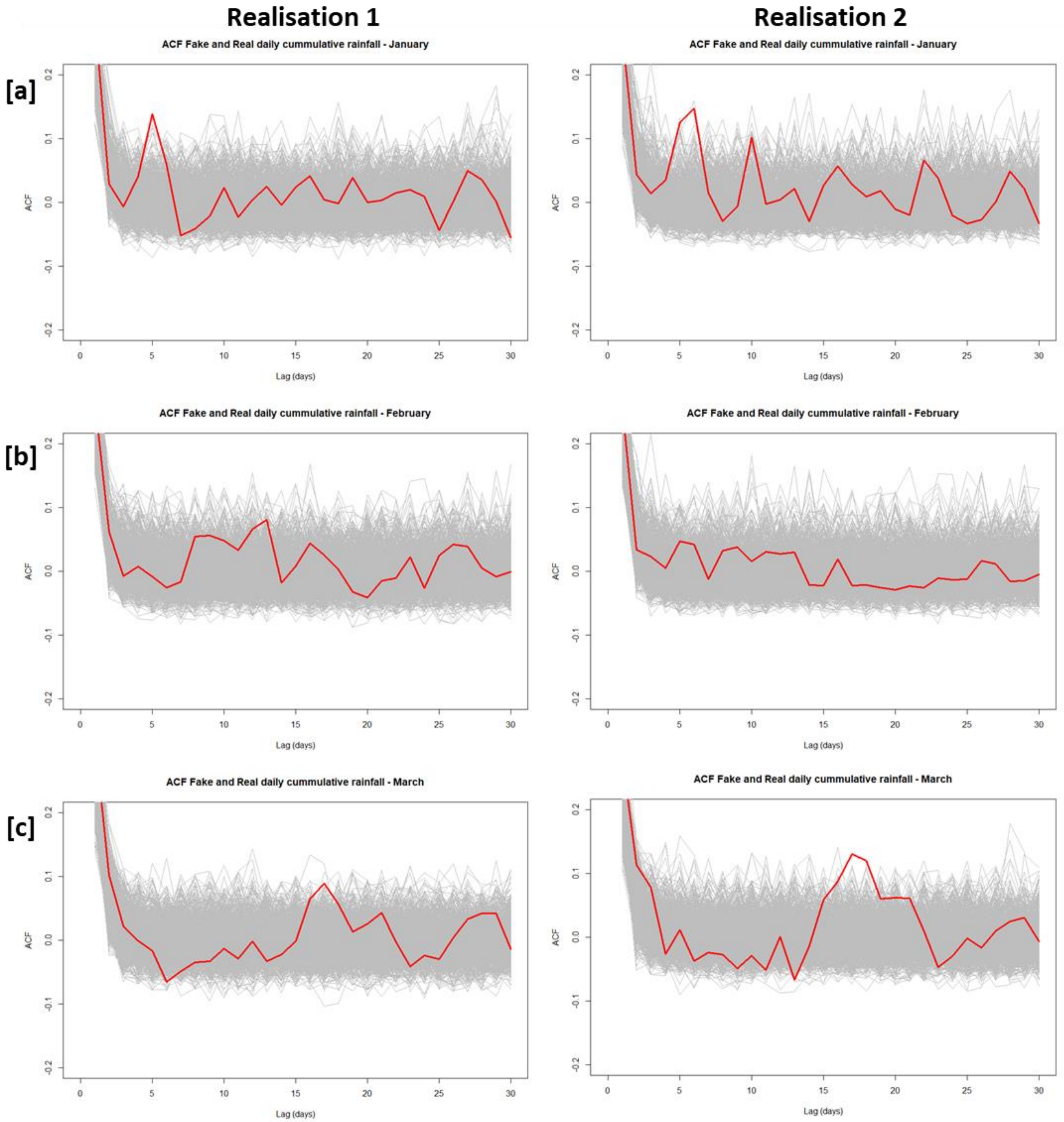


Figure S2: Autocorrelation Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (a) January, (b) February, and (c) March.

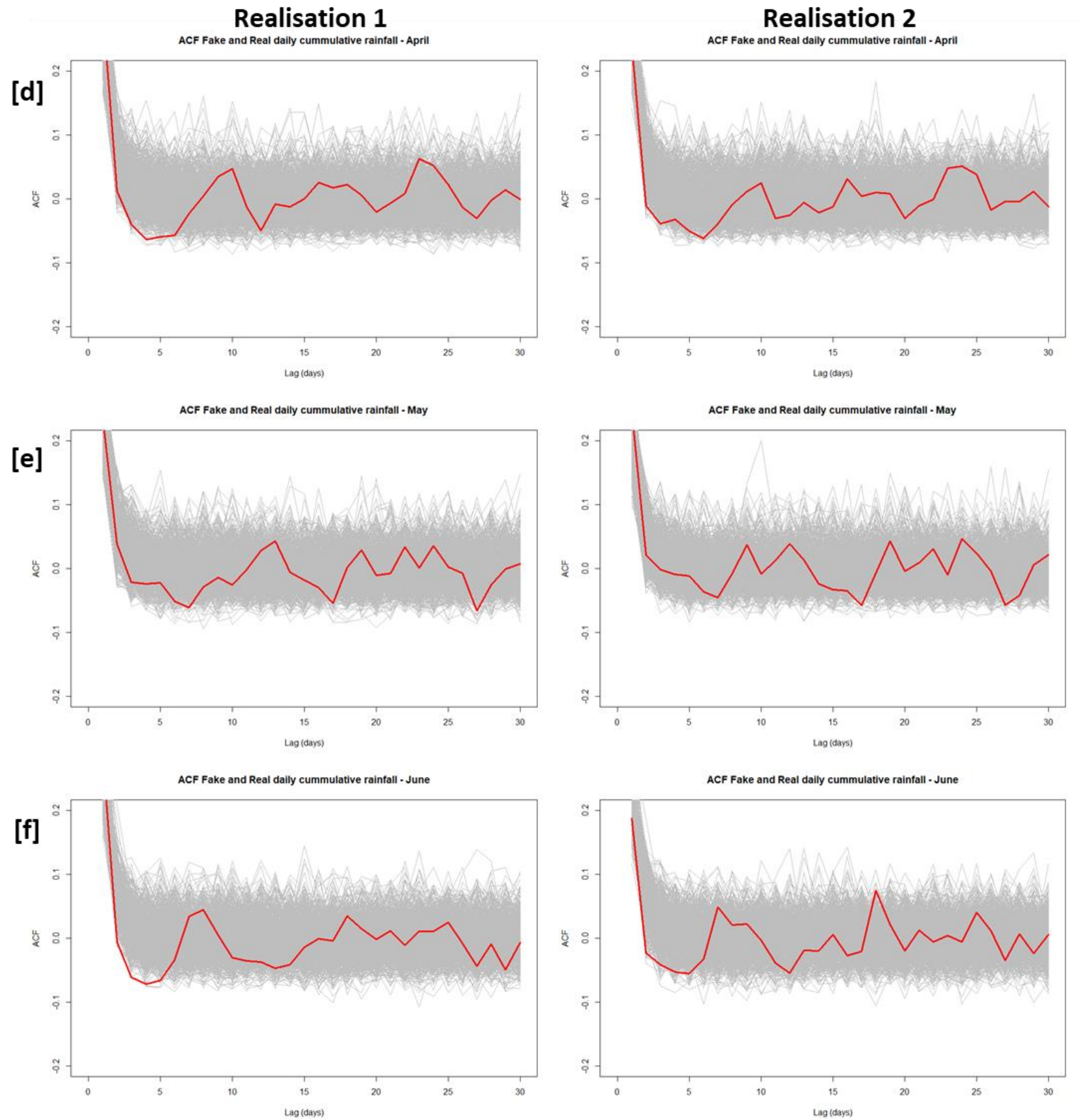


Figure S2: Autocorrelation Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2,* for (d) April, (e) May, and (f) June.

