



## Evaluation of Entropy Theory Based on Random Forest in Quality Monitoring of Ground Water Network

Fatemeh Bageri<sup>1</sup> | Keivan Khalili<sup>2</sup> | Mohammad Nazeri Tahroudi<sup>3</sup>

1. Department of Civil Engineering, Saba Institute of Higher Education, Urmia, Iran. E-mail: [bageri27553@chmail.ir](mailto:bageri27553@chmail.ir)
2. Corresponding Author, Department of Water Engineering, Urmia University, Urmia, Iran. E-mail: [k.khalili@urmia.ac.ir](mailto:k.khalili@urmia.ac.ir)
3. Department of Water Engineering, Shahrekord University, Shahrekord, Iran. E-mail: [m\\_nazeri2007@yahoo.com](mailto:m_nazeri2007@yahoo.com)

### Article Info

#### Article type:

Research Article

#### Article history:

Received: August 11, 2022

Received in revised form:

October 12, 2022

Accepted: October 23, 2022

Published online: April 14, 2023

#### Keywords:

Entropy Theory,

Information Transformation,

Quality Monitoring,

Random Forest.

### ABSTRACT

The quality monitoring of groundwater networks is of great importance due to its importance in different sectors of agriculture, drinking, industry, etc., and the necessity of its periodic use leads to the recognition of quality changes of water resources in different periods. In this study, the entropy theory based on random forest was used to quality monitoring of electrical conductivity (EC) and total dissolved solids (TDS) in the groundwater of 12 wells in the Tasouj plain located in south of Lake Urmia from 2002 through 2019. In order to investigate the interaction of wells in the aquifer area, the conventional method (multivariate regression) and the random forest algorithm were used. By comparing the performance of the two mentioned models in simulating EC and TDS values in a 12-variable mode, the results showed that the random forest model has a better performance and a lower error rate than the multi-variable regression model. On average, the random forest algorithm reduced the error rate by 40 percent and 56 percent in simulation EC and TDS values, respectively, in the studied aquifer. The ranking results of the studied wells showed that the Qara Tape well has the highest rank and the Amestjan well has the least important rank among the studied wells, which indicates the importance of the information extracted from the Qara Tape well. According to the zoning of the information transformation index in the aquifer area, the results showed that there is no limitation in monitoring of EC values in the aquifer, and the scattering of the wells is the best. There is no shortage of wells in terms of exchange of salinity information in the study area. Regarding the TDS values, a lack of wells was observed in the central areas and the eastern and western border areas.

**Cite this article:** Bageri, F., Khalili, K., & Nazeri Tahroudi, M. (2023). Evaluation of Entropy Theory Based on Random Forest in Quality Monitoring of Ground Water Network. *Journal of Water and Irrigation Management*, 13 (1), 123-139. DOI: <https://doi.org/10.22059/jwim.2022.347038.1010>



© The Author(s).

DOI: <https://doi.org/10.22059/jwim.2022.347038.1010>

Publisher: University of Tehran Press.



## ارزیابی آنتروپی تئوری مبتنی بر جنگل تصادفی در پایش کیفی شبکه آب زیرزمینی

فاطمه باقری<sup>۱</sup> | کیوان خلیلی<sup>۲</sup> | محمد ناظری تهرودی<sup>۳</sup>

۱. گروه عمران، دانشکده مهندسی، موسسه آموزش عالی صبا، ارومیه، ایران. رایانامه: [bageri27553@chmail.ir](mailto:bageri27553@chmail.ir)

۲. نویسنده مسئول، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه ارومیه، ارومیه، ایران. رایانامه: [k.khalili@urmia.ac.ir](mailto:k.khalili@urmia.ac.ir)

۳. گروه مهندسی آب، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران. رایانامه: [m\\_nazeri2007@yahoo.com](mailto:m_nazeri2007@yahoo.com)

### اطلاعات مقاله

### چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۱/۰۵/۲۰

تاریخ بازنگری: ۱۴۰۱/۰۷/۲۰

تاریخ پذیرش: ۱۴۰۱/۰۸/۰۱

تاریخ انتشار: ۱۴۰۲/۰۱/۲۵

### کلیدواژه‌ها:

انتقال اطلاعات،

پایش کیفی،

تئوری آنتروپی،

جنگل تصادفی.

پایش کیفی شبکه‌های آب زیرزمینی به دلیل اهمیت آن در بخش‌های مختلف کشاورزی، شرب، صنعت و غیره از اهمیت بالایی برخوردار بوده و لزوم استفاده دوره‌ای آن موجب شناخت تغییرات کیفی منابع آب در دوره‌های مختلف می‌شود. در این مطالعه جهت پایش کیفی مقادیر هدایت الکتریکی (EC) و کل مواد جامد محلول (TDS) در آب زیرزمینی ۱۲ حلقه چاه در محدوده دشت تسوج واقع در شمال دریاچه ارومیه از تئوری آنتروپی مبتنی بر جنگل تصادفی در دوره آماری ۹۸-۱۳۸۲ استفاده شد. جهت بررسی اثر متقابل چاه‌ها در محدوده آبخوان از روش مرسوم (رگرسیون چندمتغیره) و الگوریتم جنگل تصادفی استفاده شد. با مقایسه عملکرد دو مدل یادشده در شبیه‌سازی مقادیر EC و TDS در حالت ۱۲ متغیره، نتایج نشان داد که مدل جنگل تصادفی از عملکرد بهتر و میزان خطای کم‌تری نسبت به مدل رگرسیون چندمتغیره برخوردار است. به طور متوسط الگوریتم جنگل تصادفی موفق به کاهش ۴۰ درصدی در شبیه‌سازی مقادیر EC و کاهش ۵۶ درصدی در شبیه‌سازی مقادیر TDS در آبخوان مورد مطالعه شد. نتایج رتبه‌بندی چاه‌های مورد بررسی نشان داد که چاه قره‌تپه برترین رتبه و چاه امستجان کم‌اهمیت‌ترین رتبه را در بین چاه‌های مورد مطالعه کسب کرده‌اند که بیانگر اهمیت اطلاعات استخراج‌شده از چاه قره‌تپه می‌باشد. با توجه به پهنه‌بندی شاخص انتقال اطلاعات در محدوده آبخوان نتایج نشان داد که در مورد پایش مقادیر EC محدودیتی در آبخوان وجود ندارد و پراکندگی چاه‌ها به بهترین حالت می‌باشد. هیچ‌گونه کمبود چاه نیز از نظر تبادل اطلاعات شوری در منطقه مورد مطالعه احساس نمی‌شود. در خصوص مقادیر TDS کمبود چاه در نواحی مرکزی و نوار مرزی شرقی و غربی مشاهده شد.

**استناد:** باقری، ف.، خلیلی، ک. و ناظری تهرودی، م (۱۴۰۲). ارزیابی آنتروپی تئوری مبتنی بر جنگل تصادفی در پایش کیفی شبکه آب زیرزمینی. نشریه

مدیریت آب و آبیاری، ۱۳ (۱)، ۱۳۹-۱۲۳. DOI: <https://doi.org/10.22059/jwim.2022.347038.1010>



## ۱. مقدمه

پایش کیفی شبکه آب زیرزمینی به دلیل اهمیت آن در مدیریت آبخوان‌ها از دیرباز توجه پژوهش‌گران را به خود جلب نموده است. در این خصوص مطالعات مختلفی در نقاط مختلف دنیا صورت گرفته و مدل‌های مختلفی نیز ارائه شده است. یکی از این مدل‌ها، تئوری آنتروپی می‌باشد (Nazeri Tahroudi *et al.*, 2019). این تئوری به دلیل ارائه شاخص‌های مختلف، توانایی بالایی در شناسایی کمبود یا افزونگی اطلاعات در آبخوان دارد. رتبه‌بندی ایستگاه‌ها و چاه‌ها نیز از دیگر شاخص‌های مهم این تئوری می‌باشد که این روش را نسبت به سایر روش‌ها متمایز کرده است.

مطالعات تئوری آنتروپی در ابتدا در پژوهش‌های Jaynes (1957) و Shannon (1948) شد. از آن پس در مطالعات مختلفی مورد بررسی قرار گرفت. در سال ۱۹۴۸ نیز روش تئوری آنتروپی در بحث پایش مطرح شد (Shannon, 1948). Chapman (1986) از تئوری آنتروپی به عنوان شاخص عدم قطعیت داده‌های هیدرولوژیک استفاده کرد و با بررسی اثر کلاس‌بندی‌های مختلف بر شاخص‌های آنتروپی، به این نتیجه رسید که تئوری آنتروپی با تغییر تعداد کلاس‌ها نتایج متفاوتی ارائه می‌کند. Krstanovic (1988) از تئوری آنتروپی برای ارزیابی مکانی و زمانی شبکه ایستگاه‌های باران سنج در ایالت لوئیزیانا در آمریکا استفاده کرد. ایشان وابستگی‌های مکانی و زمانی در اطلاعات ایستگاه‌های باران سنج را به وسیله ماتریس‌های خود همبسته ارائه کرد. این روش سبب ارائه روابط ساده‌تر در خصوص تئوری آنتروپی شد. Singh (1997) تئوری آنتروپی مرزی را برای ترسیم نقشه‌های خطوط هم بارش ارائه کرد. ایشان با استفاده از تئوری آنتروپی مرزی به ارزیابی شبکه باران سنجی پرداخت. نتایج پژوهش‌های نشان‌دهنده قابلیت بالای تئوری آنتروپی در ارزیابی شبکه طراحی بود. Ozkul *et al.* (2000) از تئوری آنتروپی برای ارزیابی شبکه پایش کیفی رودخانه استفاده کردند.

