

## برآورد هدایت هیدرولیکی اشباع خاک با استفاده از برنامه‌ریزی بیان ژنی و رگرسیون ریبج (مطالعه موردی در استان آذربایجان شرقی)

عباس احمدی<sup>۱\*</sup>، پیمان پالیزوان‌زند<sup>۲</sup>، حبیب پالیزوان‌زند<sup>۳</sup>

۱. دانشیار، دانشکده کشاورزی دانشگاه تبریز

۲. دانشجوی سابق کارشناسی ارشد، دانشکده کشاورزی، دانشگاه آزاد اسلامی تبریز

۳. دانشجوی دکتری، دانشکده کشاورزی، دانشگاه تبریز

(تاریخ دریافت: ۱۳۹۵/۸/۳ - تاریخ بازنگری: ۱۳۹۶/۱/۳۰ - تاریخ تصویب: ۱۳۹۶/۲/۶)

### چکیده

هدایت هیدرولیکی خاک از ویژگی‌های مهم فیزیکی خاک است که در مدل‌سازی انتقال آب، املاح و آلاینده‌ها کاربرد دارد. اندازه‌گیری مستقیم هدایت هیدرولیکی خاک وقت‌گیر و پرهزینه بوده و گاهی اوقات به دلیل خطاهای آزمایشی و عدم یکنواختی خاک نتایج بدست آمده چندان قابل اعتماد نمی‌باشد. از طرف دیگر این پارامتر می‌تواند با استفاده از پارامترهای زودپایافت خاک برآورد شود. هدف از این پژوهش، ارائه مدل‌های برنامه‌ریزی بیان ژنی و رگرسیون خطی بر اساس ویژگی‌های زودپایافت هست. برای این منظور ۱۶۰ نمونه خاک با خصوصیات متفاوت از مناطق مختلف استان آذربایجان شرقی برداشته شد. سپس برخی ویژگی‌های فیزیکی و شیمیایی آن‌ها همانند درصد شن، سیلت، رس و مواد آلی، جرم مخصوص ظاهری، pH و EC اندازه‌گیری شد. سپس داده‌ها بطور تصادفی به دو دسته داده‌های سری آموزش (۷۵ درصد) و داده‌های سری آزمون (۲۵ درصد) تقسیم شدند. شش تابع انتقالی (PTFs) با ترکیبی از عملگرهای ریاضی متفاوت توسط برنامه‌ریزی بیان ژنی طراحی شد. در نهایت یکی از توابع که از دقت و صحت بیشتری نسبت به بقیه برخوردار بود، انتخاب گردید. همچنین از رگرسیون ریبج برای ارائه تابع انتقالی رگرسیونی استفاده شد. دقت و صحت توابع با معیارهای آماری  $R^2$ ، RMSE و MAE ارزیابی گردید. نتایج نشان داد که تابع انتقالی ارائه شده با روش برنامه‌ریزی بیان ژنی از دقت و صحت بیشتری نسبت به مدل رگرسیونی برخوردار می‌باشد. به طوری که مقادیر  $R^2$ ، RMSE ( $\text{Cm h}^{-1}$ ) و MAE ( $\text{Cm h}^{-1}$ ) برای تابع انتقالی برنامه‌ریزی بیان ژنی در داده‌های سری آموزش به ترتیب برابر ۰/۹۱، ۱/۸۲ و ۱/۲۳ و برای داده‌های سری اعتبارسنجی برابر ۰/۹۲، ۲/۲۷ و ۱/۵۹ بود. در حالی که مقادیر معیارهای فوق در مدل رگرسیونی، برای داده‌های سری آموزش به ترتیب برابر ۰/۷۰، ۳/۴۸ و ۲/۰۷ و برای داده‌های سری اعتبارسنجی برابر ۰/۷۶، ۳/۱۱ و ۱/۸۸ بود.

**واژه‌های کلیدی:** برنامه‌ریزی بیان ژنی، توابع انتقالی، مدل‌سازی

### مقدمه

به همین منظور استفاده از روش‌های تخمین غیر مستقیم مورد توجه قرار گرفته است. روش غیر مستقیم بدست آوردن داده‌های دیریافت خاک، توابع انتقالی نام‌گذاری شده است (Bouma, 1989). به دلیل این که توابع انتقالی خصوصیات دیریافت را با استفاده از خصوصیات زودپایافت خاک پیش‌بینی می‌کنند، دارای مزیت‌های عمده‌ای همچون صرفه‌جویی در وقت و هزینه هستند (Wösten et al, 2001). یکی از روش‌های ایجاد و توسعه توابع انتقالی روش رگرسیون خطی است. Campbell (1985) از پارامترهای ورودی میانگین و انحراف معیار هندسی ذرات معدنی خاک، جرم مخصوص ظاهری، درصد رس و سیلت، Cosby et al. (1984) از پارامترهای ورودی درصد شن و رس، Vereecken et al. (1990) از پارامترهای ورودی

هدایت هیدرولیکی، توانایی خاک برای انتقال آب است (Klute, 1986). هدایت هیدرولیکی اشباع یکی از مهم‌ترین ویژگی‌های فیزیکی خاک است که در تعیین سرعت نفوذ، مقدار نفوذ عمقی و طراحی زهکش‌ها و دیگر فرایندهای هیدرولوژیکی مؤثر است. اندازه‌گیری مستقیم هدایت هیدرولیکی اشباع به‌طور معمول با صرف زمان و هزینه زیاد همراه است و از طرف دیگر این پارامتر دارای تغییرات زمانی و مکانی قابل توجهی بوده و به روش اندازه‌گیری نیز بستگی دارد. لذا بایستی برای تخمین دقیق این نوع ویژگی‌ها به دنبال راه‌حل‌های مناسب، ساده و کم هزینه بود.

\* نویسنده مسئول: a\_ahmadi@tabrizu.ac.ir

هیدرولیکی اشباع (Parasuraman *et al.*, 2007)، برآورد دبی جریان رودخانه (Guven, 2009)، تبخیر و تعرق (Parasuraman *et al.*, 2007) و رطوبت خاک (Makkeasorn *et al.*, 2006) بکار برده شده است. هدف از تحقیق حاضر استفاده از روش‌های برنامه‌ریزی بیان‌ژنی و رگرسیون رنج در تخمین هدایت هیدرولیکی اشباع خاک و ارزیابی دقت و صحت این مدل‌ها است.