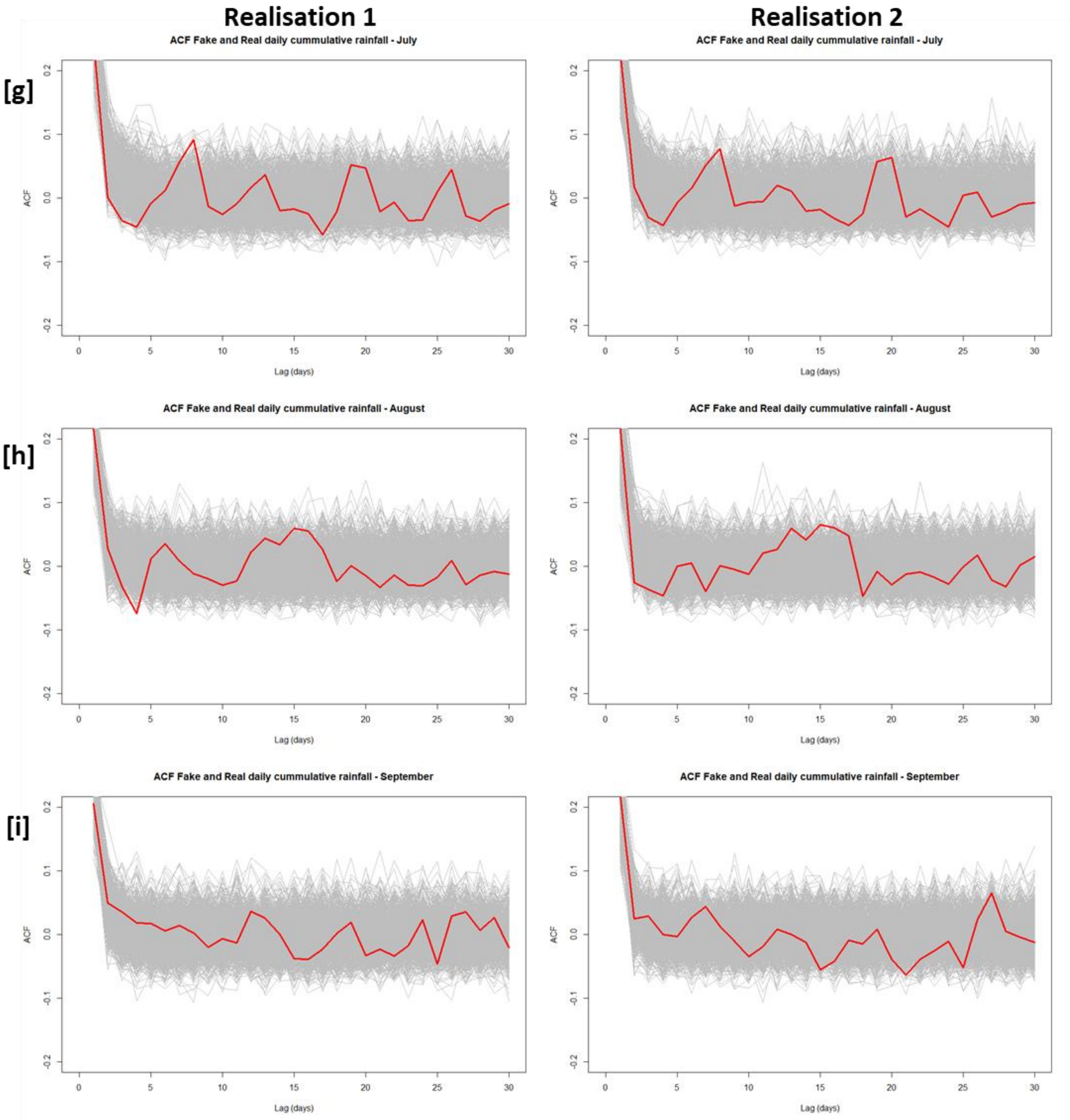


Figure S2: Autocorrelation Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (g) July, (h) August, and (i) September.

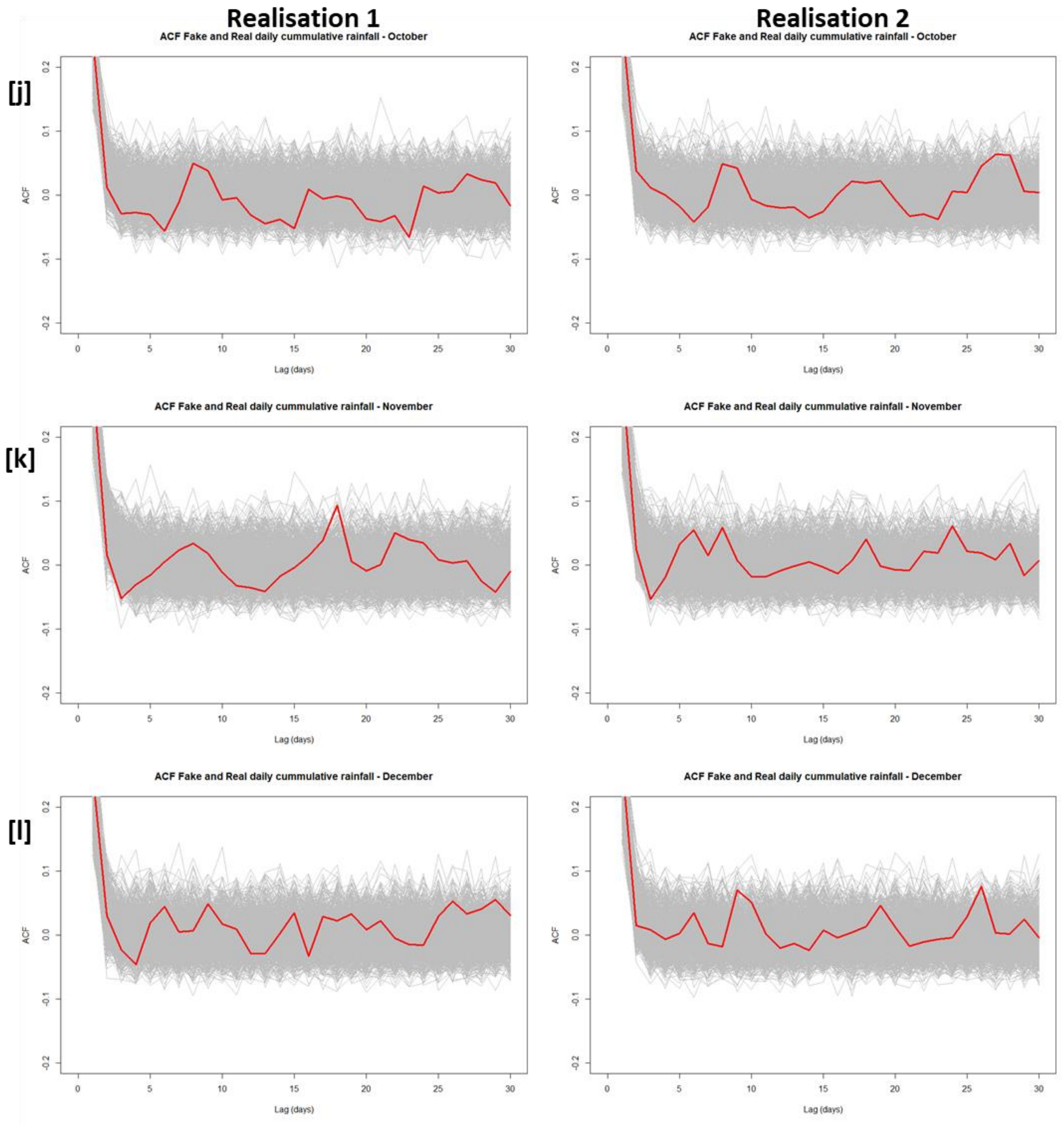


Figure S2: Autocorrelation Functions for simulated (grey) and ERA5-land (red) data for *Realisation 1* and *Realisation 2*, for (j) October, (k) November, and (l) December.

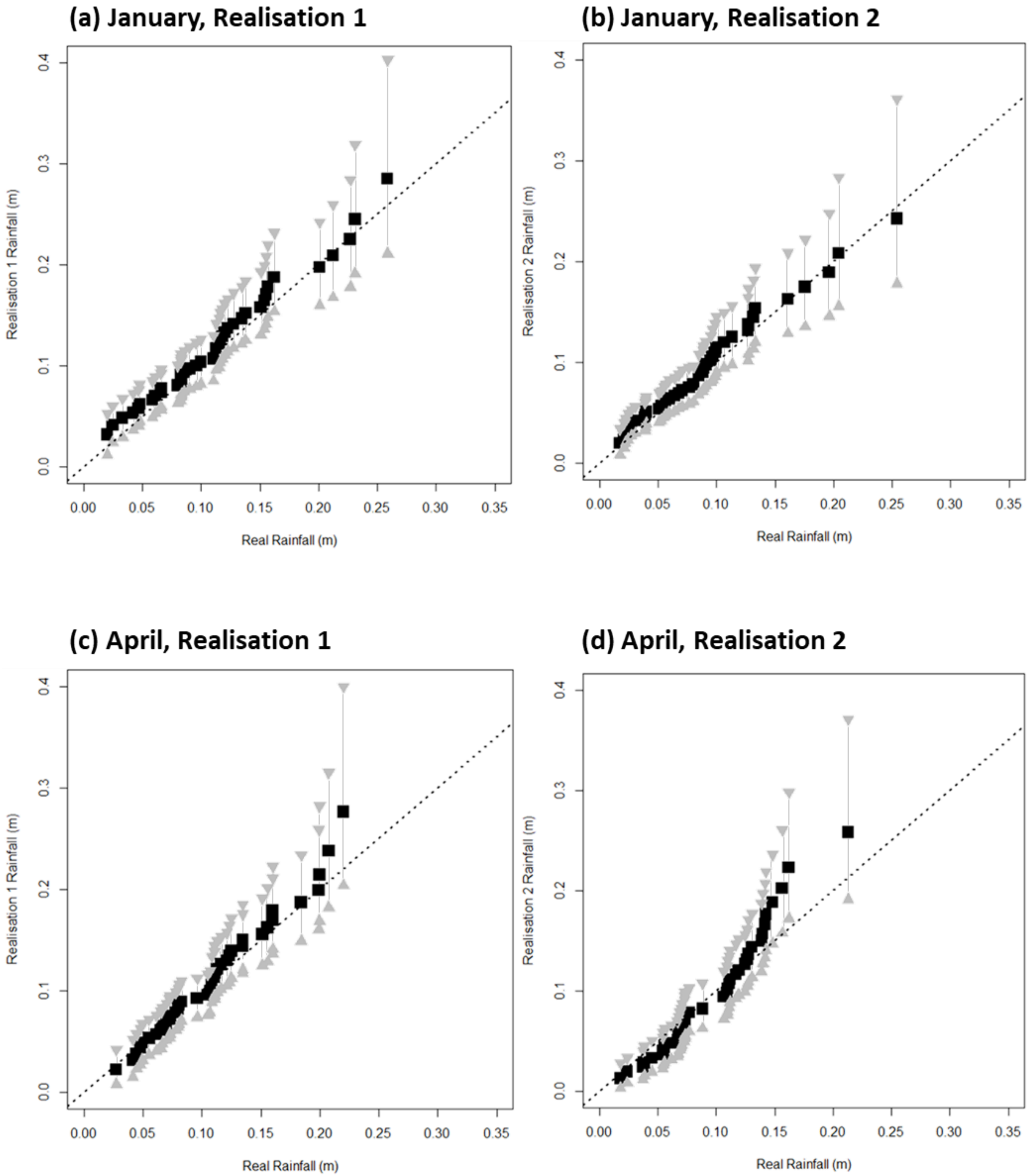


Figure S3: Comparison of observed (ERA5-land) rainfall quantiles (x-axis) and simulated rainfall quantiles (y-axis) for January (a), (b), and April (c), (d) monthly rainfalls. Black squares represent median of 999 simulations for each realisation, with grey arrows representing 95th percentiles.

