

MODELING RAINFALL AND GROUNDWATER LEVEL DATA USING TIME-VARYING COPULA MODELS

HOSSEIN ZAMANII^{1*}, ZOHREH PAKDAMANI², MARZIEH SHEKARI³

ABSTRACT. Non-stationary data are often created when the observations of a study are collected sequentially in a time-dependent structure. In such a case, there will usually be a time trend with abrupt changes in the average or/and variance of the observations, which indicates that the data is non-stationary. To describe such data using statistical distributions and fitting parameters, time-varying models are suitable. The aim of this study is to introduce and apply time-varying models in which parameters are considered as time-varying in both marginal distributions and copula models. According to the monthly collection of rainfall and groundwater level data, the nature of these data is time-varying and the trend changes in these data shows that the average of data changes abruptly over time. To describe the correlation structure between these data, marginal distributions and then time-varying copulas have been used, so that the parameter of these models is considered to be vary over time as a function of time in the form of polynomial or exponential regression functions.

1. Introduction

Data stationary is an essential prerequisite in frequency analysis. But, in practice, in various field specially in economic due to collecting data sequentially in a regular time or in hydrology, due to

Keywords: Time-varying Copula , Copula, Rainfall.

Article Type: Resaerch Paper.

Communicated by Khalil Shafie.

*Corresponding author.

Received: 01-01-2024, Accepted: 19-06-2024, Published Online: 20-07-2024.

Cite this article: H. Zamani, Z. Pakdaman and M. Shekari, Modeling Rainfall and Groundwater level data Using Time-varying copula models, *Journal of Mathematics and Society*, **9** no. 4 (2024) 23–43.

<https://dx.doi.org/10.22108/msci.2024.140292.1637> .



climate changes, droughts, geological features, or human activities data often are non-stationarity in time series [13]. Any statistical frequency analysis based on the assumption of stationarity will be invalid if non-stationarity is not investigated in the hydrological time series[7]. Groundwater level is one of the hydrological variables that has suffered a significant drop due to over-extraction and decreased rainfall in most parts of the world [17]. Therefore, frequency analysis of this non-stationary series seems necessary.

The copula theory with its multivariate nature, provides great flexibility for modeling bivariate or multivariate data. In recent years, the use of copula functions in multivariate hydrological analysis has increased significantly. Copula functions have been widely used by different researchers for multivariate analysis of hydrological, hydrogeological and hydrometeorological data in recent years. for instance, several researchers [1, 2, 3, 5, 8, 12, 14, 18, 19] performed their multivariate analysis based on copula functions although the data were time dependent with time varying structure. Groundwater is one of the sources of fresh water in arid and semi-arid regions, which provides resilience against the lack of rainfall. But in the past few decades, due to population growth and agricultural and industrial development, there is an irreparable pressure on it which has caused a drop in the underground water level in most parts of the world, especially in Iran. Rainfall is one of the variables whose changes are effective on groundwater fluctuations, but the effects of increasing or decreasing rainfall on the groundwater level are always delayed due to several factors involving aquifer characteristics. Therefore, the aim of the current research is to investigate how the groundwater level has changed and most importantly, how the rainfall changes have been effective on the correlation structure of the rainfall with the groundwater changes.

2. Main Results

In statistical literature, the maximum likelihood method is a well-known approach for estimation the parameters of the models. To perform this, it is assumed that the observations are independent and identically distributed with a distribution that contains one or several parameters that are taken to be constant [?]. In some situations, however, the data are collected sequentially in a time dependent manner. In this case, a serial dependence is observed among the data which leads to a non-stationary distribution with non-constant parameters. Copula theory provides a theoretical basis for constructing multivariate models [6, 15]. The copula applies the marginal distribution functions to construct the joint distribution of random variables in order to describe the correlation between the random variables, and reflect the dependence structure. Suppose $Y_1^t, Y_2^t, \dots, Y_n^t$ is a sequence of time dependent random variables that follows a distribution function $F^t = F(., \Theta^t)$ with a non-stationary mean and a non-constant variance. In the other word, $\Theta^t = (\theta_1^t, \dots, \theta_k^t)$ contains one or more non-constant parameters



which is a function of time $t (t = 1, 2, \dots, n)$. Various researches considered different approach to handle the time-varying parameters in the model. For instance the CAS [1, 20], SCAR [20, 14], ARIMA [16], DCC [10] methods were used by authors for handling time-varying parameters. In this study, we incorporated the time as explanatory variable in the model through two polynomial regression models as below,

$$(2.1) \quad \theta_j^t = \varphi_j(t) = \beta_0 + \sum_{k=1}^q \beta_{jk} t^k$$

or

$$(2.2) \quad \theta_j^t = \phi_j(t) = \exp(\varphi_j(t))$$

depends on whether there is a constraints on parameters to be positive or not. In equation (2.1) and (2.2), $\varphi_j(\cdot)$ and $\phi_j(\cdot)$ are the link functions and $\beta_j = (\beta_0, \beta_{j1}, \dots, \beta_{jq})^T$ are the regression parameters. In order to detect whether there is a variation in the mean and variance of the data through the time and select the degree of polynomial regression in the marginal model, the time series plot of the observation is investigated. In this study we applied Generalized gamma, Weibull, Log-normal and Gumbel probability distributions [9] for modeling the rainfall and groundwater series. These probability distributions are widely used in many studies specially in the hydrological studies. In this paper we applied special functional forms of the abovementioned models which was developed by Rigby and Stasinopoulos [19] as a generalized regression of univariate models called Generalized Addictive Models in Location, Scale and Shape (GAMLSS) with parameters $\pi = (\mu, \sigma, v)^T$. Herein, the parameters μ, σ, v are the location, scale, and shape parameters respectively. Based on the definition of the copula [18, 8], the time-varying copula can be expressed as

$$(2.3) \quad H_{Y_1^t, Y_2^t}(y_1^t, y_2^t) = C [F_1(y_1^t | \Theta_1^t), F_2(y_2^t | \Theta_2^t) | \xi^t] = C(u_1^t, u_2^t | \xi^t)$$

Where $C(\cdot)$ represents the copula function and $F_j(\cdot | \Theta_k^t) = u_j^t, (j = 1, 2)$ represents the time dependent marginal cumulative distribution functions and ξ^t is the time-varying copula parameters. Generally, a simple liner model $\xi^t = \delta_0 + \delta_1 t$, or its exponential form $\xi^t = \exp(\delta_0 + \delta_1 t)$ is suitable to describe the time-varying copula models.

3. Summary of Proofs

To investigate the proposed method practically, we conducted this study on the rainfall and groundwater level data which was collected during 1990 to 2023 from the Shamil basin in Hormozgan province. Figure 1 represents the time series plot of the rainfall and groundwater data during the study period. This figure shows that the rainfall series data changes in a quadratic form with a relatively high fluctuation around the mean. Therefore, it can be concluded that the mean of the rainfall is a quadratic

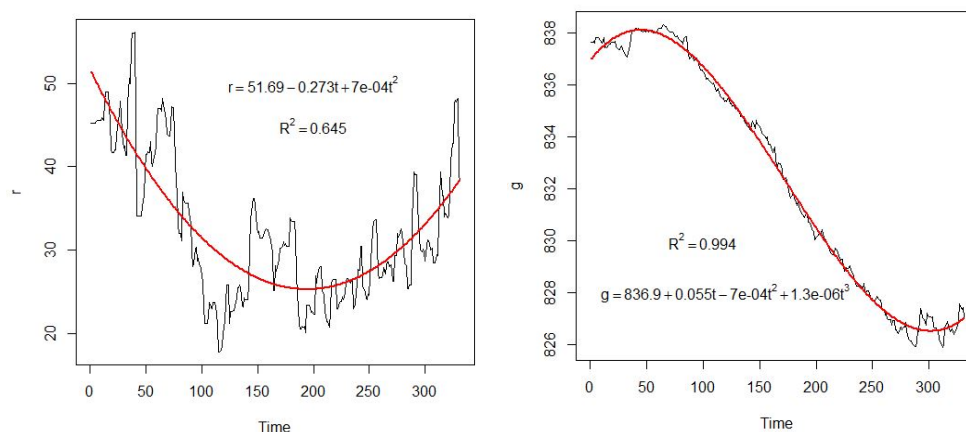


FIGURE 1. Rainfall (left) and Groundwater (right)

function of the time and its non-constant variance that can be considered as a linear function of the time. Moreover, as figure shows, the groundwater level changes in a cubic trend form whereas no fluctuation around the mean is observed. So that, we consider the mean of the groundwater to be a cubic function of the time with a constant dispersion parameter. Therefore, in both cases the time-varying distributions are distinguished to be appropriate models for fitting rainfall and groundwater series and obtaining the marginal cumulative distribution functions. Results of fitting marginal distributions to the rainfall and groundwater series with constant and time-varying parameters indicate that considering time varying parameters using appropriate polynomial regression according to Figure 1, significantly improves the results of AIC criterion. For instance, in fitting groundwater level using the fixed parameters Log-normal model the AIC=1927.44 while using the time-varying Log-normal model the AIC=187.96 which a reduction of 1739.48 has been achieved in the AIC.

The diagnostic plots of fitting time-varying Log-normal including the Q-Q plot and the Worm plot are illustrated in Figure 2. Both plots confirm the suitability of fitting time-varying Log-normal on rainfall and groundwater level series.