### مواد و روش‌ها

منطقه تبریز که در تقسیمات کشوری به‌عنوان بخشی از شهرستان تبریز و شهرستان‌های همجوار آن شامل آذرشهر، اسکو، بستان‌آباد، هریس، ورزقان، مرند و شبستر به‌حساب می‌آید در استان آذربایجان شرقی واقع شده است. این استان با مساحت ۴۵۴۹۱ کیلومترمربع در شمال غرب ایران ما بین عرض جغرافیایی ۲۶° ۳۹' - ۴۵° ۳۹' و طول جغرافیایی ۲۲° ۴۸' - ۴۵° ۰۵' قرار دارد. این استان دارای آب و هوای سرد کوهستانی بوده و کل محدوده استان را کوه‌ها و ارتفاعات تشکیل داده‌اند. به‌منظور دستیابی به اهداف تحقیق ۱۶۰ نمونه از عمق ۳۰ - ۰ سانتی‌متری سطح خاک تهیه (شکل ۱) و جهت اندازه‌گیری به آزمایشگاه انتقال داده شد.

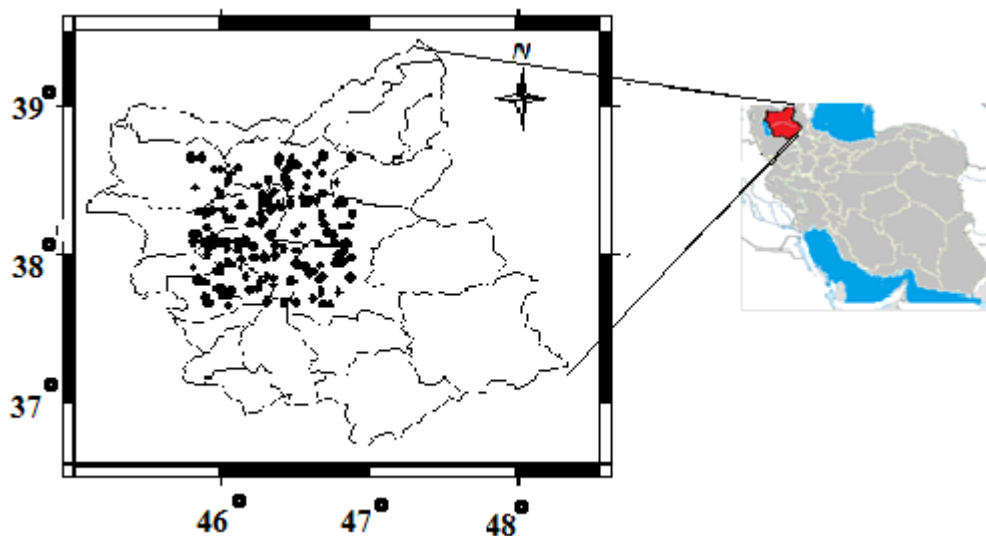
بافت خاک به روش هیدرومتری (Gee and Or, 2002)، مواد آلی به روش والکی- بلک (Nelson and Sommer, 1982)، جرم مخصوص ظاهری خاک ( $\rho_b$ ) به روش نمونه‌برداری دست نخورده سیلندر (Blake and Hartge, 1986)، هدایت هیدرولیکی اشباع خاک بعد از تهیه نمونه‌های دست‌نخورده به روش استوانه فلزی به روش بار ثابت، EC خاک در عصاره گل اشباع و pH گل اشباع با روش‌های متداول آزمایشگاهی اندازه‌گیری شدند.

به‌منظور ارائه رابطه مناسب بین ویژگی‌های زودپافت خاک و هدایت هیدرولیکی اشباع، کل مجموعه داده‌های گردآوری شده ابتدا بطور تصادفی به دو دسته تقسیم شد. دسته اول شامل ۷۵ درصد کل مجموعه داده‌ها بوده و زیرمجموعه آموزشی نامیده می‌شود. دسته دوم زیرمجموعه صحت‌سنجی شامل ۲۵ درصد کل داده‌ها می‌باشد. به این ترتیب بر اساس داده‌های آموزشی بهترین رابطه با بیش‌ترین تناسب بین داده‌ها انتخاب شد و صحت و عملکرد رابطه ارائه شده در زیرمجموعه آزمایشی مورد بررسی قرار گرفت. با این هدف که قابلیت اعتماد به رابطه ارائه شده برای مجموعه داده‌هایی که در مسئله استفاده نشده‌اند سنجیده شود.

جرم مخصوص ظاهری، درصد رس، شن و ماده آلی، Wösten (1997) از پارامترهای ورودی جرم مخصوص ظاهری، درصد رس، سیلت و ماده آلی، Jabro (1992) از پارامترهای ورودی جرم مخصوص ظاهری، درصد رس و سیلت، Dane and Puckett (1994) از پارامتر ورودی درصد رس برای ایجاد توابع انتقالی جهت تخمین هدایت هیدرولیکی بهره برده‌اند. et.al. (2001) جرم مخصوص ظاهری و میزان ماده آلی را دو پارامتر مهم ورودی در توابع انتقالی به‌منظور برآورد هدایت هیدرولیکی معرفی نمودند. از خصوصیات فیزیکی که بر روی هدایت هیدرولیکی تأثیر عمده‌ای دارند می‌توان به تخلخل کل، توزیع اندازه منافذ و هندسه منافذ اشاره کرد (Hillel, 1982). Morales *et al.* (2007) نشان دادند که افزایش هدایت هیدرولیکی اشباع خاک در لایه ۴۵ سانتی‌متری سطحی خاک مشاهده شده اما در لایه‌های پایینی تغییرات چشم‌گیری نداشته است. Merdun *et al.* (2006) در پژوهشی در کشور ترکیه، عملکرد توابع انتقالی نقطه‌ای و پارامتری را در پیش‌بینی خصوصیات هیدرولیکی خاک با استفاده از شبکه عصبی مصنوعی و رگرسیون خطی بررسی کرده و تفاوت معنی‌داری بین دو روش مذکور مشاهده نکردند.

مدل‌های رگرسیونی در حوزه‌های مختلف علمی کاربرد وسیعی دارند. مسئله جدی که می‌تواند استفاده از مدل رگرسیونی را با محدودیت مواجه کند، وجود هم‌راستایی بین متغیرهای رگرسیونی است. وجود هم‌راستایی، توانایی برآورد دقیق ضرایب مدل رگرسیون را با مشکل مواجه می‌کند. در شرایط وجود هم‌راستایی تفسیر ضرایب رگرسیون قابل صحت نخواهد بود. بنابراین برای از بین بردن هم‌راستایی، از روش رگرسیون رنج استفاده می‌شود. Li *et al.* (2010) سیستمی با دو سری داده متفاوت برای مقایسه دو روش رگرسیون خطی ساده و رگرسیون خطی رنج ارائه دادند. نتایج نشان‌دهنده عملکرد بهتر رگرسیون رنج در ایجاد مدل‌های رگرسیونی با داده‌های دارای هم‌راستایی نسبت به رگرسیون خطی ساده است.

