Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Satellite-Based Environmental Monitoring of Rainfall Accumulation and Joint Low Surface-Water Conditions in Karnataka Lakes, India

Anusha James

anusha.james@christuniversity.in

Christ University

Syam Mohan E

Christ University

Deepa B G

Christ University

Research Article

Keywords: environmental monitoring, satellite remote sensing, lake surface-water extent, CHIRPS precipitation, rainfall accumulation, copula modelling, Karnataka

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11626167/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Satellite-Based Environmental Monitoring of Rainfall Accumulation and Joint Low Surface-Water Conditions in Karnataka Lakes, India

Anusha James[1](Corresponding Author), Syam Mohan E[2], and Deepa B G[3]

[1] CHRIST (Deemed to be University), Bengaluru, Karnataka, India. Email: syam.mohan@christuniversity.in

[2] CHRIST (Deemed to be University), Bengaluru, Karnataka, India. Email: anusha.james@christuniversity.in

[3] CHRIST (Deemed to be University), Bengaluru, Karnataka, India. Email: deepa.bg@christuniversity.in

Corresponding author: Anusha James (anusha.james@christuniversity.in)

Abstract

Lake surface water is dynamic with rainfall and also due to hydrological processes influencing over various time scales. The present study investigates the relationship between rainfall and surface water of the 25 lakes of Karnataka, India, which were selected by stratified sampling, during 2000-2021. Using monthly JRC Global Surface Water observations and CHIRPS rainfall. The main analysis focuses on October–May, following a minimum observation coverage rule, and is based on 2,974 lake-month observations. Spearman rank correlation is used to compare monthly anomalies and rainfall totals with each other over 1, 3 and 6 months. Pseudo-observations of lake levels are better fitted by five bivariate copula families: Gaussian, Student-t, Clayton, Frank and Gumbel. Assessed on the basis of information criteria, the in-sample Cramér–von Mises discrepancy and the error in the lower-tail probability for the joint distribution. A longer time window (six months) gives a better rank association than shorter time windows: median paired Spearman differences are 0.206 for one month and 0.164 for three months. The median value of the lower-tail amplification is 1.708 at the 20th percentile; and 19 out of 25 lakes had amplifications greater than the independence threshold. The models chosen for 10, 7, 6, 1 and 1 lakes are the Frank, Gaussian, Student-t, Gumbel models, respectively. But in 17 lakes the difference between the best and second best models is less than two for the AIC. The overall insample discrepancy for 22 lakes is improved by fitted copulas, while the lower tail probability error is only improved by 16 lakes. Association with cumulative rainfall is found to vary among lakes and joint lows are an actual observed phenomenon in the observed months. Results cannot be readily extrapolated due to missing data from satellites, spatial and temporal dependence, and model evaluation in the sample. The monitoring relevance is associated with identifying the differences in the co-variation of rainfall and surface-water outcomes for each lake, as well as distinguishing observed joint low condition frequency from model-derived estimates; they are exploratory and not validated operational drought warnings.

Keywords: environmental monitoring; satellite remote sensing; lake surface-water extent; CHIRPS precipitation; rainfall accumulation; copula modelling; Karnataka

1. Introduction

Water bodies in the inland regions connect climate variability and ecosystem, local livelihoods and water management. Rainfall, catchment runoff, storage in the lake, evaporation from the water surface, ground water exchange, and human affects all influence the gain or loss of water in the water balance of the mapped extent of a lake and affect the extent of the map area. The relationship between rainfall and observed water in the lake is therefore a geographical problem, as well as a hydrological one. Requires observations which are comparable from place to place, an appropriate time-scale, and the ability to distinguish between average association and joint behaviour where conditions are abnormally low.

This can now be compared across regions as there are long satellite records. Pekel et al. (2016) created a global surface-water record from Landsat data that provides high-resolution surface water occurrence and change data. This time series is valuable when there is no consistent field data for several lakes. It has been noted, however, that its monthly observations are dependent upon acquisition and classification by successful satellite observations, and that the absence of an observation does not imply absence of water. For regions with monsoons, observation coverage may vary from one season to another, and this is indeed a sampling problem. Satellite monitoring of wetlands is also feasible in India (Alam et al., 2024), and multi-source datasets of lake-extent provide complementary opportunities to compare and contrast changes in surface-water mapped from satellites (Bai et al., 2025).

The latest study validates the separation of real variability of surface water from the lack of observation. Hao et al. (2024) proposed a technique to deal with missing monthly surface-water observations, and Xiao et al. (2024) explored the dynamics of intermittent water bodies. All of these studies suggest that the presence change and representation of observations in long-term monitoring via satellites needs to be explicitly accounted for.

Lakes are also not constant map features, but rather water stores that vary over time, as seen from recent satellite studies. Yao et al. (2023) found that loss of water storage in lakes is widespread across the globe, and that both climate and human factors play a role. Regional monitoring cannot be done based on these global findings and it is not reflected that the Karnataka lakes sampled in this study exhibited similar storage trends.

Lake monitoring has also been given a new perspective in the context of water resources in recent times. Changes in reservoir storage on a global scale were documented by Li et al. (2023), and Li and Rodell (2024) documented large anomalies in TWS on the terrestrial scale. The results shown here warrant further observation but cannot measure extent anomaly of surface water – it only measures extent anomaly, it does not measure water storage or water availability.

A matching climate record is provided from gridded rainfall estimates. Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) are derived from satellite data and gauge data and are used for climate monitoring and drought assessment (Funk et al., 2015). The combination of CHIRPS with satellite surface-water observations can be used to perform a regional analysis using similar data sources. But rainfall and lake extents are not always correlated. Lags and persistence can be due to catchment storage, upstream inflow, delayed runoff, infiltration and regulated withdrawals. Precipitation for one month can thus not be used as an indicator of lake water; it should be considered in combination with precipitation over several previous months.

Satellite altimetry and optical remote sensing has also enabled the reconstruction of several decades of lake water levels (Yao et al., 2024). However, water level and volume of water stored differ from the extent of mapped surface-water. This study examines the third of these, and does not estimate volume of the lake.

General association measures provide a good basis. Spearman correlation is used for monotonic rank association and does not assume a linear relationship, but the relationship of all the joint outcomes cannot be defined by a single coefficient. For instance, a lake can have only a weak overall association, but a fairly high frequency of co-occurrence of low rainfall and low surface-water anomalies. Joint low conditions can be relevant to freshwater vulnerability. However, statistical co-occurrence cannot determine cause(s) or operational drought risk.

Copulas provide a flexible description of bivariate dependence, which is achieved by decomposing the bivariate dependence into the marginal behaviour of each component variable and their dependence structure (Nelsen, 2006). Copula inference for joint probabilities and multivariate extremes has been demonstrated to be valuable via hydrological studies (Genest & Favre, 2007; Salvadori et al., 2007). The different copula families correspond to different types of dependence and can be compared by likelihood based methods. In the case of candidate fits that are close, however, the uncertainty from model selection is important, and an overall better fit does not necessarily mean more accurate estimation of rare joint events.

Previous hydrological copula studies provide a direct precedent for assessing joint low condition events. The drought duration and drought severity were modeled by Shiau (2006) using two-dimensional copulas, and Shiau et al. (2007) used copulas for hydrological drought assessment. Sajeev et al. 2021 used copulas to calculate the drought characteristics at multiple scales in different climatic regions of India. Although the drought variables, settings, and models used in these studies are not the same as the rainfall–lake surface-water variables used here, such studies support the use of flexible dependence modeling.

The study of rainfall and lake water relationship is appropriate for Karnataka as water bodies are spread across different landscape types which are affected by the monsoon period. The state is not a single entity that is a hydrological system. This work employs a stratified sample of lakes and estimates dependence separately for each lake, instead. The emphasis is placed on the relationship of rainfall anomalies and surface-water anomalies observed via satellite. The study does not attempt to attribute changes in lake water solely or primarily to the influence of rainfall.

