

## Capability of Regression and Random Forest Methods to Estimate Soil Water Retention Curve by Developing Pseudo-Continuous Pedotransfer Functions

Reza Kiani<sup>1</sup> , Hossein Bayat<sup>2</sup> 

1-Department of Soil Sciences and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran.

E-mail: [siavashkiani37@gmail.com](mailto:siavashkiani37@gmail.com)

2-Corresponding author, Department of Soil Sciences and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran. E-mail: [h.bayat@basu.ac.ir](mailto:h.bayat@basu.ac.ir) or [hbayat2001@gmail.com](mailto:hbayat2001@gmail.com)

Received: May 19, 2024

Revised: June 21, 2024

Accepted: June 24, 2024

Published: December 21, 2024

### Extended Abstract

#### Background and Objectives

Direct methods of measuring soil water retention curve (SWRC) are time-consuming and expensive, so they are not easily applicable to large scales. Therefore, researchers use pedotransfer functions (PTFs) to obtain it. Various point and parametric pedotransfer functions have been used so far, with numerous methods to estimate the SWRC, each of which has its drawbacks. However, rare methods have been used to develop pseudo-continuous pedotransfer functions. The random forest (RF) method has not been utilized in any study so far, to create pseudo-continuous pedotransfer functions. Some variables have not been used as predictors in pseudo-continuous pedotransfer functions in any research. Therefore, the objectives of this article include investigating the potential of the RF method in creating pseudo-continuous pedotransfer functions, comparing its performance with linear regression, and examining the probability of improving the performance of these functions using the geometric mean and standard deviation of particles diameter and field capacity (FC) and permanent wilting point (PWP) as predictors.

#### Methodology

A total of 120 disturbed and undisturbed soil samples were collected from two provinces of Tehran and Hamedan. Soil texture, bulk density, and soil water retention curve in the range of 0 to 15000 hPa were measured. Then pseudo-continuous pedotransfer functions were created using two methods of linear regression and random forest. The soil water matric suction, soil texture, percentage of silt and sand, bulk density, geometric mean, standard deviation of particles diameter, and moisture content at FC and PWP were used in various combinations to estimate the soil water retention curve. The accuracy and reliability of the generated functions were compared between the two methods and within each method.

#### Results

Using soil water matric suction as the only input variable for estimating moisture at different matric suctions was not effective in the RF method, and no model was created. However, in the linear regression method, a model with acceptable results was developed (with  $R^2$  values of 0.675 and 0.674 for training and validation stages, respectively), which can be utilized in situations where additional information is not available. The inclusion of soil texture in the linear regression method significantly improved the accuracy of estimates by 5.4% and 5.3% in both training and validation stages, respectively. In the third function, incorporating the percentage of clay and sand alongside soil water matric suction as predictors improved SWRC estimation by 1.5% to 25.0% in both training and validation stages for both RF and linear regression compared to the second function. In the fourth function, using bulk density as an additional predictor led to a significant improvement

in accuracy by 6.9% to 13.1%, because bulk density serves as an indicator of soil structure, enhancing the estimation of the soil water retention curve. Utilizing FC improved estimation accuracy by 3.5% to 24.4%, because FC is a point on the SWRC and enters direct information to the models. However, using the PWP as a predictor did not significantly improve estimation accuracy. Using geometric mean (dg) and geometric standard deviation (Sg) instead of percentage of clay and sand in pseudo-continuous pedotransfer functions did not lead to noticeable improvements. Error distribution across soil texture triangles in the linear regression method showed no dependence on soil texture. Because, in pedotransfer functions 1, 2, 4, 7, and 8, the highest error values were obtained in coarse-textured soils, while in pedotransfer functions 5, 6, 9, and 10, the lowest error values were associated with coarse-textured soils. Error distribution across soil texture triangles depended on the type of input variables and the method used to create pedotransfer functions. In all pseudo-continuous pedotransfer functions created by both methods, the accuracy of estimates in both training and validation stages in the RF method was significantly and noticeably higher, ranging from 22% to 46% more than those in linear regression.

### Conclusion

Using the regression method and solely relying on soil water matric suction as a predictor, an acceptable pseudo-continuous pedotransfer function was developed. Investigating the potential of establishing a similar relationship using the state-of-the-art estimation methods may lead to independence from relying on numerous soil water retention curve models. Utilizing more detailed information such as particle size distribution and FC for estimating the SWRC through pseudo-continuous pedotransfer functions is recommended. The dependence of error distribution on soil texture triangles on the type of input variables and the method used to create pedotransfer functions underscores the importance of selecting an appropriate combination of input variables and method for creating pseudo-continuous pedotransfer functions for estimating the SWRC. Given the significant superiority of the random forest method over linear regression, using soil water matric suction, percentage of clay and sand, bulk density, and FC as predictors in pseudo-continuous pedotransfer functions with the RF method yielded the best results in estimating the SWRC.

**Keywords:** Geometric mean diameter, Geometric standard deviation, Pseudo-continuous pedotransfer functions, Random forest, Soil moisture

### Data Availability Statement

Data is available on reasonable request from the authors.

### Acknowledgements

This paper is published as a part of a Ph. D. thesis supported by the Vice Chancellor for Research and Technology of the Bu-Ali Sina University, Iran. The authors are thankful to the Bu-Ali Sina University for financial support.

### Conflict of interest

The authors declare no conflict of interest.

### Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

**Cite this article:** Kiani, R. & Bayat, H. (2024). Capability of regression and random forest methods to estimate soil water retention curve by developing pseudo-continuous pedotransfer functions. *Journal of Soil and Plant Science*, 34(4), 15–36.

<https://doi.org/10.22034/sps.2024.19180>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.



Copyright © 2024 The Authors.  
Publisher: The University of Tabriz





مقاله پژوهشی

## قابلیت دو روش رگرسیون و جنگل تصادفی برای تخمین منحنی نگهداری آب خاک با ایجاد توابع انتقالی شبه‌پیوسته

رضا کیانی<sup>۱</sup>، حسین بیات<sup>۲</sup>✉

۱- گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان، ایران.

رایانامه: [siavashkiani37@gmail.com](mailto:siavashkiani37@gmail.com)

۲- نویسنده مسئول، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه بوعلی سینا، همدان، ایران.

رایانامه: [h.bayat@basu.ac.ir](mailto:h.bayat@basu.ac.ir)

تاریخ دریافت: ۱۴۰۳/۰۲/۳۰	تاریخ بازنگری: ۱۴۰۳/۰۴/۰۱
تاریخ پذیرش: ۱۴۰۳/۰۴/۰۴	تاریخ انتشار: ۱۴۰۳/۱۰/۰۱

### چکیده

تا کنون توابع انتقالی نقطه‌ای و پارامتریک با روش‌های زیادی برای تخمین منحنی نگهداری آب خاک (SWRC) استفاده شده‌اند، اما از روش جنگل تصادفی (RF) با برخی متغیرهای ورودی تا کنون در هیچ مطالعه‌ای برای ایجاد توابع انتقالی شبه‌پیوسته استفاده نشده است. تعداد ۱۲۰ نمونه خاک از دو استان تهران و همدان برداشت و ویژگی‌های فیزیکی آن‌ها اندازه‌گیری گردید. تعداد ۱۰ تابع انتقالی شبه‌پیوسته با روش‌های رگرسیون خطی و RF ایجاد شد. از متغیرهای مکش آب خاک، بافت خاک، درصد رس و شن، جرم مخصوص ظاهری، میانگین و انحراف معیار هندسی قطر ذرات، و رطوبت در ظرفیت مزرعه‌ای (FC) و نقطه پژمردگی دائم (PWP) در ترکیب‌های مختلف برای تخمین SWRC استفاده شد. استفاده از مکش خاک به‌عنوان تنها متغیر ورودی برای تخمین SWRC در روش رگرسیون خطی، مدلی با نتایج قابل قبول ایجاد کرد ( $R^2$  مراحل آموزش و معبرسازی به ترتیب ۰/۶۷۵ و ۰/۶۷۴ بود). استفاده از درصد رس و شن به‌عنوان تخمین‌گر موجب بهبود تخمین (۱/۵ تا ۲۵/۰ درصد) گردید. جرم مخصوص ظاهری موجب بهبود معنادار درستی تخمین‌ها در دامنه ۶/۹ تا ۱۳/۱ درصد گردید. بر خلاف PWP، استفاده از FC موجب بهبود درستی تخمین‌ها در دامنه ۳/۵ تا ۲۴/۴ درصد شد. توزیع خطا (RMSE) بر روی مثلث بافت خاک وابسته به نوع متغیرهای ورودی و روش ایجاد توابع بود. در تمام توابع شبه‌پیوسته، درستی تخمین‌ها، بر مبنای RMSE، در روش RF به‌طور معنادار و قابل توجهی در دامنه ۲۲ تا ۴۶ درصد بیشتر از رگرسیون خطی بود.