Mogheir and Singh (2002) برای طراحی شبکه پایش کیفی آب زیرزمینی تئوری آنتروپی انتقالی را توسعه و معرفی کردند. در نهایت با استفاده از نقشه‌های آنتروپی مرزی، ارزیابی شبکه پایش کیفی آب زیرزمینی را ارائه کردند. پژوهش‌های یادشده براساس تئوری آنتروپی ساده بنا شده و همگی به طراحی شبکه پایش پرداخته‌اند. درحالی که Mogheir *et al.* (2006) مطالعه متفاوتی در این خصوص ارائه کردند. Mogheir *et al.* (2006) پایش شبکه آب زیرزمینی را در نوار غزه با استفاده از تئوری آنتروپی انجام دادند. آن‌ها چارچوب منظمی را برای ارزیابی رابطه بین اهداف، عملکردها، داده‌ها و فعالیت‌های پایشی با استفاده از تئوری آنتروپی پیشنهاد دادند. آن‌ها شاخص‌های کاربردی در خصوص تئوری آنتروپی ارائه کردند که در مطالعات بعدی توسط پژوهش‌گران مختلف مورد استفاده قرار گرفت. Chen *et al.* (2008) از روش کریجینگ برای درون‌یابی داده‌های بارش ماهیانه چندین چاه باران سنجی در منطقه شیمن تایوان در شبکه‌ای به ابعاد  $7/5 \times 7/5$  کیلومتر استفاده کردند. در نهایت از آنتروپی انتقال اطلاعات برای اولویت‌بندی ایستگاه‌های بارش استفاده کردند. نتایج نشان داد که شبکه بهینه در این منطقه به تعداد شش ایستگاه نیاز دارد. Lee (2013) روشی را برای انتخاب بهینه نقاط پایش کیفیت آب در سیستم‌های فاضلاب براساس تئوری آنتروپی پیشنهاد کرد. مدل پیشنهادی به دنبال ارزیابی کمی از داده‌های جمع‌آوری شده از نقاط برداشت بود. نقاطی که کل اطلاعات را در بین داده‌های جمع‌آوری شده در مکان‌های مختلف به حداکثر می‌رسانند، با استفاده از الگوریتم ژنتیک (GA) برای پایش کیفیت آب انتخاب شدند. مدل پیشنهادی برای یک سیستم فاضلاب شهری کوچک نشان‌دهنده سیستم پایش منظم و کارآمد بود. Leite *et al.* (2021) جهت بررسی انتقال اطلاعات بین ایستگاه‌های بادسنجی از تئوری آنتروپی استفاده کردند. نتایج نشان داد که روش پیشنهادی می‌تواند به خوبی بین حالت‌های خطا تمایز ایجاد کند، اما محدودیت‌هایی را نیز ارائه می‌کند، که به طور عمده مربوط به پیچیدگی سیگنال از توربین‌های بادی است. براساس این نتایج، روش‌های جدیدی مکمل تحلیل فوریه برای استفاده در داده‌های توربین بادی با هدف افزایش قابلیت تشخیص خطا در چنین محیط پیچیده‌ای پیشنهاد شده است.

Shi *et al.* (2018) جهت بررسی و پایش شبکه کمی و کیفی رودخانه در مواقع اورژانسی نشت از تئوری آنتروپی استفاده کردند. نتایج نشان داد که این چارچوب می‌تواند برنامه‌های آمادگی علمی پایش اضطراری را براساس تحلیل سناریویی خطرات نشت و همچنین طراحی سریع به محض وقوع حادثه، اما آماده نشده، تولید کند. در مقایسه با رویکردهای طراحی شبکه پایش موجود، مفاهیم مدیریتی و کاربرد تعمیم یافته نیز مورد بحث قرار گرفت. Nazeri Tahroudi *et al.* (2019) جهت بازطراحی و پایش شبکه آب زیرزمینی دشت نقده در جنوب غربی دریاچه ارومیه به منظور بررسی تعداد و موقعیت چاه‌های بهینه برای انتقال اطلاعات شوری (EC) و بررسی سطح آب زیرزمینی در سفره‌های زیرزمینی از تئوری آنتروپی استفاده کردند. نتایج شاخص‌های آنتروپی در آبخوان نشان داد که کمبود شدید اطلاعات از نظر شوری در شمال غربی آبخوان وجود دارد که نیاز به حفاری چاه جدیدی در این منطقه برای پایش دقیق مقادیر EC دارد.

پژوهش‌های یادشده در خصوص استفاده از تئوری آنتروپی همگی بر پایه رگرسیون چندمتغیره که روش مرسوم در این خصوص است، بنا شده‌اند. روش رگرسیون چندمتغیره در مواردی که همبستگی بین مقادیر مورد مطالعه در هر چاه وجود نداشته باشد و یا کم باشد، نتایج مطلوبی ارائه نمی‌کند. همچنین به دلیل وجود عدد ثابت در رابطه پیشنهادی، وزن بیش‌تر شبیه‌سازی بر این عدد ثابت بنا شده است. روش رگرسیون چندمتغیره به دلیل ساده بودن روابط در بیش‌تر پژوهش‌های مربوط به شبکه‌های پایش مورد استفاده قرار گرفته است. در برخی موارد مانند (Ramezani *et al.*, 2018; Shahidi *et al.*, 2019; Khashei-Siuki *et al.*, 2020) مشاهده می‌شود که روش‌هایی هم‌چون رگرسیون بردار پشتیبان و مدل‌های سری زمانی چندمتغیره جایگزین رگرسیون چندمتغیره شده است. در این مطالعه سعی شده است که روش جنگل تصادفی نیز در خصوص شبیه‌سازی مقادیر EC و TDS در تعامل و اثر متقابل ایستگاه‌ها بررسی شود. بر این اساس در این مطالعه عملکرد تئوری آنتروپی مبتنی بر دو روش رگرسیون چندمتغیره و جنگل تصادفی در پایش کیفی آب زیرزمینی دشت تسوج در خصوص پایش اطلاعات شوری و TDS مورد بررسی قرار گرفته است.

## ۲. مواد و روش‌ها

منطقه مورد مطالعه در این پژوهش آبخوان زیرحوضه تسوج واقع در شمال غرب کشور، شمال استان آذربایجان غربی و شمال دریاچه ارومیه می‌باشد، مساحت زیرحوضه تسوج برابر با ۵۵۹ کیلومتر مربع و مساحت آبخوان آن برابر با ۲۷۷ کیلومتر مربع می‌باشد. محیط آبخوان دشت تسوج نیز برابر با ۱۰۶ کیلومتر می‌باشد. در این مطالعه از مقادیر EC و TDS اندازه‌گیری شده از تعداد ۱۲ حلقه چاه در دوره آماری ۹۸-۱۳۸۲ استفاده شد. موقعیت منطقه مورد مطالعه به شرح شکل (۱) و مشخصات آماری داده‌های مورد بررسی به شرح جدول (۱) ارائه شد.

### ۱.۲. تئوری آنتروپی

تئوری آنتروپی در لغت به معنای بی‌نظمی است. تعاریف مختلفی از تئوری آنتروپی در مطالعات مختلف ارائه شده است. اما تئوری آنتروپی برای استفاده به‌عنوان یک ابزار مدل‌سازی به‌منظور استفاده بیش‌تر از اطلاعات موجود استفاده می‌شود (Harmancioglu *et al.*, 1999). این تئوری اولین بار در مطالعات Shannon (1948) بیان شد. اولین مطالعات در خصوص استفاده از تئوری آنتروپی در بحث عدم قطعیت بود. با به‌دست آوردن اطلاعات جدید، می‌توان عدم قطعیت‌ها را کاهش داد. شاخص‌های تئوری آنتروپی را می‌توان به دو دسته آنتروپی گسسته و پیوسته تقسیم نمود (Harmancioglu and Alpaslan, 1992).

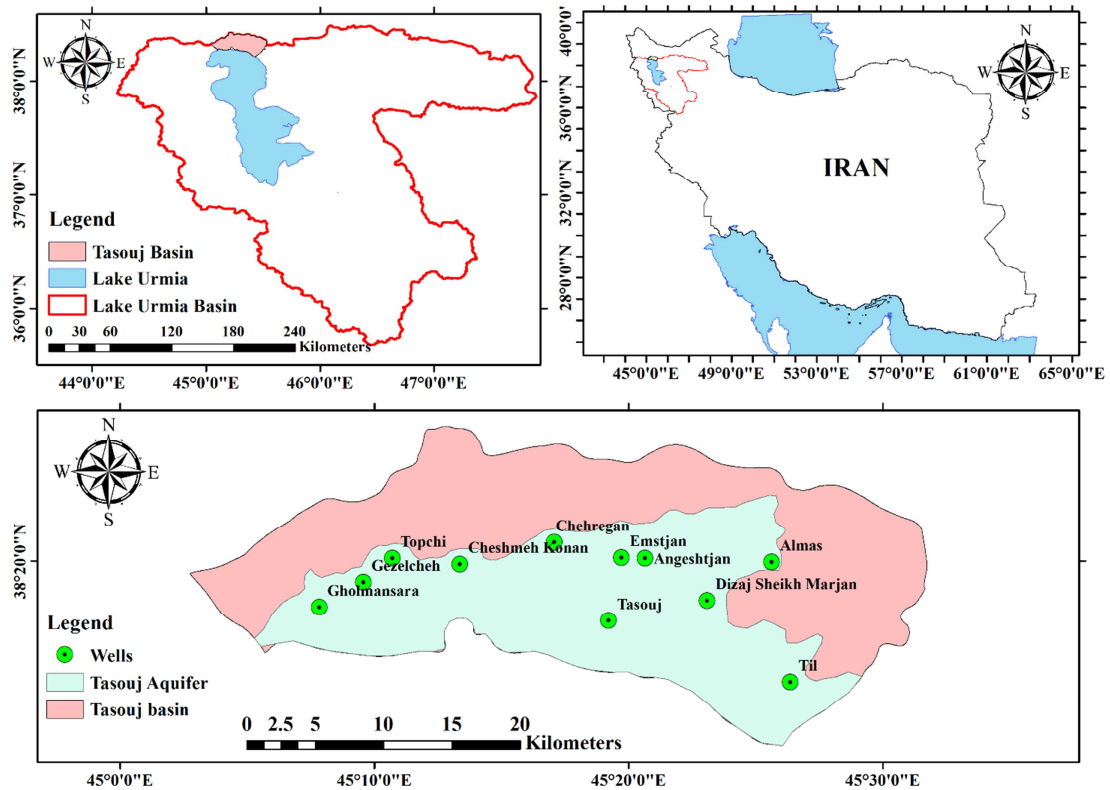


Figure 1. Location of studied wells in Iran and Lake Urmia Basin

Table 1. Statistical characteristics of the investigated data in the Tasouj aquifer