The study considers four questions: (i) is there a dependence family that appears suitable for lakes as a whole; (ii) is there a dependence family which is suitable for the individual lakes; (iii) is the dependence of low rainfall with low surface-water anomalies or independence better captured by fitted copulas; and (iv) is there a dependence family suitable for lakes as a whole and for each lake? The contribution is a lake-level comparison which combines the temporal accumulation, the copula selection, tail-specific validation, and explicit sensitivity and uncertainty checks.Three complementary diagnostic outputs are available from the analysis for environmental monitoring: comparative analysis of the antecedent rainfall windows, lake-specific empirical joint low-condition frequencies, and a test of whether the dependence models that are fitted for the window improve the probability of the event being monitored. These diagnostics do not form the basis for an attribution to cause or a validated early-warning system.

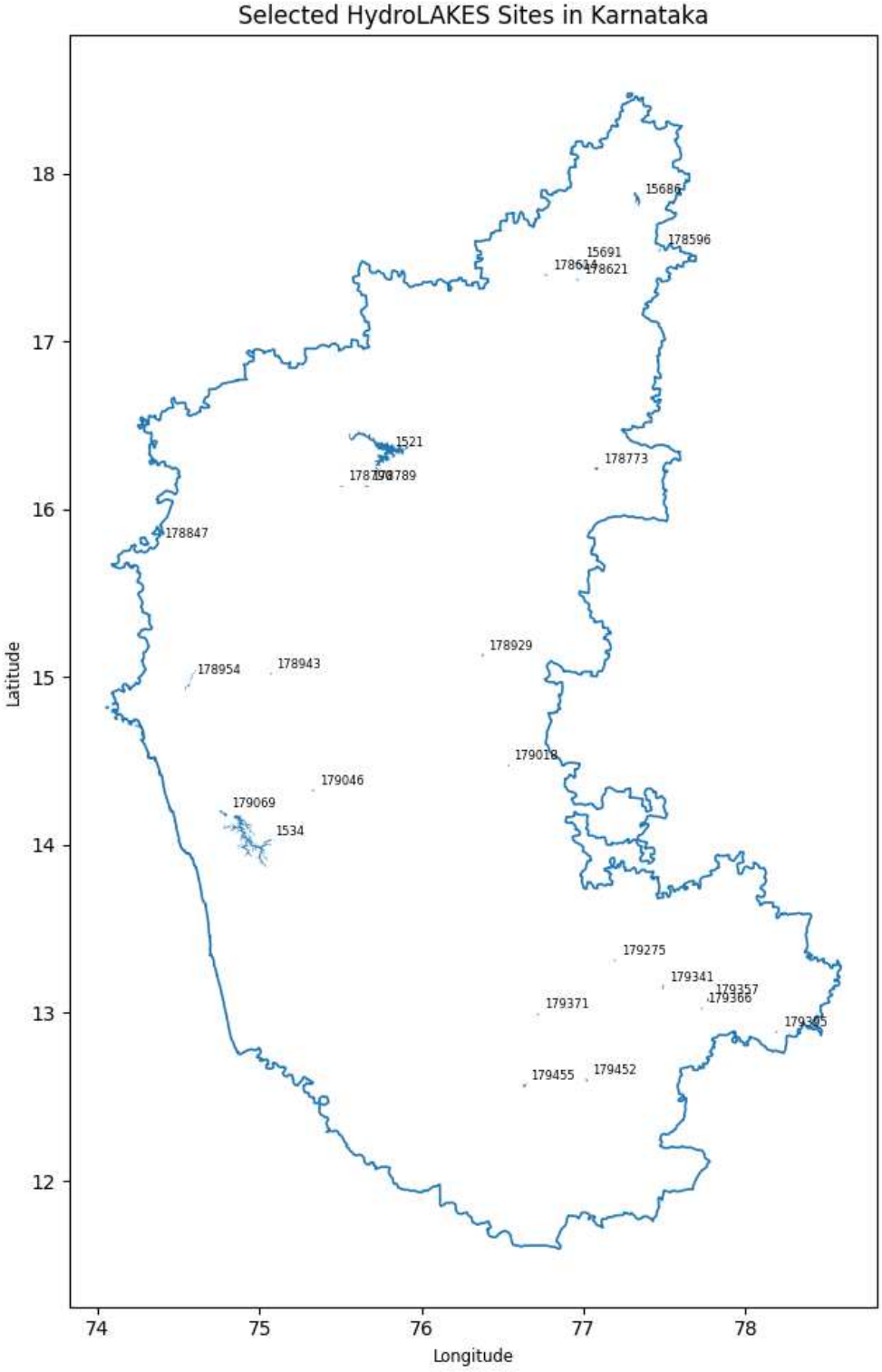
2. Study area, data and methods

2.1 Study area and sampling

The study area is Karnataka, India. The source inventory contained 156 candidate lakes with a mapped area of at least 1 km², and 25 lakes were chosen from it by stratified sampling. The unit of analysis is the lake-month. The study period runs from January 2000 to December 2021, which gives 264 possible months per lake and 6,600 possible lake-month combinations before observation-quality screening. The 25 lake identifiers are: 1521, 1534, 15686, 15691, 178596, 178614, 178621, 178773, 178789, 178790, 178847, 178929, 178943, 178954, 179018, 179046, 179069, 179275, 179341, 179357, 179366, 179371, 179395, 179452 and 179455.

Analysis for each lake was made separately. The representative points within Karnataka were used to spatially filter the HydroLAKES inventory. The number of lakes available for mapping was 156 (lakes ≥1 km²) (candidate lakes). Representative-point latitude was divided into five equal-width bands labelled South, South-Central, Central, North-Central and North. 25 lakes were sampled from 1 to 5 lakes per band (plus an additional sample when there were not enough candidates). These lake-level summaries do not represent a probability-weighted estimate for all of Karnataka's lakes; they are based upon a sample of

lakes. Selected_25_lakes.geojson and selected-site map kept the mapped boundaries and map saved at 300



dpi.

Figure 1. Study-area map of sampled Karnataka .

2.2 Satellite surface-water observations

Monthly lake surface-water observations were taken from the Joint Research Centre (JRC) Global Surface Water record (Pekel et al., 2016). A lake-month was treated as valid only if observation coverage was at

least 80%. Of the 6,600 possible lake-months, 3,415 met this coverage requirement. After rainfall pairing and anomaly construction, 3,352 paired records remained before restriction to the main analysis season. The difference between possible, valid and paired observations matters, because treating every calendar month as equally observed would hide the seasonal pattern of missing data. Validating extent from the surface-water maps derived from these Landsat products independently with other surface-water maps is still a critical step for the purpose of interpreting the extent, and other methods of mapping surfaces with Landsat can tackle some of the challenges of dynamic water boundaries (Cai et al., 2024).

The years where months are poorly observed correspond to the known deficiencies of optical satellite time series. Hao et al. (2024) identified missing data for monthly surface-water records globally and Qi et al. (2026) showed the potential of using lake monitoring with a higher temporal frequency to capture intra-annual variability. The present study is not an attempt to reconstruct missing observations of the monsoon; it rather presents results only for the months that pass its observation-coverage criterion.

2.3 Rainfall and anomaly construction

Daily precipitation from the CHIRPS time series (UCSB-CHG/CHIRPS/DAILY) was summed to provide the monthly mean precipitation for each calendar month, within each fixed HydroLAKES polygon at a nominal scale of 5,566 m. JRC Global Surface Water MonthlyHistory (JRC/GSW1_4/MonthlyHistory) water classes were processed at 30 m: class 2 water, class 1 land and class 0 no data. The water area observed was the total of the areas of all water pixels and the valid observed area was the sum of the areas of all class 1 and 2 water pixels. Valid coverage refers to the valid observed area/polygon area. The precipitation measurements were not removed and the lake-month observations below 80% coverage were masked for water analysis. Each lake had anomalies defined as the difference between 1-, 3- and 6-month accumulated rainfall and their respective historical averages.