**واژه‌های کلیدی:** انحراف معیار هندسی، توابع انتقالی شبه‌پیوسته، جنگل تصادفی، رگرسیون خطی، رطوبت ظرفیت مزرعه‌ای، میانگین هندسی قطر ذرات خاک

استناد به این مقاله: کیانی، ر. و بیات، ح. (۱۴۰۳). قابلیت دو روش رگرسیون و جنگل تصادفی برای تخمین منحنی نگهداری آب خاک با ایجاد توابع انتقالی شبه‌پیوسته. نشریه دانش خاک و گیاه، ۳۴(۴)، ۱۵-۳۶.

<https://doi.org/10.22034/sps.2024.19180>

## مقدمه

بیشتر مطالعات هیدرولوژی-کشاورزی نیازمند اطلاعاتی در مورد ویژگی‌های هیدرولیکی خاک‌های غیراشباع، مانند منحنی‌های نگهداری آب خاک<sup>۱</sup> (SWRC) و هدایت هیدرولیکی غیراشباع است. با این حال، به دلیل تغییرات زمانی و مکانی قابل توجه آن‌ها، این ویژگی‌ها به عنوان داده‌هایی کمتر دسترس‌پذیر شناخته می‌شوند. هر چند که پیشرفت‌های عمده‌ای در توسعه تکنیک‌های اندازه‌گیری آزمایشگاهی و میدانی انجام شده است، هنوز هم روش‌های مستقیم اندازه‌گیری این منحنی‌ها زمان‌بر و هزینه‌بر هستند و به‌سادگی در مقیاس‌های بزرگ قابل اجرا نیستند (Vereecken et al., 2010). از طرف دیگر پیش‌بینی خواص هیدرولیکی ممکن است به اندازه کافی دقیق باشد که بتواند در مدل‌های انتقال آب و املاح مورد استفاده قرار گیرد (Wösten et al., 2001). بنابراین، ایجاد توابع انتقالی<sup>۲</sup> (PTFs) برای برقراری ارتباط ریاضی بین داده‌های قابل دسترس و ویژگی‌های هیدرولیکی همچنان نیازمند تحقیقات است (Haghverdi et al., 2014).

تا کنون توابع انتقالی بسیاری برای تخمین منحنی نگهداری آب خاک توسعه یافته‌اند که به دو گروه اصلی تقسیم می‌شوند: PTF‌های «نقطه‌ای» که محتوای آب خاک را در یک مکش ماتریک معین پیش‌بینی می‌کند (Wösten et al., 2001) و PTF‌های «پارامتریک» که پارامترهای یک مدل منحنی نگهداری آب خاک، مانند مدل ون‌گنوختن، را پیش‌بینی می‌کنند (Cueff et al., 2021). هر یک از توابع نقطه‌ای و پارامتریک دارای معایبی هستند. در روش توابع نقطه‌ای تعداد زیادی از معادله‌های رگرسیونی برای به دست آوردن کل منحنی نگهداری آب خاک لازم است (Wösten et al., 2001). نقص‌های ذاتی PTF‌های پارامتریک نیز عبارتند از: ۱. شکل واقعی SWRC همیشه با شکل معادله انتخاب شده یکسان نیست، ۲. SWRC‌هایی که توسط خروجی‌های PTF پارامتریک تشکیل می‌شوند همیشه دارای خطای بیشتر از SWRC‌های برازش شده هستند. در هر دو مورد، یعنی برازش و تخمین SWRC، معادله هیدرولیک خاک را می‌توان توسط یک مدل شبکه عصبی مناسب جایگزین کرد (Haghverdi et al., 2012; Haghverdi et al., 2014).

اخیراً (Haghverdi et al., 2012) شکل جدیدی از توابع انتقالی برپایه شبکه‌های عصبی به نام توابع انتقالی شبه‌پیوسته<sup>۳</sup> را معرفی کردند. تقریباً این روش قادر به تعیین SWRC‌های پیوسته بدون استفاده از هیچ معادله هیدرولیکی خاک است. آن‌ها نشان دادند که توابع انتقالی شبه‌پیوسته دقیق‌تر و قابل اعتمادتر از توابع انتقالی پارامتریک مدل ون‌گنوختن بودند. توابع انتقالی شبه‌پیوسته می‌تواند برای پیش‌بینی محتوای آب در برخی از مکش‌های مشخص مانند مکش معادل رطوبت ظرفیت مزرعه‌ای مورد استفاده قرار گیرد یا برای پیش‌بینی منحنی کامل SWRC استفاده شود. با وجود عملکرد امیدوارکننده توابع انتقالی شبه‌پیوسته، هنوز برخی از سوالات اساسی در مورد چگونگی اعمال این روش جدید برای پیش‌بینی SWRC وجود دارد که بی‌پاسخ مانده است (Haghverdi et al., 2014). برخی از این سوالات عبارتند از اینکه تراکم و توزیع داده‌های جفت نقاط اندازه‌گیری شده SWRC چگونه باشد؟ و نیز اینکه از چه روش‌هایی برای ایجاد این توابع استفاده شود تا عملکرد این‌ها افزایش یابد؟ بخشی از اهداف این تحقیق یافتن پاسخ سوال دوم می‌باشد.

روش‌های زیادی برای ایجاد توابع انتقالی نقطه‌ای و پارامتریک جهت تخمین منحنی نگهداری آب خاک مورد استفاده قرار گرفته‌اند که از جمله آن‌ها می‌توان رگرسیون خطی و غیرخطی چندگانه (Gunarathna et al., 2019b; Merdun et al., 2006)، شبکه‌های عصبی مصنوعی<sup>۴</sup> (ANN) (Bayat et al., 2013a; Bayat et al., 2013b)، روش مدیریت گروهی