Figure 3 displays fitting the fix parameter and the time-varying parameter copulas to the rainfall and groundwater data. As figure shows, the plot of time-varying copulas changes over time while the plot of fix copulas are constant. Generally, results of the study indicated that applying time-varying models, for both of marginal and copula functions, significantly decreases the AIC of the models. Therefore, this approach is useful when there exists autocorrelation among the data and the presumption of the MLE estimation is violated.

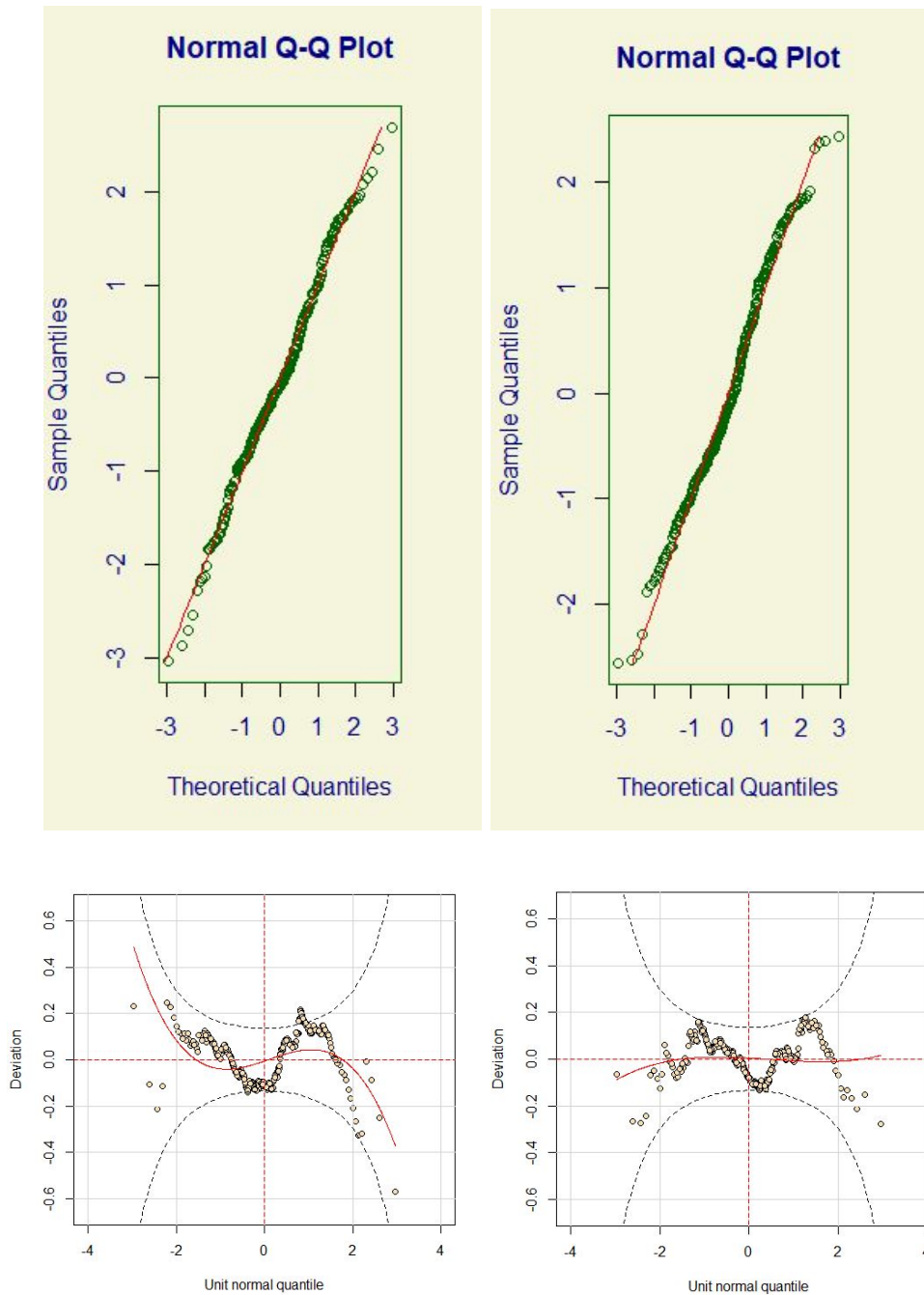


FIGURE 2. Q-Q plot and worm plot of fitting time-varying lognormal to Rainfall (left) and Groundwater (right)

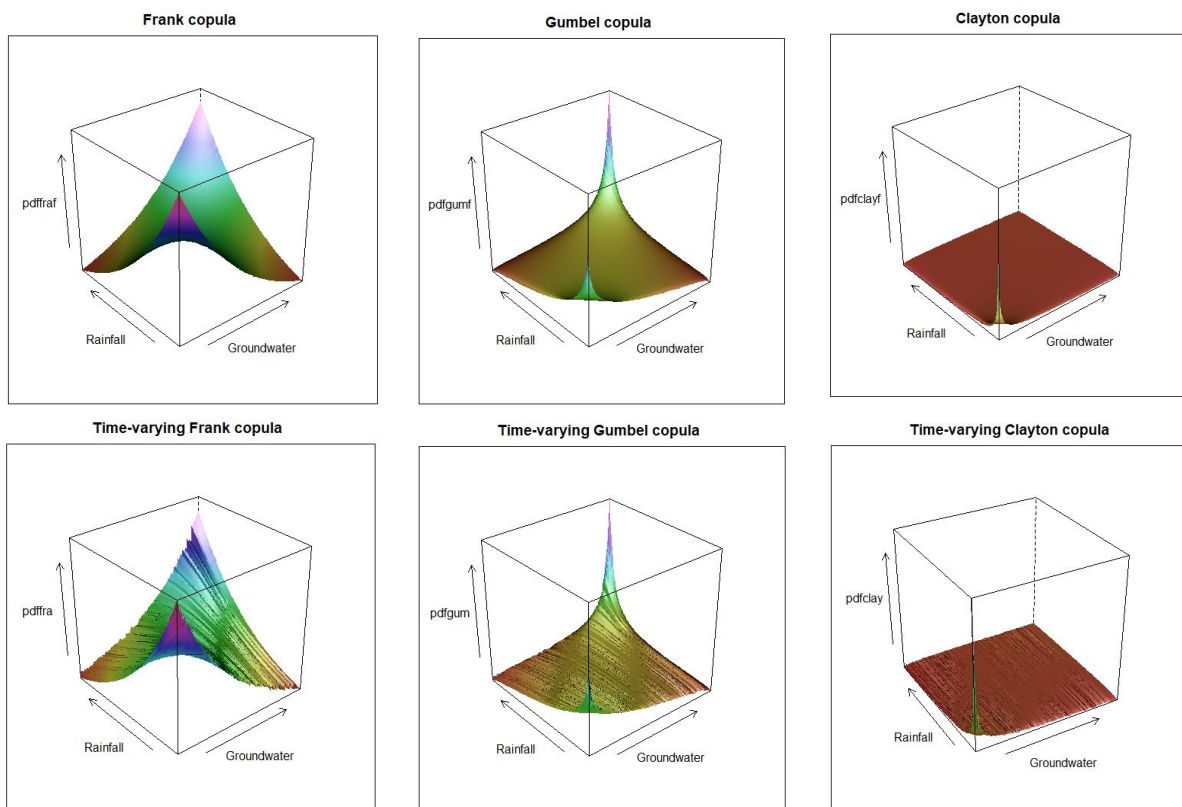


FIGURE 3. Wireframe plot of fitting fix parameters (top) and time-varying parameters (bottom) copulas to the rainfall and groundwater data

4. Conclusions

In this paper, we presented a method for fitting distributions in the case that the data are collected in a regular time structure and have autocorrelation. In such a case, the assumption of independence and identical distribution of the data is violated, and therefore it is incorrect to apply the maximum likelihood method to the data considering a fixed and identical parameter. In such a case, the time should also be considered and entered into the model through a regression function. Such models are called time-varying models. In this paper, polynomial regression and exponential function are used to enter the time factor into the model

Hossein Zamani

Department of Statistics, University of Hormozgan BandarAbbas, Iran

Email: zamani.huni@hormozgan.ac.ir

Zohreh Pakdaman

Department of Statistics, University of Hormozgan BandarAbbas, Iran

Email: `zpakdaman hormozgan.ac.ir`

Marzieh Shekari

Department of Statistics, University of Hormozgan BandarAbbas, Iran

Email: `shekarimuni hormozgan.ac.ir`

مدل‌بندی داده‌های بارش و سطح آب زیرزمینی با استفاده از توابع مفصل زمان-متغیر

حسین زمانی^{id} *، زهره پاکدامن^{id} و مرضیه شکاری^{id}

چکیده. داده‌های نایستا غالباً زمانی ایجاد می‌شوند که مشاهدات یک مطالعه به صورت پیاپی و وابسته به زمان جمع‌آوری شده باشند. در چنین حالتی معمولاً یک روند زمانی همراه با تغییرات در میانگین و یا واریانس مشاهدات وجود خواهد داشت که حاکی از نایستا بودن داده‌هاست. برای توصیف چنین داده‌هایی با استفاده از توزیع‌های آماری و برازش پارامترها، مدل‌های زمان-وابسته مناسب هستند. هدف این مطالعه معرفی و به‌کاربردن مدل‌های زمان-وابسته است که در آنها، پارامترهای مدل در توزیع‌های حاشیه‌ای و نیز در مدل‌های مفصل به صورت متغیری از زمان در نظر گرفته می‌شود. با توجه به جمع‌آوری داده‌های بارش و سطح آب زیرزمینی به صورت ماهیانه، ماهیت این داده‌ها زمان-وابسته بوده و روند تغییرات این داده‌ها نشان می‌دهد که میانگین داده‌ها در طول زمان دارای تغییرات می‌باشند. برای توصیف ساختار همبستگی میان این داده‌ها، از توزیع‌های حاشیه‌ای و سپس توابع مفصل زمان-متغیر استفاده شده است، بطوریکه پارامتر این مدل‌ها متغیر و به‌عنوان تابعی از زمان در قالب توابع رگرسیون چندجمله‌ای یا نامایی در نظر گرفته شده است.

