Hydrogeol J (2018) 26:89–108
DOI 10.1007/s10040-017-1636-7

REPORT

Application of Archimedean copulas to the impact assessment of hydro-climatic variables in semi-arid aquifers of western India

Pawan S. Wable[1] · **Madan K. Jha**[1]

Received: 20 January 2017 / Accepted: 30 June 2017 / Published online: 3 August 2017

Abstract The effects of rainfall and the El Niño Southern Oscillation (ENSO) on groundwater in a semi-arid basin of India were analyzed using Archimedean copulas considering 17 years of data for monsoon rainfall, post-monsoon groundwater level (PMGL) and ENSO Index. The evaluated dependence among these hydro-climatic variables revealed that PMGL-Rainfall and PMGL-ENSO Index pairs have significant dependence. Hence, these pairs were used for modeling dependence by employing four types of Archimedean copulas: Ali-Mikhail-Haq, Clayton, Gumbel-Hougaard, and Frank. For the copula modeling, the results of probability distributions fitting to these hydro-climatic variables indicated that the PMGL and rainfall time series are best represented by Weibull and lognormal distributions, respectively, while the non-parametric kernel-based normal distribution is the most suitable for the ENSO Index. Further, the PMGL-Rainfall pair is best modeled by the Clayton copula, and the PMGL-ENSO Index pair is best modeled by the Frank copula. The Clayton copula-based conditional probability of PMGL being less than or equal to its average value at a given mean rainfall is above 70% for 33% of the study area. In contrast, the spatial variation of the Frank copula-based probability of PMGL being less than or equal to its average value is 35–40% in 23% of the study area during El Niño phase, while it is below 15% in 35% of the area during the La Niña phase. This copula-based methodology can be applied under data-scarce conditions for exploring the impacts of rainfall and ENSO on groundwater at basin scales.

Keywords Archimedean copulas · Groundwater vulnerability · El Niño-Southern Oscillation (ENSO) · Climate change · India

✉ Pawan S. Wable
pawan.wable@gmail.com

✉ Madan K. Jha
madan@agfe.iitkgp.ernet.in

[1] Agricultural and Food Engineering Dept., IIT Kharagpur, Kharagpur 721 302, India

Introduction

Groundwater contributes to one-third of the global freshwater supply, which supports a population of over 2 billion (Gorelick and Zheng 2015). In India, groundwater is a source of water supply for more than 80% of the rural and 50% of the urban populations, and for 50% of irrigation demand, which contributes to 70–80% of irrigated production (Mall et al. 2006). There is a consensus among climate scientists that global warming will intensify, accelerate, or enhance the water cycle, which will have important consequences for the world's freshwater resources (UNESCO 2009). Although the effects of climate change on water resources are already visible worldwide, the greatest concern of water experts is its impact on groundwater as this is a more dependable source of water supply for domestic, irrigation and industrial sectors (e.g., Holman 2006; Gurdak et al. 2009; Gorelick and Zheng 2015). Groundwater is affected by climate through major hydrological processes such as precipitation, evapotranspiration, and runoff as well as through interaction with surface-water bodies. The extremes of climate (droughts and floods) are often related to the drivers of climate variability, i.e., large-scale climatic patterns/oscillations such as the North Atlantic Oscillation (NAO), Arctic Oscillation (AO), Pacific Decadal Oscillation (PDO), El Niño Southern Oscillation (ENSO), etc. These large-scale and long-term climatic cycles can have the most discernible impacts on groundwater due to slow aquifer

recharge processes and long aquifer response times (Russo et al. 2014); however, proper assessment of climate-change impacts on groundwater is challenging due to its complex relations with hydro-climatic variables (Mishra and Singh 2010). This challenge is further complicated for data-scarce developing countries.

In the recent past, several researchers ascertained the connection between groundwater and large-scale climate patterns across the world (e.g., Jones and Banner 2003; Fleming and Quilty 2006; Hanson et al. 2006; Luque-Espinar et al. 2008; Gurdak et al. 2009; Tremblay et al. 2011; Perez-Valdivia et al. 2012). Fleming and Quilty (2006) studied this link in the aquifer of southwest British Columbia, Canada, and reported that the groundwater of the area has a significant correlation with seasonal ENSO-related precipitation anomalies, i.e., groundwater levels are higher during La Niña years and lower during El Niño years. Using wavelet and coherence analysis, Tremblay et al. (2011) investigated the links of climatic oscillations (NAO, AO, PDO, Pacific-Northern-America-Pattern, and multivariate ENSO Index) with the groundwater levels of three unconfined aquifers in Canada and concluded that the inter-annual cycles observed in large-scale climatic patterns were also found in groundwater levels, thereby suggesting substantial influence of these climatic patterns on groundwater. All of these studies confirm that there exists a definite linkage between groundwater levels and long-term climatic cycles occurring thousands of kilometers away from the area/basin under study.

India receives around 70–90% of its rainfall from the southwest monsoon and its interannual variability is mostly influenced by the large-scale climatic pattern of ENSO (Revadekar et al. 2012). The ENSO is a climatic phenomenon, which affects global climate variability owing to the interaction between the tropical Pacific Ocean and its surrounding atmosphere. It has two phases, El Niño and La Niña, which are linked to sea surface temperatures and approximately alternates every 2–7 years (IRI 2017). During the warm phase (El Niño), the sea surface temperature is anomalously warm, while in the cool phase (La Niña), it is anomalously cool. In a recent study, it was found that 10 out of the past 13 droughts in India had a high correlation with the El Niño phase (Singh 2014). A study conducted in an arid region of Western Rajasthan (India) revealed that the droughts during El Niño phases were more severe than La Niña phases (Ganguli and Reddy 2013). In another study, Reddy and Ganguli (2012b) applied bivariate copulas and reported that during La Niña (El Niño), there was higher (lower) precipitation and shallow (deeper) groundwater levels in the Manjra River basin of western India.

The preceding reviews suggest that the association between hydro-climatic variables can be studied using emerging tools and techniques. One of such techniques is copula functions, which have been extensively used in hydro-meteorological studies (ICSH 2017). The copula technique has salient features, which are helpful in hydrological studies: (1) it derives joint distributions independent of the marginal, (2) along with the composite likelihood approach, it reduces uncertainty in the estimates of frequency distribution parameters, and (3) it handles non-linearity for modeling dependence between random variables (Genest and Favre 2007; Chowdhary and Singh 2010). Although the copula-based bivariate, trivariate and quadravariate analyses have been used for the frequency analysis of extreme events (Salvadori and De Michele 2004; Shiau et al. 2007; Kao and Govindaraju 2008; Karmakar and Simonovic 2009; Wong et al. 2010), the use of bivariate analysis is mostly preferred due to its simplicity (Klein et al. 2011).

It is also apparent from the aforementioned review that to date, only one study (Reddy and Ganguli 2012b) has applied copula models for the risk assessment of changes

Table 1 Copula cumulative distribution functions, their generator function φ_θ, and the relation of *Kendall's tau* (τ) with parameter θ

Family	$C(u, v)$	φ_θ	Relation of *Kendall's tau* (τ) with θ
Ali-Mikhail-Haq	$\frac{(u,v)}{[1-\theta(1-u)(1-v)]}$	$\ln\left(\frac{[1-\theta(1-t)]}{t}\right)$	$\left(\frac{3\theta-2}{3\theta}\right)-\left(\frac{3\theta-2}{3\theta}\right)\ln(1-\theta)$
Clayton	$(u^{-\theta}+v^{-\theta}-1)^{-1/\theta}$	$\left(\frac{1}{\theta}\right)\left(t^{-\theta}-1\right)$	$\frac{\theta}{(\theta+2)}$
Gumbel-Hougaard	$\exp(-[(-\ln u)^{\theta}+(-\ln v)^{\theta}]^{\frac{1}{\theta}})$	$(-\ln t)^{-\theta}$	$\frac{(\theta-1)}{\theta}$
Frank	$-\left(\frac{1}{\theta}\right)\ln\left[1+\frac{\left(e^{-\theta u}-1\right)\left(e^{-\theta v}-1\right)}{\left(e^{-\theta}-1\right)}\right]$	$-\ln\left(\frac{\left(e^{-\theta t}-1\right)}{\left(e^{-\theta}-1\right)}\right)$	$1+\left(\frac{4}{\theta}\right)\left[D_1(\theta)-1\right]$

Note: $D_k(x)$ is Debye function; for any positive integer k, $D_k(x) = k/x^k \int_0^x t^k(e^t-1)\,dt$

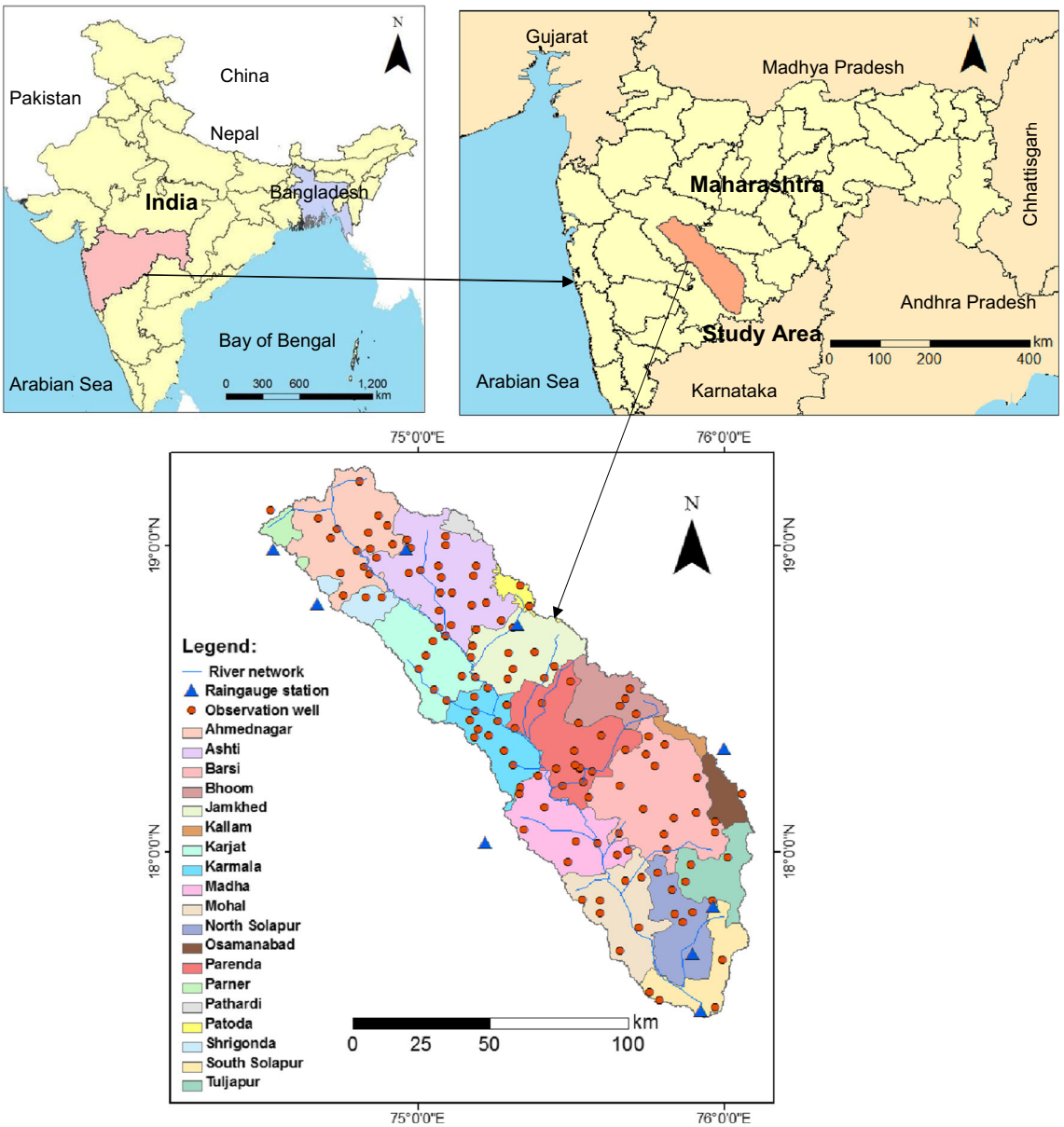

Fig. 1 Location of the study basin along with boundaries of administrative 'blocks' and locations of gauging sites

in hydro-climatic variables on groundwater; however, this study considered only one observation well as a representative for the entire river basin (143.32 km^2), which is not practically appropriate because the findings of the study may not be useful for the entire basin. To address this shortcoming, the present study was conceived to explore the applicability of the copula technique at a larger scale (e.g., basin/sub-basin scale) so as to ensure more realistic findings for the area under study. Also, the studies on groundwater linkage with the ENSO phenomenon are very limited in developing countries in general and the Indian subcontinent in particular (Reddy and Ganguli 2012b; Susilo et al. 2013; Seeboonruang 2014). Given these research gaps and increasing drought incidences in India, this study was carried out to address some of the aforementioned research gaps considering the Sina River basin as a study area, which is located in the semi-arid region of Maharashtra, western India, and it comes under the 'chronically drought-prone area' (PACS 2004). As a result, frequent droughts occur and water scarcity is a serious problem in the area (Chary et al. 2010; DTE 2016; News World India 2016). In addition, to the best of the authors' knowledge, no scientific study has been conducted so far in the study area and the present study is first of its kind in the area. The specific objectives of this study are: (1) to evaluate dependence among hydro-