<sup>۱</sup>-Soil water retention curve

<sup>۲</sup>-Pedotransfer functions

<sup>۳</sup>-Pseudo-continuous pedotransfer functions

<sup>۴</sup>-Artificial neural networks

داده‌ها (Bayat et al., 2011; Neyshaburi et al., 2015; Pachepsky & Rawls, 1999) k-نزدیک‌ترین همسایه<sup>۵</sup> (kNN) (Gunarathna et al., 2019a; Gunarathna et al., 2019b; Nguyen et al., 2017) و ماشین‌های بردار پشتیبان<sup>۶</sup> (SVM) (Khlosi et al., 2016) را نام برد؛ اما روش‌های استفاده شده برای ایجاد توابع انتقالی شبه‌پیوسته جهت تخمین منحنی نگهداری آب خاک فقط محدود به روش‌های شبکه‌های عصبی مصنوعی (Haghverdi et al., 2012)، ماشین‌های بردار پشتیبان (Haghverdi et al., 2014)، k-نزدیک‌ترین همسایه و رگرسیون خطی چندگانه (Nguyen et al., 2017) می‌باشد. جنگل تصادفی<sup>۷</sup> (RF) یا جنگل‌های تصمیم تصادفی<sup>۸</sup>، به عنوان یک رویکرد مطلوب روش یادگیری تجمعی برای پیش‌بینی و دسته‌بندی، شناخته شده است (Verikas et al., 2011). روش RF توسط (Breiman 2001) با گسترش تکنیک درخت‌های دسته‌بندی و رگرسیون<sup>۹</sup> (CART) توسعه یافته است تا عملکرد بهتری در پیش‌بینی ارائه دهد. تا کنون، روش RF به ندرت برای تخمین منحنی نگهداری آب خاک، حتی از طریق توابع انتقالی نقطه‌ای و پارامتریک، مورد استفاده قرار گرفته است. (Tóth et al. 2014) از روش RF برای تجزیه و تحلیل رابطه بین مقدار آب خاک در چهار سطح مکش ماتریک خاک (۰/۱، ۳۳ و ۱۵۰۰ کیلوپاسکال و ۱۵۰ مگاپاسکال) و اطلاعات نقشه خاک مجارستان استفاده کردند. آن‌ها متوجه شدند که اهمیت ویژگی‌های خاک در پیش‌بینی مقدار آب خاک بر اساس نوع خاک و سطح مکش ماتریک خاک متفاوت است. (Szabó et al. 2019) در ایجاد توابع انتقالی نقطه‌ای برای تخمین رطوبت خاک از روش RF استفاده کردند. محققان دیگری (Gunarathna et al., 2019a; Ließ et al., 2012) نیز کاربرد RF در تخمین خصوصیات هیدرولیکی را با روش‌های دیگری مانند ANN و kNN مقایسه کرده و برتری عملکرد RF را گزارش کردند. (Rastgou et al. 2020) قابلیت اعتماد کمتر روش RF نسبت به رگرسیون غیرخطی<sup>۱۰</sup> (NLR) در تخمین پارامتریک SWRC را گزارش کردند.

(Taftah et al. 2022) برتری روش RF نسبت به SVM و ANN در تخمین نقطه‌ای SWRC را گزارش کردند. Cueff et al. (2021) قابلیت RF، رگرسیون خطی<sup>۱۱</sup> (LR) و درخت‌های رگرسیونی را در تخمین نقطه‌ای SWRC برای زمین‌های کشاورزی حفاظتی بررسی کرده و درستی برابر RF و LR را گزارش کردند. (Kotlar et al. 2020) نیز قابلیت RF و LR را در تخمین نقطه‌ای SWRC برای یک خاک سیلتی جنگل آمازون بررسی کردند و گزارش کردند که بهترین نتیجه توسط روش RF به دست آمد. مرور منابع نشان داد که روش RF، هر چند به ندرت، برای ایجاد توابع انتقالی نقطه‌ای یا پارامتریک SWRC به کار گرفته شده است، اما از این روش تا کنون در هیچ مطالعه‌ای برای ایجاد توابع انتقالی شبه‌پیوسته استفاده نشده است. از طرف دیگر نتایج ضد و نقیضی در مورد قابلیت این روش گزارش شده است. تا کنون متغیرهای متعددی برای تخمین نقطه‌ای و پارامتریک SWRC استفاده شده است. از جمله این متغیرها اجزای بافت خاک، جرم مخصوص ظاهری و رطوبت در برخی پتانسیل‌ها (Schaap et al., 2001) و میانگین و انحراف معیار هندسی قطر ذرات خاک (Rastgou et al., 2020) را می‌توان نام برد. کارایی هر متغیر در تخمین SWRC می‌تواند وابسته به نوع تخمین (نقطه‌ای و پارامتریک) و روش ایجاد توابع انتقالی متفاوت باشد. با توجه به اندک بودن تحقیقات در مورد توابع انتقالی شبه‌پیوسته، هر چند که کارایی اجزای بافت خاک، جرم مخصوص ظاهری و ماده آلی، (Haghverdi et al.,

<sup>۵</sup>-k-nearest neighbor

<sup>۶</sup>-Support vector machines

<sup>۷</sup>-Random forest

<sup>۸</sup>-Random decision forests

<sup>۹</sup>-Classification and regression trees

<sup>۱۰</sup>-Non-linear regression

<sup>۱۱</sup>-Linear regression

(2012) و رطوبت اولیه خاک (Haghverdi et al., 2018) به عنوان تخمین‌گر در این نوع توابع مورد بررسی قرار گرفته، ولی کارایی اطلاعات نوع دیگری از توصیف بافت خاک، یعنی میانگین و انحراف معیار هندسی قطر ذرات، و رطوبت در نقاط پتانسیلی خاص مانند FC و PWP به عنوان تخمین‌گر در توابع انتقالی شبه‌پیوسته، تا کنون در هیچ تحقیقی مورد بررسی قرار نگرفته است. بنابراین، اهداف این پژوهش عبارت بود از ۱- بررسی قابلیت روش RF در ایجاد توابع انتقالی شبه‌پیوسته و مقایسه عملکرد آن با روش رگرسیون خطی، ۲- بررسی احتمال افزایش درستی و قابلیت اعتماد توابع انتقالی شبه‌پیوسته با استفاده از میانگین و انحراف معیار هندسی قطر ذرات و رطوبت در نقاط پتانسیلی خاص مانند FC و PWP به عنوان تخمین‌گر.

## مواد و روش‌ها

### منطقه مورد مطالعه و نمونه‌برداری

برای انجام این تحقیق تعداد ۱۲۰ نمونه دست‌خورده و دست‌نخورده از لایه سطحی و در برخی نقاط از لایه زیرسطحی جمع‌آوری شد. در محل‌های نمونه‌برداری برای این تحقیق عمق لایه سطحی در نقاط مختلف متفاوت بود که از ۰ - ۱۵ تا ۰ - ۴۰ سانتی‌متر متغیر بود. عمق لایه زیر سطحی نیز از ۱۵ - ۴۰ تا ۳۰ - ۶۰ سانتی‌متر متغیر بود. تعداد ۹۰ نمونه از خاک‌های مختلف استان تهران با مختصات جغرافیائی  $35^{\circ}00'0.4''$  N تا  $36^{\circ}59'10.3''$  N و  $50^{\circ}24'55.8''$  E تا  $50^{\circ}50'12.7''$  E، و تعداد ۳۰ نمونه از استان همدان از دو منطقه مزرعه آموزشی و پژوهشی دانشگاه بوعلی سینا در روستای دست‌جرد (۱۶ نمونه) با مختصات جغرافیائی  $34^{\circ}52'4.06''$  N تا  $34^{\circ}52'55.2''$  N و  $48^{\circ}28'14.3''$  E تا  $48^{\circ}29'21.18''$  E و شهرستان تویسرکان، روستای سیستانه (۱۴ نمونه) با مختصات جغرافیائی  $34^{\circ}40'11.82''$  N تا  $34^{\circ}40'17.07''$  N و  $48^{\circ}19'14.7''$  E تا  $48^{\circ}19'18.95''$  E برداشت شد. انتخاب محل‌های نمونه‌برداری در استان تهران بر اساس سری‌های غالب خاک و در استان همدان از محل‌هایی با بیشترین تفاوت در ویژگی‌های خاک به صورت نامنظم انجام شد. فاصله نمونه‌ها در استان تهران حدود ۳ تا ۴ کیلومتر، در دست‌جرد همدان حدود ۲ کیلومتر و در تویسرکان حدود ۱ کیلومتر بود. نمونه‌های دست‌نخورده با استفاده از سیلندرهای نمونه‌برداری به ارتفاع ۴/۵ و قطر ۵/۱ سانتی‌متر از وسط لایه سطحی یا در برخی نقاط زیرسطحی خاک تهیه گردید. نمونه‌های دست‌خورده از همان نقاط با بیل جمع‌آوری شد. نمونه‌ها پس از جمع‌آوری به آزمایشگاه فیزیک خاک دانشگاه بوعلی‌سینا-همدان منتقل شد. نمونه‌های دست‌خورده هواخشک شده و از الک ۲ میلی‌متری عبور داده شدند.