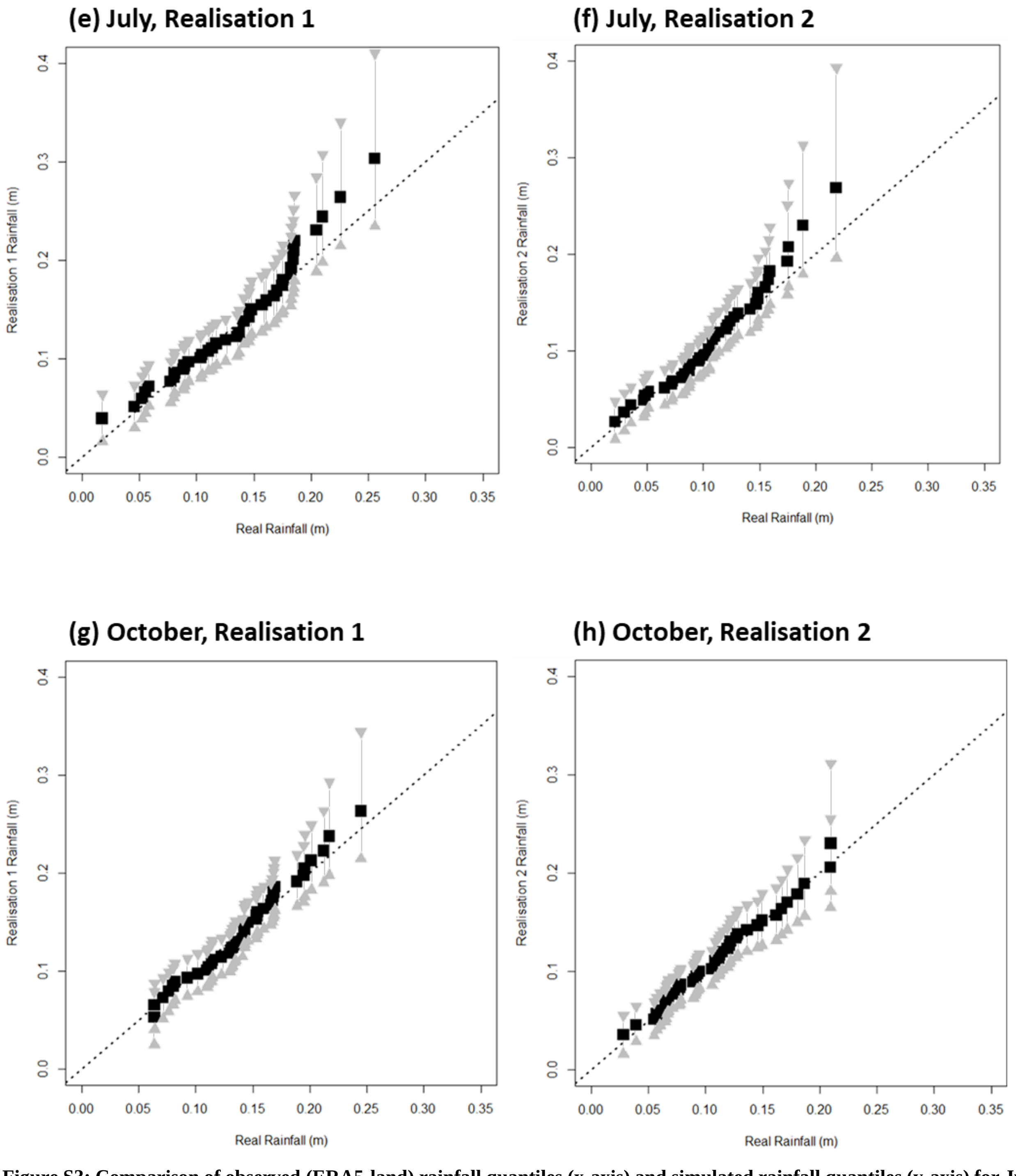


Figure S3: Comparison of observed (ERA5-land) rainfall quantiles (x-axis) and simulated rainfall quantiles (y-axis) for July (e), (f), and October (g), (h) monthly rainfalls. Black squares represent median of 999 simulations for each realisation, with grey arrows representing 95th percentiles.

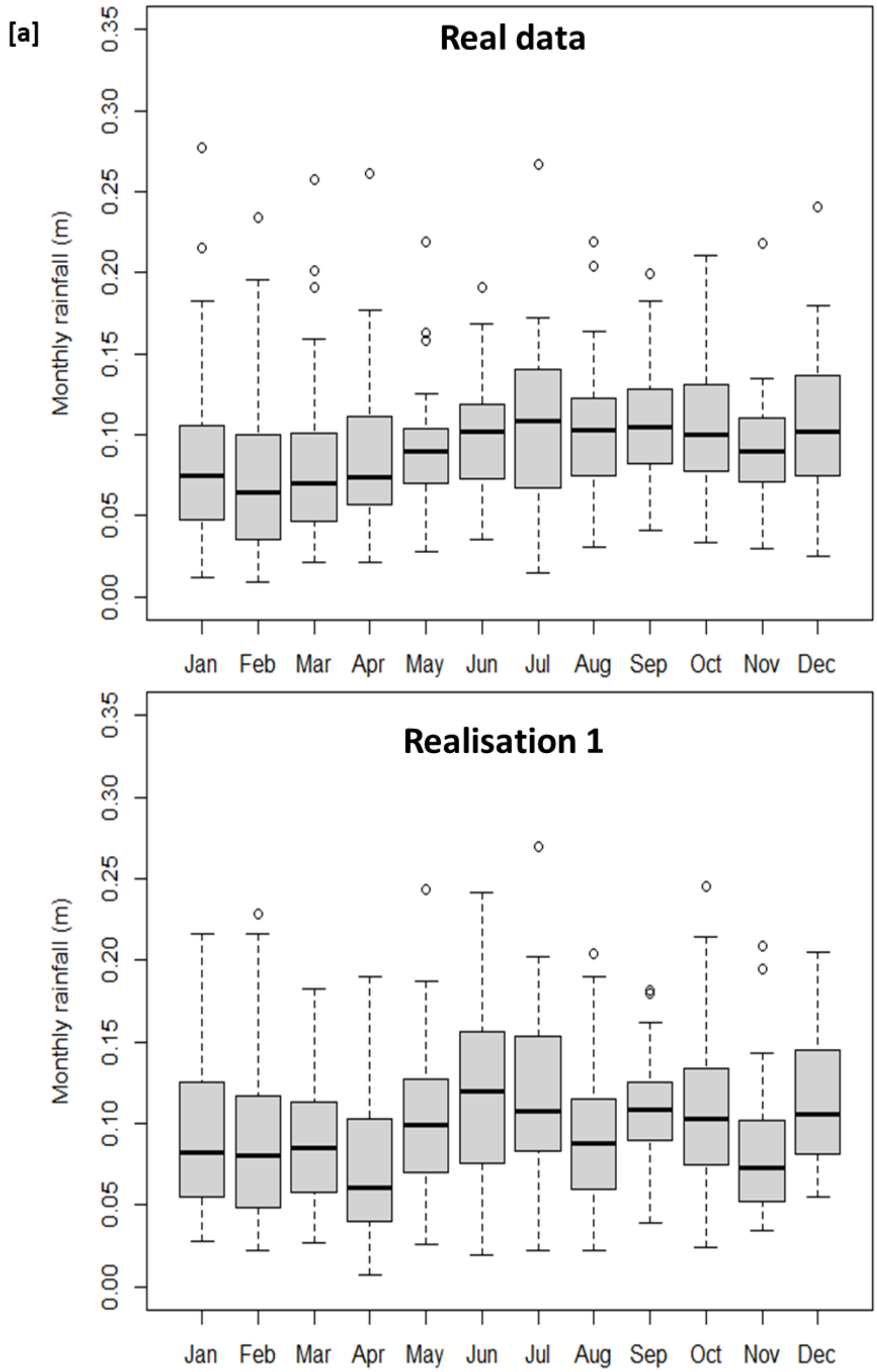


Figure S4: Comparison of monthly rainfall totals, (a) ERA5-land data and Realisation 1, comparison location shown as white box in Figure 3 (main text).

[b]