Well	Min	Max	Mean	STD
Almas	435.0	799.00	659.57	84.96
Emstjan	1018.5	1400.0	1277.9	99.32
Tasouj	962.0	1523.5	1340.40	174.18
Angeshtjan	648.0	1020.5	839.68	126.03
Topchi	1177.0	2390.3	2007.94	259.39
Til	1064.5	2564.0	1754.99	478.28
Cheshmeh Konan	1280.0	3300.0	2047.31	608.55
Chehregan	563.0	2299.7	1268.31	554.05
Dizaj Sheikh Marjan	508.5	726.0	611.09	56.94
Gholmansara	1273.5	2755.0	1850.84	559.13
Qareh Tapeh	1650.0	2930.0	2144.31	410.33
Gezelcheh	2500.0	329.0	3012.90	203.74

یک متغیر تصادفی پیوسته با تابع توزیع چگالی احتمال  $f(x)$  را در نظر بگیرید، در این صورت احتمال این که  $x$  در  $\lambda$ امین بازه باشد، عبارت است از:

$$P(x_i) = P(x_i - \frac{\Delta x}{2} \leq x_i \leq x_i + \frac{\Delta x}{2}) = \int_{x_i - \frac{\Delta x}{2}}^{x_i + \frac{\Delta x}{2}} f(x) dx \quad \text{رابطه (۱)}$$

با قراردادن  $p(x)$  در معادله  $H(x) = -k \sum_{i=1}^{\infty} p(x_i) \ln p(x_i)$  رابطه آنتروپی مرزی به صورت زیر خواهد بود (Shannon, 1948):

$$H(x) = -\int_{-\infty}^{\infty} f(x) \ln f(x) dx \quad \text{رابطه ۲}$$

در رابطه بالا،  $H$  بیان کننده مقدار عدم قطعیت است که این عدم قطعیت بیان کننده میزان کمبود اطلاعات در خصوص یک سیستم است. آنتروپی مشترک دو متغیر  $x$  و  $y$  با  $H(x, y)$  بیان می شود. اگر  $x$  و  $y$  غیر وابسته باشند برابر است:

$$H(x, y) = H(x) + H(y) \quad \text{رابطه ۳}$$

که در آن،  $H(x)$  و  $H(y)$  آنتروپی مرزی  $x$  و  $y$  می باشند. حالت وابسته آن نیز به صورت زیر می باشد:

$$H(x, y) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \ln f(x, y) dx dy \quad \text{رابطه ۴}$$

که در آن،  $f(x, y)$  چگالی احتمال مشترک  $x$  و  $y$  هست. در تئوری آنتروپی، حالت آنتروپی شرطی  $x$  به شرط  $y$  بیان کننده عدم قطعیت باقیمانده در  $x$  به شرط مشخص بودن  $y$  است:

$$H(x|y) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x|y) \ln f(x|y) dx dy \quad \text{رابطه ۵}$$

که  $f(x|y)$  تابع توزیع چگالی احتمال  $x$  به شرط  $y$  است. نوعی دیگر از آنتروپی که برابر با تفاضل مجموع آنتروپی های حاشیه ای و آنتروپی مشترک است، آنتروپی انتقال اطلاعات گویند که بین دو متغیر  $x$  و  $y$  به صورت زیر بیان می شود:

$$T(x, y) = H(x) + H(y) - H(x, y) \quad \text{رابطه ۶}$$

$$T(x, y) = T(y, x) = -\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \ln \left( \frac{f(x, y)}{f(x)f(y)} \right) dx dy \quad \text{رابطه ۷}$$

$T(x, y)$  را می توان به صورت اطلاعات مشترک  $x$  و  $y$  نیز تعریف نمود. آنتروپی حاشیه ای با توابع توزیع مختلف قابل برآورد می باشد که آنتروپی حاشیه ای توزیع های رایج نظیر نرمال، لاگ نرمال، گاما، بتا، نمایی و مقادری حدی تعمیم یافته به ترتیب به صورت زیر می باشد (Nazeri Tahroudi *et al.*, 2019):

## ۲.۲. شاخص های مهم آنتروپی

انتقال اطلاعات هم چنین می تواند با استفاده از یک شاخص انتقال اطلاعات نرمال که با  $ITI$  نشان داده می شود، بیان گردد:

$$ITI = \frac{T(x, y)}{H(x, y)} \quad \text{رابطه ۸}$$

که میزان اطلاعات انتقالی از ایستگاه به ایستگاه دیگر را نشان می دهد. این شاخص از مهم ترین شاخص های تئوری آنتروپی می باشد که میزان کمبود و افزونگی اطلاعات را در منطقه نشان می دهد. طبقه بندی این شاخص به صورت جدول (۲) می باشد.

**Table 2. Classification of information transfer index ITI (Mishra and Coulibaly, 2010)**

The degree of importance of the area	ITI (i)
Severe shortage	0 – 0.2
Shortage	0.2 – 0.4
Average	0.4 – 0.6
Above average	0.6 – 0.8
Extra	>0.8

اطلاعات دریافتی توسط ایستگاه  $x$  از  $y$  نیز به صورت زیر تعریف می‌شود:

$$R(x, y) = \frac{T(x, y)}{H(x)} \quad \text{رابطه ۹}$$

میزان اطلاعات ارسالی ایستگاه  $x$  از ایستگاه  $y$  به صورت زیر تعریف می‌شود:

$$S(x, y) = \frac{T(x, y)}{H(y)} \quad \text{رابطه ۱۰}$$

در نهایت، شاخص  $N(i)$  یا شاخص اطلاعات خالص تبادل به صورت زیر می‌باشد:

$$N(i) = S(i) - R(i) \quad \text{رابطه ۱۱}$$

شاخص  $N(i)$  ارزش هر ایستگاه را نشان می‌دهد که کم‌ترین میزان  $N(i)$  بیانگر کم‌ترین رتبه و اهمیت را در شبکه پایش نشان می‌دهد (Markus et al., 2003). لازمه اجرای تئوری آنتروپی، تخمین مقادیر مورد مطالعه در هر چاه در خصوص بررسی اثر متقابل ایستگاه‌ها می‌باشد که در حالت مرسوم، از رگرسیون چندمتغیره استفاده می‌شود. در این مطالعه سعی شده است که روش جنگل تصادفی استفاده شود. رگرسیون چندمتغیره به صورت زیر تعریف می‌شود:

$$\hat{x}(i) = a(i) + \sum_{j=1}^{I-1} y_j(i) \times b_j(i) \quad \text{رابطه ۱۲}$$

که  $x(i)$  بیان‌کننده داده‌های ایستگاه  $i$  ام و  $\hat{x}(i)$  از رابطه خطی بالا به دست می‌آید.  $y(i)$  ماتریس داده‌ها از تمام ایستگاه‌های دیگر و  $a(i)$  و  $b(i)$  پارامترهای رگرسیون می‌باشند.

### ۳.۲. مدل جنگل تصادفی

الگوریتم جنگل تصادفی توسط Breiman (2001) به عنوان یک روش یادگیری تجمعی برای مسائل مبتنی بر رگرسیون و خوشه بندی براساس توسعه درخت تصمیم، ارائه شده است. یک جنگل تصادفی، مجموعه‌ای از درخت‌های هرس نشده است که هر درخت با الگوریتم جزءبندی بازگشتی حاصل می‌شود. به عبارتی، جنگل تصادفی ترکیبی از چندین درخت تصمیم است که در ساخت آن چندین نمونه خودسازمانده از داده‌ها شرکت دارند (Friedman et al., 2001). به منظور ایجاد درخت رگرسیونی از جزءبندی بازگشتی و رگرسیون‌های چندگانه استفاده می‌شود. فرایند تصمیم در هر گره داخلی از گره ریشه، طبق قانون درختی تکرار می‌شود تا زمانی که شرط توقف قبلی تعیین شده به دست آید (Breiman, 2001).