Let $P_{i,t}$ be the average over all months in year t over lake polygon i of CHIRPS precipitation. The accumulated precipitation for k = 1, 3 and 6 are the cumulative sum of k consecutive monthly values, respectively, (where k = 6 for the 6-month window). The seasonal anomaly consists of the difference between the mean of that k-month accumulated series for the same lake and calendar month in available years. The notebook is clearly coded to roll the 6 months with pandas groupby(lake_id).rolling(6,6).sum() and then perform a groupby(lake_id,month).transform(mean)

$$.P_{i,t}^{(k)} = \sum_{j=0}^{k-1} P_{i,t-j}\,, \quad k \in \{1,3,6\} \quad (1)$$

$$A_{i,t}^{(k)} = P_{i,t}^{(k)} - \overline{P}_{i,m(t)}^{(k)} \quad (2)$$

For the lake-water series, let $W_{i,t}$ be the selected satellite-derived surface-water measure and $\bar{W}_{i,m}$ the observed mean for the same lake and calendar month across available years. The seasonally adjusted anomaly is defined as:

$$B_{i,t} = W_{i,t} - \overline{W}_{i,m(t)} \quad (3)$$

Anomalies show how far a value is from the usual value for that lake in the same calendar month. They are not absolute rainfall or water-extent deficits, so their joint behaviour may differ from that of the raw measurements. Because the rainfall windows overlap, the monthly sequences are temporally dependent, and inference and uncertainty statements must take this into account.

2.4 Rank association and copula inputs

Spearman rank correlation was used to compare monotonic dependence across the rainfall windows. Comparisons were made for each lake, and the paired differences were then summarized across the 25 lakes. The six-month window was used for the main copula comparison because it gave the strongest typical rank association in the reported sensitivity analysis. This choice does not mean that a causal six-month hydrological lag was identified.

Within each lake, paired rainfall and water anomalies were transformed to pseudo-observations using scipy.stats.rankdata(method="average"). For n paired observations, U = rank(X)/(n+1) and V = rank(Y)/(n+1). The fitting code also clipped these values to [EPS, 1−EPS] for numerical stability.

$$U_{i,t} = \frac{R_{i,t}}{n_i+1}, \qquad V_{i,t} = \frac{S_{i,t}}{n_i+1} \quad (4)$$

By Sklar's theorem, a joint cumulative distribution function F(x,y) can be written using the marginal distribution functions F_x and F_γ and a copula C. Here, C describes the dependence between the rainfall and water anomalies after ranking.

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y); \boldsymbol{\theta}) \quad (5)$$

2.5 Copula estimation and model selection

Five candidate bivariate families were compared separately for each lake: Gaussian, Student-t, Clayton, Frank and Gumbel. The Gaussian and Student-t copulas represent elliptical dependence. Clayton and Gumbel can represent different asymmetric tail structures under suitable parameter values. Frank is a symmetric Archimedean copula without asymptotic tail dependence. Choosing a family does not by itself establish a physical hydrological mechanism, and it does not show that the data have asymptotic tail dependence.

The copula parameters were estimated by maximizing the sum of the log-densities of the copulas, which is the same as minimizing the negative log-likelihood, using the scipy.optimize.minimize method L-BFGS-B. The code includes family-specific bounds on the parameters and multiple initial values, and also keeps the successful optimizations and chooses the successful start with the smallest value of the objective for each family. The Akaike and Bayesian information criteria were used: AIC = 2k − 2log L and BIC = k log(n) − 2log L, respectively. Exact bounds and convergence diagnostics should be retained in the reproducibility supplement.

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log c(U_t, V_t; \boldsymbol{\theta}) \quad (6)$$

$$\mathrm{AIC} = 2k - 2\ell\left(\widehat{\boldsymbol{\theta}}\right) \quad (7)$$

$$\mathrm{BIC} = k\ln(n) - 2\ell\left(\widehat{\boldsymbol{\theta}}\right) \quad (8)$$

The model with the lowest AIC was taken as the preferred family for each lake, and BIC was used as a secondary check. The AIC difference between the best and second-best models was kept as a practical measure of how uncertain the selection was. A difference below two was treated as weak discrimination between candidate models. It does not prove that their dependence structures are equivalent. The reported counts of best models should therefore be read together with these AIC differences.

2.6 Overall fit, independence benchmark and finite-threshold joint occurrence

The fitted copulas were compared with the independence copula $C_0(u,v) = uv$. A Cramér–von Mises (CvM) discrepancy was calculated between the empirical copula and the fitted copula. Lower values mean closer

in-sample agreement under this definition of discrepancy. The analysis counted the lakes for which the fitted model had a smaller CvM value than independence. These statistics were used descriptively and were not treated as formal goodness-of-fit p-values.

For the joint low-condition analysis, the event of interest was a rainfall pseudo-observation at or below q together with a water pseudo-observation at or below q. Under independence, the probability of this joint event is q^2. The empirical joint probability was calculated as the share of paired observations that meet both conditions, and the fitted joint probability is C(q,q). The amplification ratio then describes how far the joint probability departs from the independence benchmark at a finite threshold. The quantities are given in Equations (9)–(11).

$$\hat{p}_{\text{joint}}(q) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}\,(U_t \le q,\ V_t \le q) \quad (9)$$

$$\hat{A}_{\text{obs}}(q) = \frac{\hat{p}_{\text{joint}}(q)}{q^2} \quad (10)$$

$$\hat{A}_{\text{model}}(q) = \frac{C(q,q;\hat{\boldsymbol{\theta}})}{q^2} \quad (11)$$

An amplification above one means that joint low conditions occur more often than independence would predict at the chosen q. This finite-threshold measure must not be confused with the asymptotic lower-tail dependence coefficient $\lambda L = \lim(q\to 0^+)\ C(q,q)/q$, shown in Equation (12). That coefficient is defined as a limit and cannot be established from a few empirical thresholds.

$$\lambda_L = \lim_{q\to 0^+} \frac{C(q,q)}{q} \quad (12)$$

Tail-specific calibration was assessed by comparing two absolute errors: the difference between the empirical and modelled joint-event probabilities, and the corresponding difference under independence. For a given lake, the copula was classified as better than independence when its absolute joint-probability error was smaller. This is an in-sample calibration comparison and is not evidence of predictive skill for unseen months.

2.7 Robustness and uncertainty assessment

The workflow included sensitivity analyses for the rainfall accumulation window, the lower-tail threshold and the observation-coverage requirement. It also included lake-level checks on fitting reliability and on the treatment of tied ranks. Thresholds of q = 0.15, 0.20 and 0.25 were evaluated. The 20th percentile was kept as the main reporting threshold because lower thresholds give fewer joint observations and therefore less stable empirical estimates.

For the main threshold, a year-block bootstrap with 1,000 replicates was used to summarize uncertainty in the observed amplification. Paired comparisons of rainfall windows and thresholds across lakes were also accompanied by bootstrap confidence intervals. Lakes may share the same climatic forcing and may be spatially correlated. The lake-level bootstrap intervals and the unadjusted comparisons across lakes should therefore not be treated as fully spatially independent inference. The same observations were also used for both fitting and assessment, so the goodness-of-fit and probability-error comparisons remain in sample.

3. Results

3.1 Observation availability and analytical sample

The 25 sampled lakes gave 6,600 possible lake-month combinations over the 22-year period. Applying the minimum 80% satellite coverage criterion kept 3,415 valid JRC lake-months. After rainfall pairing and anomaly construction, 3,352 records were available. The main October–May six-month copula analysis contained 2,974 lake-month observations, and the sample size for each lake varied with observation availability. These counts show why the theoretical panel of lake-months must be kept separate from the sample that was actually analyzed.

The restriction to October–May was made because observation coverage was comparatively poor in the monsoon months. The results therefore describe the non-monsoon and transition-season conditions captured by the retained data. They should not be generalized to the full annual hydrological cycle without qualification. Figure 1 should show the study-area sampling configuration, and Figure 2 should report coverage and temporal availability. The study-area map and observation-availability figure are presented as Figures 1 and 2; their source values and cartographic details require final verification.