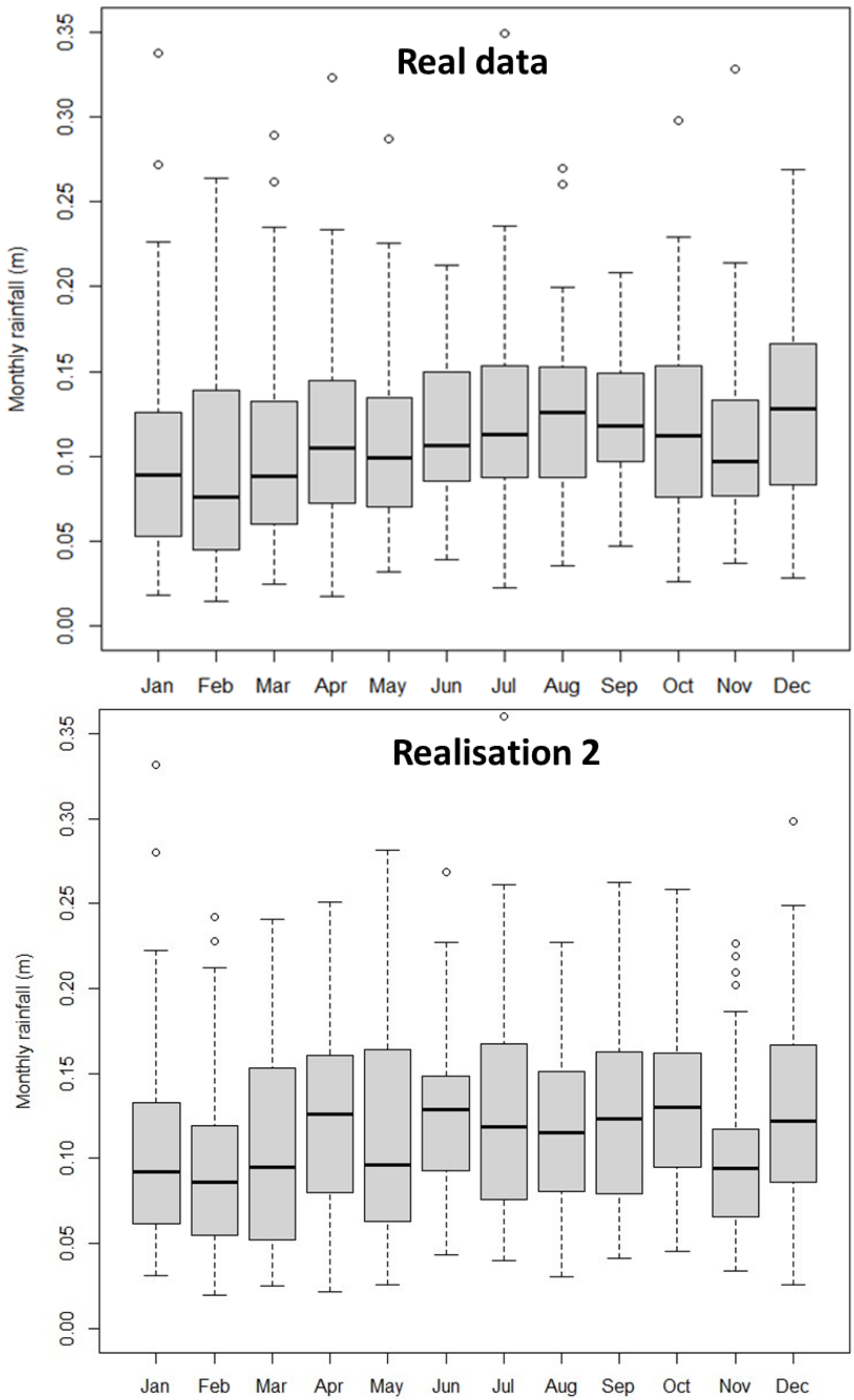


Figure S4: Comparison of monthly rainfall totals, (b) ERA5-land data and Realisation 2, comparison location shown as pink box in Figure 3 (main text).

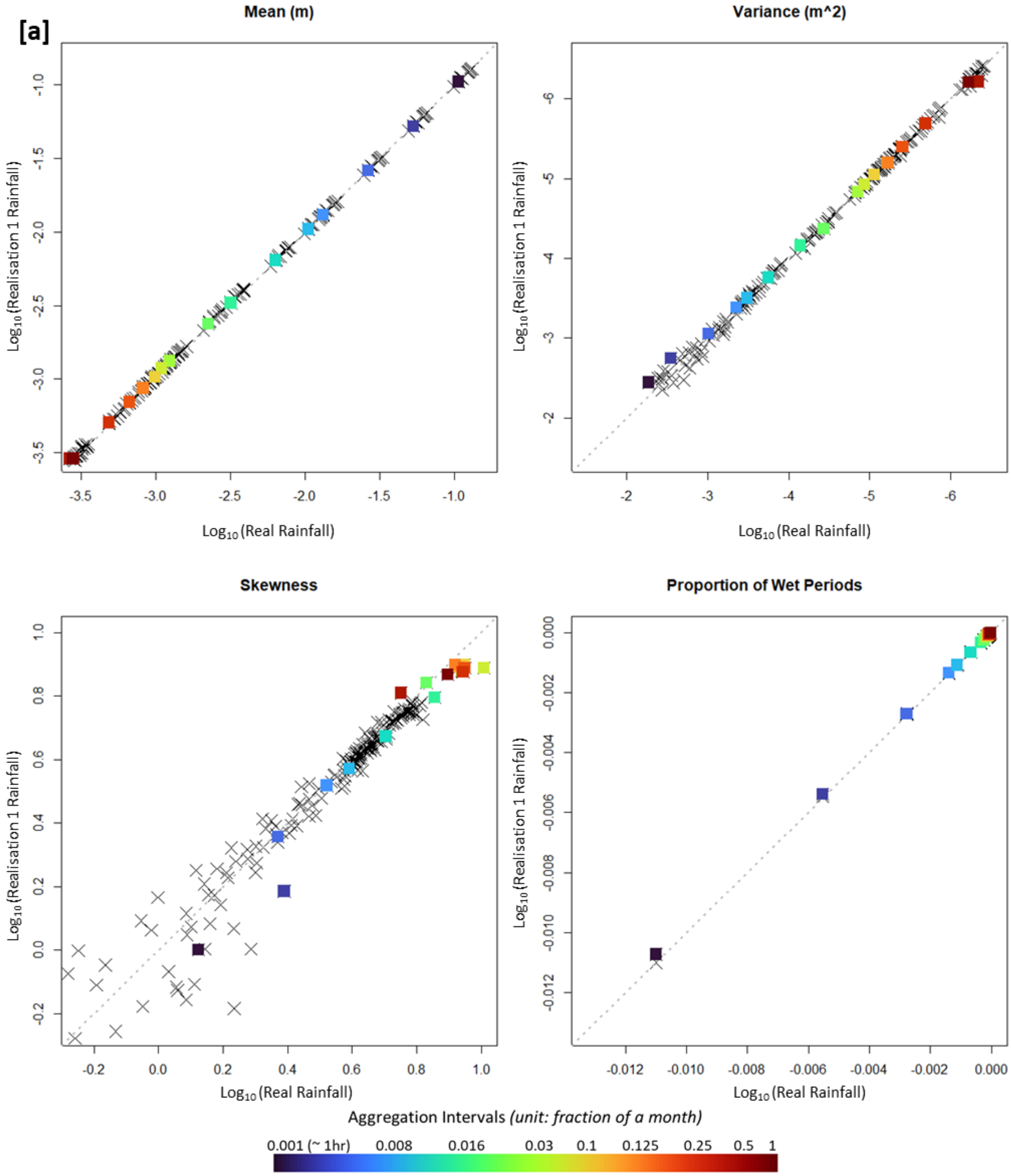
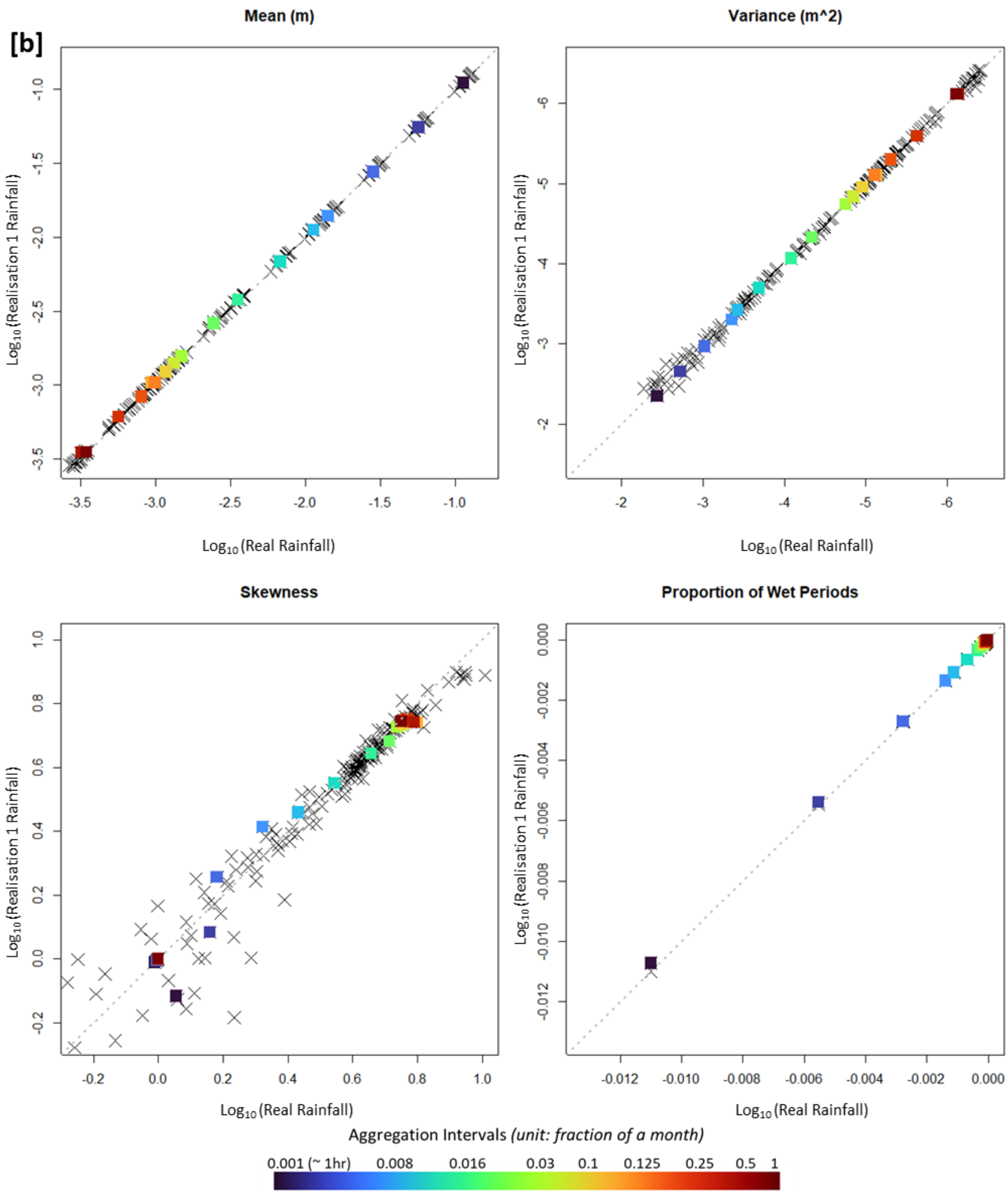


Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (a) ERA5-land data and Realisation 1. All data shown (grey crosses), January presented in aggregation intervals as fractions of a month as coloured squares.



110

Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (b) ERA5-land data and Realisation 1. All data shown (grey crosses), April presented in aggregation intervals as fractions of a month as coloured squares.