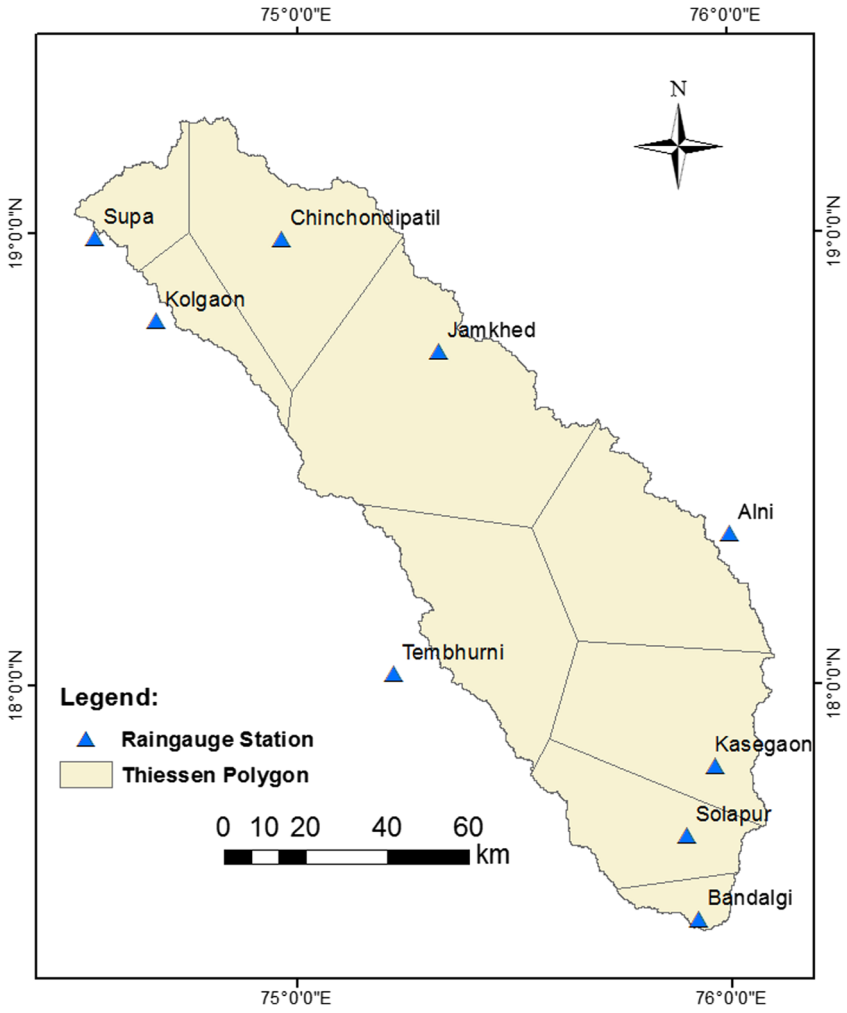

Fig. 2 Locations of raingauge stations and areas of their Thiessen polygons

climatic variables, (2) to model joint dependence between hydro-climatic variables using suitable Archimedean copulas, and (3) to explore the effect of hydro-climatic variables (rainfall and ENSO phenomena) on groundwater using copula-based conditional distributions. The methodology used to fulfill these objectives is an extension of that reported by Reddy and Ganguli (2012b).

Overview of copulas

A copula is a multivariate probability distribution having uniform marginal distribution of random variables. It can represent and model dependence between associated random variables irrespective of their marginal distributions. Sklar's theorem (Sklar 1959) states that every joint distribution F can be expressed as:

$$F_{X_1,X_2.....X_n}(x_1,x_2,\ldots,x_n) = C[F_{X_1}(x_1), F_{X_2}(x_2), \ldots., F_{X_n}] \quad (1)$$

where $F_{X_1,X_2.....X_n}(x_1,x_2,\ldots,x_n)$ stands for the joint cumulative distribution function (CDF) with continuous marginal distributions $F_{X_1}(x_1), F_{X_2}(x_2), \ldots., F_{X_n}$ of the random variables $X_1, \ldots, X_n$ and C is a copula, i.e., a CDF whose margins are uniform on the interval (0, 1).

Table 2 Number of observation wells, area and percentage of total area under the Thiessen polygons of nine raingauge stations

Name of station	No. of observation wells	Area (km^2)	Area (%)
Alni	16	1,827.75	15
Chinchondipatil	23	1,814.56	15
Jamked	39	2,992.47	24
Kasegaon	18	1,625.90	13
Tembhurni	21	1,683.57	14
Supa	4	501.48	4
Solapur	5	1,055.50	9
Kolgaon	3	507.98	4
Bandalgi	3	235.23	2

Table 3 Expressions of probability density function and parameter estimation for parametric and non-parametric distributions used in the study

Distribution	Probability density function	Parameter estimation
Parametric distributions		
Gamma	$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta};\ x > 0$ $\alpha =$ shape parameter, $\beta =$ scale parameter	$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}, \beta = \frac{1}{4A}\left[1 + \sqrt{1 + \frac{1}{4A}}\right]; \alpha = \frac{\beta}{\bar{x}}$
Lognormal	$f_X(x) = \frac{1}{x\sqrt{2\pi\overline{\sigma}^2_y}} \exp\left(-\frac{1}{x\sqrt{2\pi\overline{\sigma}^2_y}}\right); x > 0$ $Y = \ln(X)$ and $-\infty < \overline{\mu}_y < \infty$ $\overline{\mu}_y =$ mean, $\overline{\sigma}_y =$ standard deviation	$\overline{\mu}_y = \frac{\sum y_i}{n}$ $\overline{\sigma}^2_y = \frac{\sum y_i}{n}$
Weibull	$f_X(x) = \alpha x^{\alpha-1}\beta^{-\alpha}\exp[(-x/\beta)^{\alpha}]; x \geq 0, \alpha, \beta > 0$ $\alpha =$ shape parameter, $\beta =$ scale parameter	Estimated by iterative procedure
Nonparametric kernel density-based distributions		
Normal	$K(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2);\ -\infty < x < \infty$ $K(\bullet)$ = Kernel density function	Optimal bandwidth $= \sigma\left({}^{4}/_{3n}\right)^{1/5}$
Quadratic	$K(x) = \frac{3(1-x^2)}{4}$ for $\lvert x\rvert \leq 1$, otherwise $K(x) = 0$	

This study focuses on two variables at a time, i.e., $n = 2$. These variables are denoted X and Y; their joint distribution $F_{X,Y}$ can be expressed in terms of their CDFs F_X and F_Y as follows:

$$F_{X,Y}(x,y) = C\left[F_x(x), F_y(y)\right] = C\left[u, v\right] \tag{2}$$

where C is unique whenever F_X and F_Y are continuous, else uniquely estimated by range $F_X \times$ range F_Y.

Archimedean copulas

In general, the copula C in Eq. (2) is assumed to come from a parametric class. Archimedean copulas, elliptical copulas, and extreme-value copulas families have been applied in hydrological studies. However, Archimedean copulas are most frequently used due to their flexibility and simplicity (e.g., Genest and MacKay 1986; Zhang and Singh 2006; Klein et al. 2011) and, hence, this type was selected for this study. A bivariate copula C is said to be Archimedean if it can be written in the following form:

$$C(u,v) = \phi^{-1}(\phi(\mathrm{u}) + \phi(\mathrm{v})) \tag{3}$$

where the generator ϕ is a function ϕ: [0, 1] → [0, ∞] which is convex, decreasing and such that $\phi(1) = 0$. Its pseudo-inverse is denoted ϕ^{-1}. Various parametric classes of Archimedean copulas are listed in Table 1, in terms of their CDF, generator function and other properties. For more information about copulas and their application, the interested reader can refer to Salvadori and De Michele (2007), Genest and Nešlehová (2012a, b) or Genest and Chebana (2016).

Methodology

Study area

For the present study, the Sina River basin was selected as the study area. This basin is located in Maharashtra, western India (Fig. 1), between 17° 28′ N and 19° 16′ N latitude, 74° 28′ E and 76° 7′ E longitude. The basin has an area of 12,244 km^2, with the topographic elevation ranging from 420 to 964 m (above mean sea level; MSL). It comprises four districts, namely Ahmednagar, Beed, Osmanabad and Solapur, but the largest portion (42%) of the basin falls in Solapur district. The 19 smaller subdivisions, i.e., blocks, for these four administrative districts are shown by different colors in Fig. 1. The average maximum and minimum air temperatures are 40.5 °C in the month of May and 10.5 °C in the month of December, respectively. The rainy season extends from mid-June to the

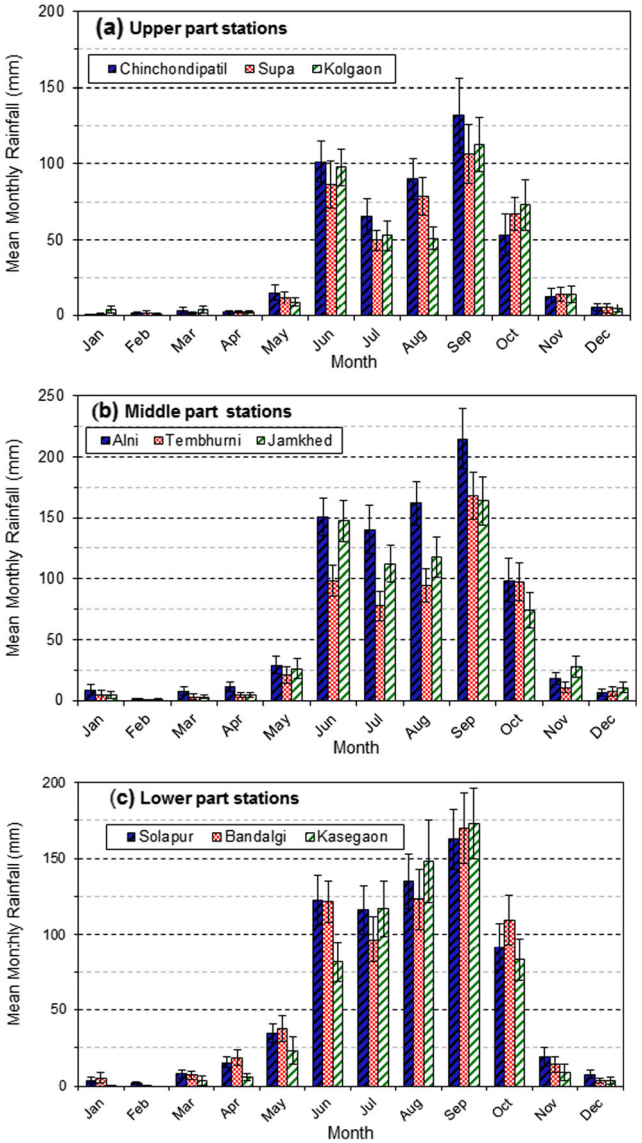

Fig. 3 Temporal variation of mean monthly rainfall for **a** upper, **b** middle, **c** lower part stations of the study area during 1990–2006 period

end of October. The average annual rainfall of the study area is 644 mm; most of the rainfall occurs due to the southwest monsoon.