### اندازه‌گیری ویژگی‌های خاک

#### توزیع اندازه ذرات خاک

برای به‌دست آوردن درصد اجزای بافت خاک، از روش هیدرومتر و بر اساس سیستم وزارت کشاورزی آمریکا (USDA) استفاده شد (Gee & Or, 2002). میانگین هندسی قطر ذرات<sup>۱۲</sup> (dg) و انحراف معیار هندسی قطر ذرات<sup>۱۳</sup> (s<sub>g</sub>) با روش (Shirazi & Boersma, 1984) محاسبه گردید. جرم مخصوص ظاهری با روش نمونه‌های دست‌نخورده اندازه‌گیری شد (Grossman & Reinsch, 2002).

برای اندازه‌گیری مقدار رطوبت در مکش‌های ۰، ۱۰، ۲۰، ۴۰ و ۶۰ هکتوپاسکال از دستگاه جعبه شن (Romano & Palladino, 2002) و در مکش‌های ۳۰۰، ۱۰۰۰، ۳۰۰۰، ۵۰۰۰ و ۱۵۰۰۰ هکتوپاسکال از دستگاه صفحات تحت فشار

<sup>۱۲</sup> . Geometric mean diameter

<sup>۱۳</sup> . Geometric standard deviation

(Dane & Hopmans, 2002) استفاده شد. اندازه گیری‌ها تا مکش ۳۰۰۰ هکتوپاسکال بر روی نمونه‌های دست‌نخورده و در مکش‌های بالاتر از آن بر روی نمونه‌های دست‌خورده انجام شد.

### ایجاد توابع انتقالی

قبل از ایجاد توابع انتقالی فرایند پیش‌پردازش داده‌ها انجام شد. بدین ترتیب که ابتدا داده‌های پرت بررسی شد، مقادیر خارج از دامنه میانگین  $\pm 2/5 \times$  انحراف معیار داده پرت محسوب گردید (Zamani, 2011). داده‌های پرت توسط مقادیر حد پایین و بالایی به ترتیب برای داده‌های کوچک‌تر و بزرگ‌تر از مقادیر حد پایین و بالایی تعیین شده، جایگزین شدند. بررسی نرمال بودن توزیع داده‌ها با استفاده از آزمون کلموگروف-اسمیرنوف<sup>۱۴</sup> در نرم افزار MiniTab انجام شد. در صورت نرمال نبودن توزیع، تبدیل‌های متعدد مانند لگاریتم، معکوس، مربع، جذر، معکوس توان دوم، کاکس باکس، جانسون، و نرمال اسکور بر روی داده‌ها انجام شد. برای هر متغیر، تبدیلی که به توزیع نرمال نزدیک‌ترین نتیجه را می‌داد، انتخاب شد. سپس همه متغیرهای ورودی (به جز مکش خاک) استاندارد شدند، بگونه‌ای که میانگین صفر و انحراف معیار یک داشته باشند. استاندارد کردن متغیرها ایجاد مدل را تسهیل و تسریع می‌نماید و احتمال رسیدن به جواب مناسب را افزایش می‌دهد. از مکش خاک لگاریتم گرفته شد. برای بررسی چند هم خطی از فاکتور عامل تورم واریانس<sup>۱۵</sup> (VIF) استفاده شد که همه متغیرها VIF کمتر از ۱۰ داشتند (Paul, 2006).

تعداد ۱۰ تابع انتقالی شبه‌پیوسته با استفاده از ترکیب‌های مختلف از متغیرهای ورودی ایجاد شد. متغیرهای ورودی مراحل مختلف در جداول مربوط به نتایج نشان داده شده است. در توابع انتقالی شبه‌پیوسته مکش خاک نیز به‌عنوان یک متغیر ورودی مورد استفاده قرار گرفته و رطوبت در مکش‌های مختلف تخمین زده می‌شود (Haghverdi et al., 2012). برای جلوگیری از ایجاد چند هم خطی متغیر سیلت به‌عنوان ورودی استفاده نشد. چرا که همبستگی آن با رطوبت در مکش‌های مختلف نیز نسبت به دیگر اجزای بافت خاک کمتر بود.

از تابع اول تا ۶ متغیرهای ورودی به ترتیب سهولت دسترسی به آن‌ها مورد استفاده قرار گرفتند. در تابع اول فقط مکش خاک به‌عنوان متغیر ورودی استفاده شد تا میزان کارایی آن در تخمین SWRC مورد بررسی قرار گیرد. در تابع دوم، علاوه بر مکش خاک، بافت خاک نیز به‌عنوان متغیر ورودی استفاده شد. لازم به ذکر است که در این مرحله به هر کلاس بافتی یک کد عددی داده شده و به‌عنوان ورودی استفاده شد. در این کدهای ابتدا کل ۱۲ کلاس بافتی سیستم وزارت کشاورزی آمریکا (USDA) به ترتیب از ریز به درشت مرتب شده و کدهای ۱ تا ۱۲ را به خود اختصاص دادند. بنابراین، به کلاس‌های رسی، رس سیلتی، لوم رسی، لوم رسی سیلتی، لوم رسی شنی، لوم، لوم شنی، شن لومی و شن به ترتیب کدهای ۱، ۲، ۴، ۵، ۶، ۹، ۱۰، ۱۱ و ۱۲ اختصاص داده شد. در تابع سوم، اجزای رس و شن به همراه مکش خاک، به‌عنوان متغیر ورودی استفاده شدند. در تابع چهارم جرم مخصوص ظاهری به همراه متغیرهای ورودی تابع سوم، به‌عنوان تخمین‌گر استفاده شد. در تابع پنجم رطوبت ظرفیت مزرع‌ای، جرم مخصوص ظاهری، اجزای رس و شن به همراه مکش خاک به‌عنوان تخمین‌گر مورد استفاده قرار گرفتند. در تابع ششم رطوبت نقطه پژمردگی دائم به همراه متغیرهای تابع پنجم به‌عنوان متغیر ورودی استفاده شدند. در توابع ۷ تا ۱۰ از میانگین (dg) و انحراف معیار (Sg) هندسی قطر ذرات به جای درصد‌های شن و رس خاک در توابع سوم تا ششم استفاده شد تا کارایی استفاده از ویژگی‌های آماری توزیع اندازه ذرات خاک به جای اجزای بافتی در توابع انتقالی شبه‌پیوسته برای پیش بینی SWRC مورد بررسی قرار گیرد. در این تحقیق برای تعیین نمونه‌های آموزش و اعتبارسنجی از روش k-Fold با k=10 استفاده

<sup>۱۴</sup>-Kolmogorov-Smirnov

<sup>۱۵</sup>-Variance inflation factor

شد (Efron & Tibshirani, 1993). برای ایجاد توابع در این تحقیق از دو روش RF و رگرسیون خطی استفاده شد. نرم افزار مورد استفاده برای هر دو روش STATISTICA بود.