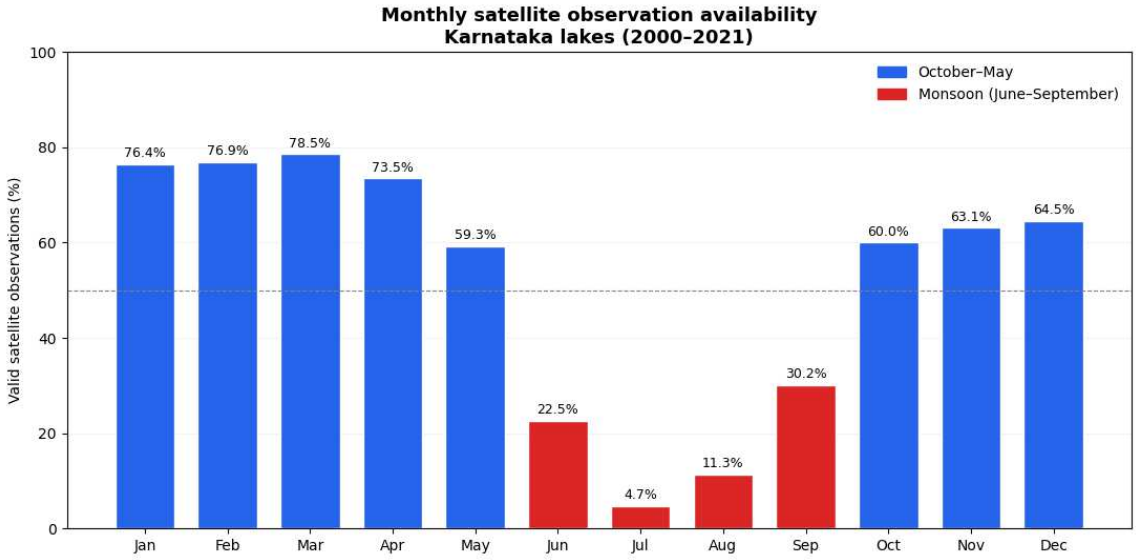


Figure 2. Monthly satellite observation availability across the study period.

3.2 Rainfall accumulation window comparison

Across the 25 lakes, the median Spearman rank association increased as the rainfall accumulation window became longer. The median lake-level correlations were 0.052 for one-month rainfall, 0.116 for three-month rainfall and 0.300 for six-month rainfall. In paired comparisons, the median advantage of the six-month window was 0.206 over one month (95% bootstrap confidence interval 0.144 to 0.325), and 22 of 25 lakes favoured six months. The median advantage over three months was 0.164 (95% interval 0.077 to 0.223), and 21 of 25 lakes favoured six months.

Finite-threshold amplification showed a similar direction but with more uncertainty. The median observed amplification was about 1.19, 1.26 and 1.71 for the one-, three- and six-month windows, respectively. However, the paired differences in amplification, six versus one month (median 0.225; 95% interval −0.018 to 1.008) and six versus three months (median 0.199; 95% interval −0.016 to 0.524), both included zero. The evidence is therefore stronger for the rank-association advantage of the six-month window than for a

similar difference in joint low-condition amplification. Figure 3 presents the paired lake-level comparisons, and Table 1 summarizes the rainfall-window results.

[Insert Figure 3 here]

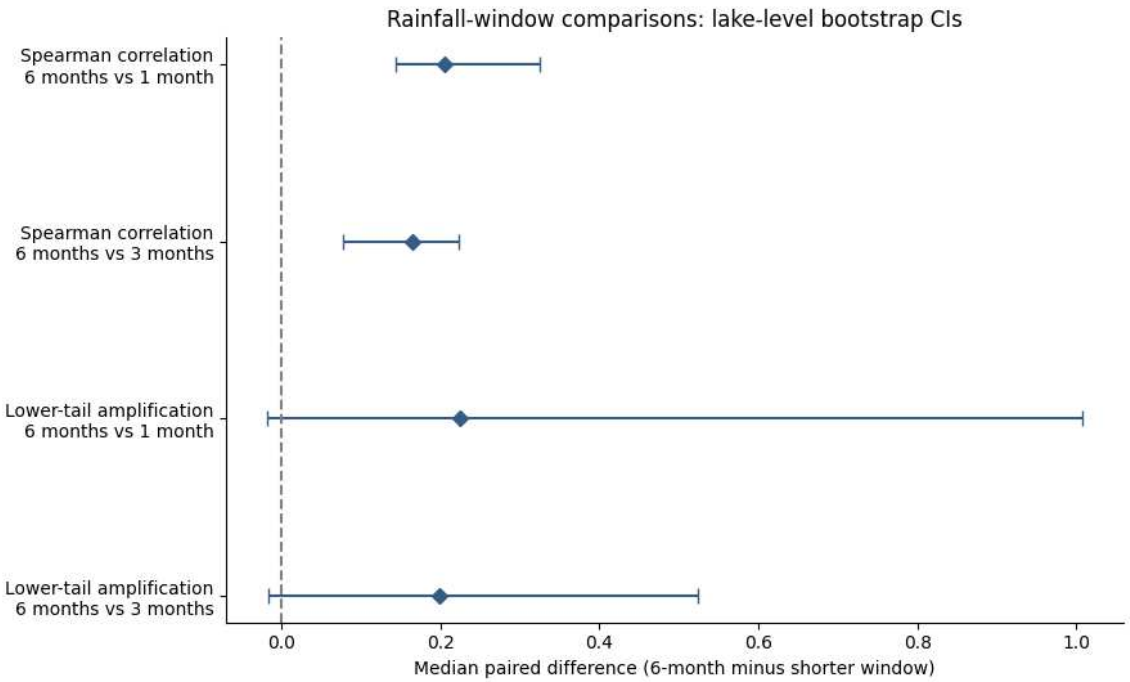


Figure 3. Paired lake-level comparisons of rainfall accumulation windows with bootstrap intervals.

[Insert Table 1 here]

Rainfall window	Median Spearman ρ	Median observed amplification
1 month	0.052	1.190
3 months	0.116	1.263
6 months	0.300	1.708

Table 1. Lake-level median association and empirical finite-threshold amplification by rainfall window. Amplification values are descriptive; see the source sensitivity files for unrounded estimates.

3.3 Copula family selection and ambiguity

Among the five candidate families, Frank was selected for 10 lakes, Clayton for seven, Gaussian for six, Student-t for one and Gumbel for one. AIC and BIC selected the same family for all 25 lakes. No single model was preferred across all lakes. However, for 17 of 25 lakes the AIC difference between the best and second-best models was below two, which shows that the preferred family was often not clearly better than the alternatives.

The variation in selected families is therefore evidence that the fitted statistical representations differ among lakes. It is not conclusive evidence that the lakes have five physically distinct dependence regimes. Figure 4 presents the family frequencies; AIC differences are summarized in the text. The full lake-level model comparison table should be kept as supplementary material. Table 2 summarizes the AIC-selected copula families.

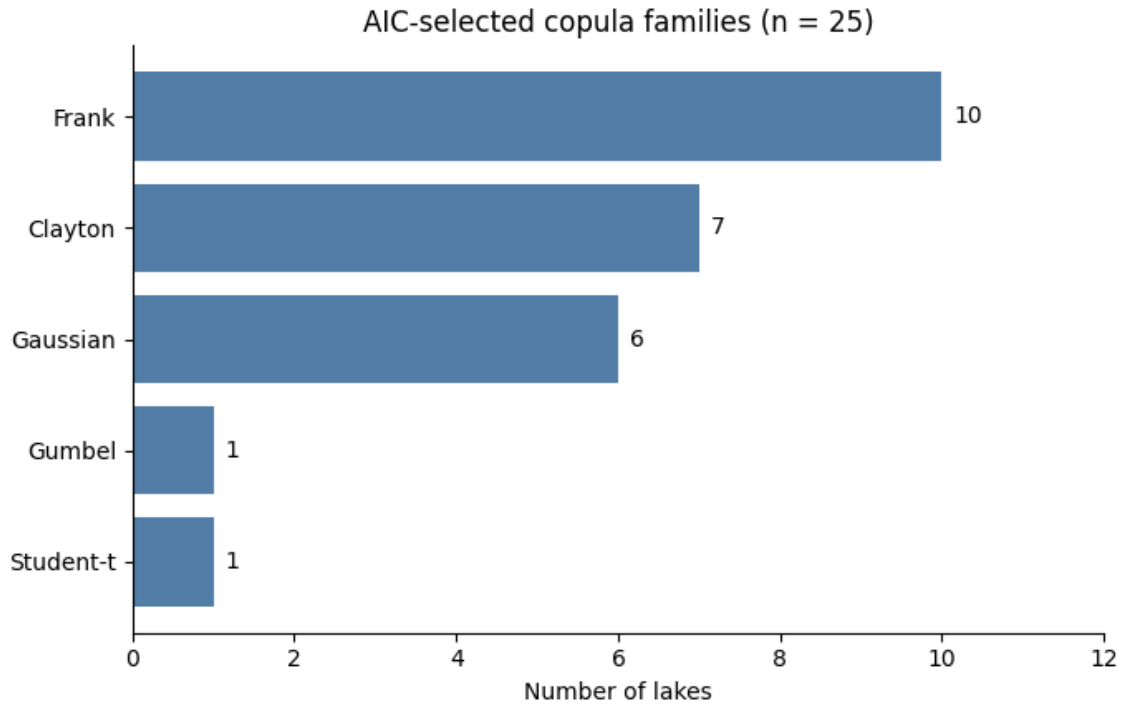


Figure 4. AIC-selected copula families across 25 lakes.