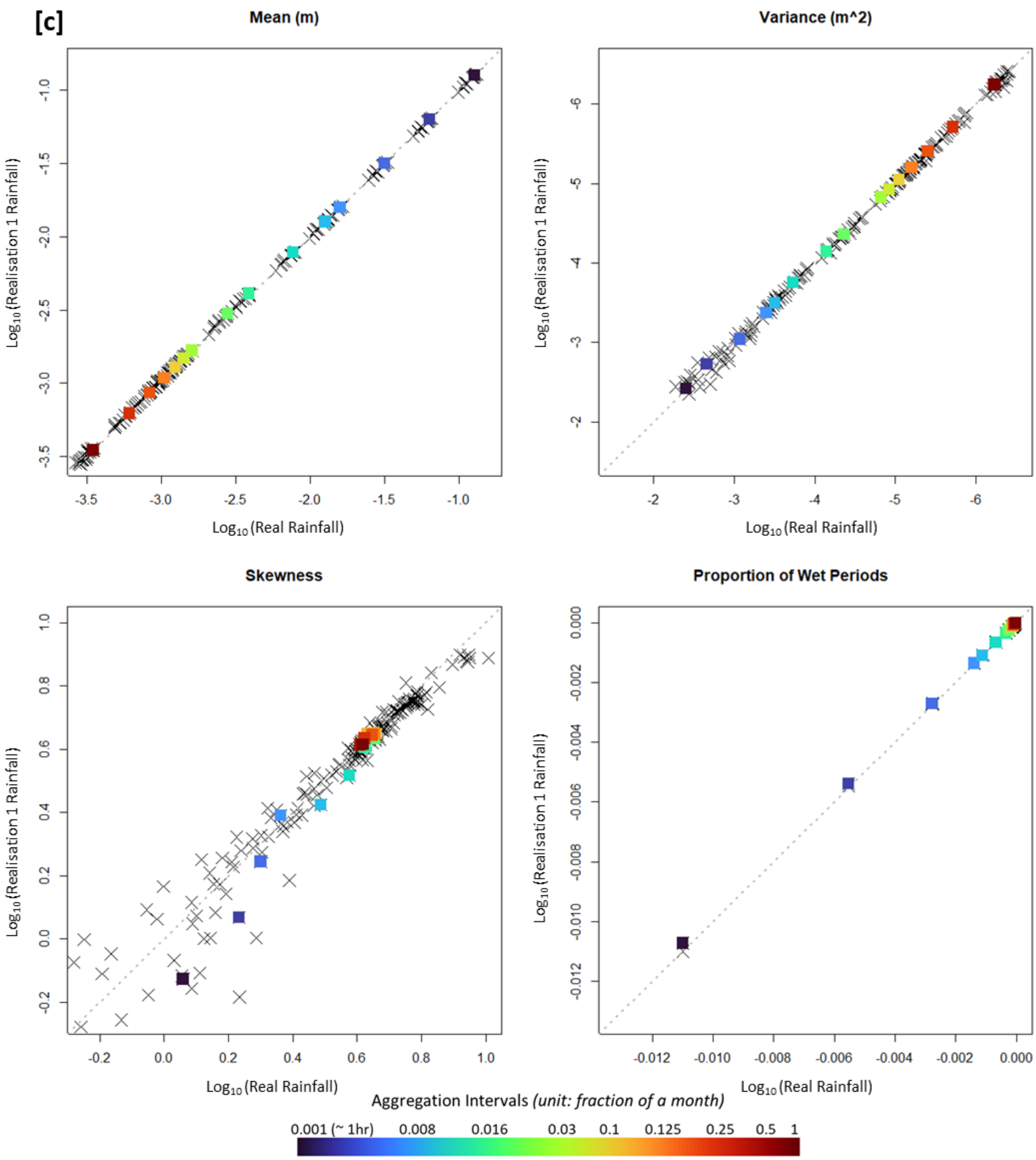
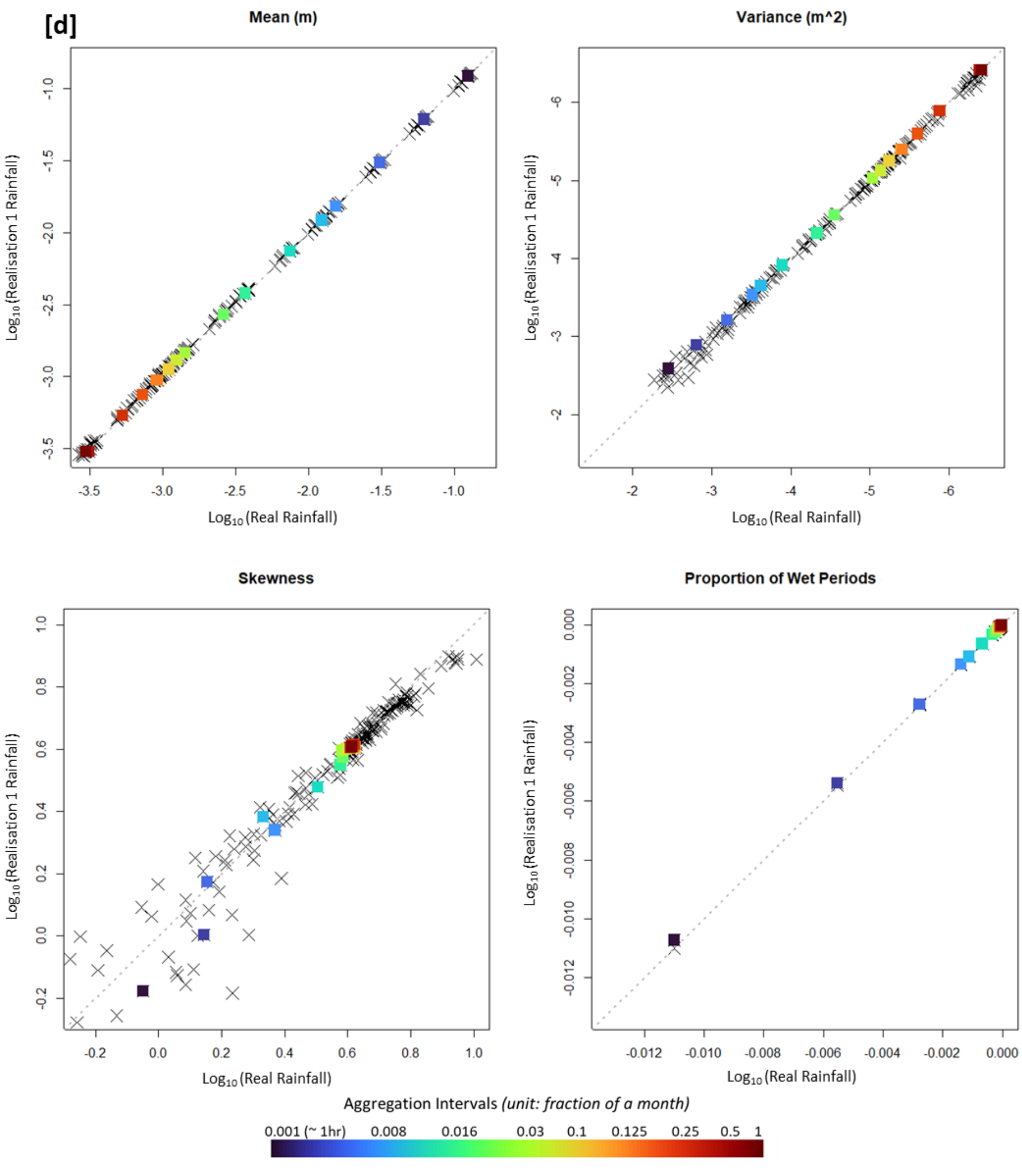


Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (c) ERA5-land data and Realisation 1. All data shown (grey crosses), July presented in aggregation intervals as fractions of a month as coloured squares.



125 **Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (d) ERA5-land data and Realisation 1. All data shown (grey crosses), October presented in aggregation intervals as fractions of a month as coloured squares.**