۱. مقدمه

یکی از پیش‌فرض‌های اصلی در تحلیل فراوانی، ایستا بودن داده‌هاست. اما در عمل در بسیاری از حوزه‌های تحقیق مانند اقتصاد به سبب اینکه داده‌ها به صورت پیاپی و در زمان‌های منظم جمع‌آوری می‌شوند، و یا در هیدرولوژی به علت تغییرات آب و هوایی، خشکسالی، ویژگی‌های زمین شناسی و یا فعالیت‌های انسانی، داده‌ها اغلب یک سری زمانی نایستا می‌باشند [۱۳]. در این صورت هر تحلیل فراوانی آماری برای چنین داده‌هایی با پیش‌فرض ایستابودن آنها غیرمعتبر بوده و بایستی ابتدا ایستایی این داده‌ها مورد بررسی قرار گیرد [۷]. سطح آب زیرزمینی یکی از متغیرهای مهم هیدرولوژیکی است که سطح آن با توجه به برداشت‌های بی‌رویه و کاهش بارندگی در بسیاری از مناطق جهان به طور معنا داری کاهش یافته است [۱۷]. بنابراین داده‌های بارش و سطح آب زیرزمینی به دلیل تغییرات مکرر در میزان بارش سالیانه و متناسب با آن تغییر در برداشت آب زیرزمینی اغلب دارای میانگین و واریانس نا ثابت بوده که سبب نایستایی سری می‌شود و تحلیل فراوانی آماری این سری نایستا در حوزه آب ضروری به نظر می‌رسد. در بسیاری از زمینه‌های تحقیقاتی از جمله هیدرولوژی، اغلب متغیرها دارای ساختار همبستگی بوده و به عبارتی پدیده‌ها از هم تأثیر می‌پذیرند. هر تغییری در زمان یا عوامل محیطی، بر روی ساختار وابستگی بین متغیرها نیز اثرگذارند. بنابراین برای تعیین مدل حاکم بر داده‌ها، در نظر گرفتن وابسته به زمان بودن توزیع‌های حاشیه‌ای و توأم متغیرهای مورد مطالعه ضروری است [۹].

عبارات و کلمات کلیدی: مفصل، متغیر زمان، کاپولا، بارش.

نوع مقاله: پژوهشی

دبیرتخصصی رابط: خلیل شفیعی

* نویسنده مسئول

تاریخ دریافت: ۱۴۰۲/۱۰/۱۱ تاریخ پذیرش: ۱۴۰۳/۰۳/۳۰ تاریخ انتشار آنلاین: ۱۴۰۳/۰۴/۳۰

ارجاع به مقاله: ح. زمانی، ز. پاکدامن و م. شکاری، مدل‌بندی داده‌های بارش و سطح آب زیرزمینی با استفاده از توابع مفصل زمان-متغیر، نشریه ریاضی و جامعه، ۹ شماره ۴ (۱۴۰۳) ۲۳-۴۳.

<https://dx.doi.org/10.22108/msci.2024.140292.1637>

نظریه مفصل^۱ با ویژگی خود، به‌عنوان یک مدل انعطاف‌پذیر در مدل‌بندی‌های دو متغیره و یا چندمتغیره داده‌ها کاربرد فراوانی دارد. در سال‌های اخیر، کاربرد توابع مفصل در تحلیل‌های چند متغیره در حوزه‌های مختلف گسترش فراوانی یافته است. در همین خصوص، بسیاری از محققان از توابع مفصل در تحلیل‌های چند متغیره داده‌های هیدرولوژی، زمین‌شناسی و هواشناسی استفاده نموده‌اند. اما اغلب نویسندگانی که در تحلیل‌های چندمتغیره خود از توابع مفصل استفاده نموده‌اند، [۱، ۲، ۳، ۵، ۸، ۱۲، ۱۴، ۱۸، ۱۹] داده‌های مورد استفاده را ایستا در نظر گرفته‌اند در حالی‌که به‌نظر می‌رسد با توجه به ماهیت زمان-وابسته بودن آنها، این داده‌ها نایستا بوده و در محاسبه توزیع‌های حاشیه‌ای یا توابع توأم آنها بایستی این عامل در نظر گرفته شود. البته برخی محققان [۹، ۱۰، ۱۱] در مدل‌بندی داده‌های خود، این موضوع را در نظر گرفته‌اند.

سطح آب زیرزمینی یکی از منابع مهم تأمین آب شیرین در مناطق خشک و نیمه‌خشک است به‌طوری‌که یک بستر مقاوم در برابر کمبود بارش ایجاد می‌کند. اما در چند دهه گذشته با توجه به رشد جمعیت، سطح زیر کشت و گسترش صنعت، تحت فشار مضاعف قرار گرفته به‌طوری‌که سطح آن در اغلب مناطق جهان به‌خصوص در ایران کاهش یافته است. بارش یکی دیگر از متغیرهای هیدرولوژیکی است که تأثیر مستقیمی بر سطوح آب زیرزمینی دارد اما تأثیر کاهش یا افزایش بارندگی بر سطح آب زیرزمینی به دلایل مختلفی از جمله خواص آبخوان‌ها با یک تأخیر همراه است.

هدف اصلی این تحقیق مدل‌بندی داده‌های بارش و سطح آب زیرزمینی است که در یک ساختار زمان-وابسته و به‌صورت ماهیانه در یک بازه زمانی جمع‌آوری شده‌اند. در مدل‌بندی این داده‌ها نیز با توجه به ماهیت آنها، از توابع حاشیه‌ای زمان-متغیر و برای بررسی ساختار وابستگی میان آنها از توابع مفصل زمان-متغیر^۲ استفاده شده است.

۲. تحلیل چند متغیره با استفاده از توابع مفصل زمان-متغیر

در مباحث آماری، برآورد درست‌نمایی ماکزیمم به‌عنوان یک روش شناخته شده در برآورد پارامترهای مدل مورد استفاده قرار می‌گیرد. اما جهت استفاده از این روش جهت برازش یک توزیع مشخص آماری بر داده‌ها، اغلب فرض می‌شود مشاهدات جمع‌آوری شده مستقل از هم هستند و همگی از یک توزیع که شامل یک یا چند پارامتر ثابت است، پیروی می‌کنند [۵]. در برخی مطالعات اما، داده‌ها ممکن است به‌صورت پیاپی و در بازه‌های زمانی منظم جمع‌آوری شده باشند که سبب ایجاد یک خودهمبستگی در میان مشاهدات خواهد شد و باعث می‌شود فرض استقلال داده‌ها نقض شود. در واقع، وقتی داده‌ها خودهمبسته باشند یعنی از هم مستقل نبوده و دارای یک همبستگی سریالی در یک توالی زمانی هستند. در چنین حالتی، معمولاً میانگین داده‌ها و یا حتی پراکندگی داده‌ها در طول زمان ثابت نبوده و تغییر می‌کند. برای چنین داده‌هایی به‌کار بردن یک توزیع آماری با پارامتر ثابت و برآورد این پارامترها با روش درست‌نمایی با فرض استقلال آنها، نادرست است. در این حالت بایستی عامل زمان به‌عنوان یک متغیر از طریق میانگین و واریانس (یا پارامتر پراکندگی) وارد مدل شود. یکی از روش‌های ورود عامل زمان به مدل از طریق مدل‌های رگرسیون سری زمانی می‌باشد که هم موضوع خودهمبستگی میان داده‌ها را پوشش می‌دهد و هم اینکه فرض متغیر بودن پارامترها را اعمال می‌کند. به چنین مدل‌هایی، مدل‌های زمان-وابسته با پارامترهای زمان-متغیر گویند. در این مقاله، برای وارد کردن پارامتر وابسته به‌زمان، به‌مدل، و برازش مدل‌های زمان-متغیر از مدل‌های رگرسیون سری زمانی چندجمله‌ای و توابع نمای استفاده شده است.

نظریه مفصل یک تابع کاربردی و انعطاف‌پذیر را برای ساخت توزیع‌های توأم و یا چندمتغیره فراهم می‌کند [۶، ۱۵]. برخلاف برخی توزیع‌های چندمتغیره مانند نرمال که لازم است توزیع‌های حاشیه‌ای همگی یکسان و از خانواده نرمال باشند، در استفاده از توابع مفصل این محدودیت وجود ندارد، بلکه توزیع‌های حاشیه‌ای می‌توانند از خانواده‌های مختلف و متفاوتی باشند. جهت ساخت توزیع توأم به‌منظور توصیف ساختار وابستگی میان دو متغیر، در گام اول توزیع‌های حاشیه‌ای هر متغیر مشخص می‌شوند که این توزیع‌ها می‌توانند از یک خانواده مشابه یا غیرمشابه باشند. در مرحله دوم از یک تابع مفصل مناسب جهت به‌دست آوردن تابع توزیع

¹copula ²time-varying copula

توام دو متغیر استفاده خواهد شد. در هر مرحله جهت انتخاب بهترین مدل برازش شده از معیارهای مناسب آماری یا آزمون‌های آماری استفاده خواهد شد.