Geologically, the study area is underlain by Deccan basalts, which are composed of vesicular amygdaloidal basalt and fraction jointed basalt (Deolankar 1980). The water-bearing formations are generally shallow unconfined or semi-confined aquifers in the cover of weathered or fractured upper portions of Deccan basalts**,** along with a patch of local alluvium. The depth of weathered/fractures zones under unconfined conditions ranges from 7.2 to 22.5 m below the ground level. Specific yield (effective porosity) of the unconfined aquifers ranges from 0.010 to 0.026, which indicates relatively low storage capability of the aquifers.

Data

Hydro-meteorological data used in this study were collected from various government organizations/agencies. Daily rainfall data of nine raingauge stations for the period of 1985–2009 were collected from India Meteorological Department (IMD), Pune and State Data Storage Center, Hydrology Project (HP), Nashik, India. It should be noted that the World Meteorological Organization recommendation of 1 rainfall station per 600–900 km^2 for plain areas could not be met, so the rainfall data were supplemented by data from stations in the vicinity (outside) the study area, to better represent the spatial average rainfall within the study area. Pre-monsoon (May month) and post-monsoon (October month) groundwater-level data of 132 sites (observation wells) over the basin for the 1985–2009 period were also acquired from the Groundwater Survey and Development Agency (GSDA), Pune, India. These data are from unconfined aquifers, which are predominant in the study area. The locations of observation wells and raingauge stations are shown in Fig. 1. Groundwater-level data for many sites and for some years are missing from the 1985–2009 dataset; this is a common problem in most developing nations of the world. As a result, the application of time-series analysis techniques under limited-data conditions becomes a challenging task for the researchers of developing nations; therefore, considering the low availability and continuity of time-series groundwater-level data in the study area, the present study was carried out under data-scarce conditions. In this study, 17 years (1990–2006) of groundwater-level and rainfall data have been used to investigate the applicability of the copula technique at larger scale. Thiessen polygons were created using the rainfall stations available in the study area (Fig. 2). The areas of the Thiessen polygons and the number of observation wells falling within each Thiessen polygon are given in Table 2. It is worth also mentioning that in some of the recent studies on copula modeling, limited datasets (15–18 years) have been used (e.g., Durocher et al. 2016; Reddy and Ganguli 2012a).

The impacts of hydro-climatic factors on groundwater are reflected in recharge and discharge processes occurring in a groundwater basin. However, detailed information about these processes are often lacking at a basin scale, especially in the developing world. Generally, groundwater level is monitored and, hence, it is easily available data compared to other components of groundwater. In fact, spatio-temporal variations of groundwater levels in a basin are the outcome of spatially and temporally varying recharge and discharge processes occurring in the basin. Given this fact and the unavailability of other groundwater-related data in the study area, the effects of hydro-climatic factors on groundwater have been explored in this study using groundwater-level data.

The ENSO phenomenon is well represented by a recent index known as ‘multivariate ENSO Index (MEI)’. MEI is defined using the first un-rotated principal component of six observed variables—sea-level pressure, zonal and meridional

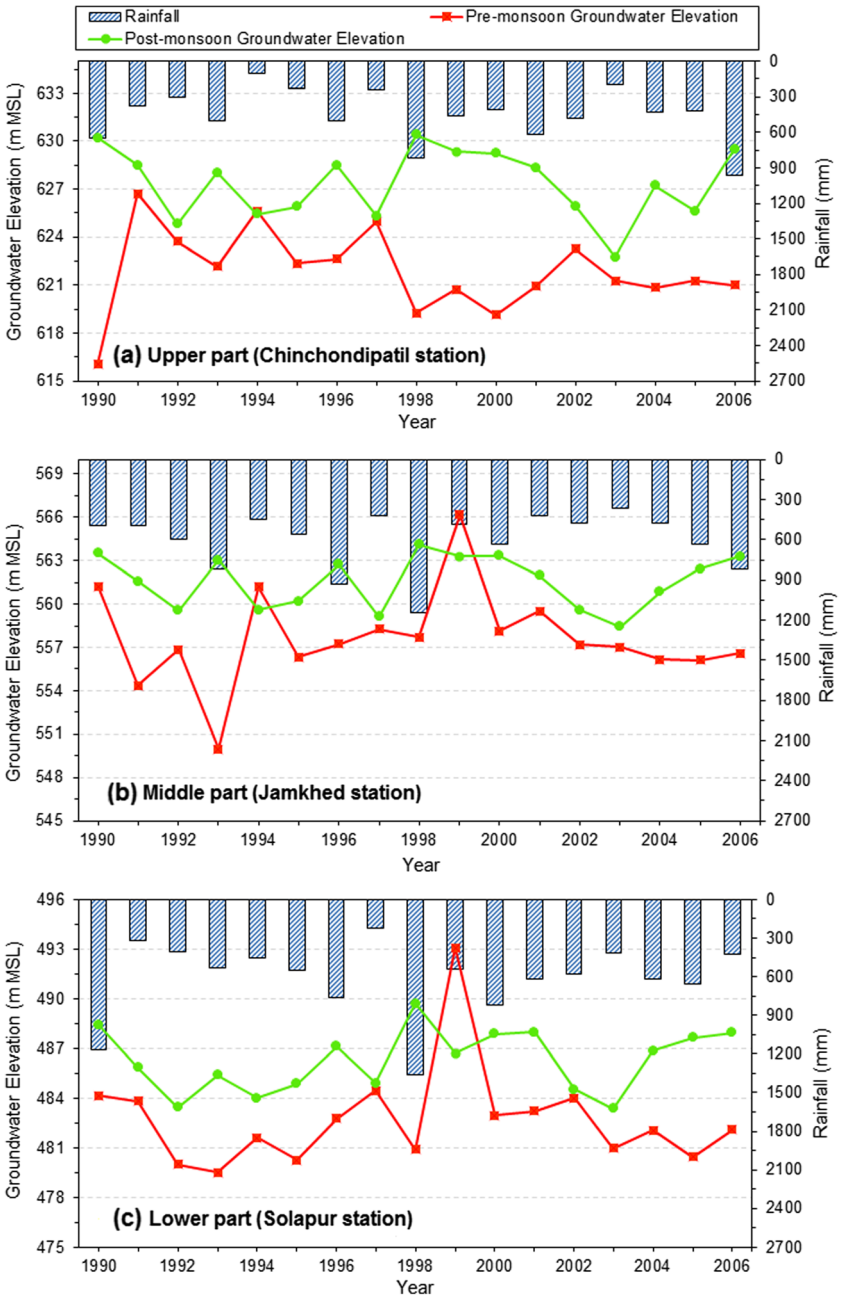

Fig. 4 Pre-monsoon and post-monsoon groundwater-level fluctuations for **a** upper, **b** middle, **c** lower parts of the study area during 1990–2006 period

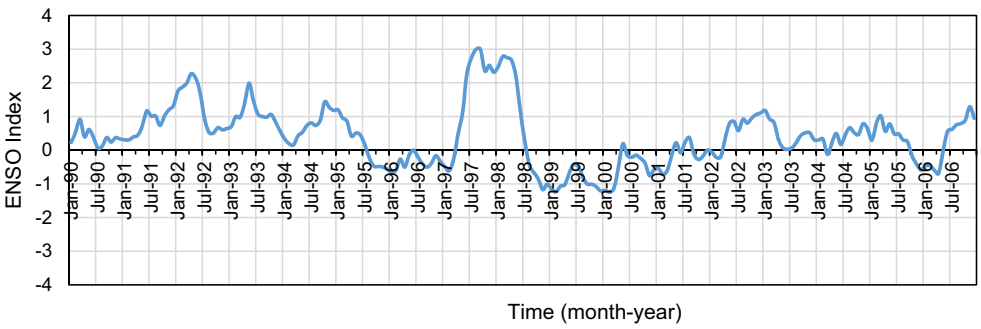

Fig. 5 Monthly variation of ENSO Index during 1990–2006 period

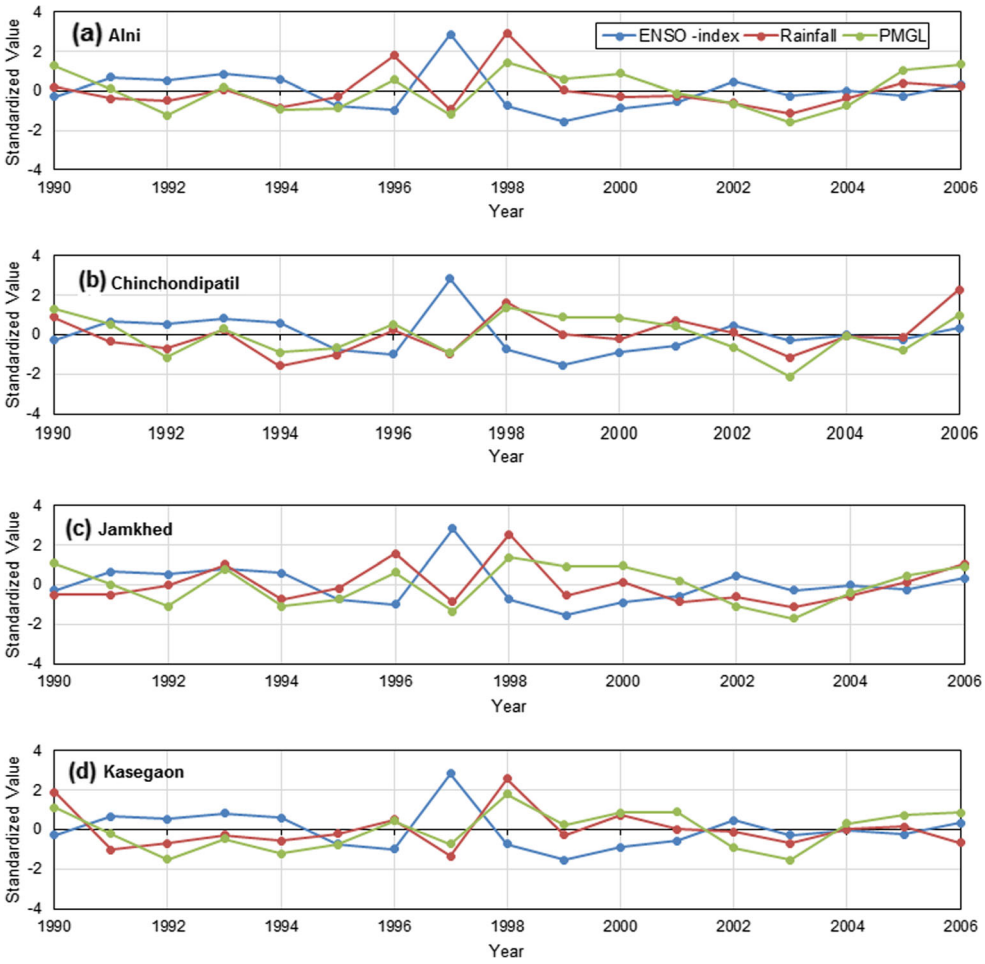

Fig. 6 Variation of standardized values of ENSO Index, post-monsoon groundwater levels (*PMGL*) and monsoon rainfall for the 1900–2006 period at rainfall stations: **a** Alni, **b** Chinchondipatil, **c** Jamkhed, and **d** Kasegaon

components of the surface wind, sea surface temperature, surface air temperature and total cloudiness fraction of the sky over the tropical Pacific (Wolter and Timlin 2011). Monthly MEI values for the study period (1990–2006) were obtained from the database provided by the National Oceanic and Atmospheric Administration (NOAA 2017).

Dependence among the hydro-climatic variables

Before evaluating dependence, the data independency in all the time series was checked by an auto-correlation test (Ljung-Box Q-test). For evaluating dependence, rank-based (and hence scale-free) measures of dependence, such as Spearman's rho (ρ) and Kendall's tau (τ) are preferred over Pearson's correlation coefficient, given that they do not rely on any assumption of linearity between the random variables and are not affected by outliers (Klein et al. 2011). In this study, Spearman's ρ was used to evaluate dependence among hydro-climatic variables at all the nine raingauge stations for the 17 years period (1990–2006). The elevation of post-monsoon groundwater levels (PMGL) was considered instead of depth to groundwater below the ground surface in order to maintain a common datum for all the groundwater-monitoring sites.