در روش RF بردار تصادفی  $X_n$  که مستقل از بردارهای تصادفی  $X_1, X_2, \dots, X_{n-1}$  بوده، برای درخت  $n$ ام، تولید می‌شود. هم‌چنین همه بردارها از توزیع مشابهی تبعیت می‌کنند. رگرسیون درختی با استفاده از مجموعه داده‌های آموزش  $X_n$  محاسبه شده مجموعه درخت‌هایی برابر با  $n$  را به شرح زیر تولید می‌نماید (Breiman, 2001):

$$X_n = \{h_1(x), h_2(x), \dots, h_n(x)\} \quad \text{رابطه ۱۳}$$

$$h_n = h(x, X_n), x = \{x_1, x_2, \dots, x_p\} \quad \text{رابطه ۱۴}$$

بردار  $p$  بعدی فوق یک جنگل را تشکیل داده و خروجی‌ها برای هر درخت به صورت زیر ارائه می‌شود:

$$\hat{y}_1 = h_1(x), \hat{y}_2 = h_2(x), \dots, \hat{y}_n = h_n(x) \quad \text{رابطه ۱۵}$$

که در رابطه فوق  $\hat{y}_n$  خروجی درخت  $n$ ام می‌باشد. برای به دست آوردن خروجی نهایی، متوسط همه پیش‌بینی‌های درخت‌ها محاسبه می‌شود. خطای پیش‌بینی نیز طبق رابطه (۱۶) محاسبه می‌شود.

$$MSE = \frac{\sum_{i=1}^n [\widehat{y}(x_i) - y_i]^2}{n} \quad \text{رابطه ۱۶}$$

در رابطه فوق  $\widehat{y}(x_i)$  مقادیر محاسباتی،  $y_i$  مقادیر مشاهداتی و  $n$  تعداد کل مشاهدات است و MSE میزان خطای بین مقادیر مشاهداتی و محاسباتی را نشان می‌دهد. در این تحقیق جهت اجرای مدل آنتروپی مبتنی بر دو روش رگرسیون چندمتغیره و جنگل تصادفی مراحل زیر در نظر گرفته شد:

- ۱- آماده‌سازی داده‌های مورد استفاده از جمله EC و TDS در منطقه مورد مطالعه
- ۱- بررسی خلأ آماری و تکمیل و تطویل آن در صورت لزوم
- ۲- اجرای روش رگرسیون چندمتغیره در خصوص شبیه‌سازی اثر متقابل ایستگاه‌ها
- ۳- اجرای روش جنگل تصادفی در خصوص شبیه‌سازی اثر متقابل ایستگاه‌ها
- ۴- ارزیابی عملکرد دو مدل‌های جنگل تصادفی و رگرسیون چندمتغیره بر اساس معیارهای ارزیابی
- ۵- اجرای روش آنتروپی مبتنی بر برترین مدل شبیه‌ساز مرحله ۵
- ۶- برآورد شاخص‌های آنتروپی و بررسی شبکه کیفی پایش آب زیرزمینی

### ۳. نتایج و بحث

در این منطقه جهت پایش کیفی آب زیرزمینی منطقه مورد مطالعه از نظر انتقال شوری (EC) و TDS در چاه‌های مورد بررسی، تئوری آنتروپی مورد بررسی قرار گرفت. در این خصوص جهت بررسی اثر متقابل چاه‌های مورد بررسی از روش مرسوم (رگرسیون چندمتغیره) و مدل جنگل تصادفی مورد بررسی قرار گرفت. از آنجاکه انتقال اطلاعات از نظر مقادیر EC و TDS ممکن است از چاهی به چاه دیگر متفاوت باشد، جهت پایش شبکه کیفی آب زیرزمینی در این منطقه از دو پارامتر EC و TDS استفاده شد.

#### ۳.۱. نتایج شبیه‌سازی مقادیر EC و بررسی اثر متقابل ایستگاه‌ها

جهت شبیه‌سازی مقادیر مقادیر EC در هر ایستگاه، از مقادیر EC سایر ایستگاه‌ها به صورت عنصر پیش‌بینی‌کننده استفاده شد. با این کار، اثر متقابل ایستگاه‌ها نسبت به یکدیگر در نظر گرفته می‌شود. با استفاده از هر دو مدل رگرسیون چندمتغیره و جنگل تصادفی، شبیه‌سازی مقادیر EC در چاه‌های موجود در زیرحوضه تسوج انجام شد. نتایج شبیه‌سازی چندمتغیره مقادیر EC در منطقه مورد مطالعه با استفاده از دو آماره جذر میانگین مربعات خطا و نش- ساتکلیف بررسی شد. نتایج بررسی مقادیر RMSE در شبیه‌سازی مقادیر EC در چاه‌های مورد مطالعه با استفاده از هر دو مدل مورد بررسی به صورت شکل (۲) ارائه شد. نتایج مقایسه مقادیر RMSE در خصوص شبیه‌سازی مقادیر EC در منطقه مورد مطالعه نشان داد که این آماره بین ۲۵ تا ۳۳۱ میکروموس بر سانتی‌متر برای مدل رگرسیون چندمتغیره و بین ۲۳ تا ۶۸ میکروموس بر سانتی‌متر برای مدل جنگل تصادفی متغیر می‌باشد. با توجه به شکل (۲) می‌توان مشاهده کرد در مدل رگرسیون چندمتغیره بیش‌ترین میزان خطای شبیه‌سازی مربوط به چاه چهرگان و در مورد مدل جنگل تصادفی مربوط به چاه چشمه‌کنان می‌باشد. کم‌ترین میزان خطا نیز در هر دو مدل مربوط به چاه شیخ مرجان می‌باشد. با توجه به شکل (۲) می‌توان مشاهده کرد که در تمامی چاه‌های مورد مطالعه میزان RMSE در شبیه‌سازی مقادیر EC ناشی از مدل جنگل تصادفی کمتر از مدل رگرسیون چندمتغیره می‌باشد. جهت بررسی کارایی دو مدل مورد بررسی در شبیه‌سازی مقادیر EC



از آماره نش- ساتکلیف استفاده شد که نتایج به صورت شکل (۳) می باشد. نتایج شکل (۳) نشان داد که کارایی مدل جنگل تصادفی در تمامی چاه های مورد بررسی بیش تر از مدل رگرسیون چندمتغیره می باشد. کارایی مدل جنگل تصادفی در شبیه سازی مقادیر EC در چاه های مورد بررسی در تمامی موارد بیش از ۸۰ درصد می باشد که نسبت به مدل رگرسیون چندمتغیره بیش تر می باشد. درصد بهبودی میزان خطا (آماره RMSE) مدل جنگل تصادفی در شبیه سازی مقادیر EC نسبت به مدل رگرسیون چندمتغیره نیز به صورت شکل (۴) قابل مشاهده می باشد. بیش ترین درصد بهبودی در نتایج RMSE ناشی از مدل جنگل تصادفی با توجه به شکل (۴) مربوط به دو چاه چهرگان و غلمانسرا می باشد. در تمامی موارد با توجه به شکل (۴)، میزان خطای شبیه سازی مقادیر EC متأثر از مدل جنگل تصادفی نسبت به مدل رگرسیون چندمتغیره بهبود یافته است. با توجه به شکل های (۲)، (۳) و (۴) می توان دقت قابل قبول مدل جنگل تصادفی را در شبیه سازی EC و بررسی اثر متقابل چاه های موجود نسبت به مدل رگرسیون چندمتغیره مشاهده کرد.

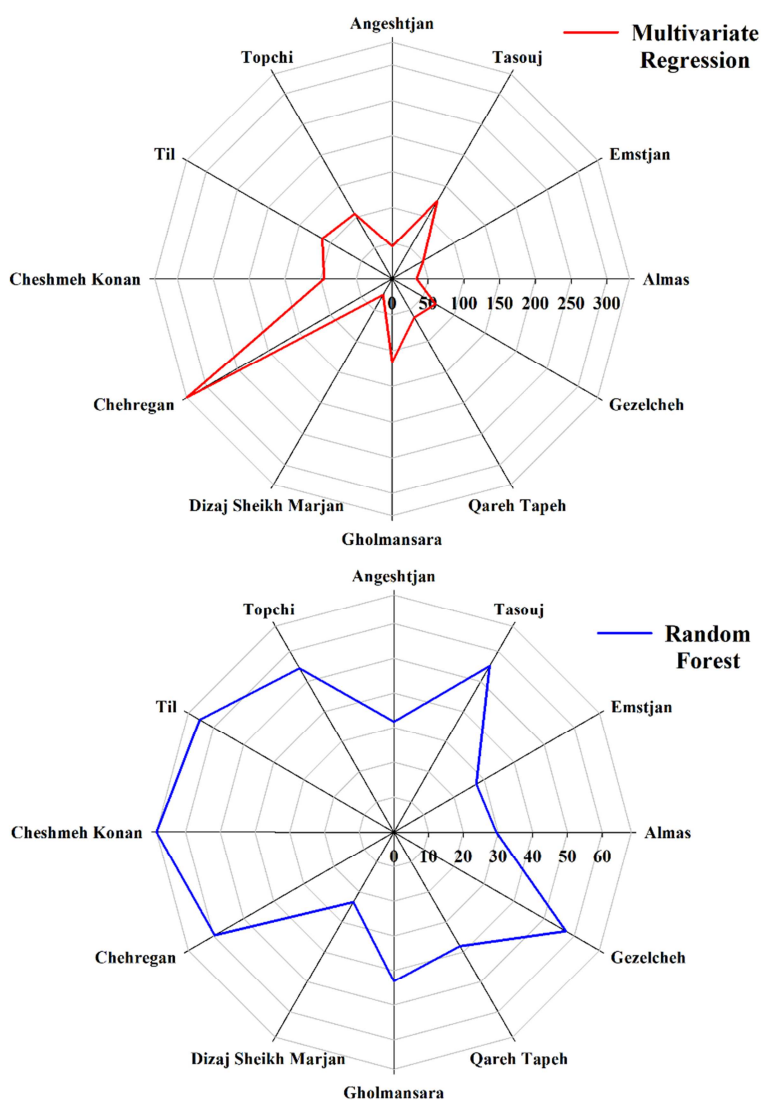


Figure 2. The results of investigating the error rate (RMSE) of random forest models and multivariate regression in the simulation of electrical conductivity values in the studied stations.