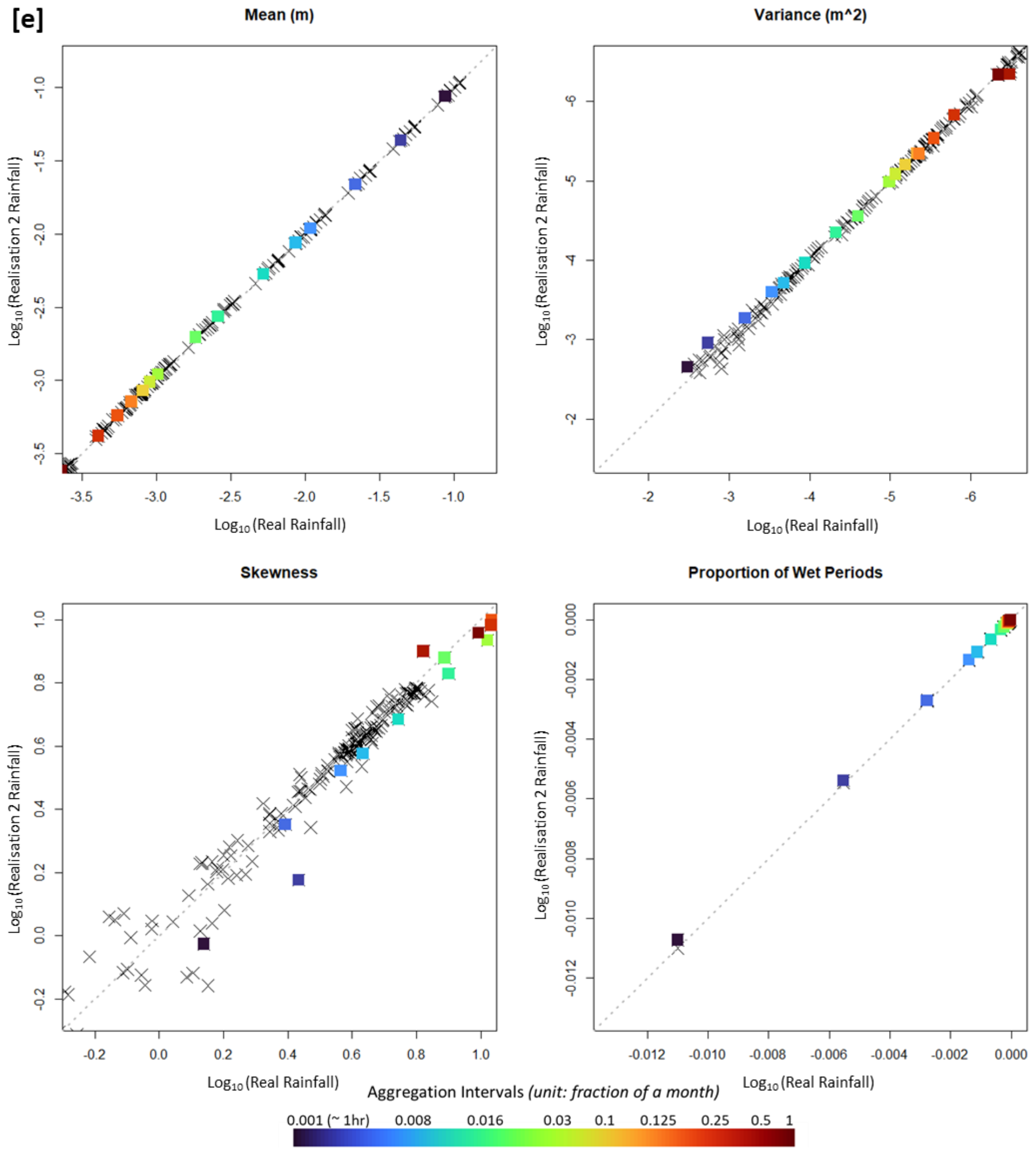
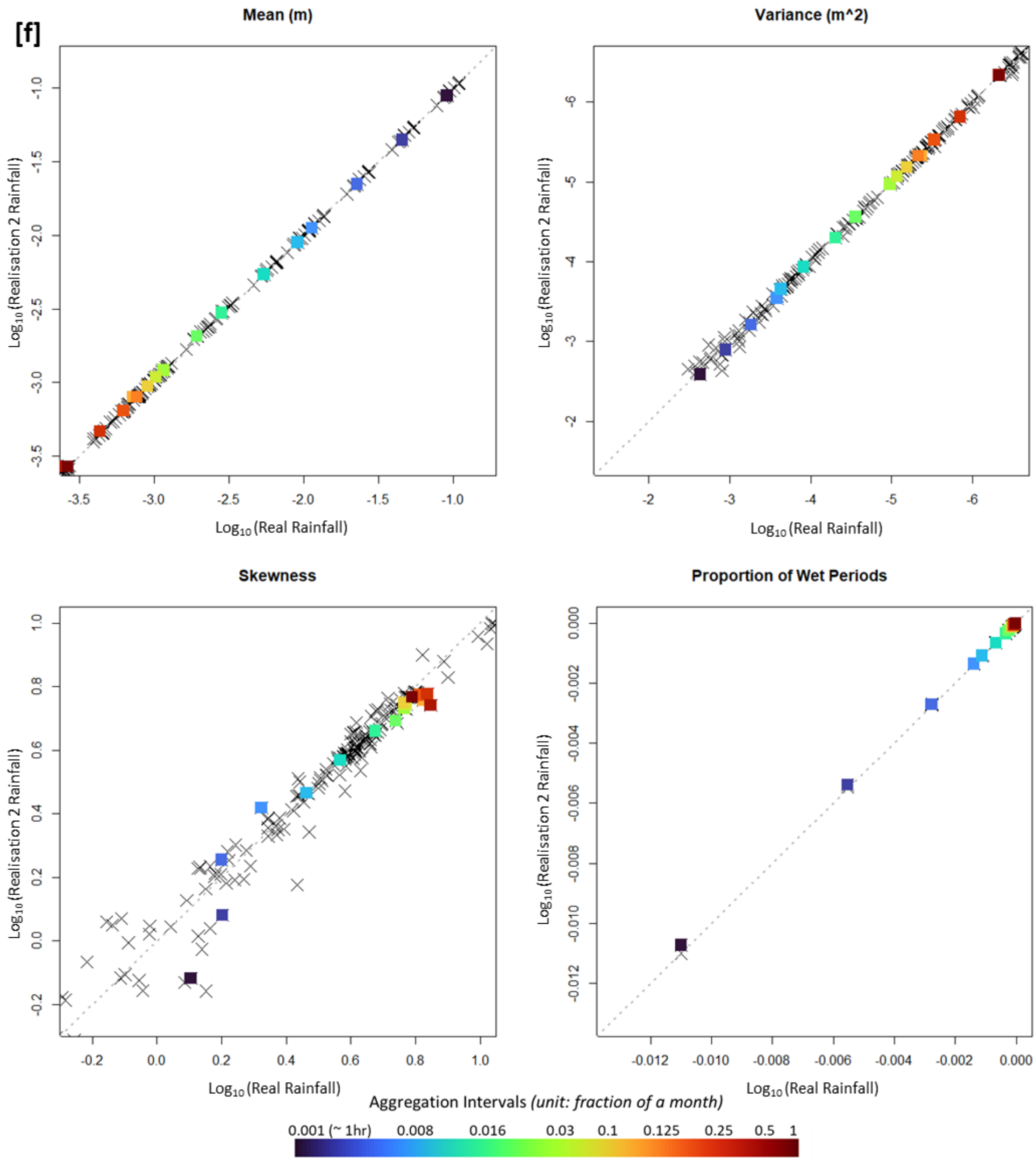


Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (e) ERA5-land data and Realisation 2. All data shown (grey crosses), January presented in aggregation intervals as fractions of a month as coloured squares.



140

Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (f) ERA5-land data and Realisation 2. All data shown (grey crosses), April presented in aggregation intervals as fractions of a month as coloured squares.

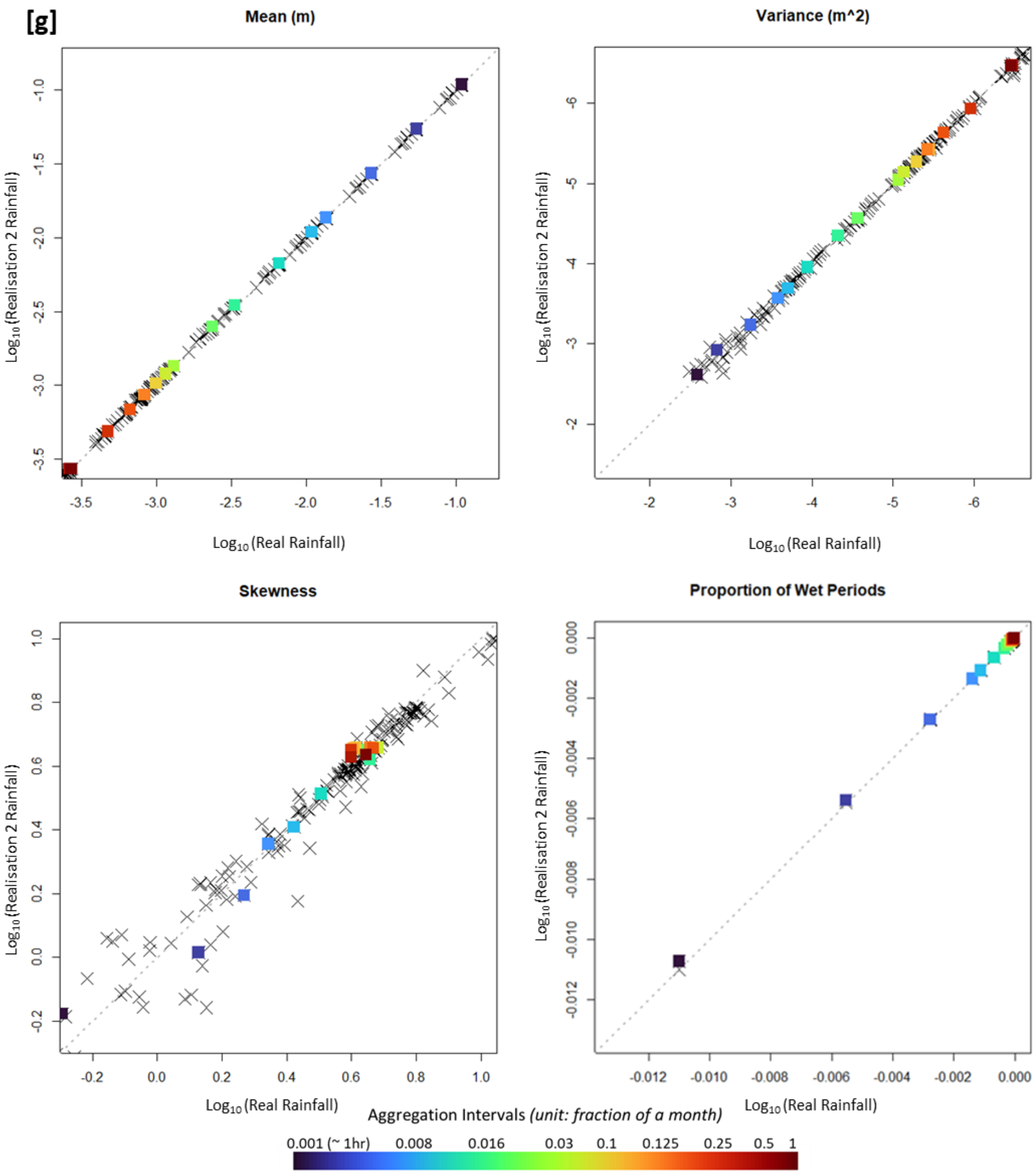


Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (g) ERA5-land data and Realisation 2. All data shown (grey crosses), July presented in aggregation intervals as fractions of a month as coloured squares.