Table 4 Summary of *Spearman's rho* (ρ) values for the pairs of *PMGL-Rainfall*, *PMGL-ENSO Index*, and *Rainfall-ENSO Index*

Name of station	Spearman's rho (ρ)		
	PMGL-Rainfall	PMGL-ENSO Index	Rainfall-ENSO Index
Alni	0.86 * ($p \leq 0.01$)	−0.43 ($p = 0.09$)	−0.51** ($p = 0.04$)
Chinchondipatil	0.78* ($p \leq 0.01$)	−0.47 ($p = 0.06$)	−0.29 ($p \geq 0.10$)
Jamked	0.68* ($p \leq 0.01$)	−0.53** ($p = 0.03$)	−0.29 ($p \geq 0.10$)
Kasegaon	0.70* ($p \leq 0.01$)	−0.47 ($p = 0.06$)	0.68* ($p \leq 0.01$)
Tembhurni	0.76* ($p \leq 0.01$)	−0.54** ($p = 0.03$)	−0.40 ($p \geq 0.10$)
Supa	0.59** ($p = 0.02$)	−0.28 ($p \geq 0.10$)	−0.58** ($p = 0.02$)
Solapur	0.63** ($p = 0.01$)	−0.53** ($p = 0.03$)	−0.32 ($p \geq 0.10$)
Kolgaon	0.86* ($p \leq 0.01$)	−0.39 ($p \geq 0.10$)	−0.45 ($p = 0.07$)
Bandalgi	0.33 ($p \geq 0.10$)	−0.66** ($p = 0.05$)	−0.17 ($p \geq 0.10$)

Note: Values *in brackets* show *p*-value; values with * indicate significant dependence at 1% level of significance; values with ** indicate significant dependence at 5% level of significance

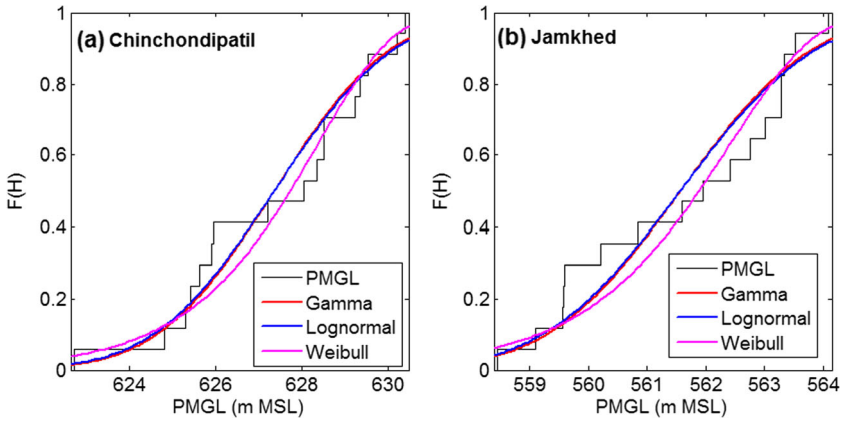

Fig. 7 Cumulative distribution function of gamma, lognormal, and Weibull distributions fitted to post-monsoon groundwater levels (*PMGL*) in the zones of rainfall stations: **a** Chinchondipatil and **b** Jamkhed

In standard climatology, variables affected by large-scale climatic patterns should be averaged over the area (Fleming and Quilty 2006); hence, to study the effect of monsoon rainfall and the ENSO phenomenon on groundwater levels, post-monsoon groundwater levels for the observation wells of a particular Thiessen polygon were averaged. To perform this analysis, the cumulative monsoon rainfall and the average of monthly MEI values for the period June to October were used. The presence of dependence between each pair of hydro-climatic variables was examined at the 1 and 5% levels of significance based on the *p*-values of the standard two-tailed t-test. It is worth mentioning that to reject the null hypothesis, the *p*-value should be less than or equal to the level of significance (α). For a visual illustration, the variation of hydro-climatic variables over their standardized value was also plotted.

Fitting marginal distributions to hydro-climatic variables

After evaluating the dependence, marginal distributions were fitted to each of the variables. For PMGL and rainfall, the most popular parametric distributions were used, namely gamma (GM), lognormal (LN) and Weibull (WB); however, for the ENSO Index, non-parametric kernel-density-based normal and quadratic distributions were considered because parametric distributions do not fit climate indices properly (Reddy and Ganguli 2012b). The probability density functions and parameter estimates for the parametric and non-parametric kernel-density-based distributions are shown in Table 3. In all cases, the estimates were obtained using the method of maximum likelihood. The best distribution was selected based on selected univariate statistical indicators—root mean square error (RMSE), Akaike information criterion (AIC), and Kolmogorov-Smirnov (KS) test—and a graphical indicator (cumulative distributive function plot).

Archimedean copulas for modeling dependence

A priori, the choice of parametric Archimedean copulas families as possible models for the dependence between hydro-climatic variables is guided by the range of association they allow. The Clayton and Gumbel-Hougaard copulas are used if the dependence is positive, whereas Ali-Mikhail-Haq and

Table 5 Performance evaluation of different probability distributions fitted to PMGL

Name of station	AIC			RMSE			KS-test		
	GM	LN	WB	GM	LN	WB	GM	LN	WB
Alni	−57.51	−57.54	*−58.09*	0.07299	0.06905	0.07817	0.1561	0.1497	0.1748
Chinchondipatil	−69.83	*−69.85*	−68.83	0.07309	0.07020	0.07425	0.1582	0.1541	0.1889
Jamked	−63.77	*−63.81*	−62.72	0.08630	0.08271	0.08328	0.1599	0.1532	0.1604
Kasegaon	−64.84	−64.87	*−66.04*	0.06652	0.06328	0.06643	0.1271	0.1218	0.1235
Tembhurni	−70.38	−70.38	*−72.75*	0.06409	0.06137	0.08117	0.1560	0.1495	0.1545
Supa	−69.67	−69.70	*−69.89*	0.06911	0.06519	0.07549	0.1348	0.1290	0.1364
Solapur	−72.67	*−72.72*	−70.80	0.04160	0.04280	0.05189	0.1159	0.1104	0.1042
Kolgaon	−66.80	*−66.83*	−66.28	0.05566	0.05361	0.05448	0.1584	0.1548	0.1051
Bandalgi	−71.60	−71.63	*−72.21*	0.06895	0.06581	0.08279	0.1699	0.1627	0.1568

Note: *KS* critical value at 5% level of significance = 0.330; *italic* value indicate the best value for particular evaluation criteria

Table 6 Parameters of the probability distributions fitted to PMGL

Name of station	GM		LN		WB	
	α	β	μ	σ	α	β
Alni	145,269.9	0.0039	6.333	0.0027	432.5	563.8
Chinchondipatil	87,448.2	0.0072	6.442	0.0035	356.8	628.4
Jamked	99,977.8	0.0056	6.331	0.0033	386.4	562.4
Kasegaon	70,408.9	0.0069	6.187	0.0039	289.4	487.2
Tembhurni	54,545.3	0.0092	6.221	0.0044	243.1	504.5
Supa	95,811.4	0.0068	6.483	0.0033	356.8	654.9
Solapur	40,973.9	0.0114	6.146	0.0051	250.1	468.1
Kolgaon	99,020.4	0.0062	6.415	0.0033	370.4	611.8
Bandalgi	40,987.5	0.0110	6.115	0.0051	228.8	453.7

Frank copulas are applied for modeling both positive and negative dependence. The Frank copula can model the entire range of dependence values [−1, +1], whereas the Ali-Mikhail-Haq family of copulas is only suitable for weakly dependent variables (Nelsen 2006). There exists a connection between a rank-based non-parametric measure of dependence called Kendall's τ and Archimedean copulas generators, which is given as follows (Genest and MacKay 1986):

$$\tau = 1 + 4\int \frac{\phi(t)}{\phi'} dt \tag{4}$$

where ϕ' denotes the derivative of ϕ with respect to t. This relation can be used to estimate the parameter θ of an Archimedean copula by the method of moments, which consists of replacing τ by an estimate thereof in Eq. (4) and solving for θ for any given choice of Archimedean copulas; thus, paired random variables can be modeled through copulas by preserving their mutual dependence. In this study, four families of Archimedean copulas (Clayton, Gumbel-Hougaard, Ali-Mikhail-Haq and Frank) were applied. The expression of the generator function for each copula family with its derivative, together with the relation of Kendall's τ with copula parameter θ, are presented in Table 1. Copula modeling was performed using MATLAB software.

Goodness-of-fit tests for selecting copulas

Goodness-of-fit tests can be used to check whether a specific copula family fits the data at hand. In this study, both graphical and statistical indicators were used to assess the fitness of Archimedean copulas.

Graphical diagnostics

In order to assess the fit of a given Archimedean copulas family C_θ, 1,000 observations were generated from C_θ after estimating its parameter. These pseudo-observations were then transformed back into the variables' original units using the inverses of the marginal distribution F_X and F_Y. The scatter plot of the resulting pairs was then visualized and compared to the original data. Algorithms to generate random pairs from different copula families (C_θ) can be found in Whelan (2004) and Genest and Favre (2007).

Statistical indicators

Apart from the graphical diagnostics, three statistical indicators for bivariate copulas were used in this study, namely RMSE; AIC; and KS goodness-of-fit test. Detailed descriptions of these statistical indicators can be found in Klein et al. (2011).

Effect of rainfall and the ENSO phenomenon on groundwater

In order to study the impacts of rainfall and the ENSO phenomenon on groundwater, the copula-based conditional distribution probabilities of PMGL ≤ $PMGL_{avg}$ for average and non-average monsoon rainfall scenarios as well as for ENSO

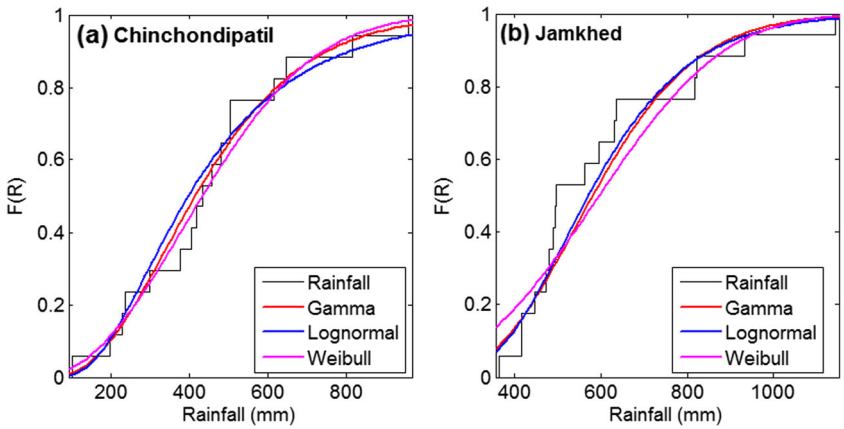

Fig. 8 Cumulative distribution function of gamma, lognormal, and Weibull distributions fitted to rainfall at stations: **a** Chinchondipatil and **b** Jamkhed

Table 7 Performance evaluation of different probability distributions fitted to rainfall

Name of station	AIC			RMSE			KS-test		
	GM	LN	WB	GM	LN	WB	GM	LN	WB
Alni	−231.3	−229.5	−235.9	0.1006	*0.08619*	0.1238	0.1733	0.1519	0.2200
Chinchondipatil	−225.3	−226.8	−225.2	*0.05393*	0.06218	0.05915	0.1337	0.1694	0.1424
Jamked	−221.3	−219.9	−225.0	0.1038	*0.09247*	0.1139	0.2152	0.2022	0.1998
Kasegaon	−231.3	−230.6	−233.7	0.07418	*0.06115*	0.09350	0.1458	0.1227	0.1805
Tembhurni	−221.1	−221.2	−222.3	0.04689	*0.04316*	0.05545	0.09304	0.1027	0.08620
Supa	−205.9	−205.3	−208.5	0.07675	*0.07048*	0.09537	0.1708	0.1510	0.2057
Solapur	−227.8	−226.7	−230.7	0.07545	*0.06410*	0.09496	0.1422	0.1251	0.1851
Kolgaon	−213.5	−214.0	−213.6	0.07007	*0.06337*	0.07461	0.1387	0.1484	0.1426
Bandalgi	−220.3	−219.3	−223.5	0.07467	*0.06638*	0.08815	0.1379	0.1266	0.1469

Note: *KS* critical value at 5% level of significance = 0.3298; *italic* values indicate the best value for particular evaluation criteria

phases were determined from the following equation (Zhang and Singh 2006; Reddy and Ganguli 2012b):

$$F_{X|\,Y\le y} = C_{U\,|\,V\le v} = \frac{C(u,v)}{v} \tag{5}$$

Further, the spatial variation of these probabilities over the study area was analyzed by generating probability maps using ArcGIS software.