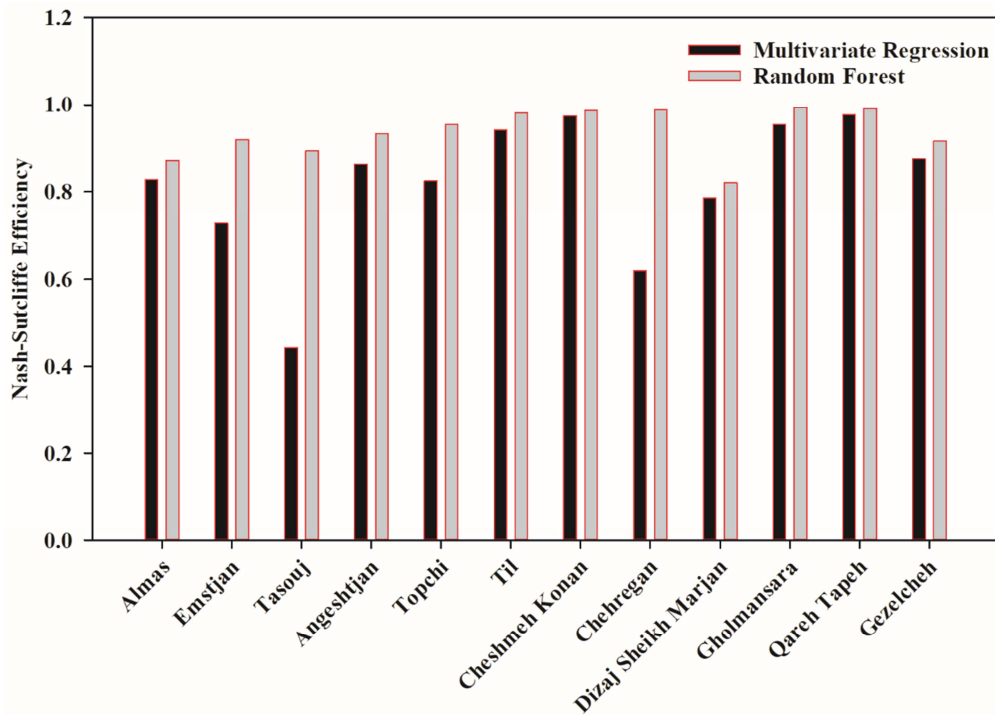


Figure 3. Performance evaluation of random forest models and multivariate regression in the simulating of EC values in the studied stations using Nash-Sutcliffe statistic

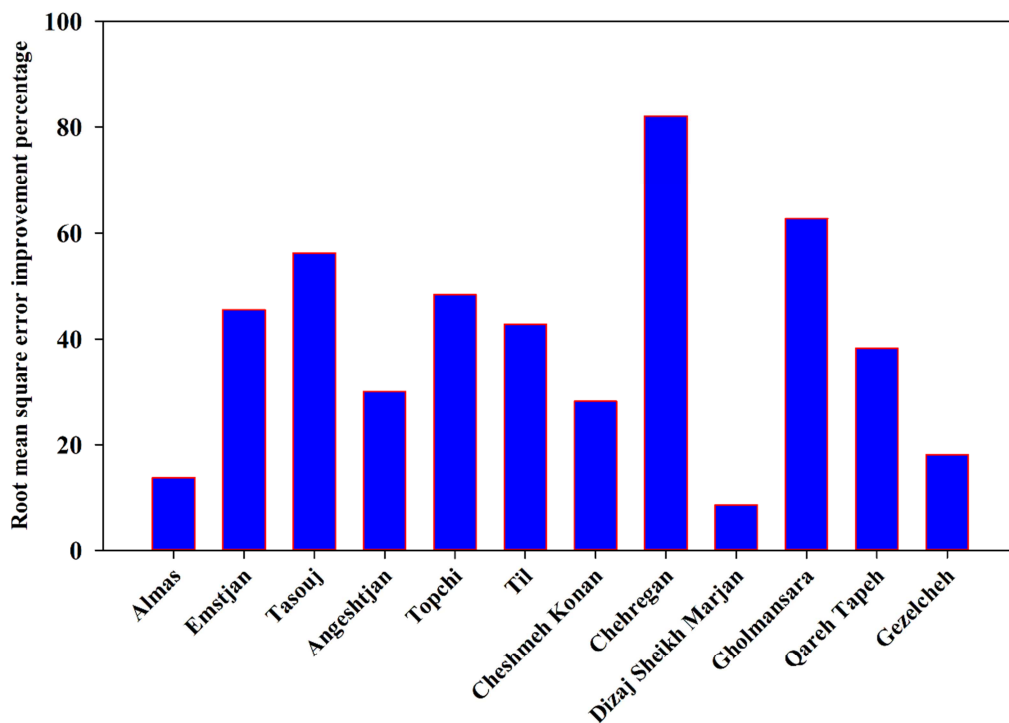


Figure 4. The percentage of improvement of the simulation results of EC values due to the random forest model compared to the multivariate regression model in the studied stations

### ۲.۳. نتایج شبیه‌سازی مقادیر TDS و بررسی اثر متقابل ایستگاه‌ها

همانند مقادیر EC، مقادیر TDS چاه‌های مورد مطالعه نیز با استفاده از دو مدل رگرسیون چندمتغیره و جنگل تصادفی جهت بررسی اثر متقابل چاه‌ها نسبت به یکدیگر شبیه‌سازی شد. نتایج شبیه‌سازی‌ها با استفاده از دو آماره RMSE و NSE مورد بررسی قرار گرفت. نتایج برآورد مقادیر RMSE در شبیه‌سازی مقادیر TDS به صورت شکل (۵) ارائه شد. با توجه به شکل (۵)، نتایج برآورد میزان خطا در شبیه‌سازی مقادیر TDS در چاه‌های مورد مطالعه نشان داد که تغییرات RMSE مدل رگرسیون چندمتغیره بین ۱۶ میلی‌گرم بر لیتر در ایستگاه درج شیخ مرجان تا ۵۰۹ میلی‌گرم بر لیتر در ایستگاه چهرگان متغیر می‌باشد. در مقابل، نتایج شبیه‌سازی مقادیر TDS با استفاده از جنگل تصادفی در چاه‌های مورد مطالعه نشان داد که میزان خطای این مدل بین ۱۳/۷ میلی‌گرم بر لیتر در چاه دیزج شیخ مرجان تا ۷۲ میلی‌گرم بر لیتر در چاه امستجان متغیر می‌باشد. با توجه به داکنه تغییرات مقادیر RMSE در دو مدل مورد بررسی، نتایج نشان داد که مدل جنگل تصادفی اختلاف معناداری با مدل رگرسیون چندمتغیره در شبیه‌سازی مقادیر TDS دارد. نتایج بررسی کارایی مدل‌های مورد بررسی در شبیه‌سازی مقادیر TDS نیز به صورت شکل (۶) ارائه شد. با توجه به شکل (۶)، متوسط مقادیر NSE در چاه‌های مورد مطالعه در مدل جنگل تصادفی ۷۵ درصد و در مدل رگرسیون چندمتغیره حدود ۳۷ درصد می‌باشد. عملکرد مدل رگرسیون چندمتغیره در شبیه‌سازی مقادیر TDS رضایت‌بخش نمی‌باشد. در واقع مدل رگرسیون چندمتغیره اثر متقابل چاه‌ها را به خوبی برآورد نکرده است. نتایج مقایسه مقادیر RMSE دو مدل رگرسیون چندمتغیره و مدل جنگل تصادفی به صورت شکل (۷) ارائه شد. درصد بهبودی مقادیر RMSE مدل جنگل تصادفی نسبت به مدل رگرسیون چندمتغیره در شکل (۷) به وضوح قابل مشاهده می‌باشد. در تمامی چاه‌های مورد مطالعه درصد بهبودی مقادیر RMSE قابل مشاهده می‌باشد. در واقع مدل جنگل تصادفی به خوبی توانسته است میزان RMSE را نسبت به مدل رگرسیون چندمتغیره کاهش دهد. بیش‌ترین کاهش مقادیر RMSE مربوط به دو ایستگاه تیل و چهرگان به ترتیب حدود ۸۷/۸۱ و ۸۷/۳۲ درصد می‌باشد. به طور متوسط نیز مدل جنگل تصادفی در مقایسه با مدل رگرسیون چندمتغیره توانست در کل محدوده مورد مطالعه میزان RMSE ناشی از شبیه‌سازی مقادیر TDS را حدود ۵۶ درصد کاهش دهد. مقایسه مقادیر شبیه‌سازی شده EC و TDS در منطقه مورد مطالعه با استفاده از دو مدل رگرسیون چندمتغیره و جنگل تصادفی نشان داد که مدل جنگل تصادفی توانسته است میزان خطای شبیه‌سازی مقادیر EC در ایستگاه‌های الماس، امستجان، تسوج، انگشتجان، توپچی، تیل، چشمه کنان، چهرگان، دیزج شیخ مرجان، غلمانسرا، قره‌تپه و قزلچه را به ترتیب حدود ۱۳/۶، ۴۵/۵، ۵۶/۳، ۳۰، ۴۸/۳، ۴۲/۸، ۲۸/۱، ۸۲، ۸/۵، ۶۲/۷، ۳۸/۲ و ۱۸/۱ درصد و میزان خطای شبیه‌سازی مقادیر TDS را به ترتیب حدود ۲۸/۴، ۴۰/۷، ۷۶/۴، ۸۱/۵، ۹/۹، ۷۸/۸، ۲۰/۵، ۸۷/۳، ۱۴/۸، ۷۸/۶، ۷۳/۱ و ۸۰/۴ درصد بهبود بخشیده است.