### جنگل تصادفی (RF)

جنگل‌های تصادفی یک روش ترکیبی بوده و بر اساس روش درخت رگرسیونی می‌باشد. این روش برای مسائل رگرسیون و طبقه‌بندی در راستای یافتن بهترین عملکرد استفاده می‌شود (Breiman, 2001). فرآیند ساخت مدل در RF مانند دیگر روش‌های درخت رگرسیون است، با این تفاوت که هرس انجام نمی‌دهد (Breiman, 1984). دیگر تفاوت RF با سایر درخت‌های رگرسیونی این است که درخت رگرسیون تنها توسط یک درخت رشد می‌کند، اما الگوریتم RF توسط جنگلی از درختان رشد می‌کند. به عبارت دیگر، در RF بر خلاف درخت رگرسیون، برای هر درخت فقط یک زیر مجموعه از متغیرهای ورودی اعمال می‌شود. تعداد ورودی‌ها در هر درخت و همچنین تعداد درختان در جنگل می‌تواند متفاوت باشد و به مجموعه داده بستگی دارد. در این پژوهش تعداد درختان بهینه با سعی و خطا ایجاد شد. در الگوریتم RF، هر درخت دارای یک نتیجه پیش‌بینی است. در نهایت، مقدار میانگین نتایج درختان به عنوان پیش‌بینی نهایی محاسبه می‌شود (Wiesmeier et al., 2011).

### آماره‌های ارزیابی درستی<sup>۱۶</sup> و قابلیت اعتماد<sup>۱۷</sup> توابع

برای بررسی درستی و قابلیت اعتماد توابع از معیارهای زیر استفاده شد.

$$R^2 = \frac{\left[ N \left( \sum_{i=1}^N Y_i Y_{pi} \right) - \sum_{i=1}^N Y_i \sum_{i=1}^N Y_{pi} \right]}{\sqrt{\left[ N \left( \sum_{i=1}^N Y_i^2 \right) - \left( \sum_{i=1}^N Y_i \right)^2 \right] \times \left[ N \left( \sum_{i=1}^N Y_{pi}^2 \right) - \left( \sum_{i=1}^N Y_{pi} \right)^2 \right]}} \quad [1]$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Y_i - Y_{pi})^2}{N}} \quad [2]$$

$R^2$ : ضریب تبیین، RMSE: مجذور میانگین مربعات خطا،  $Y_i$ : مقدار اندازه‌گیری شده رطوبت،  $Y_{pi}$ : مقدار رطوبت پیش‌بینی شده توسط توابع و  $N$  تعداد کل داده‌های مرحله آموزش یا تعمیم می‌باشد (Wösten et al., 2001). معیار اطلاعات آکایک<sup>۱۸</sup>، AIC (Akaike, 1974):

$$AIC = N \ln \left[ \sum_{i=1}^N (Y_i - Y_{pi})^2 \right] + 2n_p \quad [3]$$

$n_p$  تعداد پارامترهای مدل که باید تخمین زده شوند. هرچه AIC کوچک‌تر باشد کیفیت برازش بالاتر است. اگر انطباق کامل مقادیر مشاهده‌ای و تخمینی اتفاق بیافتد AIC به سمت منفی بی‌نهایت میل خواهد کرد.

## نتایج و بحث

### ویژگی‌های آماری

پراکنش ۱۲۰ نمونه بر روی مثلث بافت خاک در شکل ۱ نشان داده شده است. این شکل نشان می‌دهد که نمونه‌های مورد استفاده در این تحقیق دارای توزیع وسیعی بوده و از ۱۲ کلاس بافتی به جز سه کلاس رس شنی، لوم سیلتی و سیلت بقیه را پوشش داده‌اند. علائم اختصاری و توضیحات مربوط به متغیرها و آماره‌ها در جدول ۱ و ویژگی‌های

<sup>۱۶</sup>-Accuracy

<sup>۱۷</sup>-Reliability

<sup>۱۸</sup>-Akaike's information criterion



آماري همه متغيرها در جدول ۲ نشان داده شده است. غالب متغيرهاي مربوط به بافت و رطوبت خاک داراي دامنه تغييرات بالايي مي باشند که مي تواند قابليت کاربرد توابع ايجاد شده را افزايش دهد. چرا که استفاده از توابع در محدوده داده هاي مورد استفاده براي ايجاد آنها با درستي بيشتري همراه خواهد بود (Ungaro et al., 2005). اعداد جدول ۲ نيز تاييدي بر شکل ۱ است که نشان مي دهد نمونه هاي خاک از کلاس بافتي رسي تا شني توزيع شده اند، پس طبيعي است که دامنه تغييرات ويژگي هاي بافتي خاک وسيع باشد. لازم به ذکر است که جرم مخصوص ظاهري داراي دامنه تغييرات وسيعي نبود. شايد شرايط ساختماني و ماده آلي موجود در خاک موجب اصلاح اثر بافت خاک بر جرم مخصوص ظاهري خاک گرديد. چراکه ساختمان خاک اصلاح کننده اثرات نامطلوبي است که در اثر بافت خاک ايجاد شده است (Brady & Weil, 2010).

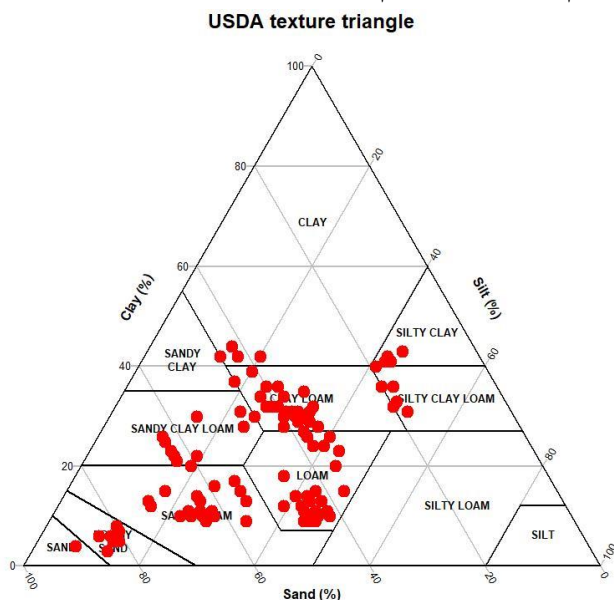
در ايجاد توابع انتقالی شبه پيوسته رطوبت در تمامی مکش ها به طور همزمان به عنوان خروجی وارد مدل شده و تخمین زده می شود. بنابراین، باید بر روی رطوبت در همه مکش ها یا تبدیل یکسان انجام شود و یا اینکه روی هیچ کدام تبدیلی انجام نشود. در این تحقیق با توجه به اینکه رطوبت در همه مکش ها به جز مکش ۵۰۰۰ هکتوپاسکال داراي توزيع نرمال بوده، و رطوبت در مکش مذکور نیز داراي  $p\text{-value}=0.04$  بود که فاصله چندانى با  $0.05$  نداشت و بزرگتر از  $0.01$  بود، لذا تبدیلی بر روی رطوبت ها انجام نشد.

جدول ۱- علائم اختصاری و توضیحات مربوط به متغيرها و آماره ها.