یکی دیگر از روش‌های ایجاد و توسعه توابع انتقالی روش برنامه‌ریزی بیان‌ژنی است. روش برنامه‌ریزی بیان‌ژنی جدیدترین شیوه از بین روش‌های الگوریتم تکاملی است که به دلیل دارا بودن دقت کافی، از کاربرد بیشتری برخوردار است (Alvisi *et al.*, 2005). در مطالعات مربوط به آب و خاک روش برنامه‌ریزی بیان‌ژنی در مدل‌سازی جریان روی یک بستر انعطاف‌پذیر (Babovic and Abbott, 1997)، فرآیندهای بارش - رواناب (Khu *et al.*, 2002; Liang *et al.*, 2001)، برآورد رواناب (Savic *et al.*, 1999)، فرآیند بارش - دبی (Hong *et al.*, 2005)، هدایت



شکل ۱- توزیع جغرافیایی نقاط نمونه برداری در منطقه تبریز واقع در استان آذربایجان شرقی

### برنامه ریزی بیان ژنی

برنامه ریزی بیان ژنی (GEP) تعمیم یافته روش الگوریتم ژنتیک می باشد که برای اولین بار توسط جان کوزا (Koza, 1992) بر اساس تئوری داروین ارائه شد. همچنین این مدل جزو روش های الگوریتم گردشی محسوب می شود. الگوریتم های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده ها به کار می گیرند و در نهایت، جواب مناسب را ارائه می نمایند. روش برنامه ریزی بیان ژنی در دهه اخیر، یکی از جدیدترین روش های الگوریتم تکاملی می باشد که به دلیل دارا بودن دقت کافی، از کاربرد بیشتری برخوردار است (Alvisi et al., 2005). تفاوت اساسی موجود بین الگوریتم ژنتیک و برنامه ریزی بیان ژنی به طبیعت هر یک از افراد برمی گردد، به نحوی که افراد در الگوریتم ژنتیک، ردیف های خطی با طول ثابت می باشند (کروموزوم ها) ولی در برنامه ریزی بیان ژنی، همان شاخه ها مجزا می باشند. همچنین در برنامه ریزی بیان ژنی بر ساختار درختی مجموعه ها تأکید می شود ولی الگوریتم ژنتیک، براساس سیستم ارقام دودویی عمل می نماید.

برنامه ریزی بیان ژنی (GEP) شامل چهار گام می باشد. در گام اول جمعیت اولیه (کروموزوم ها) ایجاد و کروموزوم ها به رایانه معرفی شدند. در گام دوم هر یک از افراد (ژن) جمعیت مذکور با استفاده از توابع برازش (شناسایی مؤثرترین افراد در ماهیت پدیده) ارزیابی شد. در گام سوم ژن های مؤثر به منظور تکثیر، جهش، جفت گیری و تولیدمثل افراد جدید با صفات اصلاح شده (فرزندان) انتخاب شد و در گام چهارم اعمال فرآیند توسعه ای، تکراری بر روی فرزندان در هر تولید تا حصول بهترین

پاسخ تکرار شد. در مرحله آموزش مدل، انتخاب جمعیت های اولیه تصادفی مختلف و تأثیرگذار در پدیده نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت الگو نیز می شود. بنابراین در الگوسازی هدایت هیدرولیکی نیز باید سعی نمود مؤثرترین پارامترهای ورودی انتخاب شود (Soltani et al., 2009). الگوهای مورد استفاده برای تخمین هدایت هیدرولیکی در جدول (۱) آورده شده است. در این پژوهش برای بدست آوردن روابط برنامه ریزی بیان ژنی برای برآورد هدایت هیدرولیکی اشباع از نرم افزار GeneXpro Tools 4.0 استفاده شد.

جدول ۱. مدل های انتخابی در برنامه ریزی بیان ژنی

شماره مدل	داده های ورودی مدل
مدل ۱	pH - EC - $\rho_b$ - رس - ماده آلی - شن - سیلت
مدل ۲	pH - EC - $\rho_b$ - ماده آلی - سیلت
مدل ۳	$\rho_b$ - رس - سیلت - شن
مدل ۴	pH - EC - ماده آلی
مدل ۵	$\rho_b$ - شن
مدل ۶	$\rho_b$ - رس

فرآیند مدل سازی هدایت هیدرولیکی اشباع خاک در این

پژوهش در گام های زیر انجام شد:

انتخاب تابع برازش مناسب: در این مطالعه تابع ریشه مربع

خطای نسبی (RRSE) به عنوان تابع برازش انتخاب گردید (Ferreira, 2006):

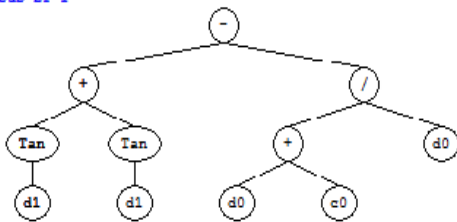
$$RRSE_i = \sqrt{\frac{\sum_{j=1}^n (P_{ij} - T_j)^2}{\sum_{j=1}^n (T_i - \bar{T})^2}} \quad (\text{رابطه ۱})$$

پس از ۱۰۰۰۰ نسل، تابع تناسب همگرایی ۷۸۴/۴۷ بدست آمد و تغییرات قابل توجهی در آن مشاهده نشد. در نتیجه درخت توصیفی از بهترین مدل تکامل یافته به دست آمده در پایان فرآیند بهینه‌سازی که شامل ۳ زیر شاخه که توسط عملگر جمع متصل شده‌اند، در شکل (۲) نشان داده شده است.