### ۳.۳. نتایج پایش کیفی مقادیر EC و TDS در منطقه مورد مطالعه

در نهایت با انتخاب مدل جنگل تصادفی به عنوان مدل برتر در زمینه شبیه‌سازی مقادیر EC و TDS و همچنین مدل برتر در خصوص بررسی اثر متقابل چاه‌ها در محدوده تسوج، اثر متقابل چاه‌ها نسبت به یکدیگر بررسی شده و تئوری آنتروپی اجرا شد. با استفاده از مقادیر مشاهداتی EC و TDS و همچنین مقادیر شبیه‌سازی شده با استفاده از مدل جنگل تصادفی، شاخص‌های آنتروپی بررسی شده و مهم‌ترین شاخص‌های آن ITI و  $N(i)$  ارائه شد. نتایج رتبه‌بندی چاه‌های مورد بررسی با استفاده از شاخص  $N(i)$  به صورت جدول (۳) ارائه شد. شاخص  $N(i)$  یکی از شاخص‌های مهم در زمینه پایش شبکه آب زیرزمینی می‌باشد. این شاخص رتبه چاه‌ها از نظر اهمیت در منطقه را نشان می‌دهد. مقادیر  $N(i)$  مثبت

نشان‌دهنده اهمیت بالای چاه موردنظر در دشت می‌باشد. با توجه به جدول (۳)، می‌توان رتبه چاه‌های موجود در منطقه مورد مطالعه را با توجه به دو پارامتر EC و TDS مشاهده کرد. با توجه به این جدول، چاه قره‌تپه رتبه ۱ را در هر دو پارامتر EC و TDS کسب کرده است. این نتایج نشان می‌دهد که در صورتی که نیاز باشد یک چاه به‌عنوان معرف دشت از نظر تبادل اطلاعات EC و TDS در منطقه معرفی شود، چاه قره‌تپه بهترین گزینه می‌باشد. در واقع ارتباط بین این چاه با سایر چاه‌های منطقه مورد مطالعه بیشتر می‌باشد. با توجه به نتایج ارائه‌شده می‌توان مقادیر EC و TDS این چاه (قره‌تپه) را به کل آبخوان تعمیم داد. چاه امستجان نیز برای هر دو متغیر کیفی EC و TDS بیش‌ترین رتبه را کسب کرده است که نشان‌دهنده ارتباط ضعیف این چاه با سایر چاه‌ها می‌باشد. با توجه به مقادیر  $N(i)$ ، چاه‌های امستجان، تسوج و الماس کم‌ترین رتبه با توجه به مقادیر EC و چاه‌های امستجان، چهرگان و قزلچه با توجه به انتقال اطلاعات TDS کم‌ترین رتبه را کسب کرده‌اند.

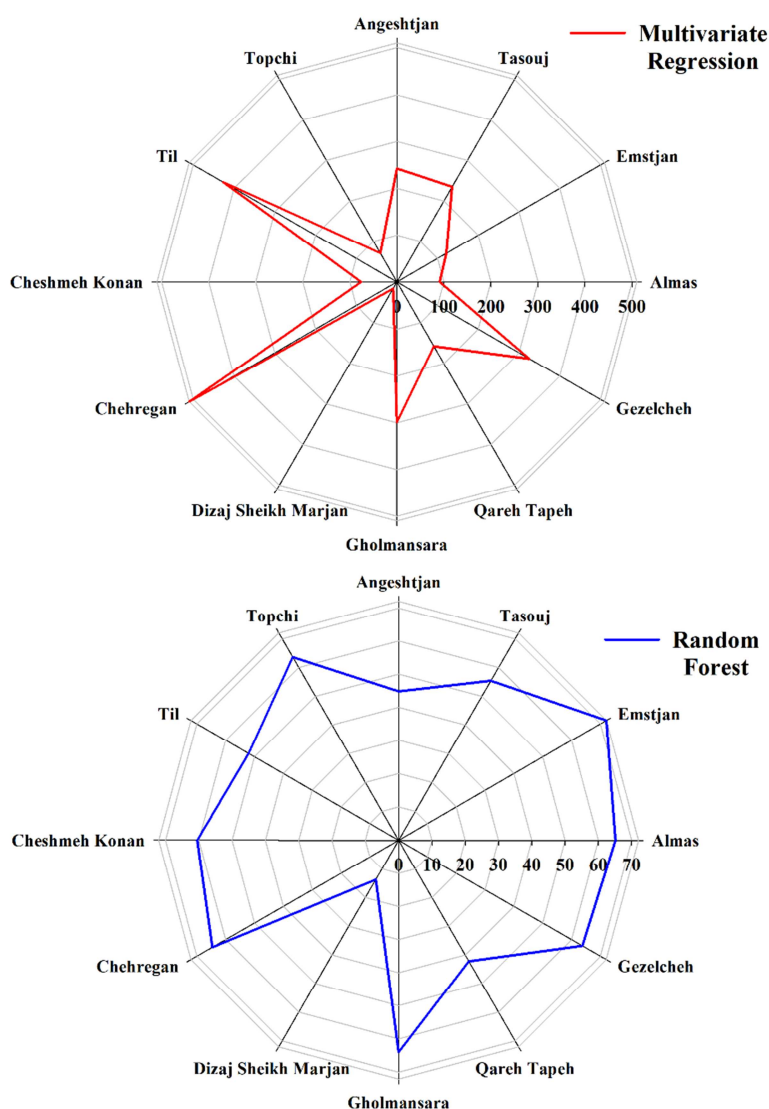


Figure 5. The results of investigating the error rate (RMSE) of random forest models and multivariate regression in the simulation of TDS values in the studied stations.

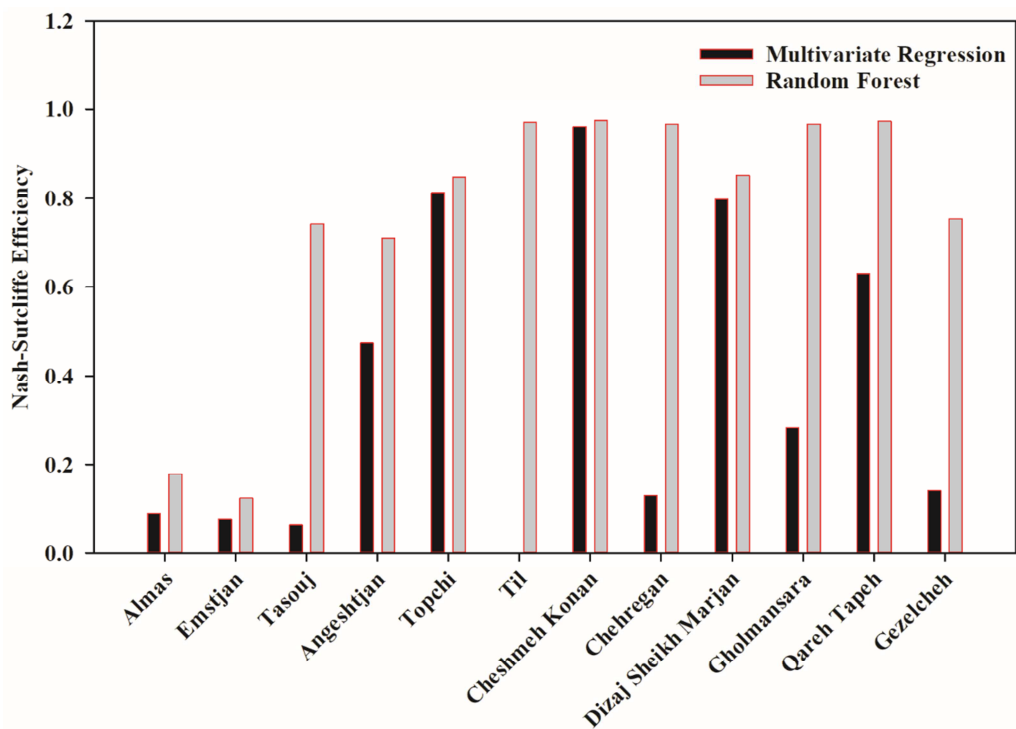


Figure 6. Performance results of random forest models and multivariate regression (Nash-Sutcliffe statistic) in the simulating of TDS values in the studied stations.