۱.۲. انتخاب توزیع‌های حاشیه‌ای. فرض کنید $Y_1^t, Y_2^t, \dots, Y_n^t$ یک دنباله از متغیرهای تصادفی متوالی و زمان-وابسته باشند که از تابع توزیع $F^t = F(\cdot, \Theta^t)$ با میانگین و واریانس متغیر در طول زمان پیروی می‌کنند. همچنین فرض کنید $\Theta^t = (\theta_1^t, \dots, \theta_k^t)$ پارامتر مدل بوده که خود شامل یک یا چند پارامتر است که هر یک ثابت یا تابعی از زمان $t (t = 1, 2, \dots, n)$ باشد. محققان مختلف، روش‌های متفاوتی را برای مدیریت پارامترهای زمان-متغیر در مدل در نظر گرفته‌اند. برای نمونه روش‌هایی مانند $CAS [1, 20]$ ، $SCAR [14, 20]$ ، $ARIMA [16]$ و $DCC [10]$ برای وارد کردن پارامترهای زمان-متغیر به مدل مورد استفاده قرار گرفته‌اند.

استفاده از مدل‌های رگرسیون سری زمانی با تابع رگرسیون خطی به‌عنوان یک روش دیگر جهت برازش مدل‌های آماری بر داده‌های زمان-وابسته (خودهمبسته) و ورود پارامترهای زمان-متغیر به مدل مورد استفاده قرار گرفته است [۹]. در این روش پارامتر مدل به‌عنوان یک متغیر پاسخ در برابر زمان به‌عنوان یک متغیر مستقل در قالب یک مدل رگرسیون خطی ساده وارد مدل می‌شود. اما استفاده از مدل رگرسیون خطی ساده در همه موارد ممکن است تا حدودی باعث بهبود برازش شود اما انتخاب یک تابع خطی در همه موارد و برای همه مدل‌ها از کارایی لازم برخوردار نیست. در این مقاله از مدل‌های رگرسیون خطی چندجمله‌ای به‌جای مدل رگرسیون خطی ساده جهت ورود پارامتر زمان-متغیر به مدل استفاده شده است. انتخاب درجه رگرسیون چندجمله‌ای نیز بر اساس تغییرات میانگین سری در بازه زمانی مورد مطالعه در نظر گرفته می‌شود. به‌عبارت دیگر اگر تغییرات میانگین سری در بازه زمانی به‌صورت تابعی خطی باشد، آنگاه میانگین توزیع تابعی خطی از زمان، و در صورتی که تغییرات میانگین سری از درجات بالاتر مانند درجه دوم و یا سوم باشد، آنگاه میانگین توزیع نیز یک مدل رگرسیون چندجمله‌ای از درجه دوم و یا سوم در نظر گرفته خواهد شد.

در این مقاله همان‌طور که اشاره شد از متغیر زمان $t (t = 1, 2, \dots, n)$ به‌عنوان یک متغیر پیشگو در برابر پارامترهای مدل به‌عنوان متغیر پاسخ در قالب مدل‌های زیر جهت ورود به مدل استفاده شده است.

$$(1) \quad \theta_j^t = \varphi_j(t) = \beta_0 + \sum_{k=1}^q \beta_{jk} t^k$$

و

$$(2) \quad \theta_j^t = \phi_j(t) = \exp(\varphi_j(t)).$$

انتخاب هر یک از مدل‌های فوق به‌عنوان پارامتر توزیع، وابسته به این است که آیا الزام و محدودیتی برای پارامتر مدل وجود دارد یا خیر. مثلاً اگر پارامتر مدل باید بزرگتر از مقدار خاص مثبتی باشد آنگاه بایستی از مدل ۲ به‌عنوان پارامتر مدل استفاده شود. در مدل‌های ۱ و ۲، $\varphi_j(\cdot)$ و $\phi_j(\cdot)$ توابع پیوند ۳ و $\beta_j = (\beta_0, \beta_{j1}, \dots, \beta_{jq})^T$ پارامترهای مدل رگرسیونی می‌باشند. جزئیات بیشتری از این مدل‌ها را می‌توان در [۴] یافت. در این مرجع، انواع مدل‌های رگرسیون سری زمانی که در آن زمان گسسته به‌عنوان متغیر پیشگو در نظر گرفته می‌شود مورد بررسی قرار گرفته است.

به‌منظور مشاهده اینکه آیا میانگین یا پراکندگی داده‌ها در طول زمان دارای تغییرات هستند و سپس انتخاب مناسب درجه رگرسیون چندجمله‌ای جهت برازش توزیع‌های حاشیه‌ای از نمودار سری زمانی داده‌ها استفاده می‌کنیم. در این مقاله از چهار توزیع گامای تعمیم یافته، وایبل، لوگ‌نرمال و گامبل که در اغلب مطالعات هیدرولوژیکی مانند بارش، روان‌آب، سیل و سطوح آب زیرزمینی به‌کار برده می‌شود، جهت برازش داده‌ها استفاده نموده‌ایم.

³link function

جدول ۰۱. فرم تابعی و مشخصات توزیع‌های به‌کار برده شده در مدل‌بندی داده‌ها

Table 1: Functional forms and characteristics of distribution used in data modeling

توزیع	تابع چگالی احتمال	گشتاورهای توزیع
گامای تعمیم یافته	$f(y \mu, \sigma, v) = \frac{\theta^v z^\theta v e^{-\theta z}}{\Gamma(\theta) y}$ $z = \left(\frac{y}{\mu}\right)^v, \theta = \sigma^{-2} v ^{-2}$ $y > 0, \mu > 0, \sigma > 0$	$E[Y] = \mu$ $Var[Y] = \sigma^2 \mu^2$
گامبل	$f(y \mu, \sigma) = \frac{1}{\sigma} \exp\left\{\left(\frac{y-\mu}{\sigma}\right) - \exp\left(\frac{y-\mu}{\sigma}\right)\right\}$ $y \in R, \mu \in R, \sigma > 0$	$E[Y] = \mu - 0.5772 \sigma$ $Var[Y] = \frac{\sigma^2 \pi^2}{6}$
لوگ نرمال	$f(y \mu, \sigma) = \frac{1}{y\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2\right\}$ $y > 0, \mu \in R, \sigma > 0$	$E[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ $Var[Y] = [\exp(\sigma^2) - 1] \exp(2\mu + \sigma^2)$
وایبل	$f(y \mu, \sigma) = \frac{\sigma y^{\sigma-1}}{\mu^\sigma} \exp\left\{-\left(\frac{y}{\mu}\right)^\sigma\right\}$ $y > 0, \mu > 0, \sigma > 0$	$E[Y] = \mu \Gamma\left(\frac{1}{\sigma} + 1\right)$ $Var[Y] = \mu^2 \left[\Gamma\left(\frac{2}{\sigma} + 1\right) - \left(\Gamma\left(\frac{1}{\sigma} + 1\right)\right)^2\right]$

برای مدیریت بهتر این موضوع و درک بهتر رابطه میان تغییرات میانگین و واریانس و پارامترهای مدل، فرم تابعی خاصی از توزیع‌های فوق که تعمیمی از رگرسیون مدل‌های تک متغیره است و با عنوان مدل‌های جمعی تعمیم یافته برای مکان، مقیاس و شکل (GAMLSS)^۴ شناخته می‌شود [۱۹] مورد استفاده قرار گرفته است. این مدل‌ها در یک بسته نرم‌افزاری با نام GAMLSS در نرم افزار R قرار گرفته‌اند که می‌توان از آنها جهت برآزش انواع مدل‌های رگرسیونی و مدل‌های خطی تعمیم یافته استفاده نمود. در این توزیع‌ها $\pi = (\mu, \sigma, v)^T$ پارامتر توزیع است که شامل پارامتر مکان μ ، پارامتر مقیاس σ و پارامتر شکل v است. در اغلب مدل‌های این بسته نرم‌افزاری، فرم تابعی توزیع‌ها طوری تنظیم شده که در آن پارامتر μ برابر میانگین توزیع و انحراف معیار توزیع نیز برابر σ یا تابعی از آن باشد. فرم تابعی توزیع‌های فوق که در بسته نرم‌افزاری GAMLSS تعریف شده در جدول ۱ آورده شده است.