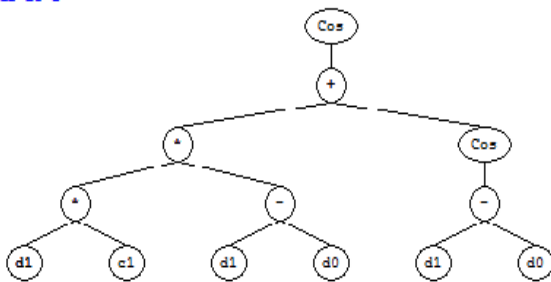
جدول ۲. پارامترهای مورد استفاده در مدل برنامه‌ریزی بیان ژن (GEP)

پارامتر	توصیف پارامتر	تنظیم پارامتر
P <sub>1</sub>	مجموعه توابع	+, -, ×, /, √, Sin X, Cos X, e <sup>x</sup> , ln, x <sup>2</sup> , x <sup>3</sup> , and Arctan x
P <sub>2</sub>	نرخ جهش	۰/۰۴۴
P <sub>3</sub>	نرخ وارونگی	۰/۱
P <sub>4</sub>	نرخ جابجایی IS	۰/۱
P <sub>5</sub>	نرخ جابجایی RIS	۰/۱
P <sub>6</sub>	نرخ جابجایی ژن	۰/۱
P <sub>7</sub>	نرخ تقاطع یک نقطه‌ای	۰/۳
P <sub>8</sub>	نرخ تقاطع دو نقطه‌ای	۰/۳
P <sub>9</sub>	نرخ تقاطع ژن	۰/۱

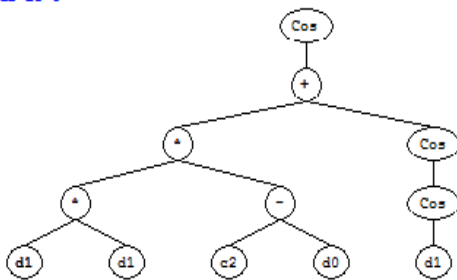
Sub-ET 1



Sub-ET 2



Sub-ET 3



شکل ۲. درخت توصیفی برای فرمول GEP که در آن d(1) و d(0) به ترتیب جرم مخصوص ظاهری و درصد رس می‌باشند.

که در آن P<sub>ij</sub> مقدار برآورد شده توسط کروموزوم i، برای تناسب z است. T<sub>j</sub> مقدار هدف (متغیر) برای تناسب z و T میانگین T هاست. پس از آن، تابع تناسب کروموزوم i (f<sub>i</sub>) برابر است با:

$$f_i = 1000 \times \frac{1}{1 + RRSE_i} \quad (\text{رابطه ۲})$$

که دامنه f<sub>i</sub> از صفر تا ۱۰۰۰ هست (۱۰۰۰ متناظر با یک کروموزوم با تناسب ایده‌آل).

انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به منظور تولید کروموزوم‌ها: در این مسئله مجموعه ترمینال‌ها متشکل از متغیرهای مستقل تشکیل‌دهنده الگوهای شش‌گانه است. همچنین از عملگرهای ریاضی و توابع مثلثاتی {+, -, ×, /, √, sin, cos, power, ln, and log} استفاده به عمل آمد.

انتخاب ساختار و معماری کروموزوم‌ها: اندازه طول هر راس ۷ و تعداد ژن‌ها ۳ عدد انتخاب شد. نتایج نشان می‌دهد وقتی که تعداد ژن و طول سر به ترتیب بزرگ‌تر از ۳ و ۷ می‌شود عملکرد مدل GEP به طور قابل توجهی در مراحل آموزش و تست بهبود نمی‌یابد. تعداد کروموزوم‌ها نیز از ۳۰ تا ۵۰ متغیر بود. مشخص شده است که بهترین تعداد ۳۰ کروموزوم می‌باشد (Ferreira, 2001).

در این بخش، برای توابع انتقالی مبتنی بر مبنای GEP توسعه یافته برای برآورد هدایت هیدرولیکی اشباع خاک از ۶ الگو استفاده شده است. مراحل مختلف برای ساخت توابع انتقالی بر اساس روش GEP بصورت خلاصه در زیر آمده است: انتخاب تابع پیوندی: در این مطالعه عمل جمع برای ایجاد پیوند بین زیرشاخه‌ها مورد استفاده قرار گرفت. انواع مختلف ارتباط توابع مانند جمع (+)، تفریق (-)، تقسیم (/) و ضرب (×) را می‌توان مورد استفاده قرار داد. در این مطالعه، زیر شاخه‌ها- ETS توسط عملگر جمع (+) پیوند یافتند. استفاده از این پیوند در مقایسه با اپراتورهای دیگر نتایج بهتری ارائه داد (Azamathulla and Jarrett, 2013; al, 2011).

عملگرهای ژنتیکی و نرخ هر یک از آن‌ها انتخاب شدند. در این مطالعه از ترکیبی از عملگرهای ژنتیکی مانند جهش و ترکیب استفاده شد که شامل سه نوع جابجایی (IS, RIS, و جابجایی ژن) و سه نوع نوترکیبی (یک نقطه، دو نقطه، و ژن نوترکیبی) است که در جدول (۲) آمده است.

پس از مراحل فوق مدل GEP به منظور برآورد هدایت هیدرولیکی اشباع اجرا شد. آموزش GEP پس از ۱۰۰۰۰ نسل به دلیل وجود تغییرات قابل اغماض در خطا خاتمه داده شد.

برای مقادیر مختلف K از صفر تا یک محاسبه کرد. مقدار ضریب K بایستی طوری باشد که حداقل کاهش را در  $R^2$  مدل و حداکثر افزایش را در شاخص تحمل داشته باشد. طبق جدول (۳) ضریب K طوری انتخاب شد تا کوچکترین شاخص تحمل مدل به حد بحرانی (۰/۲) برسد ( $K=0/1$ ).

برای بدست آوردن روابط رگرسیونی از روش گام به گام پس رو استفاده گردید. Aimrun and Amin (2009) بیان کردند بهترین رابطه رگرسیونی برای برآورد هدایت هیدرولیکی اشباع خاک‌های مزارع شالیزار یک مدل رگرسیونی گام به گام با متغیرهای ورودی درصد ماده آلی، جرم مخصوص ظاهری، درصد رس و میانگین هندسی اندازه ذرات می‌باشد. از نمایه‌های آمار توصیفی حداقل، حداکثر، میانگین، انحراف معیار و ضریب تغییرات برای پارامترهای فیزیکی و شیمیایی خاک مورد مطالعه استفاده شد. همچنین برای نشان دادن همبستگی بین پارامترها از همبستگی پیرسون استفاده گردید.