Results and discussion

Preliminary data analysis

Rainfall characteristics

The monthly variation of rainfall for stations in upper, middle, and lower parts of the study area is shown in Fig. 3a–c. Maximum amount of rainfall in the study area is confined to five monsoon months, i.e., from June to October. For all rainfall stations, among monsoon months, the maximum amount of rainfall is received in the month of September. In upper, middle, and lower parts of the study area, the maximum rainfall is received for Chinchondipatil, Alni and Solapur stations, respectively. The variation of the monthly rainfall over the period of 25 years for that particular month is shown by standard error bars in Fig. 3a–c. All stations have the highest standard error in the month of September, except Kasegaon station, which has it in month August. Minimum standard errors are found in the months of January and February at all the stations.

Table 8 Parameters estimated of the probability distributions fitted to rainfall

Name of station	GM		LN		WB	
	α	β	μ	σ	α	β
Alni	8.756	86.162	6.568	0.3359	2.658	848.1
Chinchondipatil	4.048	111.452	5.983	0.5574	2.235	510.0
Jamked	10.160	59.220	6.350	0.3153	3.019	673.3
Kasegaon	5.571	110.483	6.330	0.4401	2.323	696.7
Tembhurni	7.303	70.521	6.174	0.3910	2.868	578.5
Supa	10.338	37.331	5.906	0.3201	3.200	430.5
Solapur	7.348	85.465	6.373	0.3765	2.628	707.9
Kolgaon	5.536	65.816	5.805	0.4572	2.645	411.5
Bandalgi	10.578	56.291	6.341	0.3121	3.150	664.8

Groundwater characteristics

The pre- and post-monsoon groundwater-level elevation time series data for Chinchondipatil, Jamkhed and Kasegaon

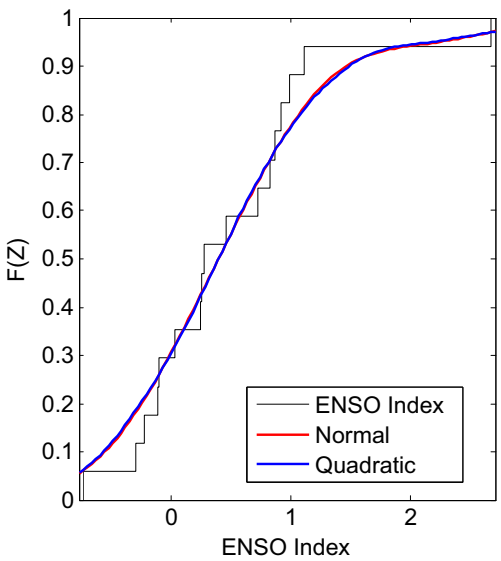

Fig. 9 Cumulative distribution function of non-parametric normal and quadratic distributions fitted to the ENSO Index

Table 9 Performance evaluation of different probability distributions fitted to ENSO Index

AIC		RMSE		KS-test	
Normal	Quadratic	Normal	Quadratic	Normal	Quadratic
−32.54	−32.44	*0.05989*	0.06200	*0.1249*	0.1297

Note: *KS* critical value at 5% level of significance = 0.3298; *italic* values indicate the best value for particular evaluation criteria

stations representing, respectively, the upper, middle, and lower parts of the study area are plotted in Fig. 4a–c along with the annual rainfall time series data. The pre-monsoon groundwater-level elevations for Chinchondipatil, Jamkhed and Kasegaon stations are 616.09–626.73, 549.99–566.16, 479.52–493.07 m MSL, and the post-monsoon groundwater-level elevations are in the range of 622.73–630.39, 558.45–564.09, 483.38–489.71 m MSL, respectively. These values clearly show the response of post-monsoon groundwater-level elevation to the variation of rainfall, i.e., post-monsoon groundwater-level elevation increases with increase in rainfall and vice-versa. In case of pre-monsoon groundwater-level elevation, a sudden peak is observed in year 1999 for Jamkhed and Kasegaon stations, which is attributed to recharge from the maximum rainfall in the previous year (i.e., 1998).

ENSO Index

The monthly variation of the ENSO index during 1990–2006 period is shown in Fig. 5. For the ENSO phases during the considered period, the top 30th percentile of ENSO index values represents El Niño years, whereas the bottom 30th percentile of ENSO Index values represents La Niña years, and the remaining as neutral years (Wolter and Timlin 2011). Accordingly, 1995–1996 and 1998–2000 years indicates La Niña years; 1990 and 2001–2006 years denote neutral years and the remaining 5 years, i.e., 1991–1994, 1997, were El Niño years.

Table 10 Parameters of different copulas fitted to PMGL-Rainfall pairs

Name of station	Clayton	Frank	Gumbel-Hougaard
Alni	4.476	11.019	3.238
Chinchondipatil	2.973	7.871	2.486
Jamked	2.075	5.907	2.038
Kasegaon	2.387	6.602	2.194
Tembhurni	2.857	7.624	2.429
Supa	2.000	5.736	2.000
Solapur	2.000	5.736	2.000
Kolgaon	4.182	10.411	3.091
Bandalgi	0.519	1.920	1.259

Evaluating dependence among the hydro-climatic variables

The auto-correlation test revealed that there is no significant time autocorrelation at the 5% level of significance, thereby suggesting that each time series is independent during the study period. However, it is apparent from Fig. 6a–d that all the hydro-climatic variables, i.e., PMGL, rainfall and ENSO Index are cross-correlated to one another. For brevity, the graphs of four selected stations are shown in Fig. 6a–d as an example and the dependence measured using Spearman's ρ is presented in Table 4. The evaluation of dependence indicated that there is positive dependence in the PMGL-Rainfall pair, which means increase in rainfall increases the PMGL. On the other hand, PMGL-ENSO Index and Rainfall-ENSO Index are negatively associated with each other. These relationships suggest that there will be a decrease in the PMGL as well as rainfall with increase in the ENSO Index values.

For the PMGL-Rainfall pair, a high level of dependence was found to be significant for all the stations, except at the Bandalgi station, which is located in the downstream portion of the study area. This lower dependence between PMGL and Rainfall at the Bandalgi station could be attributed to concentrated runoff (overland flow) at the downstream end. For the PMGL-ENSO Index pair, only Jamkhed, Tembhurni, Solapur and Bandalgi stations, which cover 49% of the study area (6,185 km^2), exhibited statistically significant negative dependence (Table 4). There is no statistically significant dependence for the remaining five stations and, hence, in the areas covered by these stations, the relationship between PMGL and ENSO can only be used for qualitative predication (high or low PMGL). This insignificant dependence may be attributed to other climatic oscillations (Jones and Banner 2003). For the Rainfall-ENSO Index pair, only three stations (Alni, Kasegaon and Supa) that cover 32% of the study area showed statistically significant negative dependence; hence, this pair was not considered in subsequent analyses.

Identifying marginal distribution for fitting hydro-climatic variables

The performance evaluation for the distribution fitting of PMGL at all the stations was carried out using cumulative distribution function (CDF) plots and statistical indicators as shown in Fig. 7a,b and Table 5, respectively. Table 6 summarizes the estimated parameters of GM, LN, and WB distributions. Upon visually assessing CDF fit for the PMGL time series (Fig. 7a,b) for different stations and AIC criteria (Table 5), it can be seen that the WB distribution provides a better fit than GM and LN distributions. For the rainfall time series, CDF plots (Fig. 8a,b) and RMSE values (Table 7) suggest that it is better represented by the LN distribution compared to WB and GM distributions. The parameter estimates

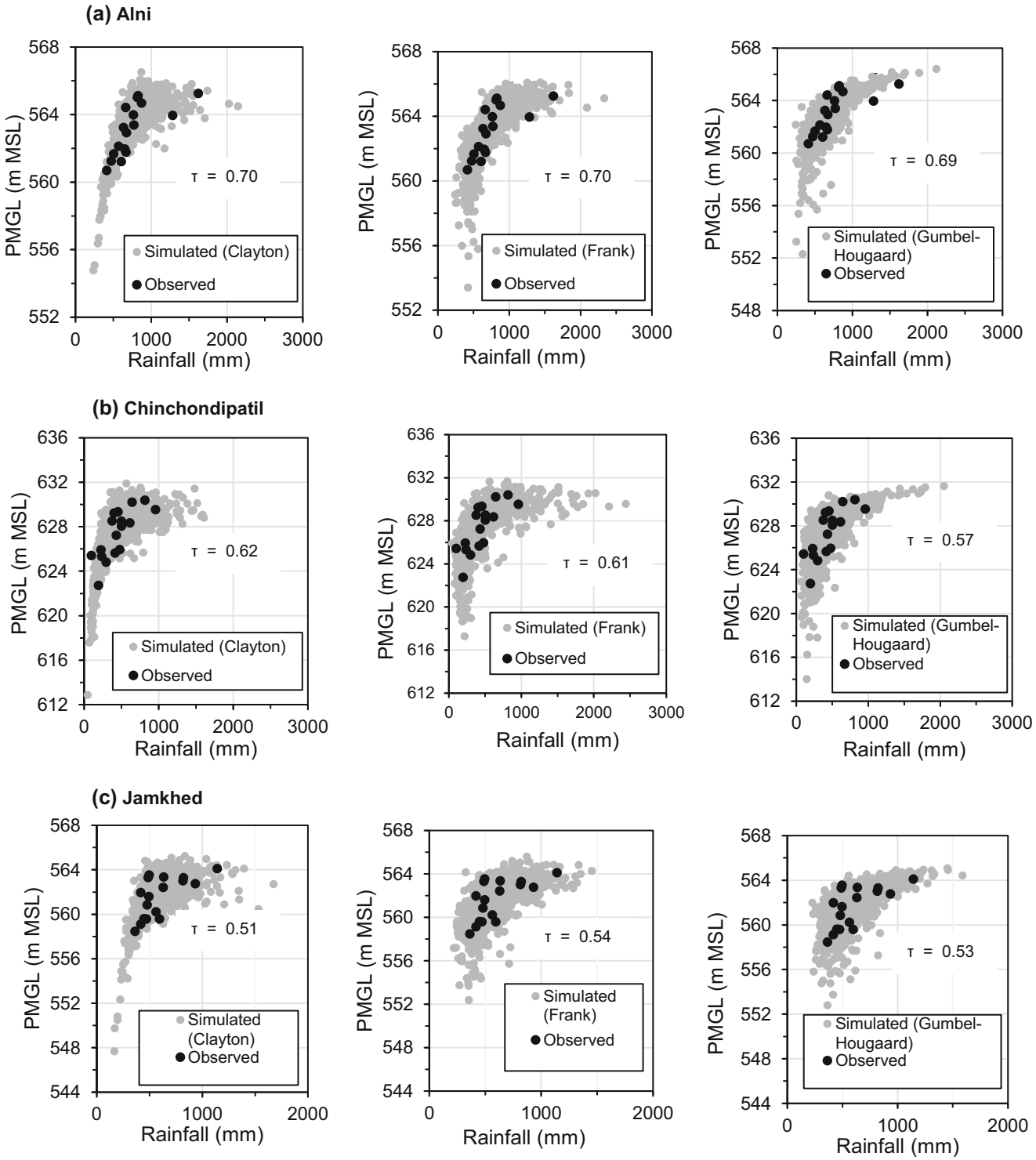

Fig. 10 Scatter plots of observed (*black dots*) versus 1,000 simulated (*gray dots*) samples using Clayton, Frank and Gumbel-Hougaard copulas for the PMGL-Rainfall pair at **a** Alni, **b** Chinchondipatil, and **c** Jamkhed stations. In this figure, Kendall's tau (τ) value is shown for simulated samples

for the GM, LN, and WB distributions, fitted rainfall time series are given in Table 8.