[Insert Table 2 here]

Preferred family	Lakes
Frank	10
Clayton	7
Gaussian	6
Student-t	1
Gumbel	1

Table 2. Lowest-AIC copula families for the 25 sampled lakes. Seventeen lakes had $\Delta AIC < 2$ between their two leading candidates.

3.4 Overall in-sample fit and independence

Fitted copulas had a lower in-sample CvM discrepancy than the independence copula for 22 of 25 lakes, with a reported median improvement of about 0.1335. The three lakes where the fitted copula did not improve the discrepancy were 178954, 179366 and 179069. This indicates that a flexible dependence model generally followed the empirical copula more closely than independence did, within the observations used for fitting.

[Insert Figure 5 here]

Because the same observations were used for estimation and evaluation, the improvement in fit is descriptive. It should not be presented as a significant goodness-of-fit test or as evidence of out-of-sample forecasting ability. A model that fits well overall can also perform worse in one region of the joint

distribution. This is why the lower-tail joint probabilities were evaluated separately. Figure 5 displays the lake-level in-sample CvM discrepancy improvements relative to independence.

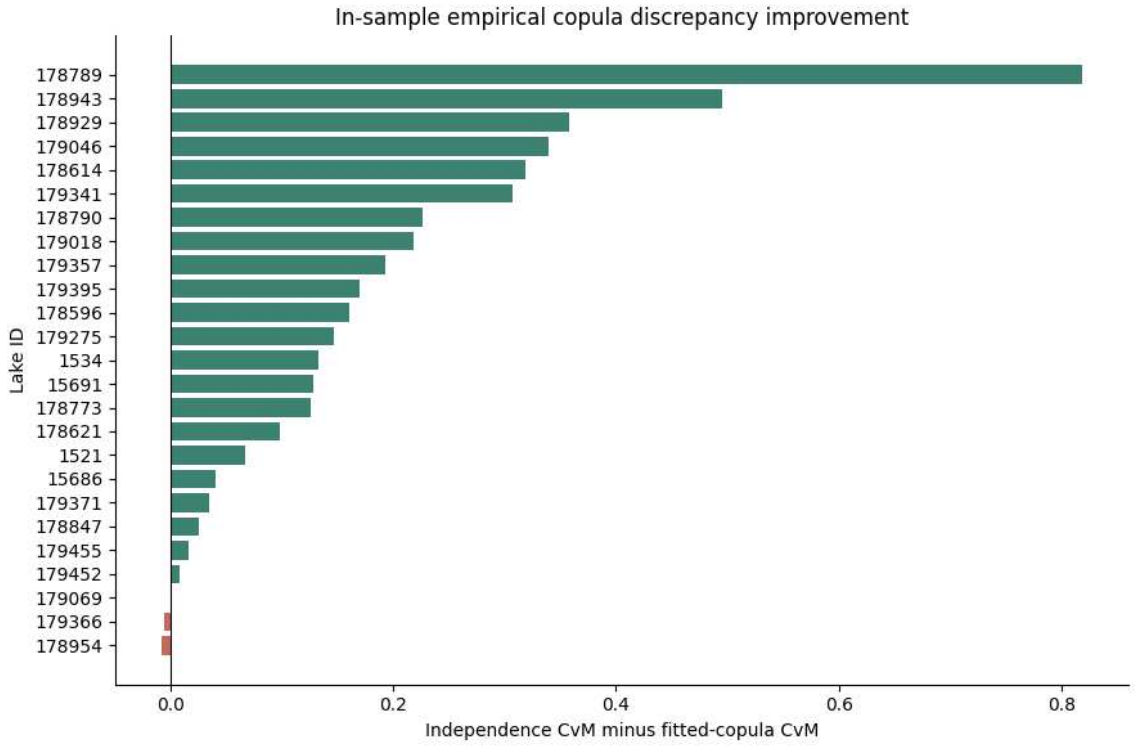


Figure 5. In-sample Cramér–von Mises discrepancy improvement relative to independence.

3.5 Lower-tail amplification and probability calibration

The median observed lower-tail amplification was 1.7083 and the median fitted amplification was 1.7523 at q = 0.20. Twenty-five lakes recorded empirical amplification above one with 19 of which being above one. This indicates that at this finite threshold, low rainfall and low surface water anomalies were more likely to occur together than they would be if they were independent events. In 16 of 25 lakes, the fitted copula performed better than independence in terms of absolute error in estimating the empirical joint-event probability. The median absolute copula probability error was about 0.00860.

The lakes observed with an amplification above one were 178614, 178789, 179046 and 179341, and these lakes had a fitted copula that was better than independence. The empirical amplification values were approximately 2.000, 2.069, 2.925 and 2.579 for each of them. The corresponding fitted values were about 1.844, 2.413, 2.181 and 2.289. These examples demonstrate that the joint occurrence can be greater or less than the modelled amplification. Lake 179046 was also assigned to a family type that was only marginally different from AIC with a difference of approximately 1.188, so the family type should be interpreted with a grain of salt.

Year-block bootstrap intervals were above the independence threshold for six lakes when each lake was analysed separately without spatial or multiplicity adjustment. These intervals allow statements to be made carefully for individual lakes. They do not provide an estimate of the prevalence of the pattern statewide.

Observed and fitted amplification are compared in figure 6. Table 3 shows the four illustrative lakes.