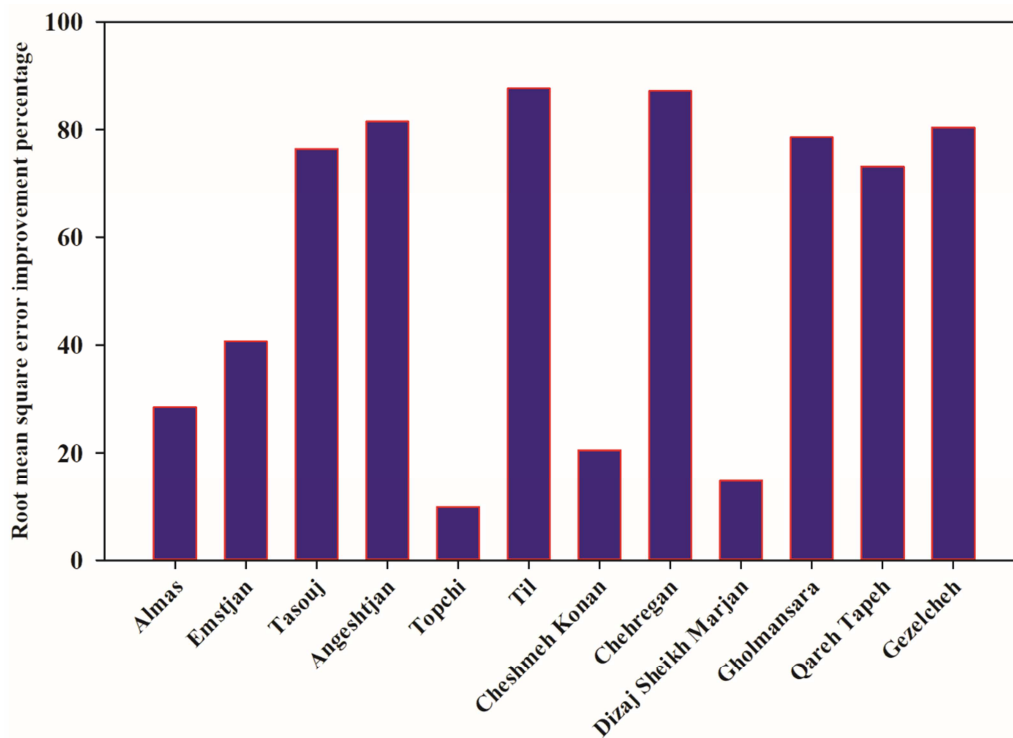


Figure 7. The percentage of improvement of the simulation results of TDS values due to the random forest model compared to the multivariate regression model in the studied stations

Table 3. The ranking results of the studied wells based on the quality values of EC and TDS

Well	N(i)-TDS	N(i)-EC	Rank based on EC	Rank based on TDS
Almas	0.010	-0.0242	10	4
Emstjan	-0.281	-0.3294	12	12
Tasouj	-0.002	-0.0334	11	8
Angeshtjan	0.018	-0.0125	7	2
Topchi	-0.002	0.0000	5	7
Til	0.000	0.0001	3	5
Cheshmeh Konan	0.000	0.0001	4	6
Chehregan	-0.017	-0.0230	9	11
Dizaj Sheikh Marjan	-0.003	-0.0192	8	9
Gholmansara	0.011	0.0002	2	3
Qareh Tapeh	0.041	0.0002	1	1
Gezelcheh	-0.008	-0.0002	12	10

یکی دیگر از شاخص‌های مهم در تئوری آنتروپی، شاخص انتقال اطلاعات یا ITI می‌باشد. این شاخص کمبود یا افزونگی اطلاعات EC و TDS را در منطقه نشان می‌دهد. این شاخص، اطلاعات مهمی در زمینه پایش مقادیر کیفی در منطقه در اختیار قرار می‌دهد. مناطقی که چاه مازاد داشته و یا کمبود چاه احساس می‌شود، توسط این شاخص بیان می‌شود. جهت تصمیم‌گیری بهتر و همچنین نمایش بهتر مقادیر ITI در منطقه مورد مطالعه، مقادیر ITI مربوط به دو پارامتر EC و TDS پهنه‌بندی و به صورت شکل‌های (۸) و (۹) به ترتیب برای مقادیر EC و TDS ارائه شد. با توجه به شکل (۸) در خصوص پهنه‌بندی مقادیر EC در سطح آبخوان دشت تسوج می‌توان مشاهده کرد که کل منطقه مورد مطالعه در حالت متوسط و بیش از متوسط قرار گرفته و تنها قسمت‌های کوچکی از نقاط مرز شرقی آبخوان کمبود چاه از نظر انتقال اطلاعات شوری مشاهده می‌شود. نواحی مرکزی و شمالی آبخوان مورد نظر از نظر انتقال اطلاعات شوری در وضعیت متوسط قرار گرفته است، اما کمبود چاه در این منطقه وجود ندارد. به طور کلی از نظر پارامتر EC، در منطقه مورد مطالعه پایش مناسبی وجود داشته و پراکندگی چاه‌ها به خوبی می‌باشد. با چیدمان و پراکندگی فعلی چاه‌ها در منطقه مورد مطالعه، می‌توان اطلاعات مناسبی از وضعیت EC در منطقه برآورد کرد. نتایج بررسی‌ها نشان داد که در منطقه مورد مطالعه، احداث چاه جدید جهت پایش بهتر ضرورتی ندارد. مناطق جنوبی منطقه مورد مطالعه با وجود عدم حضور چاه نیز کمبود چاه از نظر انتقال اطلاعات EC مشاهده نمی‌شود. یکی از کاربردهای مهم تئوری آنتروپی نیز همین موضوع است. این نتیجه نشان می‌دهد با وجود کمبود چاه در این منطقه، انتقال اطلاعات شوری بین تیل و تسوج به خوبی برقرار است. البته در این نواحی سطح آب زیرزمینی به دلیل نزدیکی به دریاچه ارومیه بالا بوده و EC این نواحی نیز بیش‌تر از نواحی شمالی می‌باشد. در مورد پارامتر TDS این شرایط متفاوت می‌باشد. با توجه به شکل (۹) می‌توان مشاهده کرد که قسمتی از نواحی مرکزی، نوار مرزی شرقی و همچنین نواحی مرزی غربی آبخوان تسوج از نظر تبادل اطلاعات TDS دچار کمبود چاه می‌باشد. نواحی مرکزی آبخوان مورد مطالعه نیز در حالت متوسط از نظر پراکندگی چاه‌ها قرار دارد. نواحی مرزی شمالی منطقه مورد مطالعه و همچنین نواحی جنوب شرقی آبخوان در حالت مازاد قرار دارد. جهت پایش بهینه آبخوان منطقه مورد مطالعه در دشت تسوج از نظر پایش اطلاعات TDS، نیاز به احداث چاه جدید در نواحی قرمز رنگ می‌باشد. در صورتی که هدف از پایش بهینه‌کردن و بهبود شبکه کیفی آب زیرزمینی باشد، می‌توان براساس نتایج موجود گمانه‌های جدید حفر کرد و یا گمانه‌هایی که انتقال اطلاعات کم‌تری دارند را از سیستم حذف کرد. در شبکه‌هایی که چاه مازاد موجود است، می‌توان چاه‌هایی که رتبه بالایی دارند حذف کرد و در مورد شبکه‌هایی که تعداد چاه مناسب در آن‌ها موجود نیست، بهتر است براساس پیشنهاد Mogheir et al. (2006) از چاه‌های متروکه و یا چاه‌های شرب جهت پایش شبکه استفاده کرد.

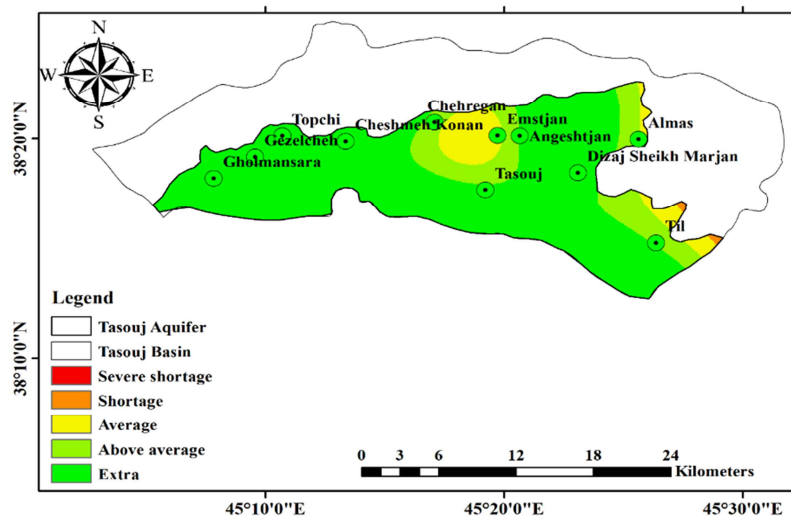


Figure 8. Zoning of ITI index values regarding the qualitative monitoring of EC values

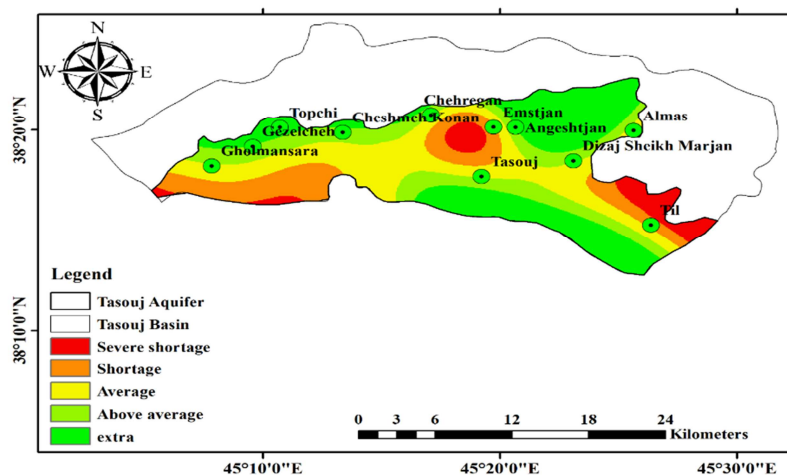


Figure 9. Zoning of ITI index values regarding the qualitative monitoring of TDS values