Moreover, the CDF plot for the ENSO Index is depicted in Fig. 9, which reveals that both the non-parametric kernel-based normal and quadratic distributions performed nearly the same; the KS-test also supports both the distributions (Table 9). However, the statistical evaluation confirmed that the ENSO Index is best fitted by the 'non-parametric kernel-

Table 11 *Kendall's tau (τ)* values for the PMGL-Rainfall and PMGL-ENSO Index pairs

Name of station	Kendall's tau (τ)	
	PMGL-Rainfall	PMGL-ENSO Index
Alni	0.6912	−0.2500
Chinchondipatil	0.5978	−0.3235
Jamked	0.5092	−0.3529
Kasegaon	0.5441	−0.2941
Tembhurni	0.5882	−0.3382
Supa	0.5000	−0.2206
Solapur	0.5000	−0.3529
Kolgaon	0.6765	−0.2500
Bandalgi	0.2059	−0.4412

based normal distribution' with lower values of AIC (−32.54) and RMSE (0.05989) as shown in Table 9. The optimal bandwidth is the only parameter for the two non-parametric kernel-based distributions and its value is estimated as 0.4455.

Selecting suitable copula for modeling dependence

As mentioned in the previous section, the PMGL and rainfall time series followed different distributions and hence, the traditional bivariate distribution cannot be used for dependence modeling. Even if in the case of same marginal distribution for the PMGL and rainfall time series, copula function is preferred to the traditional bivariate distribution due to its better performance (Ganguli and Reddy 2012). Hence, the dependence between PMGL and rainfall is modeled using a copula function, because it does not need the condition of random variables to follow the same marginal distribution family. As the pair of PMGL-Rainfall exhibited highly positive dependence ($p < 0.01$), an attempt was made to capture their dependence using Clayton (Cl), Frank (Fr) and Gumbel-Hougaard (GH) copula models. The estimates of copula model parame-

Table 12 Results of goodness-of-fit tests for different the PMGL-Rainfall copula models

Name of station	AIC			RMSE			KS test		
	Cl	Fr	GH	Cl	Fr	GH	Cl	Fr	GH
Alni	22.59	18.22	*11.97*	*0.02386*	0.02858	0.02981	0.1521	0.1533	0.1575
Chinchondipatil	*10.01*	16.79	16.11	*0.03309*	0.03913	0.03875	0.2240	0.2352	0.2477
Jamked	12.40	*10.74*	11.77	*0.01791*	0.02699	0.03126	0.2695	0.2708	0.2798
Kasegaon	*7.176*	13.74	17.43	*0.03895*	0.04136	0.04255	0.1164	0.1256	0.1366
Tembhurni	13.00	12.94	13.28	*0.02709*	0.03420	0.03516	0.1569	0.1588	0.1716
Supa	*9.875*	10.36	11.85	*0.02740*	0.03018	0.03161	0.1458	0.1681	0.1826
Solapur	13.45	*10.46*	15.42	*0.02820*	0.03759	0.03945	0.1832	0.1861	0.1995
Kolgaon	24.61	26.50	*23.18*	*0.02547*	0.03718	0.04262	0.1710	0.1792	0.1810
Bandalgi	*2.592*	4.044	5.986	0.03326	*0.03045*	0.02741	0.1544	0.1639	0.1736

Note: *KS* critical value at 5% level of significance = 0.3298; *italic* values indicate the best values for corresponding criterion

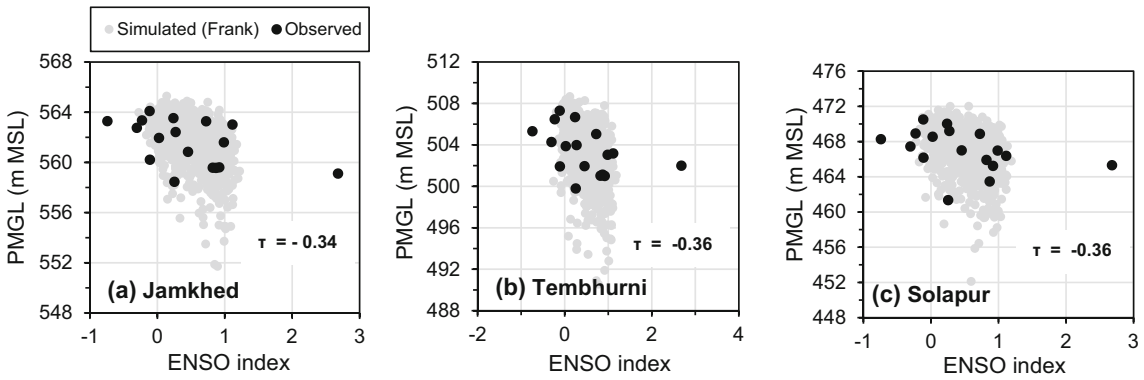

Fig. 11 Scatter plots of observed (*black dots*) versus 1,000 simulated (*gray dots*) samples using Frank copula for the PMGL-ENSO index pair at **a** Jamkhed, **b** Tembhurni, and **c** Solapur stations

Table 13 Estimated parameters and KS-test values of the Frank copula for PMGL-ENSO Index pairs

Name of station	Parameter estimated	KS-test
Alni	−2.372	0.1280
Chinchondipatil	−3.189	0.1712
Jamked	−3.545	0.2193
Kasegaon	−2.851	0.1535
Tembhurni	−3.365	0.1636
Supa	−2.068	0.2631
Solapur	−3.545	0.1136
Kolgaon	−2.372	0.1882
Bandalgi	−4.758	0.1510

Note: *KS* critical value at 5% level of significance = 0.3298

ters for the PMGL-Rainfall pair are shown in Table 10. The scatter plots of observed and simulated data from the three fitted copula models are shown in Fig. 10a–c, together with the Kendall's τ values computed from simulated samples for the three copula models. It is evident that the random pairs generated by all three copula models (shown as gray dots) are well intertwined with the observed data (shown as black dots). Furthermore, the values of Kendall's τ for the simulated data are close to those of the observed data (Fig. 10a–c and Table 11). However, the Clayton copula better simulates the trend of the observed data compared to the other two copulas. In addition, all the statistical indicators (Table 12) also confirm that the Clayton copula is a better choice among the three copula families considered. It should be noted from Fig. 10a–c that the upper bound appears for groundwater levels and rainfall dependence. For this, the estimated non-parametric upper tail dependence coefficient for all stations in the study area is found to be varying from 0.30 to 0.72. Also, at a certain threshold of high rainfall, very weak dependency exists between PMGL and rainfall. In fact, the rate of recharge is a function of depth to the water table. Hence, when the water table reaches the threshold value, the recharge rate is drastically reduced, which causes less dependency on rainfall.

For the PMGL-ENSO Index pair, a negative dependence was found (Table 4). Therefore, they were only modeled by using the Frank copula, which is applicable to the entire range of dependence [−1, +1]. The scatter plots show a good overlap and close Kendall's τ values between the observed data and

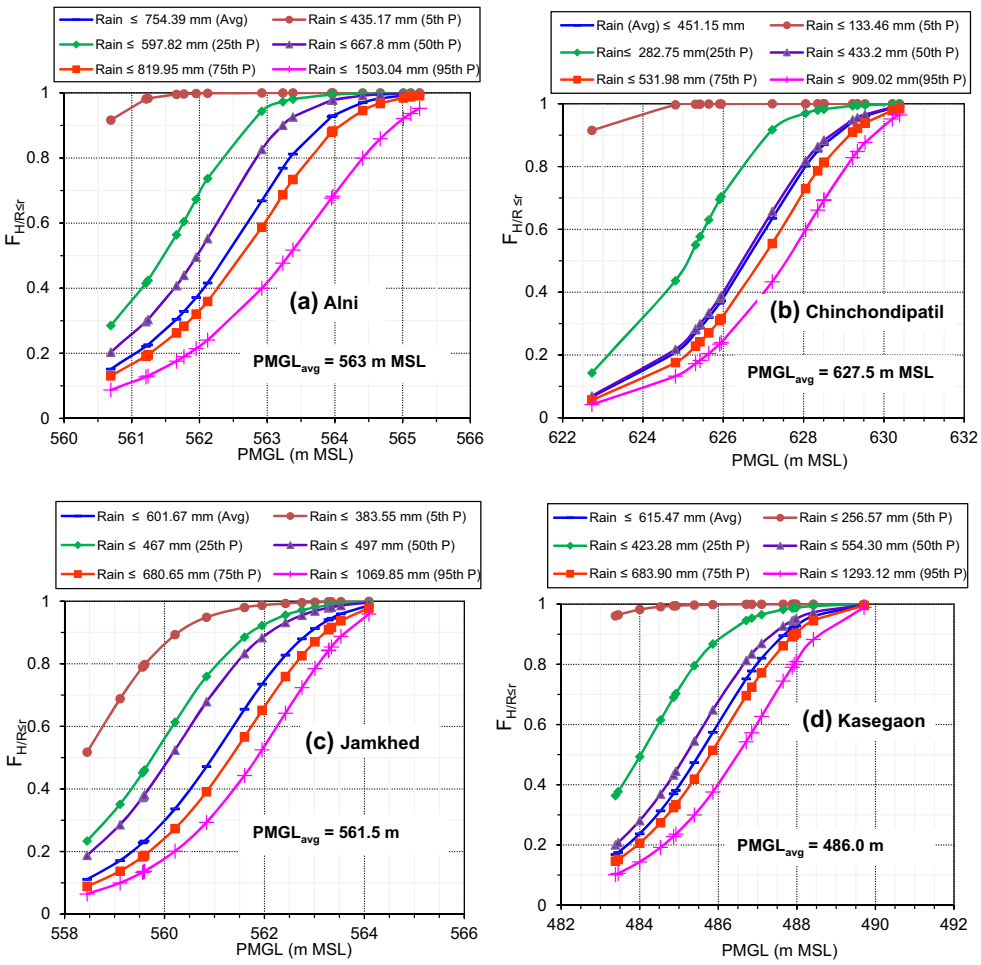

Fig. 12 Conditional probabilities of post-monsoon groundwater levels (*PMGL*) for given average and non-average rainfall scenarios (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles) at four stations: **a** Alni, **b** Chinchondipatil, **c** Jamkhed, and **d** Kasegaon. In this figure, *P* denotes percentile of rainfall

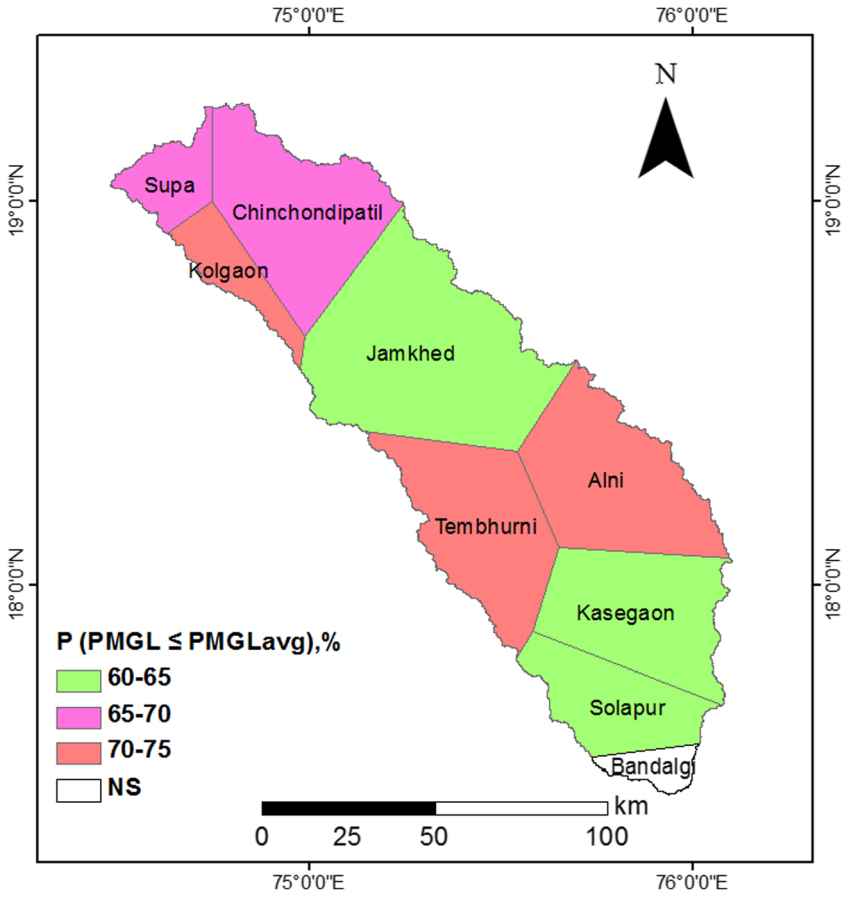

Fig. 13 Probability (P) map of $PMGL \leq PMGL_{avg}$ for a given average rainfall scenario. *NS* non-significant dependence

the pseudo-sample generated by the Frank copula for all the stations (Fig. 11a–c and Table 11). The results of the KS-test (Table 13) also suggest that the dependence in the PMGL-ENSO Index pair is adequately captured by the Frank copula. The estimates of the parameters of the Frank copula fitted to the PMGL-ENSO Index pair are shown in Table 13. This choice of copula model corroborates the earlier study reported by Reddy and Ganguli (2012b) in which depth-to-groundwater data were considered instead of PMGL.