ردیف	علامت	شرح	واحد	نوع تبدیل متغير	p-value <sup>†</sup>
۱	C	مقدار رس	(%)	Normal scores <sup>‡</sup>	>0.15
۲	Si	مقدار سيلت	(%)		
۳	S	مقدار شن	(%)	Normal scores	>0.15
۴	BD	جرم مخصوص ظاهري	$g\ cm^{-3}$	-	0.09
۵	dg	ميانگين هندسي قطر ذرات خاک	mm	Normal scores	>0.15
۶	Sg	انحراف معيار هندسي قطر ذرات خاک	-	Normal scores	>0.15
۷	T0	رطوبت حجمي در مکش ۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۸	T10	رطوبت حجمي در مکش ۱۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۹	T20	رطوبت حجمي در مکش ۲۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۰	T40	رطوبت حجمي در مکش ۴۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۱	T60	رطوبت حجمي در مکش ۶۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۲	T300 (FC)	رطوبت حجمي در مکش ۳۰۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۳	T1000	رطوبت حجمي در مکش ۱۰۰۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۴	T3000	رطوبت حجمي در مکش ۳۰۰۰ هکتوپاسکال	$m^3\ m^{-3}$	-	>0.15
۱۵	T5000	رطوبت حجمي در مکش ۵۰۰۰ هکتوپاسکال	$m^3\ m^{-3}$	-	0.04
۱۶	T15000 (PWP)	رطوبت حجمي در مکش ۱۵۰۰۰ هکتوپاسکال	$m^3\ m^{-3}$	-	0.07
۱۷	AIC	معیار اطلاعات آکايک	وابسته به متغير		
۱۸	R <sup>2</sup>	ضريب تبیین	-		
۱۹	RMSE	مجذور میانگین مربعات خطا	$m^3\ m^{-3}$		
۲۰	h	مکش خاک	hPa		

†. در این ستون p-value آزمون نرمال بودن داده ها با استفاده از آزمون کلموگروف-اسمیرنوف برای متغيرها، یا در صورت لزوم برای متغيرهاي تبدیل شده، آورده شده است. نوع تبدیل لازم نیز در ستون مربوطه ذکر شده است. hPa: هکتوپاسکال.

‡. تبدیل با روش نرمال اسکور در نرم افزار مینی تب انجام شد.



شکل ۱- توزیع بافت خاک نمونه‌های مطالعه شده بر روی مثلث بافت خاک USDA.

جدول ۲- ویژگی‌های آماری متغیرهای مورد مطالعه در این تحقیق.

متغیر - آماره	T0	BD (g cm <sup>-3</sup> )	Sg	dg (mm)	سیلت (%)	شن (%)	رس (%)	
میانگین	۳۴/۰	۴۴/۱	۰/۱۴	۱۰/۰	۴/۳۰	۲/۴۸	۵/۲۱	
ضریب تغییرات (%)	۲۰	۵	۳۱	۷۷	۳۸	۳۵	۵۳	
حداقل	۱۶/۰	۲۸/۱	۰/۵	۰/۱۰	۶/۶	۵/۱۳	۹/۲	
حداکثر	۵۰/۰	۵۸/۱	۳/۲۵	۲۸/۰	۷/۵۱	۲/۸۹	۰/۴۴	
میانه	۳۴/۰	۴۳/۱	۸/۱۲	۰/۸	۹/۳۹	۵/۴۴	۳/۲۰	
متغیر - آماره	T20	T40	T60	T300 (FC)	T1000	T3000	T5000	T15000 (PWP)
میانگین	۳۲/۰	۲۹/۰	۲۷/۰	۲۲/۰	۱۹/۰	۱۷/۰	۱۰/۰	۰/۸
ضریب تغییرات (%)	۲۲	۲۳	۲۵	۳۰	۳۴	۳۴	۳۵	۳۹
حداقل	۱۴/۰	۱۲/۰	۱۰/۰	۰/۸	۰/۵	۰/۴	۰/۲	۰/۱
حداکثر	۴۸/۰	۴۴/۰	۴۲/۰	۳۷/۰	۳۵/۰	۳۲/۰	۱۹/۰	۱۶/۰
میانه	۳۲/۰	۲۹/۰	۲۷/۰	۲۱/۰	۱۹/۰	۱۶/۰	۱۱/۰	۰/۸

† کليه علائم اختصاری در جدول ۱ معرفی شده است.

### همبستگی پیرسون بین متغیرهای ورودی و خروجی

همبستگی بین متغیرهای ورودی و خروجی در جدول ۲ نشان داده شده است. همه متغیرهای ورودی (بافت، درصد رس و شن، dg، Sg و جرم مخصوص ظاهری) همبستگی معناداری با رطوبت خاک در تمام مکش‌ها داشتند (به جز همبستگی بافت، درصد شن و dg با رطوبت در مکش ۱۵۰۰۰ هکتوپاسکال). بافت خاک و ساختمان خاک از مهمترین عوامل مؤثر بر قابلیت خاک برای نگهداشت آب در خاک می‌باشند (Dexter et al., 2008; Hillel, 2004; Nemes et al., 2006). بافت خاک، درصد شن و dg دارای همبستگی منفی و معنی‌دار با مقادیر رطوبت در تمام مکش‌ها به جز مکش ۱۵۰۰۰ هکتوپاسکال بودند. با توجه به اینکه در کدهای به کلاس بافت خاک، کلاس‌های بافتی از ریز به درشت مرتب

شدند و کد ۱ به بافت رس و به ترتیب کد ۱۲ به کلاس شن اختصاص داده شد، پس طبیعی است که رفتار آن شبیه به ذرات درشت بوده و با افزایش آن مقدار رطوبت کاهش یابد. dg نیز میانگینی از اندازه ذرات خاک است که با بزرگتر شدن ذرات خاک قابلیت نگهداشت آب کاهش می‌یابد. همبستگی منفی و معنادار درصد شن و dg با رطوبت اشباع خاک توسط Rastgou et al. (2020) نیز گزارش شده است.

جدول ۳- همبستگی پیرسون بین متغیرهای ورودی (ستون سمت چپ) و خروجی (ردیف بالایی که رطوبت حجمی اندازه‌گیری شده در مکش‌های مختلف می‌باشند)، (N = 120).

متغیر	T15000 (PWP)	T5000	T3000	T1000	T300 (FC)	T60	T40	T20	T10	T0 <sup>†</sup>
بافت	-./۱۷۱	-./۳۷۸**	-./۳۲۹**	-./۴۴۳**	-./۴۳۰**	-./۴۲۴**	-./۴۱۰**	-./۴۰۲**	-./۳۸۲**	-./۳۸۵**
رس	./۲۶۵**	./۴۵۰**	./۴۰۹**	./۴۷۷**	./۴۶۷**	./۴۷۱**	./۴۵۶**	./۴۴۳**	./۴۲۹**	./۴۳۰**
شن	-./۰۷	-./۱۸۵*	-./۳۴۰**	-./۴۰۰**	-./۳۷۲**	-./۳۷۸**	-./۳۶۴**	-./۳۵۹**	-./۳۵۲**	-./۳۶۹**
سیلت	-./۱۱	-./۱۳	./۲۰۱*	./۱۷	./۱۵	./۱۷	./۱۶	./۱۶	./۱۸	./۲۰۱*
dg	-./۱۱	-./۲۷۶**	-./۳۷۷**	-./۴۴۸**	-./۴۲۳**	-./۴۱۷**	-./۳۹۷**	-./۳۸۶**	-./۳۷۴**	-./۳۸۶**
Sg	./۳۳۲**	./۵۰۷**	./۳۸۰**	./۴۶۴**	./۴۷۹**	./۵۰۷**	./۵۱۲**	./۵۱۱**	./۴۹۴**	./۴۷۹**
BD	-./۲۴۹**	-./۳۷۳**	-./۵۹۶**	-./۶۱۲**	-./۶۰۷**	-./۵۶۶**	-./۵۴۷**	-./۵۲۵**	-./۴۹۷**	-./۴۸۹**

†: کلیه علائم اختصاری در جدول ۱ معرفی شده است.