#### ۴. بحث و نتیجه گیری

در این مطالعه کارایی تئوری آنتروپی در رتبه بندی و پایش منطقه ای شبکه کیفیت آب زیرزمینی آبخوان تسوج مورد ارزیابی قرار گرفت. در این خصوص از مقادیر EC و TDS آب زیرزمینی ۱۲ حلقه چاه در دوره آماری ۹۸-۱۳۸۲ استفاده شد. در تئوری آنتروپی مرسوم جهت بررسی اثر متقابل چاه ها نسبت به یکدیگر، از روش رگرسیون چندمتغیره استفاده می شود. همچنین کارایی و دقت الگوریتم جنگل تصادفی در بررسی اثر متقابل چاه ها در کنار روش رگرسیون چندمتغیره بررسی شد. نتایج بررسی ها نشان داد که مدل جنگل تصادفی به خوبی توانسته است اثر متقابل چاه ها را برآورد کرده و میزان کارایی مدل جنگل تصادفی در مقایسه با مدل رگرسیون چندمتغیره را بهبود بخشد. در بررسی اثر متقابل ایستگاه ها، روش جنگل تصادفی موفق شد به طور متوسط میزان خطا (RMSE) در شبیه سازی اثر متقابل ایستگاه ها در پارامتر EC و TDS به ترتیب حدود ۵۶ و ۴۰ درصد نسبت به روش مرسوم کاهش دهد. با انتخاب مدل جنگل تصادفی به عنوان مدل برتر، شاخص های تئوری آنتروپی برآورد شد. نتایج بررسی شاخص ITI مربوط به پایش کیفی مقادیر EC نشان داد که وضعیت فعلی آبخوان مورد مطالعه از نظر پراکندگی مناسب بوده و با چیدمان فعلی چاه های موجود، انتقال

اطلاعات شوری آب زیرزمینی به خوبی پایش می‌شود. با توجه به وضعیت فعلی آبخوان دشت تسوج، جهت پایش دقیق مقادیر EC نیاز به حفر چاه جدید وجود ندارد. طبقه‌بندی غالب در خصوص مقادیر ITI پارامتر EC، در دسته مازاد قرار دارد. چاه قره‌تپه و چاه امستجان به ترتیب با اهمیت‌ترین و کم اهمیت‌ترین چاه در خصوص پایش مقادیر EC در آبخوان دشت تسوج می‌باشد. نتایج پایش مقادیر TDS در آبخوان دشت تسوج با توجه به شاخص ITI نشان داد که در برخی نقاط مرکزی و نوار مرزی جنوب غرب و شرق آبخوان کمبود چاه احساس می‌شود. به نظر می‌رسد جهت پایش دقیق تر و بررسی انتقال اطلاعات TDS در آبخوان دشت مورد بررسی نیاز به حفر چاه‌های جدید در منطقه می‌باشد که می‌توان از مقادیر TDS در چاه‌های مشاهداتی و یا شرب نیز استفاده کرد. همانند مقادیر EC، چاه‌های قره‌تپه و امستجان به ترتیب با اهمیت‌ترین و کم‌اهمیت‌ترین چاه‌ها در آبخوان می‌باشند. این موضوع نشان می‌دهد که برای پایش و تعمیم مقادیر EC و TDS به آبخوان تسوج بهتر است از مقادیر EC و TDS چاه قره‌تپه استفاده شود. در مورد چاه‌هایی که رتبه بالایی کسب کردند، ادامه فعالیت آن‌ها مستلزم تجدیدنظر جدی بوده و داده‌های این چاه‌ها از اعتبار کم‌تری برخوردار می‌باشند. چاه‌هایی مثل قره‌تپه که رتبه کم‌تری کسب کردند، مفیدترین اطلاعات در شبکه پایش را تولید می‌نمایند و حفظ این چاه‌ها در شبکه آب زیرزمینی ضروری می‌باشد. با توجه به تغییرات اقلیمی و همچنین افزایش برداشت از سفره‌ها، پایش کمی و کیفی آب زیرزمینی می‌تواند به‌عنوان سیستم هشدار در خصوص کمیت و کیفیت آب‌های زیرزمینی عمل کند که روش آنتروپی این امکان را فراهم کرده است. با توجه به تغییرات اقلیمی و تغییرات کمی و کیفی آب زیرزمینی سال‌های اخیر در نقاط مختلف ایران به‌ویژه حوضه دریاچه ارومیه، لزوم پایش کمی و کیفی آب زیرزمینی بیش از پیش احساس می‌شود، تا با مدیریت صحیح بتوان منابع آب موجود را مدیریت کرد. نتیجه کاربردی این پژوهش ارائه رویکردی بهینه جهت پایش کیفی آب زیرزمینی می‌باشد که بتوان اثر متقابل ایستگاه‌ها را با دقت و قطعیت بالاتری تخمین زد.

## ۵. پی‌نوشت‌ها

### 1. Information Transfer Index

## ۶. تعارض منافع

هیچ‌گونه تعارض منافی بین نویسندگان وجود ندارد.

## ۷. منابع

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Chapman, T. G. (1986). Entropy as a measure of hydrologic data uncertainty and model performance. *Journal of Hydrology*, 85(1-2), 111-126.
- Chen, Y. C., Wei, C., & Yeh, H. C. (2008). Rainfall network design using kriging and entropy. *Hydrological Processes: An International Journal*, 22(3), 340-346.
- Hastie, T., Friedman, J., Tibshirani, R., Hastie, T., Friedman, J., & Tibshirani, R. (2001). Unsupervised learning. *The elements of statistical learning: Data mining, inference, and prediction*, 437-508.
- Harmancioglu, N. B., & Alpaslan, N. (1992). Water quality monitoring network design: a problem of multi-objective decision making 1. *JAWRA Journal of the American Water Resources Association*, 28(1), 179-192.
- Harmancioglu, NB., Fistikoglu, O., Ozkul, S D., Singh, VP., & Alpaslan N. (1999). Water quality Monitoring Network Design. Kluwer, Boston, USA, 1999; 299pp.



- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review*, 106(4), 620.
- Khashei-Siuki, A., Shahidi, A., Ramezani, Y., & Nazeri Tahroudi, M. (2020). Forecasting the groundwater monitoring network using hybrid time series models (Case study: Nazlochai). *Water and Soil Conversation*, 27(3), 85-103 (in Persian).
- Krstanovic, P. F. (1988). Application of entropy theory to multivariate hydrologic analysis. (Volumes I and II) (Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College).
- Lee, J. H. (2013). Determination of optimal water quality monitoring points in sewer systems using entropy theory. *Entropy*, 15(9), 3419-3434.
- Leite, G. D. N. P., da Cunha, G. T. M., dos Santos Junior, J. G., Araújo, A. M., Rosas, P. A. C., Stosic, T., ... & Rosso, O. A. (2021). Alternative fault detection and diagnostic using information theory quantifiers based on vibration time-waveforms from condition monitoring systems: Application to operational wind turbines. *Renewable Energy*, 164, 1183-1194.
- Markus, M., Knapp, H. V., & Tasker, G. D. (2003). Entropy and generalized least square methods in assessment of the regional value of streamgages. *Journal of hydrology*, 283(1-4), 107-121.
- Mishra, A. K., & Coulibaly, P. (2010). Hydrometric network evaluation for Canadian watersheds. *Journal of Hydrology*, 380(3-4), 420-437.
- Mogheir, Y., & Singh, V. P. (2002). Application of information theory to groundwater quality monitoring networks. *Water Resources Management*, 16(1), 37-49.
- Mogheir, Y., Singh, V. P., & De Lima, J. L. M. P. (2006). Spatial assessment and redesign of a groundwater quality monitoring network using entropy theory, Gaza Strip, Palestine. *Hydrogeology Journal*, 14(5), 700-712.
- Nazeri Tahroudi, M., Khashei Siuki, A., & Ramezani, Y. (2019). Redesigning and monitoring groundwater quality and quantity networks by using the entropy theory. *Environmental monitoring and assessment*, 191(4), 1-17.
- Ozkul, S., Harmancioglu, N. B., & Singh, V. P. (2000). Entropy-based assessment of water quality monitoring networks. *Journal of hydrologic engineering*, 5(1), 90-100.
- Ramezani, Y., Pourreza-Bilondi, M., Yaghoobzadeh, M., & Nazeri Tahroudi, M. (2018). Qualitative Monitoring of Drinking Water Using Entropy Indices (Case Study: Central Aquifer of Birjand Plain). *Iranian journal of irrigation and drainage*, 12(3), 556-568. (in Persian).
- Shahidi, A., Khashei-Siuki, A., & Nazeri Tahroudi, M. (2019). Designing Monitoring Network for Rain Gauge Stations Using Irregularity Theory (Case Study: Urmia Lake Basin). *Iranian journal of irrigation and drainage*, 13(2), 296-308. (in Persian).
- Shannon, CE. (1948). A mathematical theory of communication, bell System technical Journal, 27, 379-423 and 623-656. Mathematical Reviews (MathSciNet), MR10, 133e.
- Shannon, C. E., & Weiner, W. (1948). A mathematical theory of communication. Publ. Urbana, IL: University of Illinois Press.
- Shi, B., Jiang, J., Sivakumar, B., Zheng, Y., & Wang, P. (2018). Quantitative design of emergency monitoring network for river chemical spills based on discrete entropy theory. *Water research*, 134, 140-152.
- Singh, V. P. (1997). The use of entropy in hydrology and water resources. *Hydrological processes*, 11(6), 587-626.