#### ارزیابی مدل

برای ارزیابی صحت و اعتبار برآورد مدل از شاخص‌های مختلفی از جمله ضریب تبیین  $R^2$ ، ریشه میانگین مربعات خطا<sup>۳</sup> (RMSE) و میانگین مطلق خطا<sup>۴</sup> (MAE) استفاده شد؛ به این ترتیب هر چه  $R^2$  مدل بیشتر، RMSE و MAE کمتر باشد مدل ایجاد شده صحیح‌تر و معتبرتر است.

$$R^2 = 1 - \frac{\sum_1^N (Y_i - \hat{Y}_i)^2}{\sum_1^N (Y_i - \bar{Y})^2} \quad (\text{رابطه ۵})$$

$$RMSE = \sqrt{\frac{\sum_1^N (Y_i - \hat{Y}_i)^2}{N}} \quad (\text{رابطه ۶})$$

$$MAE = \frac{\sum_1^N |Y_i - \hat{Y}_i|}{N} \quad (\text{رابطه ۷})$$

که در آن‌ها  $Y_i$ ،  $\bar{Y}$ ، و  $\hat{Y}_i$  به ترتیب متغیر وابسته اندازه‌گیری شده، میانگین آن و متغیر وابسته برآورد شده و N تعداد مشاهدات است.

#### نتایج و بحث

برخی نمایه‌های آمار توصیفی پارامترهای فیزیکی و شیمیایی خاک مورد مطالعه در جدول (۴) نشان داده شد. بر پایه ضریب تغییرات مقادیر EC، هدایت هیدرولیکی اشباع، مواد آلی، رس، سیلت و شن در میان خاک‌های مورد استفاده در این تحقیق زیاد است. ضریب تغییرات زیاد پارامترهای اندازه‌گیری شده بیانگر تنوع زیاد خاک‌ها بوده و بنابراین مدل ارائه شده از تعمیم‌پذیری زیادی برخوردار خواهد بود.

#### رگرسیون چندگانه

در این پژوهش برای بدست آوردن روابط رگرسیون خطی چندگانه برای برآورد هدایت هیدرولیکی اشباع از نرم‌افزار Statistica استفاده شد. ابتدا داده‌ها از لحاظ نرمال بودن تست شدند. همه پارامترها به جزء پارامترهای EC و هدایت هیدرولیکی اشباع نرمال بودند. برای نرمال کردن پارامترهای EC و هدایت هیدرولیکی اشباع از روابط تبدیل زیر استفاده شد:

$$EC^* = EC^{0.5} \quad (\text{رابطه ۳})$$

$$K_s^* = \ln(K_s) \quad (\text{رابطه ۴})$$

یکی از مفروضاتی که در ایجاد توابع رگرسیون مد نظر قرار می‌گیرد، استقلال خطاها (تفاوت بین مقادیر واقعی و مقادیر برآوردی توسط برآورد روابط رگرسیون) از یکدیگر است. در صورتی که فرضیه استقلال خطاها رد شود و خطاها با یکدیگر همبستگی داشته باشند امکان استفاده از رگرسیون وجود ندارد. به منظور بررسی استقلال خطاها از یکدیگر از آزمون دوربین-واتسون استفاده شد. از آنجایی که مقدار آماره دوربین-واتسون (۱/۶۷) در فاصله ۱/۵ تا ۲/۵ قرار دارد فرض عدم وجود همبستگی بین خطاها رد می‌شود. فرض دیگر رگرسیون خطی بررسی هم‌راستایی بین متغیرهای مستقل است. بدین منظور از شاخص استقلال یا تحمل<sup>۱</sup> استفاده گردید. عکس شاخص تحمل را شاخص تورم واریانس<sup>۲</sup> (VIF) می‌نامند. به همین دلیل نیز هر چه شاخص هم‌راستایی بزرگتر شود، شاخص تحمل کمتر و خطای استاندارد ضرایب رگرسیون متناظر با آن بزرگ‌تر می‌شود و در نتیجه توان آزمون برای آزمون معنی‌داری ضرایب رگرسیون کاهش می‌یابد (Rezaei and Soltani, 2004). به منظور بررسی هم‌راستایی، بین متغیرهای مستقل از شاخص تحمل استفاده گردید (جدول ۴). اگر شاخص تورم واریانس برابر ۵ تا ۱۰، یا شاخص تحمل برابر ۰/۱ تا ۰/۲ باشد هم‌راستایی حتماً پیش خواهد آمد (Mansoorfar, 2008). طبق این جدول چون متغیرهای درصد شن، سیلت و رس دارای شاخص تحمل کمتر از ۰/۲ می‌باشند بنابراین دارای هم‌راستایی هستند. در این صورت نمی‌توان به صحت رگرسیون چندگانه اعتماد کرد. برای رفع این مشکل ابتدا بر روی اعداد اجزای قطر اصلی ماتریس  $X'X$  عدد کوچکی به نام K را اضافه می‌کنند که اصطلاحاً به آن اریب ریج می‌گویند. در رگرسیون ریج مهمترین مسئله به دست آوردن مقادیر مناسب برای K است (Guilkey and Murphy, 1975). برای محاسبه ضریب K، باید مقادیر شاخص تحمل را

3- Root Mean Squared Error  
4- Mean Absolute Error

1- Tolerance indices  
2- Variance inflation factor

جدول ۳. مقادیر شاخص تحمل (tolerance) برای متغیرهای مستقل به ازای مقادیر ضریب رگرسیون ریح (K)

متغیر مستقل							مقادیر ضریب رگرسیون ریح (K)
رس	سیلت	شن	$\rho_b$	OM	EC	pH	
۰/۰۰۰۷	۰/۰۰۰۳	۰/۰۰۰۱	۰/۱۵	۰/۵۳	۰/۷۸	۰/۸۱	۰/۰۰۰۱
۰/۲۶	۰/۱۸	۰/۱۱	۰/۲۴	۰/۶۴	۰/۸۵	۰/۸۹	۰/۰۶۵
۰/۴۱	۰/۳۳	۰/۲۰۴	۰/۳۳	۰/۷۴	۰/۹۳	۰/۹۶	۰/۱

OM و  $\rho_b$  به ترتیب ماده آلی و جرم مخصوص ظاهری خاک می‌باشند.