دو نکته قابل توجه در مورد جدول همبستگی وجود دارد. اول اینکه همبستگی همه متغیرهای ورودی با رطوبت در مکش ۱۵۰۰۰ هکتوپا سکال کمترین مقدار را در مقایسه با رطوبت در سایر مکش‌ها داشته است. به گونه‌ای که این همبستگی با بافت، درصد شن و dg معنادار نگردید. تأثیر بافت خاک بر نگهداشت آب در مکش‌های بالا کاملاً تأیید شده است (Brady & Weil, 2010; Hillel, 1998)، با این وجود متغیرهای مربوط به بافت خاک همبستگی بالایی با رطوبت در مکش ۱۵۰۰۰ هکتوپا سکال ندا شدند. بنابراین، عواملی به جز بافت خاک نیز بر نگهداشت آب در مکش‌های بالا به ویژه رطوبت در مکش ۱۵۰۰۰ هکتوپا سکال مؤثر هستند. این عوامل می‌توانند مقدار ماده آلی (Vereecken et al., 1989)، نوع کانی‌های خاک (Williams et al., 1983)، مقدار رس‌های قابل انبساط (Williams et al., 1983)، ظرفیت تبادل کاتیونی، سطح ویژه (Warkentin, 1972)، اکسیدهای آهن و آلومینیم و کربنات کلسیم و منیزیم (Ebrahimi et al., 2016) باشند. همچنین با دقت در جدول ۲، مشخص می‌شود که ضریب تغییرات رطوبت خاک با افزایش مکش خاک افزایش می‌یابد، شاید این نکته یکی دیگر از علت‌های کاهش همبستگی متغیرهای ورودی با رطوبت در مکش‌های بالا باشد که منجر به غیرمعنادار شدن همبستگی رطوبت در مکش ۱۵۰۰۰ هکتوپا سکال با بافت، درصد شن و dg گردید. البته افزایش ضریب تغییرات با افزایش مکش، با توجه به تنوع بافتی، بدلیل تأثیر بافت خاک بر تغییرات رطوبتی در مکش‌های بالا نیز می‌باشد.

نکته دوم جدول همبستگی این است که همبستگی تقریباً تمام ۷ متغیر با رطوبت خاک، در مکش‌های میانی بیشترین مقدار را دارد، و در مکش‌های بالاتر و پایین‌تر کمتر است. همبستگی کمتر متغیرها در مکش‌های بالا، جلوتر بحث شد؛ اما همبستگی کم در مکش‌های پایین نیز تأثیر عواملی غیر از متغیرهای ورودی در این تحقیق بر نگهداشت رطوبت را نشان می‌دهد. شاید عواملی مانند بقایای گیاهی، ویژگی‌های ساختمانی خاک و خاکدانه‌های درشت (Bayat et al., 2013a; Bayat et al., 2013b) جزء متغیرهای مؤثر باشند. یکی از مسئله‌های حائز اهمیت دو نکته فوق این است که برای تخمین رطوبت در محدوده ظرفیت مزرعه‌ای، متغیرهای بافت خاک به همراه جرم مخصوص ظاهری کفایت می‌کند.

## نتایج توابع انتقالی ایجاد شده

نتایج توابع انتقالی شبه‌پیوسته ایجاد شده برای تخمین رطوبت با روش‌های رگرسیون خطی و RF به‌ترتیب در جداول ۴ و ۵ نشان داده شده است. در جداول ۴ و ۵ هر تابع با یک تابعی مقایسه شده است که تابع مقایسه شونده در ستون مربوطه نشان داده شده است. مقایسه معناداری بر اساس آماره AIC انجام شده است. در صورت وجود تفاوت معنادار سلول مربوط به آکایک هایلایت شده است.

در این تحقیق از روش k-fold با  $k=10$  برای تقسیم‌بندی داده‌های آموزش و اعتبارسنجی استفاده شد. یعنی داده‌ها به ۱۰ قسمت تقسیم شد و مدل ۱۰ بار اجرا شد. در هر اجرا ۹ قسمت داده‌ها برای آموزش و ۱ قسمت برای معتبرسازی استفاده شد، بگونه‌ای که تمام نمونه‌ها، هر یک یک‌بار برای معتبرسازی استفاده شدند. پس از هر اجرا سه آماره آکایک،  $R^2$  و RMSE برای هر دو مرحله به‌صورت جداگانه محاسبه شد. در نهایت میانگین آماره‌ها در ۱۰ اجرا به‌عنوان آماره نهایی گزارش شد. بنابراین، احتمالاً این نحوه اجرای مدل باعث نزدیک شدن نتایج آماره‌های  $R^2$  و RMSE در دو مرحله گردید که اعتبار بالای مدل‌های ایجاد شده را نشان می‌دهد؛ اما علت تفاوت بالای آماره آکایک در دو مرحله آموزش و معتبرسازی، فرمول مربوطه می‌باشد. چون در فرمول این آماره از مجموع مربعات خطا استفاده می‌کند که مقدار آن تحت تأثیر تعداد نمونه است. همچنین پس از آنکه از مجموع مربعات خطا لگاریتم طبیعی گرفته می‌شود، عدد حاصل در تعداد مشاهدات ضرب می‌گردد. پس هرچه تعداد مشاهدات بیشتر باشد آکایک بیشتر است. چون در این تحقیق تعداد نمونه‌های آموزش در هر اجرا ۹ برابر تعداد نمونه‌های معتبرسازی است، بنابراین، تفاوت زیادی بین آن‌ها وجود دارد.

جدول ۴- نتایج تخمین SWRC با روش رگرسیون خطی برای ده تابع انتقالی شبه‌پیوسته.

تابع	متغیرهای ورودی	آموزش			معتبرسازی			مدل مقایسه شونده
		RMSE ( $\text{cm}^3\text{cm}^{-3}$ )	$R^2$	AIC	RMSE ( $\text{cm}^3\text{cm}^{-3}$ )	$R^2$	AIC	
۱	$h^\dagger$	۰/۰۷۱	۰/۶۷۵	۲۱۴۳	۰/۰۷۱	۰/۶۷۴	-۴۳	
۲	hTexture	۰/۰۶۷	۰/۷۱۰	۲۰۱۱	۰/۰۶۷	۰/۷۱۱	-۵۸	۱
۳	hSC	۰/۰۶۶	۰/۷۱۹	۱۹۷۱	۰/۰۶۶	۰/۷۱۸	-۶۲	۲
۴	hSCBD	۰/۰۶۲	۰/۷۵۶	۱۸۰۱	۰/۰۶۲	۰/۷۵۳	-۸۲	۳
۵	hSCBDFC	۰/۰۴۷	۰/۸۶۱	۱۱۳۸	۰/۰۴۷	۰/۸۶۲	-۱۵۴	۴
۶	hSCBDFCPWP	۰/۰۴۶	۰/۸۶۴	۱۱۰۹	۰/۰۴۶	۰/۸۶۶	-۱۵۶	۵
۷	hdgSg	۰/۰۶۵	۰/۷۳۲	۱۹۱۶	۰/۰۶۵	۰/۷۳۲	-۶۸	۳
۸	hdgSgBD	۰/۰۶۱	۰/۷۶۱	۱۷۷۹	۰/۰۶۱	۰/۷۶۱	-۸۳	۴
۹	hdgSgBDFC	۰/۰۴۷	۰/۸۶۱	۱۱۳۴	۰/۰۴۷	۰/۸۶۱	-۱۵۴	۵
۱۰	hdgSgBDFCPWP	۰/۰۴۶	۰/۸۶۴	۱۱۱۰	۰/۰۴۶	۰/۸۶۴	-۱۵۷	۶

$^\dagger$  کلیمه علائم اختصاری در جدول ۱ معرفی شده است.

دو نکته کلیدی برای بهبود درستی تخمین‌ها از جمله تخمین رطوبت خاک، در هر مسئله‌ای وجود دارد. نکته اول انتخاب بهترین مجموعه متغیرهای ورودی است. یعنی یافتن مجموعه متغیرهایی که استفاده از آن‌ها به‌عنوان تخمین‌گر منجر به بالاترین درستی تخمین متغیر خروجی گردد (Vereecken et al., 2010). نکته دوم یافتن بهترین روش برای ایجاد توابع انتقالی است. در واقع یکی از روش‌های مهم جهت بهبود درستی و قابلیت اطمینان توابع، استفاده از روش‌های

داده کاوی<sup>۱۹</sup> است (Wösten et al., 2001). با توجه به نتایج گزارش شده در منابع، هنوز در مورد بهترین مجموعه متغیرهای ورودی و بهترین روش تخمین رطوبت خاک بین محققان اختلاف نظر وجود دارد.