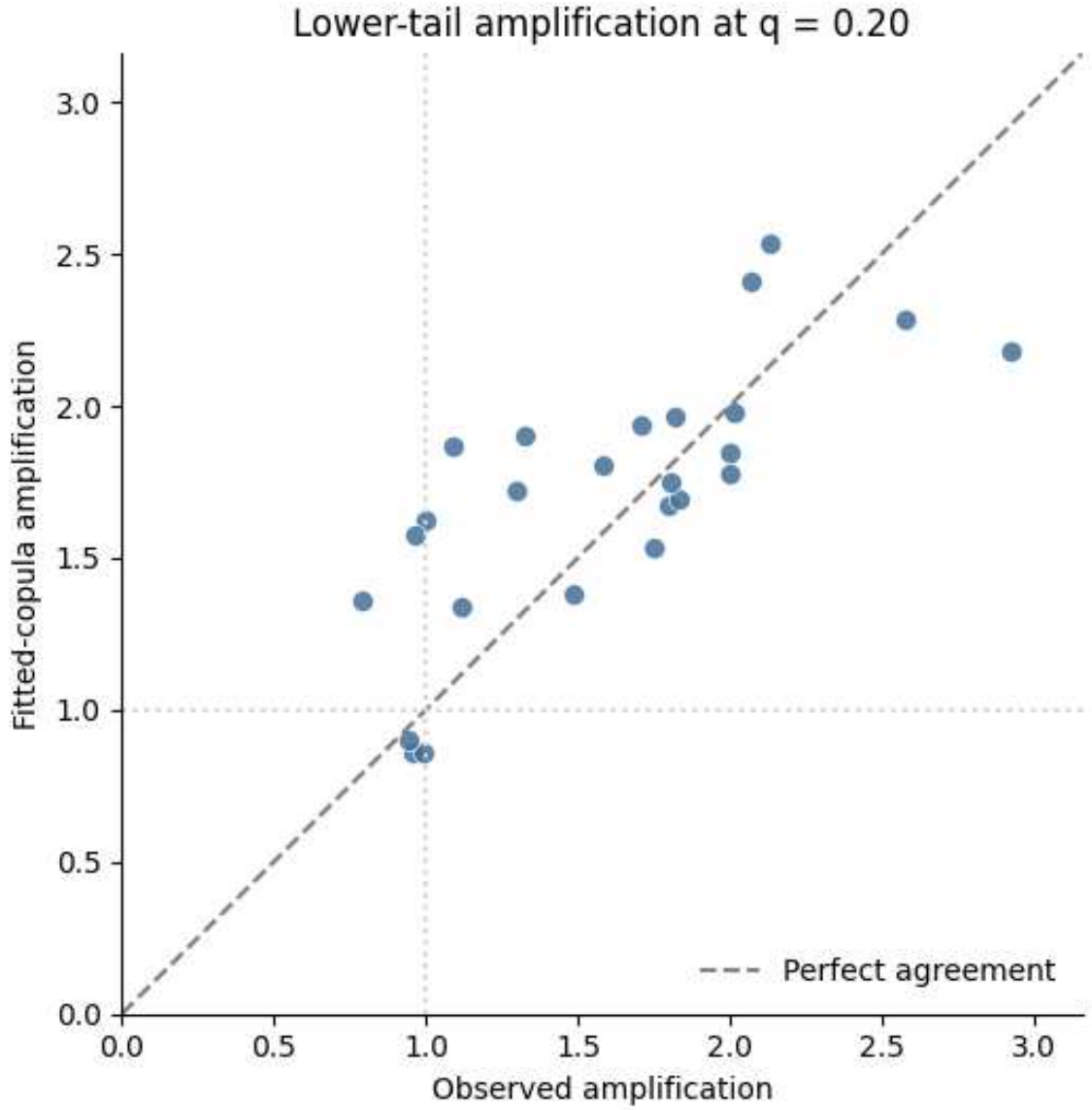


Figure 6. Empirical versus modelled finite-threshold lower-tail amplification at q = 0.20.

Lake ID	Observed A(0.20)	Fitted A(0.20)	Preferred family
178614	2.000	1.844	Frank
178789	2.069	2.413	Frank
179046	2.925	2.181	Frank
179341	2.579	2.289	Student-t

Table 3. Illustrative lakes with empirical amplification above one and lower fitted probability error than independence. Values are rounded.

3.6 Threshold sensitivity

[Insert Figure 7 here]

The median empirical amplification was 1.8595 at q = 0.15, 1.7083 at q = 0.20 and 1.4640 at q = 0.25. The median fitted amplification was about 2.0106, 1.7523 and 1.5757 at these thresholds. The number of lakes with observed amplification above one was 21, 19 and 22, respectively. The number of lakes where fitted copulas beat independence in absolute joint-probability error was 18, 16 and 19.

The apparent fall in median amplification as q increases is descriptive. Paired threshold comparisons generally had intervals that included zero. The comparison of the 15th and 25th percentiles only narrowly excluded zero in an unadjusted, exploratory interval. In addition, the median number of empirical joint events per lake was about five at q = 0.15, seven at q = 0.20 and eleven at q = 0.25. The apparent amplification at the lowest threshold should therefore be interpreted with particular caution. Figure 7 shows the threshold-sensitivity comparison, and Table 4 reports the corresponding numerical summaries.

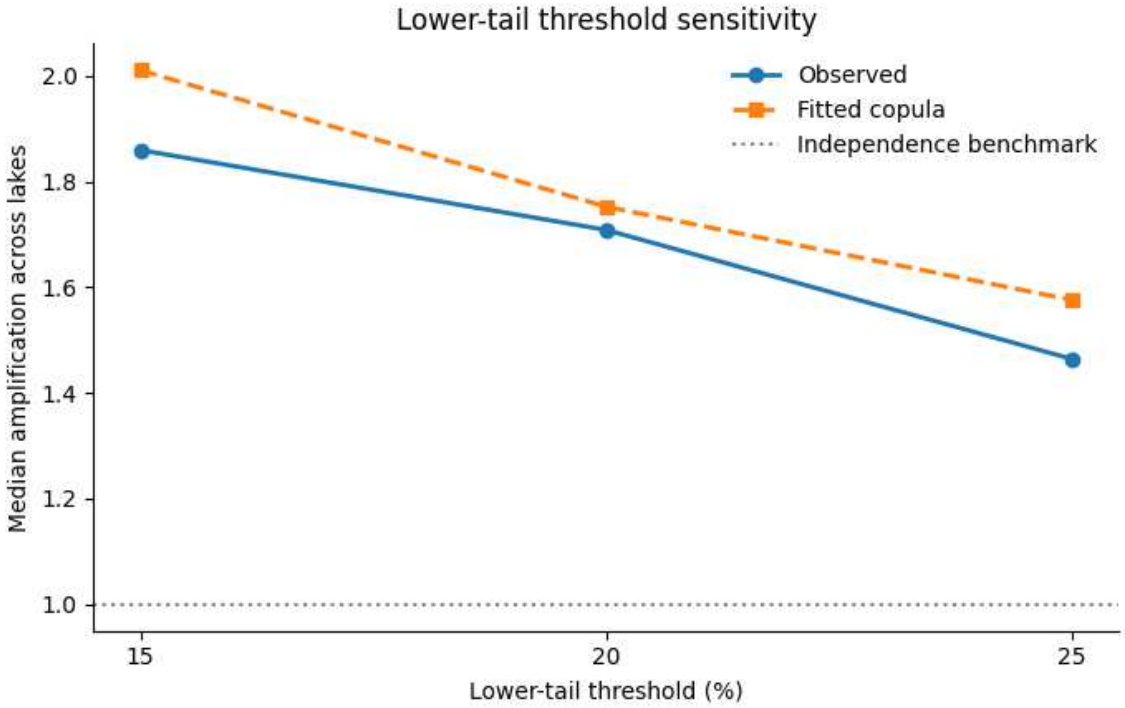


Figure 7. Sensitivity of median observed and fitted amplification to the lower-tail threshold.

[Insert Table 4 here]

Threshold q	Observed median A(q)	Fitted median A(q)	Lakes observed A>1	Copula better
0.15	1.860	2.011	21/25	18/25
0.20	1.708	1.752	19/25	16/25
0.25	1.464	1.576	22/25	19/25

Table 4. Threshold sensitivity of joint lower-tail amplification. Model comparisons are in sample.

4. Discussion

4.1 Cumulative rainfall association and hydrological interpretation

The most clear cut comparative result is found in the greater rank association with six month accumulated rainfall. A potential explanation is that the surface-water conditions in the lake sum up the rainfall inputs and catchment processes over a time scale of more than one calendar month. Earlier rainfalls may increase soil and catchment storage, runoff, water availability in the lake, and water levels may not reflect earlier wet or dry conditions. These processes provide a conceptual basis for why it may be more informative to use cumulative rainfall than rainfall for the same month.

But the analysis is not able to separate the mechanism, nor single out a time scale of physical memory. A rolling series of six months may be more highly correlated due to smoothing, persistence, undercorrected seasonality, and/or other temporal structure. The serial dependence is also increased by overlapping windows. The observed association with 6-month anomalies should therefore be described as a stronger observed association, rather than a 6-month causal lag. Again, the lower-tail window differences' confidence intervals caution against stating that six months is definitely the best option for joint extremes.

4.2 Heterogeneous lake-level dependence and model uncertainty

The set of preferred copula families are still compatible with the scenario that the relationship between rainfall and lake water varies among the sampled sites. The statistical relationship could vary due to differences in catchment size, catchment connectivity, landscape setting, irrigation or reservoir operation, groundwater exchange and human withdrawals. These factors were not measured and not controlled for in this study; thus, their influence is speculative. If a geographical explanation was needed, the lake-level dependence estimates would need to be correlated with independent catchment and management indicators.

Importantly, for most lakes, the most suitable copula family was not easily distinguished from the second best family. Fifteen of the AIC differences were very small, so the number of preferred families of Frank, Clayton etc. should not be used to establish firm regional regimes. A better future analysis should average over candidate families, incorporate model-selection uncertainty into the event estimates, or leverage held-out periods to assess if the differences between families persist.

4.3 Joint low conditions versus asymptotic tail dependence

When the median amplification is greater than one at the main threshold, a concrete description of the joint occurrence of low rainfall and low surface-water anomalies among the retained observations is obtained. It is different from rank correlation as it considers only one selected area of the bivariate distribution. The likelihood of combined low conditions could be helpful in screening lakes that have low rainfall and low surface water anomalies.