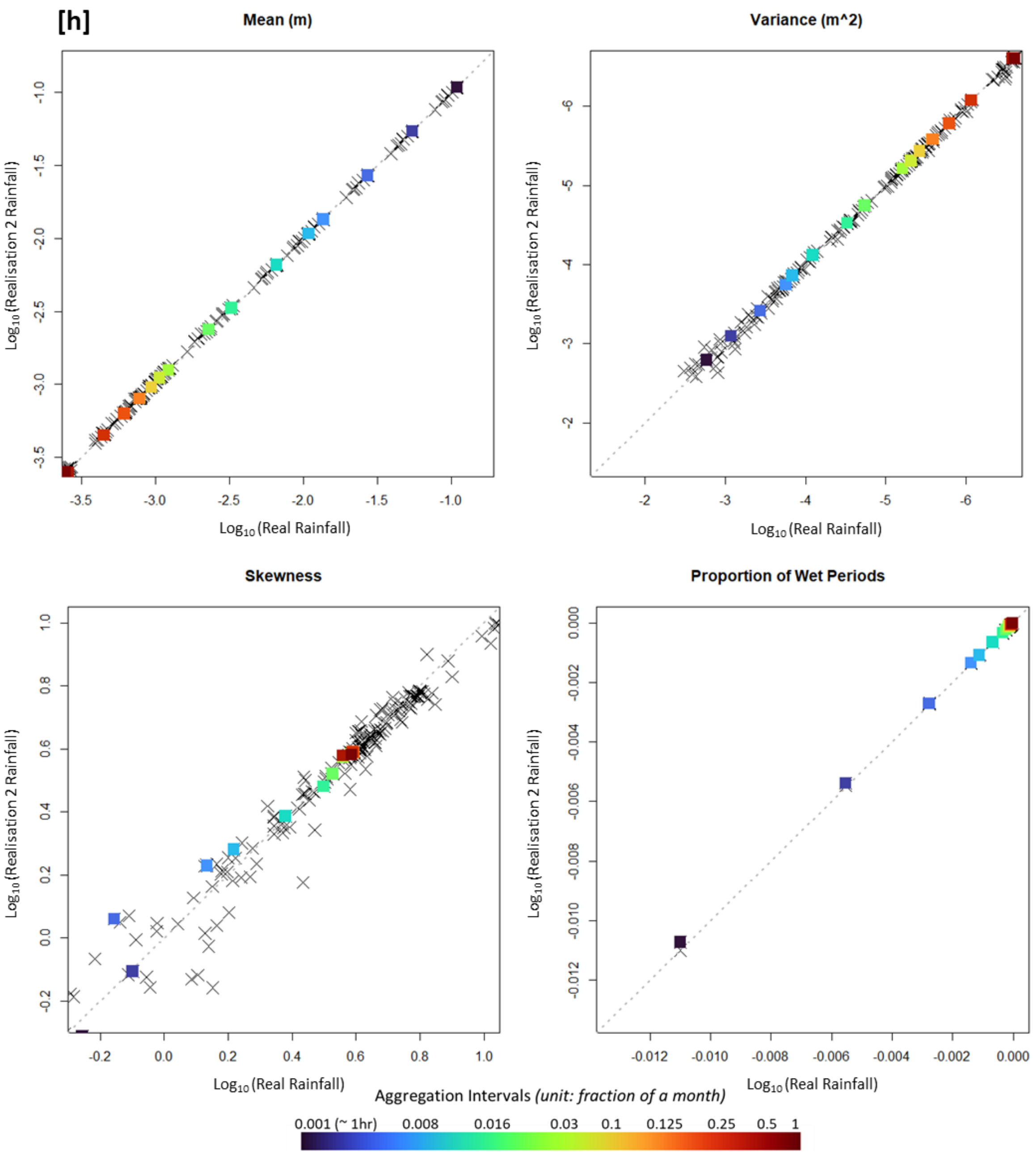


Figure S5: Mean, variance, skewness, and proportion of wet periods of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods (> (1/24)/1000 m), (h) ERA5-land data and Realisation 2. All data shown (grey crosses), October presented in aggregation intervals as fractions of a month as coloured squares.

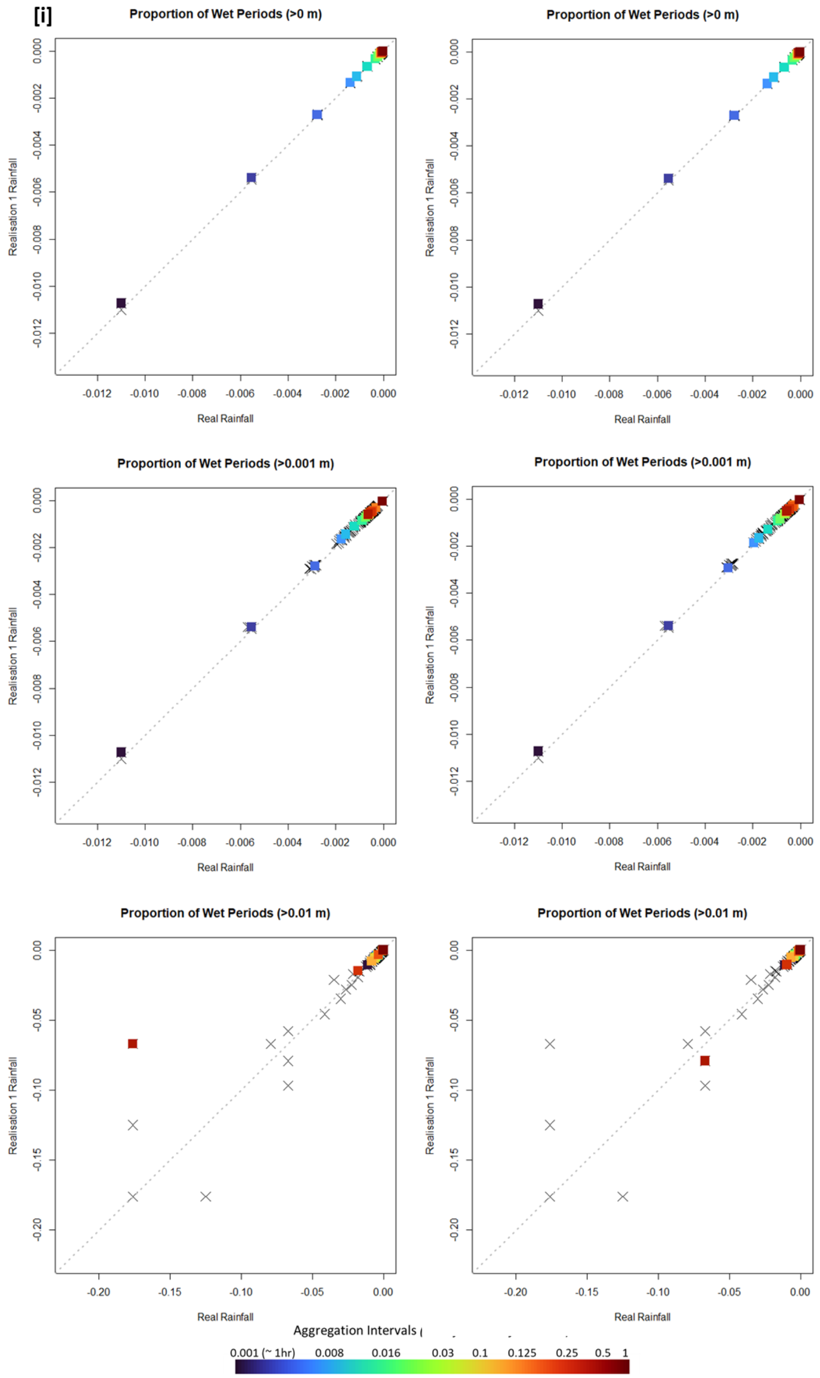


Figure S5: Variation in proportion of wet periods by with change in "wet" definition of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods, (i) ERA5-land data and Realisation 1. All data shown (grey crosses), January (left column), April (right column), presented in aggregation intervals as fractions of a month as coloured squares, Proportions (top to bottom), are {> 0 m, > 0,001 m, > 0.01 m}.