جدول ۴. آمار توصیفی برخی از پارامترهای فیزیکی و شیمیایی خاک مورد مطالعه

خصوصیات خاک	حدافل	حداکثر	میانگین	انحراف معیار	ضریب تغییرات (%)
هدایت الکتریکی	۰/۰۴۴	۱۹/۱	۳/۶۴	۳/۲۵	۸۹/۴۳
pH گل اشباع	۷/۵	۸/۳۵	۷/۹۶	۰/۱۷	۲/۲۱
ماده آلی	۰/۰۴	۲/۸	۰/۷۴	۰/۴۶	۶۲/۰۲
جرم مخصوص ظاهری	۱/۱۲	۱/۴۹	۱/۳۰	۰/۰۹۸	۷/۵۵
رس	۴	۴۸	۱۸/۵۳	۸/۹۶	۴۸/۳۳
سیلت	۴	۵۰	۲۶/۲۳	۱۱/۵۸	۴۴/۱۶
شن	۱۶	۹۰	۵۵/۲۳	۱۸/۴۰	۳۳/۳۲
هدایت هیدرولیکی اشباع	۱/۱	۲۹/۱	۷/۶۰	۵/۶۰	۷۳/۷۱

آمد که نشان‌دهنده تأثیر توزیع اندازه ذرات معدنی خاک بر هدایت هیدرولیکی اشباع می‌باشد. الگوی ۴ که دارای کم‌ترین  $R^2$  و بیشترین RMSE و MAE در میان الگوها می‌باشد پارامترهای رس، شن، سیلت و وزن مخصوص ظاهری حذف شده است که نشان‌دهنده تأثیر بافت و ساختمان خاک در هدایت هیدرولیکی اشباع می‌باشد. در مرحله آموزش مدل، انتخاب جمعیت‌های اولیه تصادفی مختلف و تأثیرگذار در پدیده نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت الگو نیز می‌شود (Soltani et al., 2009). چون هدف این پژوهش صرفه‌جویی در وقت و هزینه می‌باشد بنابراین الگوی ۶ که شامل خصوصیات سهل‌الوصول رس و جرم مخصوص ظاهری است برای مدل برنامه‌ریزی ژنتیک انتخاب شد. بطوری‌که این مدل با کمترین ورودی، دارای صحت بالایی در برآورد هدایت هیدرولیکی اشباع خاک است. رابطه (۸) معادله بدست آمده تابع انتقالی برنامه‌ریزی بیان ژنی را نشان می‌دهد:

(رابطه ۸)

$$Ks = (\tan \text{clay} + \tan \text{clay}) - ((\rho_b + 8.49) / \rho_b) + \cos(((1.08 \text{clay})(\text{clay} - \rho_b) + (\cos \text{clay} - \rho_b))) + \cos((\text{clay} \times \text{clay})(3.07 - \rho_b)) + \cos(\cos(\text{clay}))$$

با توجه به جدول (۵) همه متغیرهای مستقل به جزء pH، EC و ماده آلی همبستگی خطی معنی‌داری در سطح احتمال ۱ و ۵ درصد با ضریب هدایت هیدرولیکی اشباع خاک نشان دادند. همبستگی خطی منفی و معنی‌داری در سطوح احتمال ۱ و ۵ درصد بین هدایت هیدرولیکی اشباع و جرم مخصوص ظاهری بدست آمد. با افزایش تراکم و جرم مخصوص ظاهری، هدایت هیدرولیکی اشباع خاک کاهش می‌یابد (Bybord, 2006; Alizadeh, 2004). همچنین تراکم خاک به عواملی چون میزان تراکم، نوع خاک، وضعیت رطوبت و سیستم زراعی بستگی دارد (Sillon et al, 2003). طبق جدول (۵)، ماده آلی همبستگی ضعیف مثبت و EC همبستگی ضعیف منفی با هدایت هیدرولیکی اشباع نشان دادند. ماده آلی و EC از عوامل تأثیرگذار در پایداری ساختمان خاک می‌باشند (Bybord, 2006; Alizadeh, 2004).

جدول (۶) مقادیر بدست آمده آماره‌های  $R^2$ ، RMSE و MAE الگوهای مدل برنامه‌ریزی بیان ژنی را نشان می‌دهد. با توجه به این جدول در الگوی ۱ که دارای بیش‌ترین  $R^2$  و کمترین RMSE و MAE می‌باشد، تمام پارامترهای اندازه‌گیری شده وارد مدل شده‌اند. در الگوی ۲ با حذف پارامترهای رس و شن  $R^2$ ، RMSE و MAE به ترتیب ۰/۶۷، ۲/۹۵ و ۲/۰۸ بدست

(تخلخل مؤثر) و اندازه ذرات از مهم‌ترین متغیرهای پیش‌بینی کننده هدایت هیدرولیکی اشباع هستند. همچنین Azamathulla and Jarrett (2013) نشان دادند که استفاده از مقدار رس به‌جای میانگین هندسی قطر ذرات برای برآورد هدایت هیدرولیکی اشباع خاک‌های استرالیا شاخص مناسب‌تری است.

که در این رابطه  $K_s$ , Clay و  $\rho_b$  به ترتیب هدایت هیدرولیکی اشباع خاک (Cm/h)، درصد رس (%) و جرم مخصوص ظاهری ( $gr/Cm^3$ ) است. Jarvis *et al.* (2002) برای تخمین هدایت هیدرولیکی اشباع از پارامترهای توزیع اندازه ذرات خاک، میانگین هندسی قطر ذرات، جرم مخصوص ظاهری، مقدار مواد آلی و تخلخل مؤثر استفاده کردند و نتایج آنها نیز نشان داد که پارامتر ساختمانی

جدول ۵. ماتریس ضریب همبستگی خطی پیرسون پارامترهای فیزیکی و شیمیایی خاک مطالعه شده

K <sub>s</sub>	Silt	Sand	Clay	$\rho_b$	OM	EC	pH
۱							
	۱						
		۱					
			۱				
				۱			
					۱		
						۱	
							۱

\*\*عدم اختلاف معنی‌دار / \*اختلاف معنی‌دار در سطوح احتمال یک و پنج درصد

جدول ۶. ارزیابی دقت و صحت مدل برنامه‌ریزی بیان‌ژنی در شبیه‌سازی هدایت هیدرولیکی اشباع خاک

مدل	داده‌های آموزشی			داده‌های صحت‌سنجی		
	MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>
۱	۱/۷۱	۱/۷۱	۰/۹۳	۱/۱۷	۱/۸۸	۰/۹۴
۲	۲/۹۵	۲/۰۸	۰/۶۷	۲/۷۳	۳/۸۵	۰/۶۳
۳	۱/۹۲	۱/۴۱	۰/۹۲	۱/۶۷	۲/۷۳	۰/۹۰
۴	۵/۱۰	۳/۱۰	۰/۳۴	۴/۸۰	۵/۹	۰/۲۷
۵	۱/۸۸	۱/۳۸	۰/۸۹	۱/۶۱	۲/۳۳	۰/۹۰
۶	۱/۸۲	۱/۲۳	۰/۹۱	۱/۵۹	۲/۲۷	۰/۹۲