This joint-event approach is similar to that of the multivariate drought literature (Shiau, 2006; Shiau et al., 2007; Sajeev et al., 2021) but it is not the same as a drought duration–severity frequency analysis. In this case, the low condition is specified by the pair-specific anomaly rank in the fixed finite value and no hydrologic return period is assumed.

Nor does finite-threshold amplification imply an asymptotic dependence nor does it imply drought impacts or future risk. The 20 percentile is not restricted to the extremes of observations, but also contains moderately low observations, and the water anomaly is a satellite-derived surface measurement and not a direct measurement of usable storage. The copula families are also characterised by different limiting tail properties, which cannot be confirmed with the limited number of joint events observed. This is why the manuscript does not claim to show asymptotic tail dependence but does show finite-threshold joint occurrence throughout the manuscript.

4.4 Why overall model fit and tail calibration differ

In particular, the difference between the improvement of 22 of 25 lakes in overall CvM discrepancy and the improvement of 16 of 25 in tail-event probability error is at the core of the methodological interpretation. Even when a model reproduces most of the empirical dependence pattern, it can be inadequate in the small joint lower tail region. Overall fit and tail-specific calibration are thus answering related but different questions. AIC selection also minimizes a global likelihood criterion, and is not based on the accuracy of any given joint-event probability.

The selection of the lowest AIC should not be the only criterion for selecting a dependence model. The three model diagnostics – the empirical amplification, the fitted amplification and the independence benchmark – provide a convenient oversight of whether the complexity of the model increases the quantity of interest. The following comparison is sample. The fitted models would need to be temporally validated independently before they could be recommended for forecasting or for screening lakes to future risk.

4.5 Threshold choice, event scarcity and robustness

The median value of empirical amplification is greater than one at all three thresholds tested, thus corroborating a general descriptive pattern of increased joint low conditions. The effect is seen to increase at lower q; however, this is not a robust monotonic response. For a typical lake, $q = 0.15$, it contributed just a few joint events and a few observations can lead to a large difference in the estimated amplification. Further limitations on the strength of inference come from the paired confidence intervals as well as the many exploratory comparisons.

A middle-ground compromise between the extremes of focusing on low conditions and not enough empirical events is reached at the 20th percentile main threshold. It is more useful to provide sensitivity analysis to each of the 15%, 20% and 25% thresholds than to just provide one amplification estimate without context. The sensitivity checks to the coverage and tied rank should also be included in the supplementary material to illustrate the impact of the choices made in the observations and statistics on conclusions.

4.6 Geographic relevance and limitations

The geographic significance of the study is the demonstration that there is not a good single coefficient for the rainfall–water relationship or a single universal family of copulas that adequately summarizes the relationship. The sample-level summaries indicate common patterns; the lake-level estimates indicate variation which may warrant some field investigation. The sample of 25 lakes does not represent all the lakes of Karnataka without a specific sampling and weighting design. Similarly, the primary conclusions are not necessarily representative of the dependence throughout the year because of the lack of regular observations during the monsoon season. The water quality assessment is not an isolated assessment as both are different indicators for the environment, and satellite observations have already been used for other remote sensing assessments of Indian lakes e.g., Vembanad lake, Sundar & Kundapura (2023).

Environmental monitoring implications. The primary use of this study is to determine lakes that are sampled where the rainfall level is consistently correlated with the satellite observed surface-water anomalies, thus providing additional observation needs. For instance, the four lakes listed below, 178614, 178789, 179046 and 179341, are examples of lakes that are candidates for follow-up that had empirical amplifications above one and fitted joint probability errors below the independence benchmark. These are not standalone high-risk lakes – the management ranking should also take into account the information pertaining to the water demand, water storage, catchment conditions and local interventions.

The significance of lake change monitoring is similar to global evidence of lake-storage losses reported by Yao et al. (2023). But without matching measurements, this study can't be used to explain changes in individual lakes caused by evaporation or water withdraw or other processes.

Some limitations of the study are: The time dependence in the repeated monthly observations and overlapping water accumulation in the lakes, the possibility of space dependence between lakes, uncertainty in the water that is detected from satellite and lack of direct covariate of the hydrological and human-management processes. Lake or year-block based bootstrap intervals might not reflect all the spatial forcing. The design was mainly an observation one and did not allow causal attribution. Recommendations for future work include the use of spatially structured resampling, catchment attributes, management records, and improved monsoon observation coverage and temporally blocked model evaluation.

4.7 Implications for environmental monitoring and assessment

A defensible monitoring interpretation is used to distinguish between observations and operational decisions. Associations reported are evaluated within the domain that is constrained by the 80% valid-observation threshold and the October-May restriction. The six-month rainfall comparison provides a possible context indicator for lake-water anomalies, and empirical finite-threshold amplification determines a sample set of lakes where anomalies of low rainfall coincide with anomalies of low lake levels. It should be noted that for 17 of 25 copula selections, it is unclear whether there is any distinction (using AIC), and for 16 of 25, there is no better in-sample joint-probability error, which is why the selected family should not be regarded as a solid alert model. Next, it would be logical to include the existing diagnostics and repeated satellite observations, lake-level field measurements and catchment and withdrawal data, and blocked temporal validation, to establish alert thresholds or determine priority.

5. Conclusions

The effect of rainfall accumulation on variability of surface-water of lakes in Karnataka, India was studied for 25 selected lakes from 2000 to 2021. The research showed that combining satellite-derived surface-water observations from the Joint Research Centre (JRC) Global Surface Water dataset and CHIRPS precipitation records enabled them to investigate the relationship between the accumulation of rainfall in various time windows and observed conditions in lake water. The differences in dependence between lakes and the co-occurrence of low rainfall and low surface-water anomalies were also analysed. The results help to understand the spatio-temporal nature of the rainfall–lake-water relationship in a monsoon-influenced area.

The main result is that 6-month total rainfall was more closely correlated with surface-water anomalies of lakes than one-month and three-month totals. For the 1-month window, the median Spearman correlation was 0.052, while the window of 3 months produced a median correlation of 0.116 and 6-months window a median correlation of 0.300. Median correlation further benefits of 0.206 and 0.164 were obtained for the 6-month window compared to the one-month and three-month windows, respectively, through paired comparisons. The results shown indicate that lake surface-water conditions may be influenced by rainfall over a number of months, rather than solely by rainfall received in a single month. The behaviour is similar to that that might be driven by the effects of antecedent moisture conditions, catchment runoff, and hydrological storage. However, these mechanisms were not measured directly, and the observed relationships are not assumed to be due to a specific causal response time.

The other key finding that comes from this is the co-occurrence of low rainfall and low lake surface-water conditions. The median empirical lower-tail amplification was 1.708 at the primary 20th percentile threshold, suggesting that the number of joint low-condition events was about 1.71 times the number of events expected if the two were statistically independent at the median lake sampled. Amplification values greater than one were observed in 19 of the 25 lakes. This results demonstrates the need to consider the joint effects of environment, and not the effects of rainfall and surface-water variability separately. The results indicate that the lakes with low rainfall and low surface-water anomalies occur more frequently than would be predicted by assuming that these two variables are statistically independent. But this measure only represents the co-occurrence at a given threshold and it is not a measure of asymptotic tail dependence, hydrological drought effects, or future risk.

The study also found a significant variety of the statistical dependence structures used in specific lakes. Frank was most preferred for ten lakes, Clayton for seven, Gaussian for six, and Student-t and Gumbel each for one lake. This variation suggests that a single statistical dependence structure is not valid for all sampled lakes for rainfall–surface-water relationships. Such heterogeneity may be due to variations in the catchment's characteristics, hydrological connectivity, interactions with ground water, water withdrawals, and practices of lake management. But it is important that these potential influences be further investigated with separate environmental and management data. Furthermore, for 17 lakes a difference of less than two between the two most preferred models was observed, suggesting that the preferred model family was not always clearly preferential as compared to alternatives.

The comparison of the performance of models represents another methodological contribution. For 22 of the 25 lakes, the fitted dependence models outperformed the independence model in terms of overall in-sample Cramér–von Mises discrepancy. The improvement in the estimation of joint lower-tail probabilities was noted in 16 lakes, however. This difference shows that a model that better represents the overall dependence structure does not necessarily have more accurate estimates of the probabilities of having joint low-condition events. Therefore, model evaluation of hydrological dependence should consider not only good overall model fit, but also the accuracy of the probabilities determined for the conditions of interest in the environment. Due to the fact that the model estimation and evaluation were carried out on the same observations, these results should be considered in-sample rather than predictive.