Impacts of rainfall and the ENSO phenomenon on groundwater

Impacts of rainfall

In order to study the effects of rainfall on groundwater, the graphs of the Clayton-copula-based conditional distribution probabilities of PMGL for given average and non-average (5^{th}, 25^{th}, 50^{th}, 75^{th} and 95^{th} percentiles) rainfall conditions were prepared for four rainfall stations as an example (Fig. 12a–d). Obviously, for a given average rainfall, the probability of PMGLs of lower magnitudes is smaller, whereas that of PMGLs of higher magnitudes is greater. This can be explained by considering Chinchondipatil station as shown in Fig. 12b. The PMGL values of 624 and 628 m are respectively lower and higher magnitude in the zone/area represented by this station. For a given average rainfall, the probability of PMGL being less than or equal to 624 m (MSL) is 15%, whereas that of PMGL less than or equal to 628 m (MSL) is nearly 80% (Fig. 12b). These probability values are not symmetrical when two different values of PMGL are considered. However, if only one value of PMGL is considered, then symmetry (x% and 100–x%) of probability values will exist—for example, the probability of PMGL being less than or equal to 624 m (MSL) is 15%, whereas that of PMGL being greater than 624 m (MSL) is 85%.

Furthermore, based on Fig. 12a–d, a spatial map of the probability of $PMGL \leq PMGL_{avg}$, i.e., probability of non-exceedance for a given average rainfall scenario, is generated as shown in Fig. 13. It can be seen from Fig. 13 that for a given average rainfall, the conditional probability of $PMGL \leq PMGL_{avg}$ is above 70% for the areas/zones covered by Alni, Tembhurni and Kolgaon stations, which encompass about 33% of the study area (4,019 km^2). It indicates that PMGL in these areas (Barshi and Madha blocks, and some parts of nearby blocks) will be much lower than its average value and,

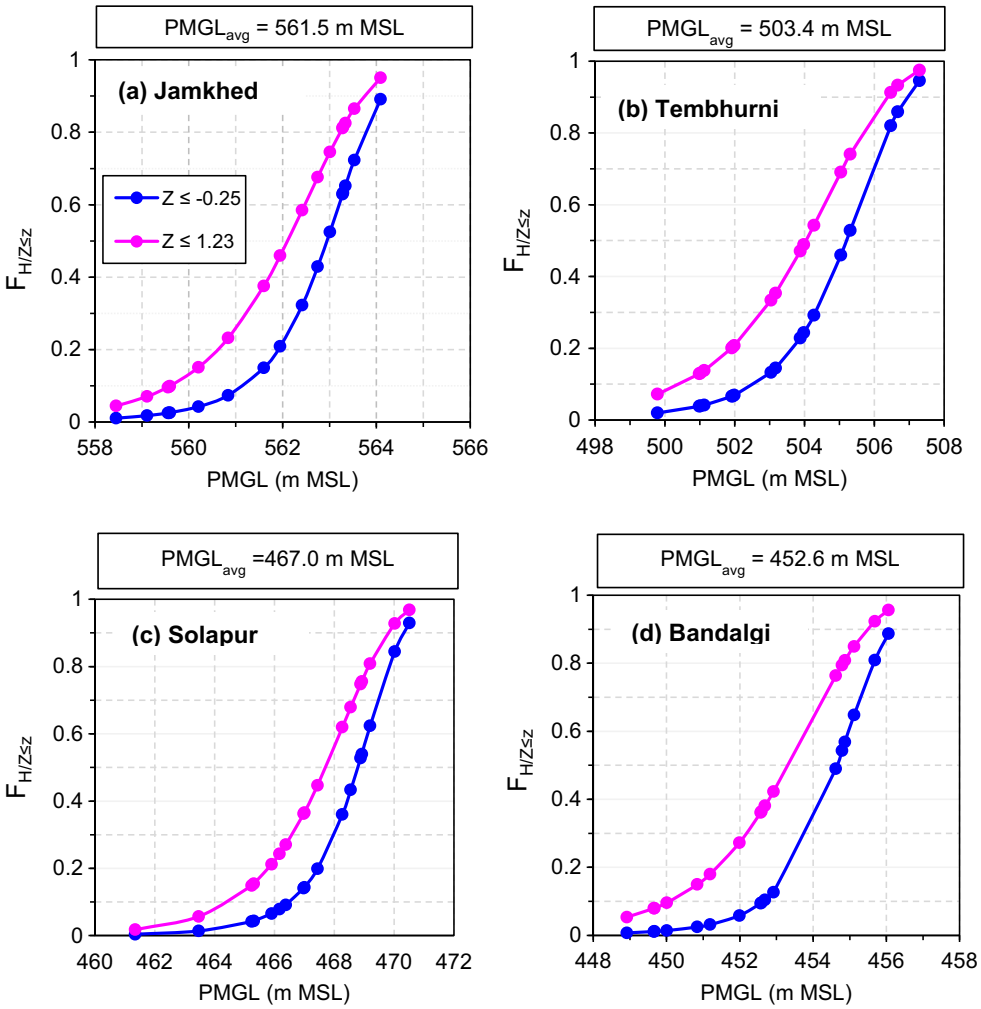

Fig. 14 Conditional probability of PMGL ≤ PMGL$_{avg}$ for the ENSO Index $Z \leq -0.25$ for La Niña phase and $Z \leq 1.23$ for El Niño phase at stations **a** Jamkhed, **b** Tembhurni, **c** Solapur, and **d** Bandalgi

hence, the groundwater of these areas should be managed with a high priority or an alternative water source should be utilized. In addition, it is recommended to propose rainwater harvesting and artificial recharge structures in these areas. The conditional probabilities of PMGL ≤ PMGL$_{avg}$ for a given average rainfall are found in the range of 65–70% for the areas (2,316 km^2) covered by Chinchondipatil and Supa stations, which suggests a moderate groundwater scenario under average rainfall conditions. The groundwater extraction from these areas should be carefully monitored to protect them from falling into higher conditional probability areas. Furthermore, the conditional probability values for a given average rainfall vary from 60 to 65% in the zones covered by Jamkhed, Kasegaon and Solapur stations, indicating that post-monsoon groundwater levels in the southern and central parts of the study area (46% of the area; 5,674 km^2) would be close to their average values under average rainfall conditions and, hence, these areas are most favorable zones for groundwater extraction for domestic and irrigation needs as compared to other parts of the study area.

Impacts of ENSO phenomenon

For evaluating the effects of ENSO phenomenon on groundwater, the Frank-copula-based conditional distributions of PMGL for different phases of ENSO were plotted, which are illustrated in Fig. 14a–d. For this study, the average of the top 30th percentile of ENSO Index values (1.23) for 1990–2006 period was considered as representative of the El Niño phase, whereas the average of the bottom 30th percentile of the ENSO Index values (−0.25) was deemed as representative of the La Niña phase. This figure reveals that with an increase in ENSO Index, the probability of PMGL for a particular interval increases at a lower magnitude of PMGL, but it decreases at a higher magnitude of PMGL. The probability of occurrence of higher PMGL is greater for a negative ENSO Index (La Niña phase) than for a positive ENSO Index (El Niño phase)—for example, at the Jamkhed station (Fig. 14a) for the ENSO Index value of $Z \leq 1.23$ (El Niño phase), the chance of occurrence of PMGL less than 563 m (above MSL) is 74%, whereas it is about 52% in the La Niña phase ($Z \leq -0.25$).

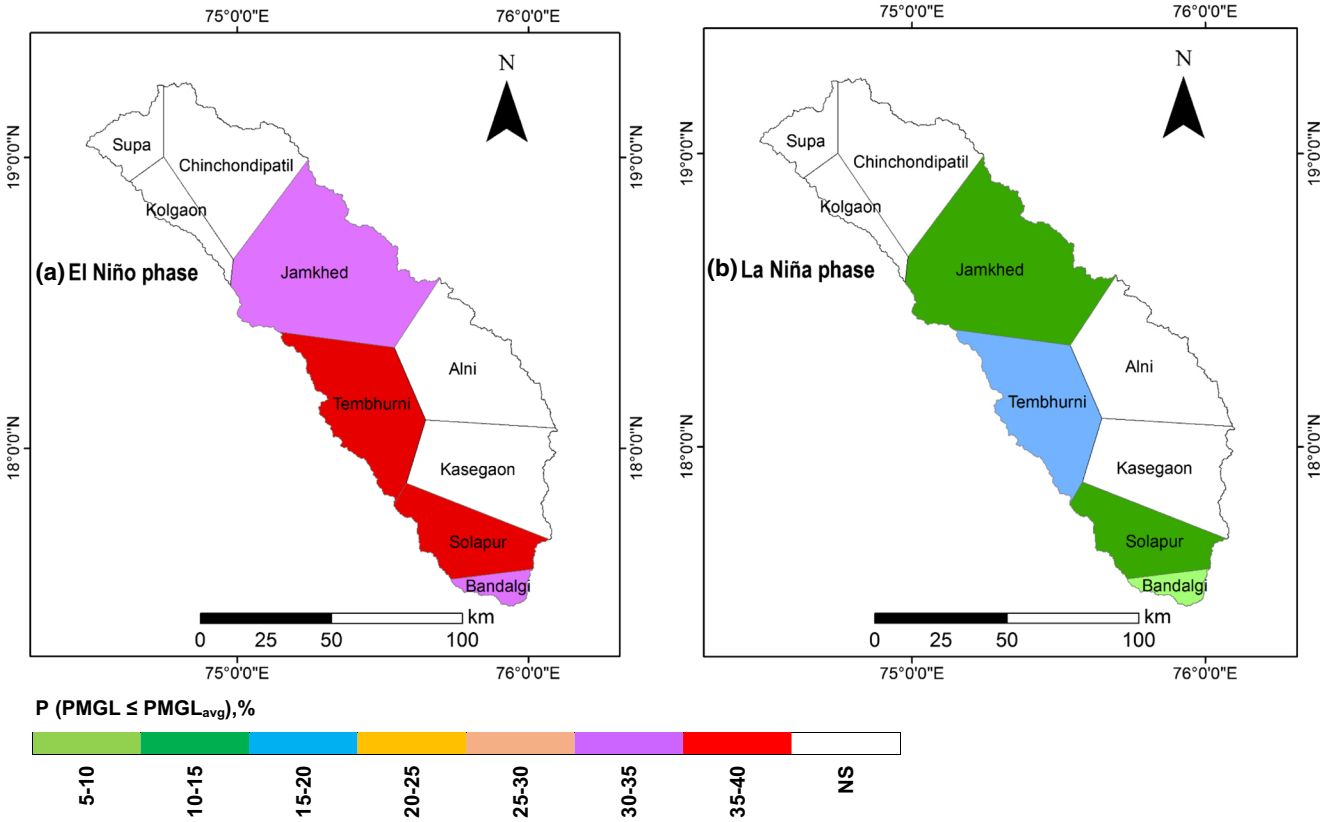

Fig. 15 Probability (*P*) maps of PMGL ≤ $PMGL_{avg}$ for **a** El Niño and **b** La Niña phases. *NS* non-significant dependence

The conditional probability (PMGL ≤ $PMGL_{avg}$) values during ENSO phases are determined for only Jamkhed, Tembhurni, Solapur and Bandalgi stations (Fig. 15a,b) where dependence is statistically significant (Table 4). It is found that during El Niño phase (Fig. 15a), the Tembhurni and Solapur stations (covering southwestern portions of the study area) show a higher (35–40%) non-exceedence probability of PMGL with respect to its average. This suggests that the blocks under maximum conditional probability (PMGL ≤ $PMGL_{avg}$) will be more severely affected during El Niño years than the other parts of the study area. The affected blocks will be Madha, Mohal, North-Solapur blocks and some portions of Parenda and Karmala blocks encompassing an area of 2,739 km^2 (23% of the study area). It is also apparent from Fig. 15a that the minimum probability of non-exceedence of PMGL during El Niño phase is in the range of 30–35% for Jamkhed and Bandalgi stations (central part of the study area), which cover 3,228 km^2 (26% of the study area). On the other hand, in the La Niña phase (Fig. 15b), the non-exceedance probability of PMGL less than or equal to its average value is found below 15% for Jamkhed, Bandalgi and Solapur stations covering an area of 4,283 km^2 (35% of the study area); thus, the central and southern portions of the study area will benefit by increased PMGL during La Niña years.