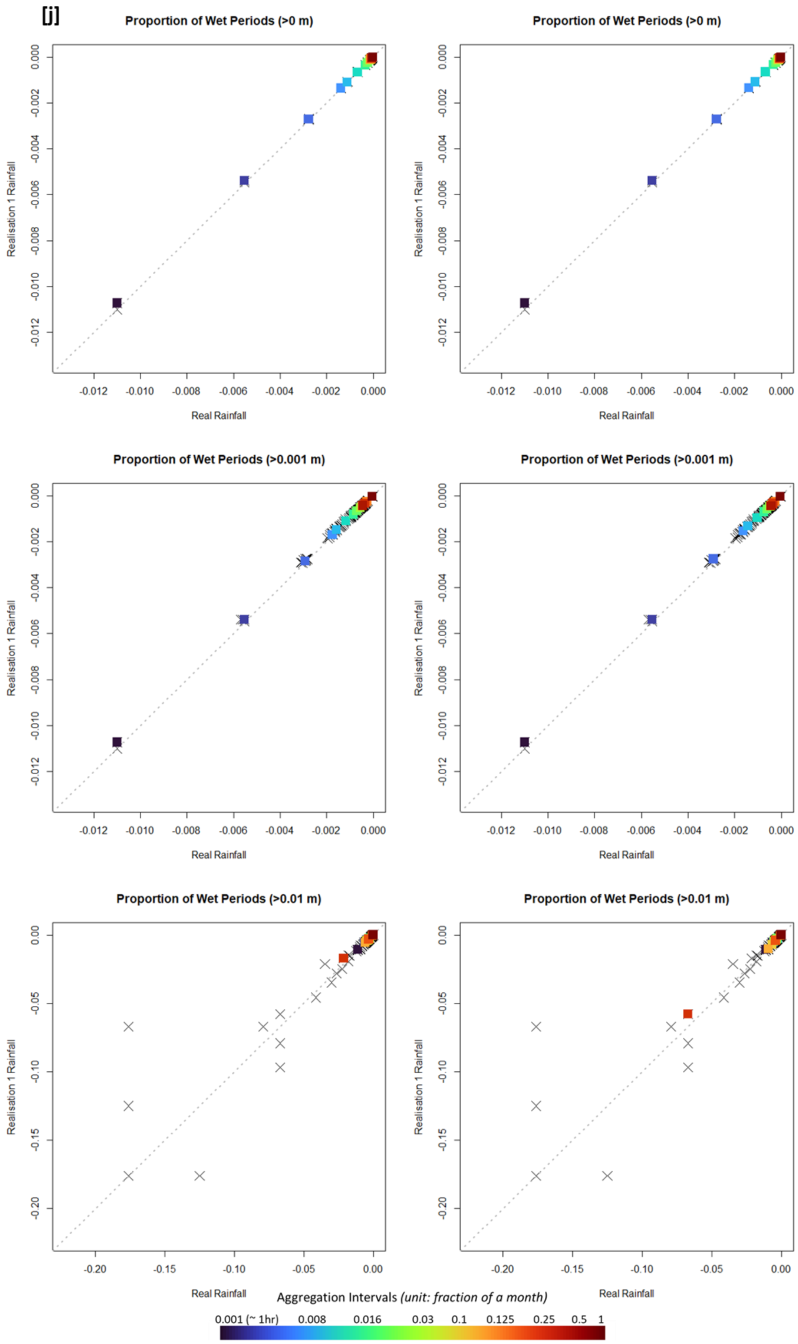


Figure S5: Variation in proportion of wet periods by with change in "wet" definition of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods, (j) ERA5-land data and Realisation 1. All data shown (grey crosses), July (left column), October (right column), presented in aggregation intervals as fractions of a month as coloured squares, Proportions (top to bottom), are {> 0 m, > 0,001 m, > 0.01 m}.

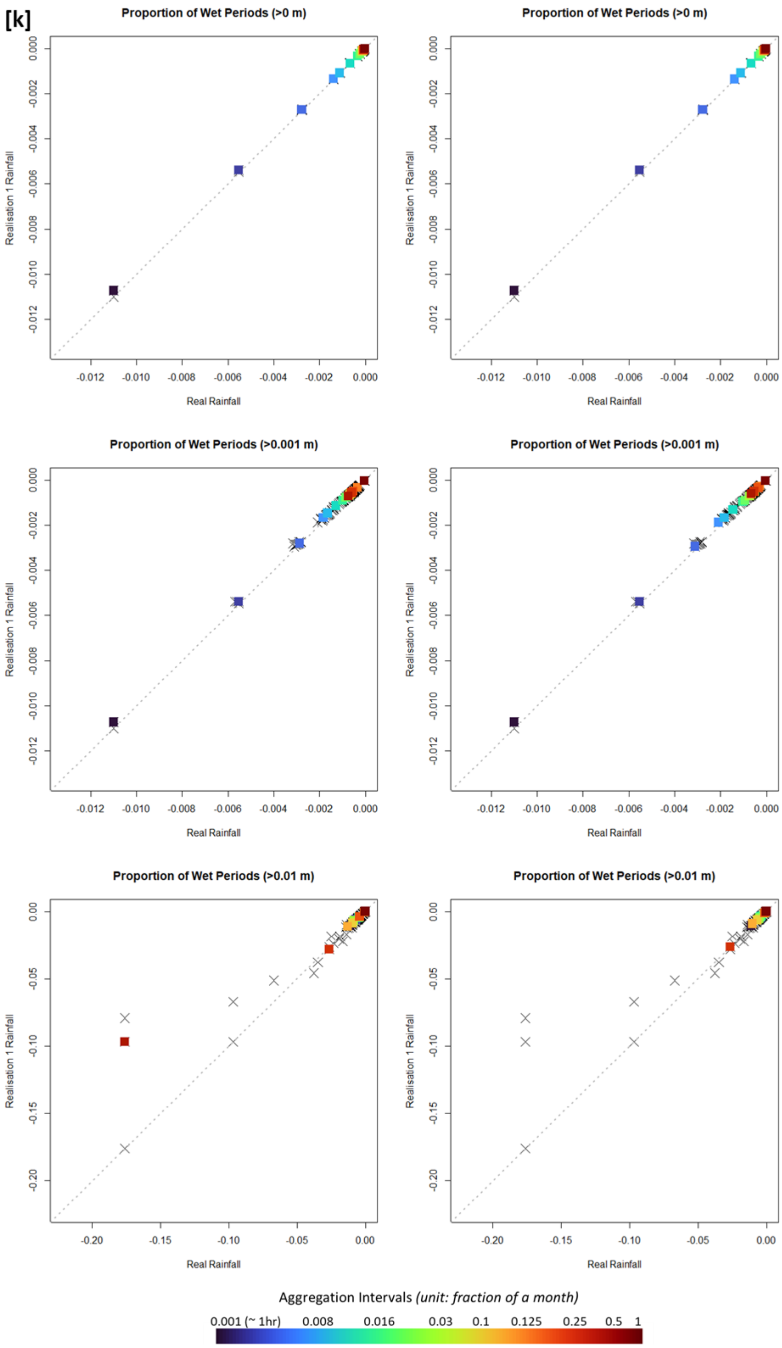


Figure S5: Variation in proportion of wet periods by with change in "wet" definition of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods, (k) ERA5-land data and Realisation 2. All data shown (grey crosses), January (left column), April (right column), presented in aggregation intervals as fractions of a month as coloured squares, Proportions (top to bottom), are {> 0 m, > 0,001 m, > 0.01 m}.

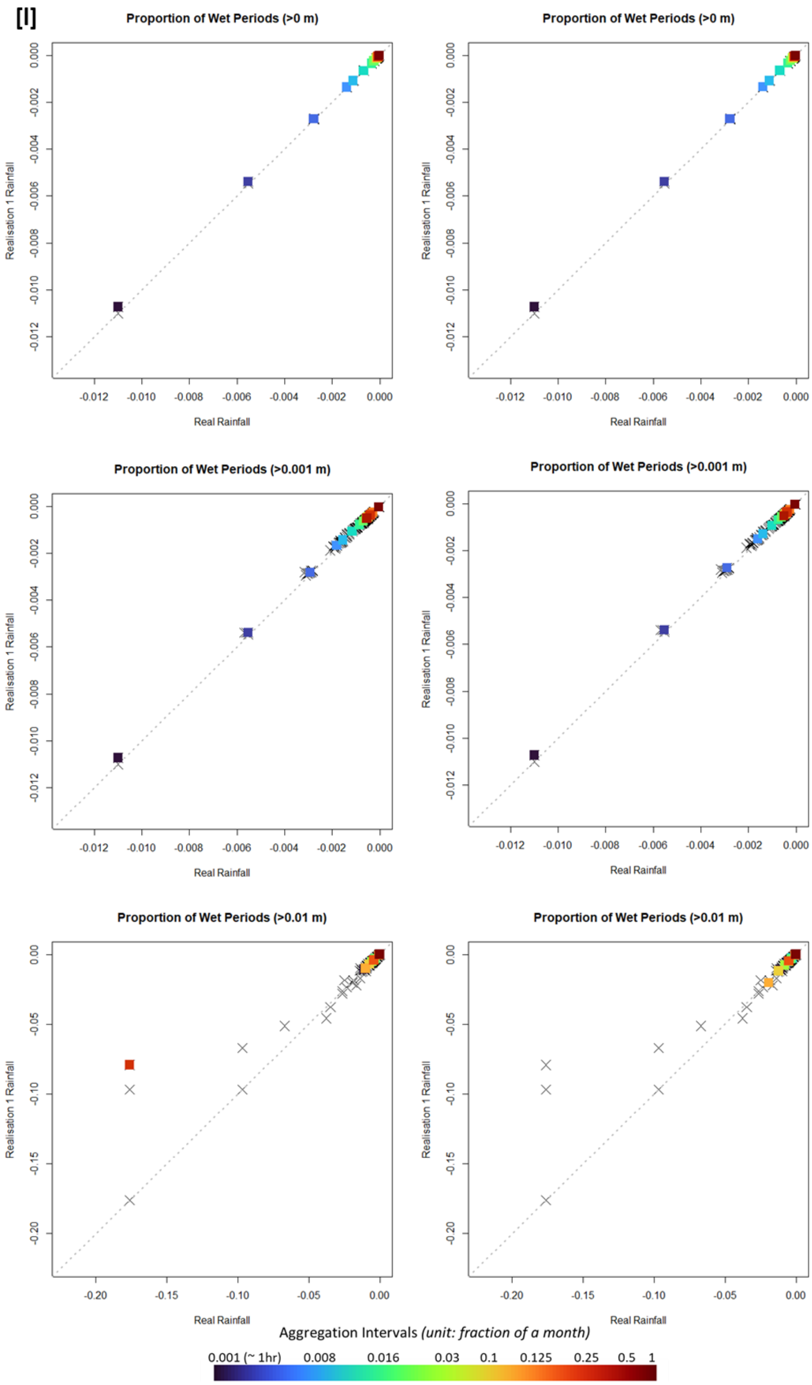


Figure S5: Variation in proportion of wet periods by with change in "wet" definition of the observed (x-axis) and simulated (y-axis) rainfall time series wet periods, (l) ERA5-land data and Realisation 2. All data shown (grey crosses), July (left column), October (right column), presented in aggregation intervals as fractions of a month as coloured squares, Proportions (top to bottom), are {> 0 m, > 0,001 m, > 0.01 m}.

References

Fox. J. and Weisberg S.: An R Companion to Applied Regression, Third edition, Sage Publications, Thousand Oaks CA, USA, 576 pp., https://socialsciences.mcmaster.ca/jfox/Books/Companion/, 2019.

Royston, P.: An extension of Shapiro and Wilk's W test for normality to large samples. Appl. Stat., 31, 115–124, https://doi.org/10.2307/2347973, 1982.

Student.: The probable error of a mean, Biometrika, 1–25, 1908.

Venables, W.N. and Ripley, B. D.: Modern Applied Statistics with S. Fourth Edition, Springer Series in Statistics and Computing, Springer-Verlag, New York, NY, 498 pp., https://doi.org/10.1007/978-0-387-21706-2, 2002.