The sensitivity analyses further support the interpretation and emphasize the significance of uncertainty. Lower tail amplification was found to have continued to exceed the independence value at the 15th, 20th, and 25th percentile thresholds, with values of 1.860, 1.708, and 1.464, respectively. The differences between thresholds were not always demonstrated by the exploratory bootstrap comparisons, though the results indicate that the joint appearance of low conditions is not limited to a particular threshold. The smaller thresholds also had fewer joint observations contributing to the estimation of empirical probability, adding uncertainty to the estimation process. The results should thus be considered as an indication of finite-threshold occurrence of the joints and not necessarily as proof of increasing dependence towards more extreme conditions.

Geographically and from the water resource management point of view, the study shows that long-term satellite observations and precipitation data can be combined to explore the variability of freshwater resources in several lakes. The better correlation with cumulative rainfall implies that current lake water conditions may be related to past rainfall and thus that antecedent moisture is a useful index. Simultaneously, the variability in results between individual lakes shows that some local variation should be considered in water-resource assessments, rather than statewide average ones. The results could be used to identify lakes that require more frequent hydrological monitoring, as often happens when rainfall is low and surface-water levels are low. But more information on lake storage capacity, water demand, catchment conditions and human interventions would be needed for practical management applications.

There are a number of caveats to be noted. The chief analysis was limited to October–May since the coverage from the satellites was relatively poor during the monsoon months. Thus, the results are not necessarily applicable when extending to the entire annual hydrological cycle without qualification. Moreover, using overlapping accumulation windows of rainfall and repeated observations on a monthly basis will cause a temporal dependence, and common climate conditions could lead to spatial dependence between lakes. The number of joint low condition observations is also relatively low at stricter thresholds, which also restricts the accuracy of tail probability estimates. Last, the study focused on surface-water variation as observed by the satellite and did not examine lake-water volume or usable water storage, nor did it allow for causation to be inferred.

The framework should be expanded in the future to include enhanced monitoring data for the monsoon season, catchment data, groundwater data, lake evaporation data, and lake management and water use data. The reliability of dependence estimates would be further enhanced by temporally independent model evaluation and spatially informed uncertainty assessment. The study could be extended to a larger and suitably weighted sample of lakes to identify the patterns in the geographical regions in Karnataka. Combining water extent information derived from satellites with water-level and storage data would give a more complete picture of water availability.

In general, the study highlights the importance of rainfall accumulation timescale, lake-specific dependence characteristics, and the joint occurrence of low conditions when using satellite observations of lake surface-water variability. It is its major finding that the cumulative rainfall relationships and the concurrent low environmental conditions may differ greatly between lakes, and that the statistical models should be evaluated for their capacity to simulate the specific joint conditions of interest. The results offer a basis for additional geographically based studies of lake-water variation and freshwater monitoring in the monsoon-affected area of the state of Karnataka as well as in the rest of the state.

The results indicate that, for monitoring programmes, not only water-extent anomalies but also the degree of coverage of satellite observations should be reported and that antecedent rainfall windows should be considered separately from overall model fit and also separately for joint-event calibration. Data beyond the scope of the samples or the observation period (October to May) should be viewed with caution and needs to be supplemented by data.

Funding

The authors received no external funding for this research.

Conflict of interest

The authors declare no competing interests.

Data and code availability

JRC Global Surface Water and CHIRPS datasets are publicly available. Processed datasets and analysis code are available from the corresponding author upon reasonable request.

References

Alam, A., Bhat, M. S., Ahsan, S., Taloor, A. K., & Farooq, H. (2024). Earth observation satellite data-based assessment of wetland dynamics in the Kashmir Himalaya. Environmental Monitoring and Assessment, 196, 32. https://doi.org/10.1007/s10661-023-12185-7

Bai, B., Mu, L., & Tan, Y. (2025). A global lakes/reservoirs surface extent dataset (GLRSED): An integration of multi-source data. Geoscience Data Journal, 12(1), e285. https://doi.org/10.1002/gdj3.285

Cai, Y., Shi, Q., & Liu, X. (2024). Spatiotemporal mapping of surface water using Landsat images and spectral mixture analysis on Google Earth Engine. Journal of Remote Sensing, 4, 0117. https://doi.org/10.34133/remotesensing.0117

Funk, C., Peterson, P., Landsfeld, M., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data, 2, 150066. https://doi.org/10.1038/sdata.2015.66

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 12(4), 347–368. https://doi.org/10.1061/(ASCE)1084-0699(2007)12:4(347)

Hao, Z., Cai, X., Ge, Y., Foody, G., Li, X., Yin, Z., Du, Y., & Ling, F. (2024). Resolving data gaps in global surface water monthly records through a self-supervised deep learning strategy. Journal of Hydrology, 640, 131673. https://doi.org/10.1016/j.jhydrol.2024.131673

Li, B., & Rodell, M. (2024). Terrestrial water storage in 2023. Nature Reviews Earth & Environment, 5, 247–249. https://doi.org/10.1038/s43017-024-00545-x

Li, Y., et al. (2023). Diminishing storage returns of reservoir construction. Nature Communications, 14, 3203. https://doi.org/10.1038/s41467-023-38843-5

Mukherjee, R., Policelli, F., Wang, R., Arellano-Thompson, E., Tellman, B., Sharma, P., Zhang, Z., & Giezendanner, J. (2024). A globally sampled high-resolution hand-labeled validation dataset for evaluating surface water extent maps. Earth System Science Data, 16, 4311–4323. https://doi.org/10.5194/essd-16-4311-2024

Nelsen, R. B. (2006). An introduction to copulas (2nd ed.). Springer. https://doi.org/10.1007/0-387-28678-0

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540, 418–422. https://doi.org/10.1038/nature20584

Qi, T., Li, X., Duan, H., & Yuan, W. (2026). MODIS-Landsat fusion reveals two-decade 8-day lake dynamics with critical intra-annual regime shifts. Journal of Hydrology, 665, 134702. https://doi.org/10.1016/j.jhydrol.2025.134702

Sajeev, A., Deb Barma, S., Mahesha, A., & Shiau, J.-T. (2021). Bivariate drought characterization of two contrasting climatic regions in India using copula. Journal of Irrigation and Drainage Engineering, 147(3), 05020005. https://doi.org/10.1061/(ASCE)IR.1943-4774.0001536

Salvadori, G., De Michele, C., Kottegoda, N. T., & Rosso, R. (2007). Extremes in nature: An approach using copulas. Springer. https://doi.org/10.1007/1-4020-4415-1

Shiau, J.-T. (2006). Fitting drought duration and severity with two-dimensional copulas. Water Resources Management, 20, 795–815. https://doi.org/10.1007/s11269-005-9008-9

Shiau, J.-T., Feng, S., & Nadarajah, S. (2007). Assessment of hydrological droughts for the Yellow River, China, using copulas. Hydrological Processes, 21(16), 2157–2163. https://doi.org/10.1002/hyp.6400

Sundar, P. K. S., & Kundapura, S. (2023). Spatiotemporal variation in the water quality of Vembanad Lake, Kerala, India: A remote sensing approach. Environmental Monitoring and Assessment, 195, 1097. https://doi.org/10.1007/s10661-023-11746-0

Xiao, Z., Li, R., Ding, M., Cai, P., Guo, J., Fu, H., Zhang, X., & Song, X. (2024). Unveiling the hidden dynamics of intermittent surface water: A remote sensing framework. Remote Sensing of Environment, 311, 114285. https://doi.org/10.1016/j.rse.2024.114285

Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Crétaux, J.-F., Wada, Y., & Berge-Nguyen, M. (2023). Satellites reveal widespread decline in global lake water storage. Science, 380(6646), 743–749. https://doi.org/10.1126/science.abo2812

Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Yang, K., Crétaux, J.-F., Wang, C., & Minear, J. T. (2024). Leveraging ICESat, ICESat-2, and Landsat for global-scale, multi-decadal reconstruction of lake water levels. Water Resources Research, 60(2). https://doi.org/10.1029/2023WR035721

•