۲.۲. انتخاب تابع مفصل زمان-متغیر. براساس تعریف تابع مفصل [۱۸، ۸] مفصل زمان-متغیر را می‌توان به صورت زیر

تعریف نمود.

$$(۳) \quad H_{Y_1^t, Y_2^t}(y_1^t, y_2^t) = C[F_1(y_1^t | \Theta_1^t), F_2(y_2^t | \Theta_2^t) | \xi^t] = C(u_1^t, u_2^t | \xi^t)$$

که در آن $C(\cdot)$ تابع مفصل، و $u_j^t = F_j(y_j^t | \Theta_j^t)$ تابع توزیع تجمعی حاشیه‌ای زمان-متغیر و ξ^t نیز پارامتر تابع مفصل زمان-متغیر می‌باشد. همانند انتخاب پارامتر زمان-متغیر برای توزیع‌های حاشیه‌ای، در تابع مفصل زمان-متغیر نیز پارامتر مفصل یعنی ξ^t تابعی از زمان t در نظر گرفته می‌شود بطوریکه میان این پارامتر و متغیر پیشگوی t براساس یکی از توابع پیوند ۱، ۲، رابطه برقرار می‌شود. برای ورود این پارامتر به مدل غالباً از فرم مدل رگرسیون خطی ساده یعنی $\xi^t = \delta_0 + \delta_1 t$ و یا فرم نمایی آن

⁴Generalized Additive Models for Location, Scale and Shape

جدول ۲. فرم تابعی و مولد توابع مفصل

Table 2: Functional form and corresponding generator of copula functions

مفصل	تابع چگالی مفصل	تابع مولد
گامبل	$C(v, \nu; \theta) = \exp \left[- \left[(-\log v)^\theta + (-\log \nu)^\theta \right]^{\frac{1}{\theta}} \right]$, $\theta \in [1, +\infty)$	$\varphi(t) = (-\ln t)^\theta$
فرانک	$C(v, \nu; \theta) = \frac{1}{\theta} \log \left(1 + \frac{(e^{\theta v} - 1)(e^{\theta \nu} - 1)}{e^\theta - 1} \right)$, $\theta \in (-\infty, +\infty)$	$\varphi(t) = \ln \left[\frac{\exp(\theta t) - 1}{\exp(\theta) - 1} \right]$
کلیتون	$C(v, \nu; \theta) = (v^{-\theta} + \nu^{-\theta} - 1)^{-\frac{1}{\theta}}$, $\theta \in (0, +\infty)$	$\varphi(t) = t^{-\theta} - 1$

یعنی $\xi^t = \exp(\delta_0 + \delta_1 t)$ استفاده می‌شود. در رابطه ۳ احتمالات حاشیه‌ای u_1^t و u_2^t بایستی مقادیری از توزیع یکنواخت بر بازه $[0, 1]$ باشند.

در این مقاله براساس روابط ۱، ۲ و ۳ چند سناریوی مختلف می‌توان برای مدل‌بندی داده‌ها با تابع مفصل زمان-متغیر در نظر گرفت. این سناریوها عبارتند از (۱) پارامترهای توزیع‌های حاشیه‌ای را ثابت و پارامتر تابع مفصل را زمان-متغیر در نظر گرفت. (۲) حداقل یکی از پارامترهای توزیع‌های حاشیه‌ای را زمان-متغیر و پارامتر تابع مفصل را ثابت در نظر گرفت. (۳) حداقل یکی از پارامترهای توزیع حاشیه‌ای را زمان-متغیر و پارامتر تابع مفصل نیز زمان-متغیر در نظر گرفت. برازش تابع مفصل رابطه ۳ نیز خود شامل دو مرحله است. در مرحله اول بهترین توزیع‌های حاشیه‌ای برازش شده بر داده‌ها با استفاده از معیار اطلاع آکائیکه استخراج و در مرحله دوم از میان توابع مفصل برازش داده شده بر داده‌ها، بر اساس معیارهای آماری یکی انتخاب خواهد شد. در این مطالعه از سه تابع مفصل گامبل ۵، مفصل کلیتون ۶ و مفصل فرانک ۷ جهت مدل‌بندی داده‌های بارش و سطح آب زیرزمینی استفاده شده است. فرم تابعی این توابع در جدول ۲ آورده شده است.

۳. تحلیل داده‌های بارش و سطح آب زیرزمینی با استفاده از مدل‌های زمان-متغیر

به‌منظور استفاده کاربردی از روش‌های پیشنهاد شده در تحلیل داده‌ها، در این مطالعه از داده‌های بارش و سطح آب زیرزمینی که به‌صورت ماهیانه در بازه زمانی سال ۱۳۶۹ تا ۱۴۰۰ از حوضه آبخیز شمیل در استان هرمزگان جمع‌آوری شده مورد استفاده قرار گرفته است. این حوضه آبخیز دارای مساحتی مشتمل بر ۱۵۷۰ کیلومتر مربع است که متوسط بارش سالیانه در آن کمتر از ۲۰۰ میلیمتر بوده و متوسط تبخیر در آن نیز ۲۶۴۰ میلیمتر در سال است [۳]. میزان بارش سالیانه در این حوضه آبخیز در بازه زمانی مذکور، بر اساس نمودار آب‌نگاری، یک روند کاهشی معنی‌داری را نشان می‌دهد.

متغیرهای این مطالعه شامل بارش ماهیانه بر حسب میلیمتر و سطح آب زیرزمینی بر حسب مترمکعب است که خلاصه‌ای از آمار توصیفی آن در جدول ۳ آورده شده است. با توجه به اینکه مشاهدات در بازه‌های زمانی منظم جمع‌آوری شده بنابراین ابتدا باید استقلال داده‌ها را با استفاده از نمودار خودهمبستگی مورد بررسی قرار دهیم. در صورت وجود خودهمبستگی میان داده‌ها می‌توان نتیجه گرفت مشاهدات مستقل و هم توزیع نبوده و به‌عبارتی از یک توزیع با پارامتر یکسان پیروی نمی‌کنند. برای بررسی وجود خودهمبستگی میان مشاهدات از نمودار acf استفاده نموده‌ایم. این نمودار در شکل ۱ نمایش داده شده است.

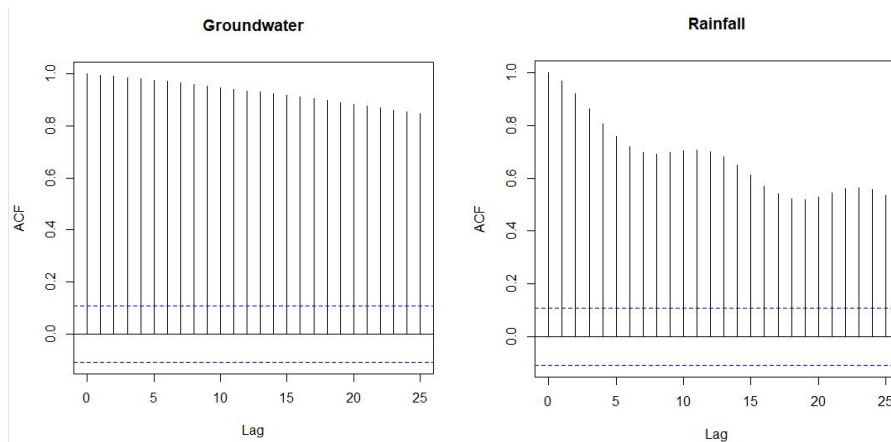
⁵Gumbel ⁶Clayton ⁷Frank

جدول ۳. مشخصه‌های توصیفی داده‌های بارش و سطح آب زیرزمینی

Table 3: Descriptive characteristics of rainfall and groundwater level data

مشخصه‌های آماری	بارش (میلی‌متر)	سطح آب زیرزمینی (مترمکعب)
میانگین	۳۲/۲۱	۸۳۲/۵
انحراف معیار	۸/۵۴	۴/۴۲
ضریب چولگی	۰/۶۵۰	- /۰۹
ضریب کشیدگی	۲/۵۰۶	۱/۴۰۹

همان‌طور که نمودار نشان می‌دهد هر دو مجموعه داده دارای خودهمبستگی درونی هستند. برای بررسی بیشتر، همچنین از آزمون روند من-کندال^۸ برای هر دو مجموعه داده استفاده نمودیم. آماره آزمون من-کندال برای داده‌های سطح آب زیرزمینی برابر $\tau = -۰/۸۶$ ($p < ۰/۰۰۰۱$) مقدار) و برای داده‌های بارش برابر $\tau = -۰/۲$ ($p < ۰/۰۰۰۲$) مقدار) می‌باشد که نشان‌دهنده وجود روند معنی‌دار در داده‌هاست. نمودار روند مشاهدات برای هر دو متغیر در شکل ۲ نشان داده شده است.

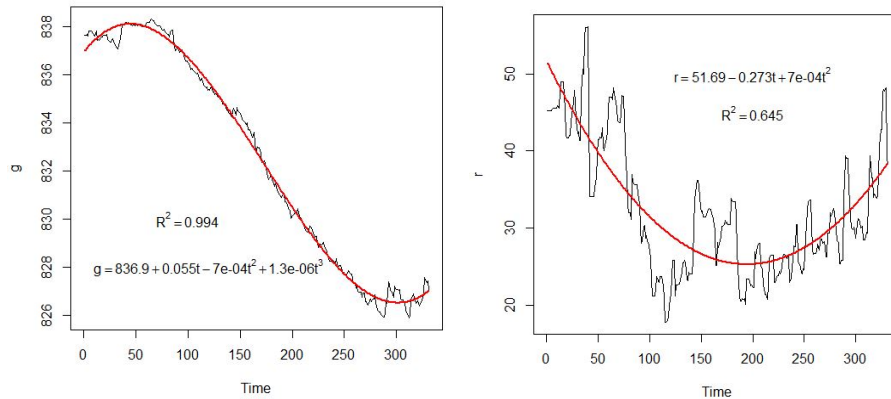


شکل ۱. نمودار خودهمبستگی داده‌های بارش (راست) و سطح آب زیرزمینی (چپ).