اندازه‌گیری شده و برآوردی هدایت هیدرولیکی اشباع خاک توسط مدل‌های برنامه‌ریزی بیان ژنتیک و مدل رگرسیون ریح را نشان می‌دهد. هر اندازه پراکندگی داده‌ها در اطراف این خط کمتر باشد مقادیر برآوردی توسط مدل به مقادیر اندازه‌گیری شده نزدیک‌تر خواهد شد. طبق این شکل‌ها در مدل برنامه‌ریزی بیان ژنتیک داده‌ها به خط یک به یک نزدیک‌تر هستند که نشان دهنده صحت بیشتر مدل برنامه‌ریزی بیان ژنتیک است. (Parasuraman *et al.*, 2007b) از برنامه‌ریزی بیان ژنی برای برآورد تبخیر و تعرق استفاده نمودند و دریافتند که استفاده از برنامه‌ریزی ژنتیک در برآورد تبخیر و تعرق سبب بروز خطای کمتر و دقت بالاتر می‌گردد. در شکل (۳) چنانچه مشاهده می‌شود در مقادیر اندک هدایت هیدرولیکی اشباع مدل حاصل

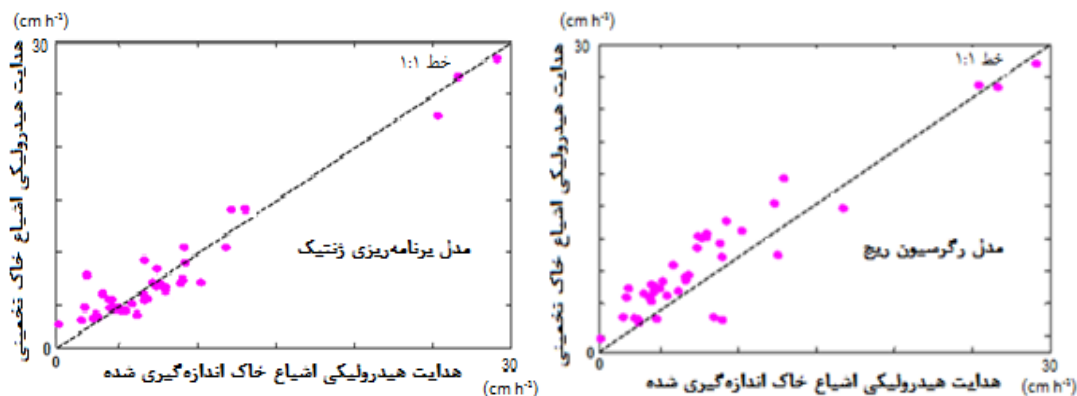
معادله رگرسیونی هدایت هیدرولیکی اشباع برای خاک‌های منطقه مورد مطالعه طبق رابطه (۹) بدست آمد: (رابطه ۹)  $K_s = 0.28 + 0.87 \text{ Sand} + 0.56 \text{ Clay} - 0.44 \rho_b$  جدول (۷) مقادیر آماره‌های مورد استفاده برای ارزیابی مدل‌ها را نشان می‌دهد. همان‌طور که مشاهده می‌شود، ضریب تبیین (R<sup>2</sup>) برای مدل برنامه‌ریزی بیان‌ژنی بیشتر از مدل رگرسیون ریح می‌باشد. همچنین مقادیر RMSE و MAE برنامه‌ریزی بیان‌ژنی کمتر از مدل رگرسیون ریح است. بنابراین تمامی شاخص‌های یاد شده نشان‌دهنده مزیت و کارایی بیشتر مدل برنامه‌ریزی بیان ژنتیک نسبت به مدل رگرسیون ریح می‌باشد.

شکل (۳) به ترتیب نمودار خط یک به یک بین مقادیر

جدول ۷. آماره‌های ارزیابی مدل‌های انتخابی برنامه‌ریزی بیان‌ژنی و

رگرسیون رنج						
داده‌های آموزش			داده‌های صحت‌سنجی			مدل
MAE	RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	
۱/۵۹	۲/۲۷	۰/۹۲	۱/۲۳	۱/۸۲	۰/۹۱	برنامه‌ریزی بیان‌ژنی
۱/۸۸	۳/۱۱	۰/۷۶	۲/۰۷	۳/۴۸	۰/۷۰	رگرسیونی

رگرسیون رنج اندکی دارای بیش تخمینی می‌باشد. دلیل این امر شاید خطی بودن رابطه رگرسیونی بدست آمده باشد و اگر مدل حاصله غیرخطی باشد شاید این بیش برآوردی مشاهده نشود. بنابراین از مزیت‌های روش برنامه‌ریزی ژنتیک آن است که تمامی روابط خطی و غیرخطی محتمل موجود را بررسی نموده و در نهایت مناسب‌ترین آن‌ها را به عنوان خروجی ارائه می‌نماید.



شکل ۳. خط یک به یک بین مقادیر صحت‌سنجی اندازه‌گیری شده و برآوردی هدایت هیدرولیکی اشباع خاک

برآورد هدایت هیدرولیکی اشباع خاک، مناسب‌ترین الگو شامل تخمین‌گرهای رس و وزن مخصوص ظاهری بود. تابع ایجاد شده در این روش از دقت و صحت بیشتری نسبت به تابع انتقالی رگرسیونی برخوردار بود.

## نتیجه‌گیری

در تابع انتقالی رگرسیونی تنها پارامترهای مقدار شن، رس و جرم مخصوص ظاهری به عنوان تخمین‌گر وارد مدل شدند. در روش برنامه‌ریزی بیان‌ژنی از بین ۶ الگوی ایجاد شده برای