Conclusions

In this report, Archimedean copulas were applied under limited data conditions to assess the effects of the ENSO phenomenon and rainfall on the groundwater resource of a semi-arid river basin of western India. With regard to the availability and continuity of hydro-climatic time-series data in the study area, the dataset used in this study comprised monsoon rainfall of nine stations, post-monsoon groundwater levels (PMGL) at 132 sites, and the ENSO Index for the 1990–2006 period. Based on the salient goodness-of-fit criteria, marginal distributions were selected to formulate copula-based joint distributions for modeling dependence between hydro-climatic variables. Thereafter, out of the four Archimedean copula families, the best-performing copula was used to derive conditional probability distributions of groundwater-level time series with respect to rainfall events and ENSO phases.

The analysis of the results of this study revealed that the dependence for the PMGL-Rainfall pair is positive, whereas that for the PMGL-ENSO Index pair is negative. The PMGL and rainfall time series are best represented respectively by the parametric Weibull and lognormal distributions, whereas the ENSO Index time series is best represented by the non-parametric kernel-based normal distribution. The performance

evaluation of the Archimedean copulas family indicated that the Clayton copula is the best for modeling dependence between PMGL and Rainfall, while the Frank copula is the best for the PMGL-ENSO Index pair. The spatial variation of the probability of PMGL ≤ $PMGL_{avg}$ for a given mean rainfall in the study area suggests that for managing groundwater, the areas having above 70% probability (33% of the study area in the eastern and western portions) should be given higher priority. In addition, the results of the probability of PMGL ≤ $PMGL_{avg}$ during ENSO phases indicated that 23% of the study area in the southwestern portion will be severely affected during El Niño years, but 35% of the study area in central and southern portions will benefit by increased PMGL (greater than $PMGL_{avg}$) during La Niña years.

Finally, it can be concluded that the copula-based approach is very useful for understanding the impacts of environmental factors on vital groundwater resources at a basin or sub-basin scale. The methodology demonstrated in this study can be replicated for the effective planning and management of water resources at a basin scale under data-scarce condition, particularly in the drought-prone regions of Indian subcontinent and other parts of the world.

Acknowledgements Our sincere thanks are due to the India Meteorology Department, Pune (Maharashtra), State Data Storage Centre, Nashik (Maharashtra), and Groundwater Survey Development Agency, Pune, for providing hydro-meteorological data to carry out this study. We are very grateful to Prof. Christian Genest (McGill University, Montréal, Canada) for his technical discussions and meticulous comments on this manuscript. In addition, the help rendered by Mr. Ankit Shekhar and the useful discussion with Dr. Poulami Ganguli (McMaster University, Hamilton, Canada) are gratefully acknowledged.

References

Chary GR, Vittal KPR, Venkateswarlu B, Mishra PK, Rao GGSN, Pratibha G, Rao KV, Sharma KL, Rao GR (2010) Drought hazards and mitigation measures. In: Jha MK (ed) Natural and anthropogenic disasters: vulnerability, preparedness and mitigation. Springer, The Netherlands, pp 197–236

Chowdhary H, Singh VP (2010) Reducing uncertainty in estimates of frequency distribution parameters using composite likelihood approach and copula-based bivariate distributions. Water Resour Res 46:W11516. doi:10.1029/2009WR008490

Deolankar SB (1980) The Deccan basalts of Maharashtra, India: their potential as aquifers. Ground Water 18(5):434–437

DTE (2016) Down to Earth. Centre for Science and Environment, New Delhi, India, 1–15 May 2016, 35 pp

Durocher M, Chebana F, Ouarda TB (2016) On the prediction of extreme flood quantiles at ungauged locations with spatial copula. J Hydrol 533:523–532

Fleming SW, Quilty EJ (2006) Aquifer responses to El Niño-Southern Oscillation, southwest British Columbia. Ground Water 44(4):595–599

Ganguli P, Reddy MJ (2012) Risk assessment of droughts in Gujarat using bivariate copulas. Water Resour Manag 26(11):3301–3327

Ganguli P, Reddy MJ (2013) Analysis of ENSO-based climate variability in modulating drought risks over western Rajasthan in India. J Earth Sys Sci 122(1):253–269

Genest C, Chebana F (2016) Copula modeling in hydrologic frequency analysis. In: Singh VP (ed) Handbook of applied hydrology. McGraw-Hill, New York

Genest C, Favre AC (2007) Everything you always wanted to know about copula modeling but were afraid to ask. J Hydrol Eng ASCE 12(4): 347–368

Genest C, MacKay RJ (1986) The joy of copulas: bivariate distributions with uniform marginals. Am Stat 40(4):280–283

Genest C, Nešlehová J (2012a) Copula modeling for extremes. In: El-Shaarawi AH, Piegorsch WW (eds) Encyclopedia of environmetrics, 2nd edn., vol 2. Wiley, Chichester, UK, pp 530–541

Genest C, Nešlehová J (2012b) Copulas and copula models. In: El-Shaarawi AH, Piegorsch WW (eds) Encyclopedia of environmetrics, 2nd edn, vol 2. Wiley, Chichester, UK, pp 530–541

Gorelick SM, Zheng C (2015) Global change and the groundwater management challenge. Water Resour Res 51:3031–3051. doi:10.1002/2014WR016825

Gurdak JJ, Hanson RT, Green TR (2009) Effects of climate variability on groundwater resources of the United States. US Geol Surv Fact Sheet 2009-3074

Hanson RT, Dettinger MD, Newhouse MW (2006) Relations between climatic variability and hydrologic time series from four alluvial basins across the southwestern United States. Hydrogeol J 14(7): 1122–1146

Holman IP (2006) Climate change impacts on groundwater recharge-uncertainty, shortcomings, and the way forward? Hydrogeol J 14(5):637–647

ICSH (2017) International commission on statistical hydrology. ICHS, Viterbo, Italy. http://www.stahy.org/Activities/STAHYReferences/ReferencesonCopulaFunctiontopic/tabid/78/Default.aspx. Accessed 4 May 2017

IRI (2017) ENSO essentials. International Research Institute for Climate and Society, New York. http://iri.columbia.edu/our-expertise/climate/enso/enso-essentials/. Accessed 15 May 2017

Jones IC, Banner JL (2003) Hydrogeologic and climatic influences on spatial and interannual variation of recharge to a tropical karst island aquifer. Water Resour Res 39:1253. doi:10.1029/2002WR001543

Kao SC, Govindaraju RS (2008) Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. Water Resour Res 44:W02415. doi:10.1029/2007WR006261

Karmakar S, Simonovic SP (2009) Bivariate flood frequency analysis, part 2: a copula-based approach with mixed marginal distributions. J Flood Risk Manag 2(1):32–44

Klein B, Schumann AH, Pahlow M (2011) Copulas: new risk assessment methodology for dam safety. In: Schumann AH (ed) Flood risk assessment and management. Springer, The Netherlands, pp 149–185

Luque-Espinar JA, Chica-Olmo M, Pardo-Igúzquiza E, García-Soldado MJ (2008) Influence of climatological cycles on hydraulic heads across a Spanish aquifer. J Hydrol 354(1):33–52

Mall RK, Gupta A, Singh R, Singh RS, Rathore LS (2006) Water resources and climate change: an Indian perspective. Curr Sci 90(12): 1610–1626

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391(1):202–216

Nelsen RB (2006) An introduction to copulas. Springer, New York

News World India (2016) Only 1 per cent water remains in Marathwada Dams. http://indianexpress.com/article/india/india-news-india/only-1-per-cent-water-remains-in-marathwada-dams-2815112/. Accessed 24 May 2016

NOAA (2017) MEI ranks. National Oceanic and Atmospheric Administration, Maryland, United States. https://www.esrl.noaa.gov/psd/enso/mei/table.html. Accessed 10 Dec 2014

PACS (2004) Drought in India: challenges and initiatives. Report of poorest areas civil society (PACS) programme 2001–2008. PACS, New Delhi, India

Perez-Valdivia C, Sauchyn D, Vanstone J (2012) Groundwater levels and teleconnection patterns in the Canadian prairies. Water Resour Res 48:W07516. doi:10.1029/2011WR010930

Reddy MJ, Ganguli P (2012a) Bivariate flood frequency analysis of upper Godavari River flows using Archimedean copulas. Water Resour Manag 26(14):3995–4018

Reddy MJ, Ganguli P (2012b) Risk assessment of hydroclimatic variability on groundwater levels in the Manjara basin aquifer in India using Archimedean copulas. J Hydrol Eng ASCE 17(12):1345–1357

Revadekar JV, Borgaonkar HP, Kothawale DR (2012) Temperature extremes over India and their relationship with El Niño-Southern Oscillation. In: Jha MK (ed) Natural and anthropogenic disasters: vulnerability, preparedness and mitigation. Springer, The Netherlands, pp 275–292

Russo T, Lall U, Wen H, Williams M (2014) Assessment of trends in groundwater levels across the United States. Columbia Water Center White Paper, Earth Institute, Columbia University, New York, 20 pp

Salvadori G, De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. Water Resour Res 40:W12511. doi:10.1029/2004WR003133

Salvadori G, De Michele C (2007) On the use of copulas in hydrology: theory and practice. J Hydrol Eng ASCE 12(4):369–380

Seeboonruang U (2014) An empirical decomposition of deep groundwater time series and possible link to climate variability. Glob Nest J 16(1):87–103

Shiau JT, Feng S, Nadarajah S (2007) Assessment of hydrological droughts for the Yellow River, China using copulas. Hydrol Process 21(16):2157–2163

Singh J (2014) Pray before you sow. Down to Earth, New Delhi. http://www.downtoearth.org.in/coverage/pray-before-you-sow-44955. Accessed 19 January 2017

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges [N-dimensional distribution functions and their margins]. Pub l'Institut Statist l'Univ Paris 8:229–231

Susilo GE, Yamamoto K, Imai T (2013) Modeling groundwater level fluctuation in the tropical peatland areas under the effect of El Nino. Procedia Environ Sci 17:119–128

Tremblay L, Larocque M, Anctil F, Rivard C (2011) Teleconnections and interannual variability in Canadian groundwater levels. J Hydrol 410(3):178–188

UNESCO (2009) Water in changing world. United Nations World Development Report 3. UNESCO, Paris

Whelan N (2004) Sampling from Archimedean copulas. Quant Finan 4(3):339–352

Wolter K, Timlin MS (2011) El Niño/southern oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI. Ext). Int J Climatol 31(7):1074–1087

Wong G, Lambert MF, Leonard M, Metcalfe AV (2010) Drought analysis using trivariate copulas conditional on climatic states. J Hydrol Eng ASCE 15(2):129–141

Zhang L, Singh VP (2006) Bivariate flood frequency analysis using the copula method. J Hydrol Eng 11(2):150–164