Figure 1: Autocorrelation plot of rainfall (right) and ground water level (left) data

با توجه به نمودارهای شکل ۲، تغییرات میانگین بارش ماهیانه در بازه زمانی مورد مطالعه به صورت یک تابع درجه دوم از زمان بوده و با توجه به نوساناتی که حول میانگین وجود دارد می‌توان نتیجه گرفت که واریانس مشاهدات نیز ثابت نبوده و دارای تغییرات است. میانگین بارش ماهیانه، از جمع بارش‌های روزانه در طول یک ماه و تقسیم آن بر تعداد روزهای آن ماه محاسبه می‌شود. همچنین برای داده‌های سطح آب زیر زمینی نیز می‌توان نتیجه گرفت که میانگین داده‌ها تابعی درجه سوم از زمان بوده اما با توجه به اینکه نوساناتی حول میانگین مشاهده نمی‌شود می‌توان نتیجه گرفت واریانس داده‌ها ثابت است. بنابراین در هر دو مجموعه داده به نظر می‌رسد که مدل‌های زمان-متغیر انتخاب مناسبی جهت برازش بر داده‌ها به منظور به دست آوردن مقادیر توزیع تجمعی حاشیه‌ای باشند و مدل‌های با پارامتر ثابت انتخاب مناسبی نخواهد بود.

⁸Mann-Kendall trend test



شکل ۲. نمودار روند تغییرات داده‌های بارش (راست) و سطح آب زیرزمینی (چپ) و رگرسیون سری زمانی برازش داده شده بر آنها.

Figure 2: Trend chart of rainfall (right) and groundwater level (left) data and the fitted time series regression

جدول ۴. برازش توزیع‌های حاشیه‌ای با پارمتر ثابت

Table 4: Fitting marginal distributions with constant parameters

متغیر	گامای تعمیم یافته	گامبل	وایبل	لوگ نرمال
سطح آب زیرزمینی	$\mu = 833.45$ $\sigma = 0.0051$ $v = 86.43$ $AIC = 1926.89$	$\mu = 834.67$ $\sigma = 3.75$ $AIC = 1925.04$	$\mu = 834.66$ $\sigma = 221.76$ $AIC = 1924.82$	$\mu = 6.72$ $\sigma = 0.0053$ $AIC = 1927.44$
بارش	$\mu = 29.62$ $\sigma = 0.248$ $v = -1.58$ $AIC = 2318.30$	$\mu = 36.72$ $\sigma = 9.15$ $AIC = 2457.29$	$\mu = 35.52$ $\sigma = 3.97$ $AIC = 2374.21$	$\mu = 3.43$ $\sigma = 0.257$ $AIC = 2321.53$

نتایج برازش توزیع‌های حاشیه‌ای با پارمتر ثابت و توزیع‌های زمان-متغیر بر داده‌های بارش و سطح آب زیرزمینی به ترتیب در جدول ۴ و ۵ داده شده است. کدهای R جهت برازش مدل گامای تعمیم یافته‌ی جدول ۵ در ضمیمه ۱ آورده شده است. جهت مقایسه مدل‌های برازش شده از معیار اطلاع آکائیکه^۹ (AIC) که به صورت $AIC = 2k - 2\ln(L)$ تعریف می‌شود استفاده شده است. در این رابطه، k تعداد پارامترهای مدل و L مقدار تابع درستنمایی مدل برازش شده است. هر چه مقدار AIC یک مدل کمتر باشد، نشان می‌دهد آن مدل عملکرد بهتری داشته است. همان‌طور که نتایج نشان می‌دهد، در نظر گرفتن پارامترهای زمان-وابسته و برازش مدل‌های زمان-متغیر باعث بهبود عملکرد مدل‌ها شده است و مقدار AIC را به طور معنی‌داری کاهش داده است. مثلاً برای

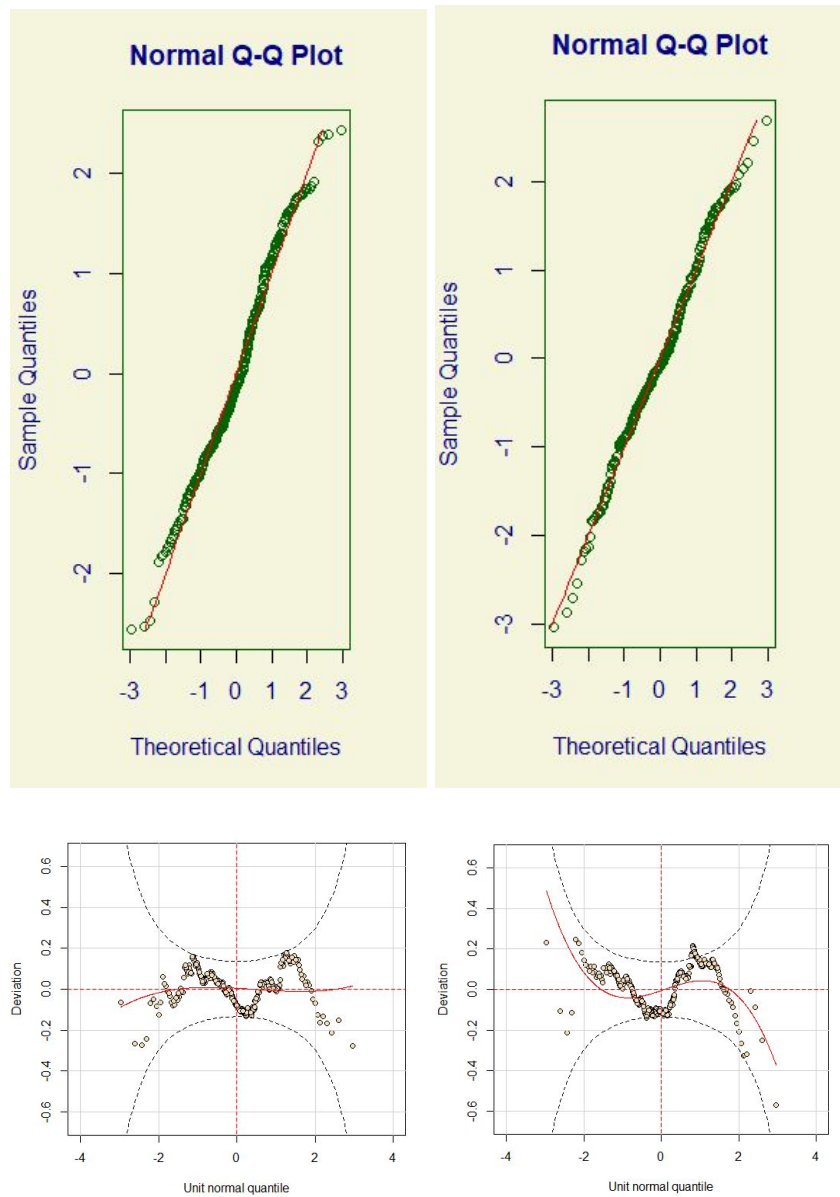
^۹akaike information criterion

جدول ۵. برازش توزیع‌های حاشیه‌ای زمان-متغیر

Table 5: Fitting marginal distributions with time-varying parameters

متغیر	توزیع	پارامتر	AIC
سطح آب زیرزمینی	گامای تعمیم یافته	$\mu = \exp(6.73 + (10^{-5})6.6t - (10^{-7})8.5t^2 + (10^{-9})1.6t^3)$ $\sigma = \exp(-7.87)$ $v = 71.08$	۱۸۹/۲۷
	گامبل	$\mu = \exp(8.372 + 0.053t - (10^{-4})7.05t^2 + (10^{-6})1.37t^3)$ $\sigma = \exp(-1.19)$	۲۲۴/۱۷
	وایبل	$\mu = \exp(6.37 + (10^{-5})6.4t - (10^{-7})8.4t^2 + (10^{-9})1.64t^3)$ $\sigma = \exp(7.91)$	۲۲۶/۲۴
	لوگ نرمال	$\mu = \exp(6.73 + (10^{-5})6.62t - (10^{-7})8.53t^2 + (10^{-9})1.64t^3)$ $\sigma = \exp(-7.84)$	۱۸۷/۹۶
بارش	گامای تعمیم یافته	$\mu = \exp(4.14 + (10^{-3})8.33t - (10^{-5})2.2t^2)$ $\sigma = \exp(-1.52 - 0.002t)$ $v = -0.38$	۲۰۰۳/۵۶
	گامبل	$\mu = \exp(5.714 - 0.31t - (10^{-4})7.98t^2)$ $\sigma = \exp(1.90 - 0.0017t)$	۲۰۸۰/۵۰
	وایبل	$\mu = \exp(4.10 + (10^{-3})8.1t - (10^{-5})2.09t^2)$ $\sigma = \exp(1.674 + 0.0013t)$	۲۰۶۰/۹۸
	لوگ نرمال	$\mu = \exp(4.02 + (10^{-3})8.3t - (10^{-5})2.2t^2)$ $\sigma = \exp(-1.52 + 0.002t)$	۲۰۰۱/۷۲

داده‌های سطح آب زیرزمینی مقدار معیار اطلاع آکائیکه در توزیع لوگ نرمال برای مدل‌های پارامتر ثابت برابر $AIC = 1927.44$ بوده است درحالی‌که این معیار برای توزیع لوگ نرمال زمان-متغیر برابر $AIC = 187.96$ به‌دست آمده است و این نشان می‌دهد که مقدار کاهش این معیار کاملاً معنی‌دار بوده و باعث بهبود برازش مدل شده است. همچنین این معیار برای داده‌های بارش نیز کاهش قابل توجهی داشته هرچند میزان کاهش آن نسبت به داده‌های سطح آب زیرزمینی کمتر بوده است. بنابراین استفاده از مدل‌های رگرسیون چندجمله‌ای برای پارامترهای مکان و مقیاس که در فرم تابعی مدل‌های به‌کار برده شده در این مقاله، این پارامترها همان میانگین و پراکندگی می‌باشند، براساس منحنی روند تغییرات سری زمانی داده‌ها، هم پیش‌فرض‌های روش برآورد درست‌نمایی را فراهم می‌کند و هم باعث افزایش کارایی مدل‌های برازش شده خواهد شد. بررسی نتایج جدول ۴ و ۵ نشان می‌دهد که توزیع لوگ نرمال زمان-متغیر در برازش هر دو متغیر بارش و سطح آب زیرزمینی عملکرد بهتری نسبت به سایر مدل‌های دیگر داشته است. برای بررسی بیشتر عملکرد این مدل، نمودارهای Q-Q Plot و Worm Plot در شکل ۳ رسم شده است.