## REFERENCES

- Aimrun, W. and Amin, M.S.M. (2009) Pedo-transfer function for saturated hydraulic conductivity of lowland paddy soils. *Paddy and Water Environment*, 7, 217-225.
- Alizadeh, A. (2004) *Soil physics*. Mashhad: Ferdowsi University of Mashhad Publications.
- Alvisi, S., Mascellani, G., Franchini, M. and Bardossy, A. (2005) Water level forecasting through fuzzy logic and artificial neural network approaches. *J. Hydrol. Earth Sys. Sci.* 2, 1107-1145.
- Azamathulla, H. M. and Jarrett, R. D. (2013) Use of gene-expression programming to estimate Manning's roughness coefficient for high gradient streams. *Water Resources Management*, 27, 715-729.
- Babovic, V. and Abbott, M. B. (1997) Evolution of equation from hydraulic data. Part 1: Theory. *J. Hydraul. Res.* 35, 1-14.
- Baybordi, M. (2006). *Soil Physics*. Tehran: University of Tehran Publications.
- Blake, G. R. and Hartge, K. H. (1986) Bulk density. In: Klute, A. Eds. *Methods of soil analysis. part 1*. (pp. 363-375). 2<sup>nd</sup> ed. Agron. Monogr. 9. ASA. Madison. WI.
- Bouma, J. (1989) Using soil survey data for qualitative land evaluation. *Advances in Soil Science*, 9, 177-213.
- Campbell, G.S. (1985). *Soil physics with basic*. Elsevier. New York.
- Cosby, B. J., Hornberger, G. M., Clapp, R. B. and Ginn, T. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20, 682-690.
- Dane J. H. and Puckett, W. (1994). Field soil hydraulic properties based on physical and mineralogical information. In: *Proceeding of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*. University of California, Riverside.
- Ferreira, C. (2001) Gene Expression Programming in Problem Solving. Invited Tutorial of the 6th Online World Conference on Soft Computing in Industrial Applications, September 10-24.
- Ferreira, C. (2006) Gene expression programming: mathematical modeling by an artificial intelligence. Springer. Berlin, Heidelberg New York. USA. 478p.
- Gee G. W. and Or D. (2002) Particle-size analysis. In: Warren, A.D. Eds. *Methods of Soil Analysis. Part 4. Physical Methods*. (pp.255-295). Soil Sci. Soc. Am. Inc.



- Guilkey, D. K. and Murphy, J. L. (1975) Directed ridge regression techniques in case of multicollinearity. *Journal of the American Statistical Association*. 70, 769-775.
- Güven, A. (2009) Linear genetic programming for time-series modeling of daily flow rate. *J. Earth Syst. Sci.* 118, 157-173.
- Hashmi, M. Z., Shamseldin, A. Y., Melville, B. W. (2011) Statistical downscaling of watershed precipitation using gene expression programming (GEP). *Environ. Modell. Softw.* 26, 1639-1646.
- Hillel, D. (1982) *Introduction to soil physics*. Academic Press, Inc. San Diego, California.
- Hong Y. S. White P. A. and Scott D. M. (2005) Automatic rainfall recharge model induction by evolutionary computational intelligence. *Water Resour. Res.* 41:W08422.
- Jabro, J. D. (1992) Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data. *Transactions of the ASAE*, 35, 557-560.
- Jarvis, N. J., Zavattaro, L. K., Reynolds, W. D., Olsen, P. A., McGechan, M., Mecke, M., Mohanty, B., Leeds-Harrison, P. B. and Jacques, D. (2002). Indirect estimation of near-saturated hydraulic conductivity from readily available soil information. *Geoderma*, 108, 1-17.
- Khu, S. T., Liong, S. Y., Babovic, V., Madsen, H. and Muttill, N. (2001) Genetic programming and its application in real-time runoff forecasting. *Journal of the American Water Resources Association*, 37, 439-451.
- Klute, A. (1986) *Methods of Soil Analysis. Part 1, Physical and Mineralogical Methods*. Madison, Wisconsin, USA.
- Koza, J. (1992) *Genetic Programming: on the Programming of Computers by Means of Natural Selection*. MIT Press.
- Li, Y. F., Min, X. and Thong-ngee, G. (2010) Adaptive ridge regression system for software cost estimating on multi-collinear datasets. *The Journal of System and Software*. 83, 2332-2343.
- Liong, S. Y., Gautam, T. R., Khu, S. T., Babovic, V., Keijzer, M. and Muttill, N. (2002). Genetic programming: A new paradigm in rainfall runoff modeling. *J. Am. Water Res. Assoc.* 38, 705-718.
- Makkeasorn, A., Chang, N. B., Beaman, M., Wyatt, C. and Slater, C. (2006). Soil moisture estimation in a semiarid watershed using RADARSAT- 1 satellite imagery and genetic programming. *Water Resour. Res.* 42, W09401.
- Mansoorfar, K. (2008) *Advanced statistics methods*. Tehran: University of Tehran Publications.
- Merdun, H., Cinar, O., Meral, R. and Apan, M. (2006). Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil and Tillage Research*. 90: 108-116.
- Morales, A., Clup, G. L. and Johnson, P. R. (2007) The soil hydrological behavior to irrigation with drainage water. pp. 129-134.
- Nelson, D. W. and Sommer, L. E. (1982) Total carbon, organic carbon, and organic matter. In: Sparks, D. L., Page, A. L., Helmke, P. A., Loeppert, R. H., Soltanpour, P. N., Tabatabai, M. A., Johnston, C. T., Sumner, M. E. Eds. *Methods of soil analysis: part 3. chemical and microbiological properties*. pp. 539-579. Soc. Agron., Madison.
- Parasuraman, K., Elshorbagy, A. and Carey, S. K. (2007a). Modeling the dynamics of the evapotranspiration process using genetic programming. *Hydrol. Sci. J.* 52, 563-578.
- Parasuraman, K., Elshorbagy, A. and Si, B. C. (2007b). Estimating Saturated Hydraulic Conductivity Using Genetic Programming. *Soil Sci. Soc. Am. J.* 71, 1676-1684.
- Rezaei, A. and Soltani, A. (2004) *An introduction to applied regression analysis*. Isfahan University of Technology Publications. 294p.
- Savic, D.A., Walters, G. A. and Davidson, J. W. (1999) Genetic programming approach to rainfall-runoff modelling. *Water Resour. Manage.* 13, 219-231.
- Sillon, J. F., Richard, G. and Cousin, I. (2003) Tillage and traffic effects on soil hydraulic properties and evaporation. *Geoderma*. 116, 29-46.
- Soltani, A., Gorbani, M. A. Fakhri- Fard A., Darbandi, S. and Farsadzadeh, D. (2011). Genetic programming and its application in rainfall-runoff modeling. *Water and Soil Science*. 20, 61-71.
- Vereecken, H., Maes, J. and Feyen, J. (1990) Estimating unsaturated hydraulic conductivity from easily measured soil properties. *Soil Science*, 149, 1-12.
- Wagner, B., Tarnawski, V.R., Hennings, V., Muller, U., Wessolek, G. and Plagge, R. (2001) Evaluation of pedotransfer function for unsaturated soil hydraulic conductivity using an independent data set. *Geoderma*, 102, 275-297.
- Wösten, J. H. M. (1997) Pedotransfer functions to evaluate soil quality. In: Gregorich, E. G., Carter, M. R. (eds.) *Developments in Soil Science*, Elsevier. 25, 221-225.
- Wösten J. H. M., Pachepsky, Y. A. A. and Rawls, W. J. (2001) Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 251, 123-150.