شکل ۳. نمودارهای نیکویی برازش توزیع لوگ نرمال زمان-متغیر بر داده‌های بارش (چپ) و سطح آب زیرزمینی (راست)

Figure 3: Goodness of fit test plot for fitted time-varying log-normal distribution to rainfall (left) and ground water level (right) data

همان‌طور که هر دو نمودار نشان می‌دهد توزیع لوگ نرمال زمان-متغیر عملکرد قابل قبولی در برازش هر دو مجموعه داده داشته است.

پس از انتخاب بهترین مدل حاشیه‌ای و محاسبه مقادیر توزیع تجمعی براساس بهترین مدل انتخاب شده، از توابع مفصل جهت مدل‌بندی داده‌ها و محاسبه توزیع توأم آنها استفاده می‌کنیم. برای این منظور از توابع مفصل ۲ استفاده شده است. همانند توزیع‌های

جدول ۶. نتایج برازش توابع مفصل پارامتر ثابت بر داده‌ها

Table 6: Results of fitting copulas with constant parameter to datas

مفصل	θ_c	AIC_c	LL
گامبل	۱/۲۴۴	-۱۹/۴۹	۳۶/۹۸
فرانک	۱/۹۴۴	-۲۳/۵۱	۴۵/۰۲
کلیتون	۰/۲۵۵	-۱۶/۸۵	۳۱/۷

جدول ۷. نتایج برازش توابع مفصل متغیر زمان بر داده‌ها

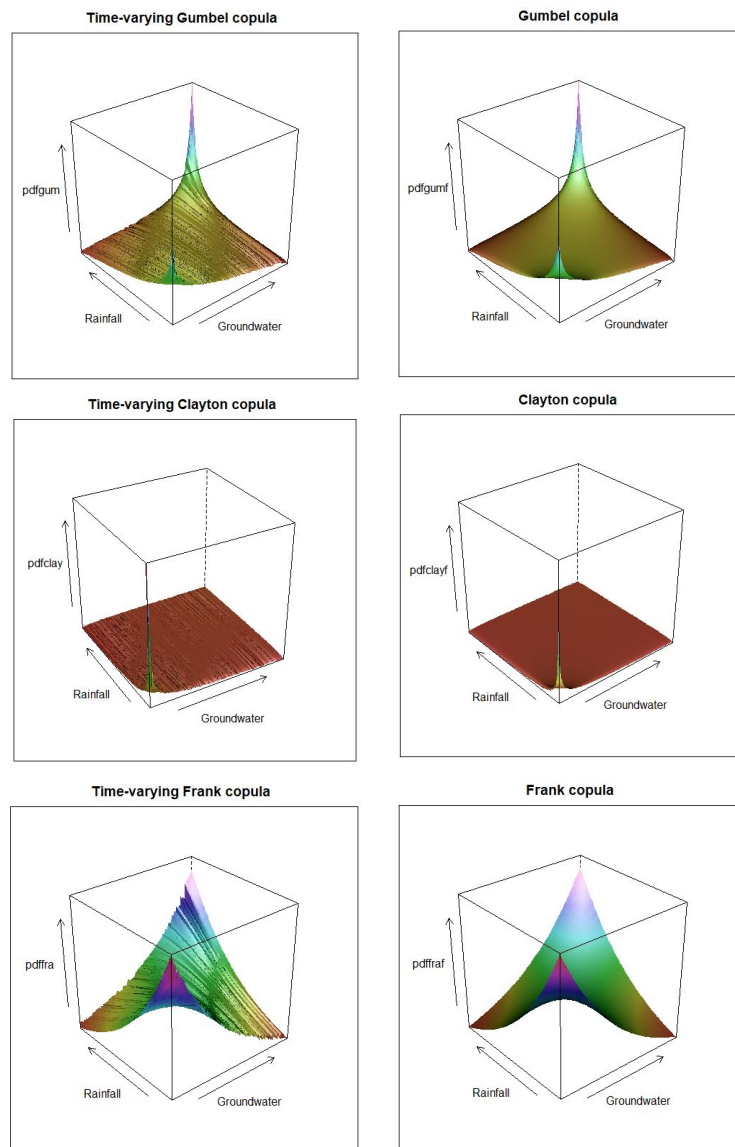
Table 7: Results of fitting copulas with time-varying parameter to data

مفصل	θ_t	AIC_t	LL
گامبل	$exp(0.189 + 0.00017t)$	-۲۱/۶۰	۳۹/۹
فرانک	$1.706 + 0.0014t$	-۲۵/۶۳	۴۷/۲۶
کلیتون	$0.160 + 0.0006t$	-۱۹/۴۲	۳۴/۸۴

حاشیه‌ای در این حالت نیز هر دو مفصل پارامتر ثابت و مفصل زمان-متغیر بر داده‌ها برازش داده شده است. نتایج برازش مفصل‌های گامبل، کلیتون و فرانک پارامتر ثابت و زمان-متغیر به ترتیب در جدول‌های ۶ و ۷ داده شده است.

همان‌طور که نتایج دو جدول ۶ و ۷ نشان می‌دهد، از میان توابع مفصل مختلف پارامتر ثابت و زمان-متغیر، مفصل زمان-متغیر فرانک دارای کمترین مقدار AIC نسبت به سایر مدل‌ها بوده است و به‌عنوان بهترین مدل برازش شده بر داده‌ها انتخاب می‌شود. بنابراین این مفصل زمان-متغیر می‌تواند به‌عنوان مدل مورد نظر جهت تحلیل ساختار همبستگی میان داده‌های بارش و سطح آب زیرزمینی و همچنین محاسبه ریسک و دوره‌های بازگشت و سایر مفاهیم کاربردی در علوم هیدرولوژی مورد استفاده قرار گیرد. جهت مقایسه گرافیکی مفصل‌های پارامتر ثابت و زمان-متغیر، نمودار سه بعدی این مفصل‌ها براساس داده‌های بارش و سطح آب زیرزمینی رسم شده است که در شکل ۴ رسم شده است.

همان‌طور که از مقایسه منحنی‌های شکل ۴ مشخص است، تفاوت آشکاری میان منحنی مفصل‌های پارامتر ثابت و زمان-متغیر وجود دارد. تغییر پارامترهای مفصل برحسب زمان به‌خوبی در منحنی‌ها نمایش داده شده است. به‌عبارت دیگر در مفصل‌های پارامتر ثابت یک پیوستگی یکسان در ابعاد منحنی‌ها وجود دارد درحالی‌که در مفصل‌های زمان-متغیر هر چند ابعاد کلی شکل حفظ شده و تفاوتی با مفصل‌های پارامتر ثابت ندارد اما منحنی به‌صورت لایه‌هایی دیده می‌شود که نشان می‌دهد، با تغییر زمان، چگونه منحنی تغییر می‌کند. جهت بررسی میزان هماهنگی تغییرات در توابع توزیع تجمعی و همچنین تابع توزیع تجمعی مفصل داده‌های بارش و سطح آب زیرزمینی که از توابع زمان-متغیر محاسبه شده، نمودار آنها در شکل ۵ رسم شده است. همان‌طور که نمودار نشان می‌دهد، تغییرات هر سه تابع در طول زمان هماهنگ بوده و به‌عبارتی تابع مفصل برازش داده شده به خوبی وابستگی میان دو متغیر را توصیف می‌کند.

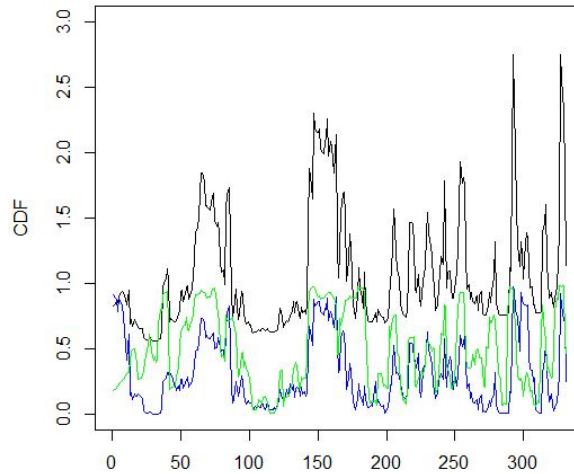


شکل ۴. نمودارهای توابع مفصل پارامتر ثابت و زمان-متغیر برازش داده شده بر داده‌های بارش و سطح آب زیرزمینی

Figure 4: Goodness of fit test plot for fitted time-varying log-normal distribution to rainfall (left) and ground water level (right) data

۴. جمع‌بندی

در این مقاله روشی برای برازش توزیع‌ها در حالتی که داده‌ها در یک ساختار زمانی منظم جمع‌آوری شده و دارای خودهمبستگی باشند ارائه نمودیم. در چنین حالتی فرض مستقل و هم توزیع بودن داده‌ها نقض می‌شود و بنابراین به‌کاربردن توزیع‌های آماری با پارامترهای ثابت برای برازش داده‌ها نادرست است. برای چنین داده‌هایی باید عامل زمان نیز در نظر گرفته شود و از طریق یک تابع رگرسیونی مناسب و بر اساس یک تابع پیوند، وارد مدل شوند. به چنین مدل‌هایی، مدل‌های زمان-متغیر گفته می‌شود. نتایج نشان



شکل ۵. نمودار سری زمانی تابع توزیع تجمعی مفصل زمان-متغیر (نمودار بالا)، تابع توزیع تجمعی متغیر زمان بارش (وسط) و تابع توزیع تجمعی متغیر زمان سطح آب زیرزمینی (پایین)

Figure 5: Plots of fixed parameter and time-varying parameter copulas fitted to rainfall and ground water data

داد که در نظر گرفتن یک تابع رگرسیونی چندجمله‌ای مناسب جهت ارتباط میان پارامتر مدل و متغیر زمان می‌تواند به صورت قابل توجهی نتایج برازش را بهبود بخشد. این تابع رگرسیونی را می‌توان بر اساس منحنی سری زمانی داده‌ها که در آن روند تغییرات میانگین و واریانس سری قابل مشاهده است انتخاب نمود. نتایج این مطالعه می‌تواند در مطالعات اقتصادی، هیدرولوژی، هواشناسی، بیمه و سایر رشته‌هایی که در آنها داده‌ها اغلب در بازه‌های زمانی معین گردآوری شده و دارای خودهمبستگی هستند مورد استفاده قرار گیرد.

ضمیمه ۱

کدهای R جهت برازش مدل گامای تعمیم یافته‌ی جدول ۵.

```
\label{ver1}
> library(gamlss)
> y=data$rainfall # number of observations=331
> t=c(1:331)
# fitting generalized gamma, a ploynomial of degree 3 for "mu" and degree 1 for "sigma"
> fit=gamlss(y~t+I(t^2)+I(t^3) , sigma.fo=~t , family=GG)
> summary(fit)
> AIC(fit)
> plot(fit)
> rqres.plot(fit,howmany=1,type=c("QQ")) # QQ plot
> wp(fit) # Worm plot
```

مراجع

- [1] C. Almeida, C. Czado and H. Manner, Modeling high-dimensional time-varying dependence using dynamic D-vine models, *Appl. Stoch. Models Bus. Ind.*, **32** no. 5 (2016) 621–638.
- [2] S. Amini, R. Z. Bidaki, R. Mirabbasi and M. Shafaei, Flood risk analysis based on nested copula structure in Armand Basin, Iran, *Acta Geophys.*, **70** (2022) 1385–1399.
- [3] Z. Azhdari, O. Bazrafshan, H. Zamani, M. Shekari and V. P. Singh, Hydro-meteorological drought risk assessment using linear and nonlinear multivariate methods, *Physics and Chemistry of the Earth, Parts A/B/C*, **123** (2021).
- [4] B. L. Bowerman, R. T. O’Connell and A. B. Koehler *Forecasting, time series, and regression: an applied approach*, Thomson Brooks/Cole, Business & Economics, 2005, 686 pp.
- [5] G. Casella and R. L. Berger, *Statistical inference*, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990.
- [6] H. Chowdhary, L. A. Escobar and V. P. Singh, Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data, *Hydrology Research*, **1** (2011) 193–216.
- [7] D. J. Dupuis, Using copulas in hydrology: Benefits, cautions, and issues, *J. Hydrol. Eng.*, **12** no. 4 (2007) 381–393.
- [8] P. Jaworski, F. Durante, W. K. Hardle and T. Rychlik, *Copula theory and its applications*, Berlin, Springer, 2010.
- [9] C. Jiang, L. Xiong, C. Y. Xu and S. Guo, Bivariate frequency analysis of nonstationary low-flow series based on the time-varying copula, *Hydrological Processes*, **15** no. 29 (2015) 1521–1534.
- [10] C. M. Hafner and H. Manner, Dynamic stochastic copula models: Estimation, inference and applications, *J. Appl. Econometrics*, **27** no. 2 (2012) 269–295.
- [11] S. Hesarkazzazi, R. Arabzadeh, M. Hajibabaei, W. Rauch, T. R. Kjeldsen, I. Prosdociami, A. Castellarin and R. Sitzenfrei, Stationary vs non-stationary modelling of flood frequency distribution across northwest England, *Hydrological Sciences Journal*, **66** no. 4 (2021) 729–474.
- [12] S. C. Kao and R. S. Govindaraju, A copula-based joint deficit index for droughts, *Journal of Hydrology*, **380** no. 1-2 (2010) 121–134.
- [13] M. J. Machado, B. A. Botero, J. López, F. Francés, A. Díez-Herrero and G. Benito, Flood frequency analysis of historical flood data under stationary and non-stationary modelling, *Hydrol. Earth Syst. Sci.*, **19** no. 6 (2015) 2561–2576.
- [14] H. Manner and O. Reznikova, A survey on time-varying copulas: specification, simulations, and application, *Econometric Rev.*, **31** no. 6 (2012) 654–687.
- [15] A. A. Pathak and B. M. Dodamani, Connection between meteorological and groundwater drought with copula-based bivariate frequency analysis, *J. Hydrol. Eng.*, **26** no. 7 (2021) 05021015.
- [16] A. J. Patton, Modelling asymmetric exchange rate dependence, *Internat. Econom. Rev.*, **47** no. 2 (2006) 527–556.
- [17] M. Rodell, I. Velicogna and J. S. Famiglietti, Satellite-based estimates of groundwater depletion in India, *Nature*, **460** (2009) 999–1002.

- [18] A. Sklar, Random variables, joint distribution functions, and copulas, *Kybernetika (Prague)*, **9** (1973) 449–460.
- [19] D. M. Stasinopoulos and R. A. Rigby, Generalized additive models for location scale and shape (GAMLSS) in R, *Journal of Statistical Software*, **23** (2008) 1-46.
- [20] M. Xing, D. Xu and J. He, Modeling method of failure dependent system based on time varying copula function, *Vibroengineering Procedia*, **21** (2017) 76-81.

حسین زمانی

گروه آمار، دانشکده علوم پایه، دانشگاه هرمزگان، هرمزگان، بندرعباس، ایران
zamani.huni@hormozgan.ac.ir

حسین زمانی متولد آبان ماه ۱۳۵۵ در لارستان است. وی در سال ۱۳۷۷ وارد مقطع کارشناسی رشته آمار دانشگاه رازی کرمانشاه شد و در سال ۱۳۸۱ مقطع کارشناسی ارشد رشته آمار ریاضی خود را در دانشگاه شیراز آغاز نمود. او در سال ۱۳۸۴ به‌عنوان عضو هیئت علمی در دانشگاه هرمزگان استخدام شد و در سال ۱۳۸۷ به‌عنوان مأمور به تحصیل وارد مقطع دکتری رشته آمار در دانشگاه ملی کشور مالزی شد. او هم‌اکنون در دانشگاه هرمزگان به‌عنوان عضو هیئت علمی گروه آمار فعالیت می‌کند.



زهره پاکدامن

گروه آمار، دانشکده علوم پایه، دانشگاه هرمزگان، هرمزگان، بندرعباس، ایران
zpakdaman@hormozgan.ac.ir

زهره پاکدامن متولد اردیبهشت ۱۳۶۷ در مشهد است. وی در سال ۱۳۸۶ وارد مقطع کارشناسی رشته آمار دانشگاه فردوسی مشهد شد و در سال ۱۳۹۰ مقطع کارشناسی ارشد رشته آمار ریاضی خود را در همین دانشگاه آغاز نمود. او در سال ۱۳۹۲ مقطع دکتری خود در دانشگاه فردوسی مشهد را آغاز نمود و هم‌اکنون در دانشگاه هرمزگان به‌عنوان عضو هیئت علمی گروه آمار فعالیت می‌کند.



مرضیه شکاری

گروه آمار، دانشکده علوم پایه، دانشگاه هرمزگان، هرمزگان، بندرعباس، ایران
shekarimuni@hormozgan.ac.ir

مرضیه شکاری متولد تیرماه ۱۳۵۲ در بندرعباس است. وی در سال ۱۳۷۰ وارد مقطع کارشناسی رشته آمار دانشگاه فردوسی مشهد شد و در سال ۱۳۷۵ مقطع کارشناسی ارشد رشته آمار ریاضی خود را در دانشگاه شهید بهشتی آغاز نمود. او در سال ۱۳۸۵ مقطع دکتری خود در دانشگاه شیراز را آغاز نمود. او هم‌اکنون در دانشگاه هرمزگان به‌عنوان عضو هیئت علمی گروه آمار فعالیت می‌کند.