جدول ۵- نتایج تخمین SWRC با روش جنگل تصادفی برای ده تابع انتقالی شبه‌پیوسته.

تابع	متغیرهای ورودی	آموزش			معتبرسازی			مدل مقایسه شونده
		RMSE (cm <sup>3</sup> cm <sup>-3</sup> )	R <sup>2</sup>	AIC	RMSE (cm <sup>3</sup> cm <sup>-3</sup> )	R <sup>2</sup>	AIC	
۱	h <sup>†</sup>							
۲	hTexture	۰/۰۵۱	۰/۸۳۲	۱۳۶۰	۰/۰۵۲	۰/۸۲۸	-۱۲۶	۱
۳	hSC	۰/۰۳۸	۰/۹۰۶	۶۷۷	۰/۰۴۲	۰/۸۸۹	-۱۸۵	۲
۴	hSCBD	۰/۰۳۳	۰/۹۲۹	۳۴۴	۰/۰۳۷	۰/۹۱۴	-۲۱۹	۳
۵	hSCBDFC	۰/۰۳۲	۰/۹۳۵	۲۴۰	۰/۰۳۵	۰/۹۲۱	-۲۲۸	۴
۶	hSCBDFCPWP	۰/۰۳۲	۰/۹۳۵	۲۳۷	۰/۰۳۶	۰/۹۲۰	-۲۲۷	۵
۷	hdgSg	۰/۰۳۷	۰/۹۱۴	۵۷۴	۰/۰۴۰	۰/۸۹۸	-۱۹۵	۳
۸	hdgSgBD	۰/۰۳۳	۰/۹۲۹	۳۳۱	۰/۰۳۷	۰/۹۱۴	-۲۱۸	۴
۹	hdgSgBDFC	۰/۰۳۲	۰/۹۳۵	۲۳۷	۰/۰۳۶	۰/۹۲۰	-۲۲۷	۵
۱۰	hdgSgBDFCPWP	۰/۰۳۲	۰/۹۳۵	۲۳۵	۰/۰۳۶	۰/۹۲۰	-۲۲۶	۶

†: کلیه علائم اختصاری در جدول ۱ معرفی شده است.

استفاده از مکش خاک به‌عنوان تنها متغیر ورودی برای تخمین رطوبت در مکش‌های مختلف در روش RF نتیجه بخش نبود و مدلی ایجاد نشد، چراکه استفاده از یک متغیر در این روش باعث بروز خطا در ماتریس محاسبات شد؛ اما در روش رگرسیون خطی، مدلی با نتایج قابل‌قبول ایجاد شد (جدول ۴). نتایج این تابع بهتر از نتایج برخی توابع انتقالی شبه‌پیوسته‌ای است که توسط (Haghverdi et al. (2018) با استفاده از متغیرهای ورودی بیشتر به‌دست آمد. همچنین نتایج این تحقیق قابل مقایسه با نتایج برخی توابع انتقالی شبه‌پیوسته‌ای است که توسط (Haghverdi et al. (2014) با متغیرهای ورودی رس-سیلت-شن-جرم مخصوص ظاهری-ماده آلی به‌دست آمد. بنابراین، هر چند نویسندگان بر این باورند که در صورت در دسترس بودن متغیرهای ورودی اضافی بهتر است از توابع دیگری برای به‌دست آوردن رطوبت خاک استفاده شود، ولی در صورت عدم دسترسی به اطلاعات مازاد، این تابع می‌تواند یک تخمین حدودی از رطوبت را در اختیار کاربران قرار دهد.

در تابع دوم بافت خاک علاوه بر مکش خاک به‌عنوان تخمین‌گر وارد مدل‌ها شد. استفاده از بافت خاک در هر دو روش RF و رگرسیون خطی موجب بهبود معنادار درستی تخمین‌ها در هر دو مرحله آموزش و معتبرسازی گردید. نتایج این مدل با هر دو روش RF و رگرسیون خطی بهتر از نتایج (Schaap et al. (2001) است که با توسعه نرم افزار رزتا بر اساس شبکه‌های عصبی مصنوعی برای تخمین پارامتریک SWRC با استفاده از بافت خاک به‌عنوان تخمین‌گر به‌دست آوردند. (Schaap et al. (2001) مجذور میانگین مربعات خطا برابر با  $0/078 \text{ cm}^3\text{cm}^{-3}$  را گزارش کرد که نسبت به مقادیر به‌دست آمده در این تحقیق برای تابع شبه‌پیوسته دوم با هر دو روش (جدول ۴ و ۵) بیشتر بود.

در تابع سوم از اجزای بافت خاک (درصد رس و شن) در کنار مکش آب خاک به‌عنوان تخمین‌گر برای تخمین رطوبت استفاده شد. مقایسه درستی تخمین در این تابع با تابع دوم نشان داد که در هر دو روش RF و رگرسیون خطی استفاده

از اجزای بافت موجب بهبود تخمین رطوبت در هر دو مرحله آموزش و معتبرسازی نسبت به تابع دوم گردید، و این بهبود نتایج به‌جز مرحله آموزش روش رگرسیون در بقیه موارد معنادار بود (جداول ۴ و ۵). نتایج به‌دست آمده در این تابع شبه‌پیوسته با هر دو روش RF و رگرسیون خطی نسبت به نتایج تابع شبه‌پیوسته گزارش شده توسط (Haghverdi et al., 2018) با متغیرهای ورودی مشابه که با روش شبکه‌های عصبی مصنوعی ایجاد شده بودند بهتر بود. آن‌ها مجذور میانگین مربعات خطا برابر با  $0.129 \text{ cm}^3 \text{ cm}^{-3}$  را گزارش کردند که نسبت به مقادیر به‌دست آمده در این تحقیق برای تابع شبه‌پیوسته سوم با هر دو روش (جداول ۴ و ۵) بیشتر بود. (Rawls et al., 2003) و (Schaap et al., 2001) گزارش کردند که استفاده از اجزای بافت خاک باعث بهبود درستی تخمین‌ها نسبت به استفاده از کلاس بافت به‌عنوان تخمین‌گر می‌شود. در واقع هرچقدر اطلاعات بیشتر و جزئی‌تری برای تخمین استفاده شود، درستی تخمین‌ها نیز بهبود خواهد یافت.

در تابع چهارم استفاده از جرم مخصوص ظاهری به‌عنوان تخمین‌گر به‌همراه مکش آب خاک و درصد رس و شن موجب بهبود معنادار درستی تخمین‌ها در هر دو مرحله آموزش و معتبرسازی هر دو روش RF و رگرسیون خطی گردید. متغیر BD در مطالعات زیادی به‌عنوان شاخصی از ساختمان خاک برای بهبود تخمین منحنی نگهداری آب خاک در توابع انتقالی نقطه‌ای و پارامتریک مورد استفاده واقع شده است (Bayat et al., 2011; Pachepsky et al., 1998). (Bayat et al., 2013b) گزارش کردند که BD می‌تواند باعث بهبود درستی و قابلیت اعتماد توابع تخمین رطوبت خاک گردد. چراکه نگهداشت رطوبت در مکش‌های ماتریک پایین تحت تأثیر ساختمان خاک است (Pachepsky et al., 2006). (Touil et al., 2016) نیز گزارش کردند که BD تحت تأثیر ساختمان بوده و یک اندازه‌گیری غیرمستقیم از فضای منافذ خاک است.

در تابع پنجم از رطوبت در ظرفیت مزرعه‌ای به‌همراه متغیرهای تابع چهارم (مکش آب خاک، درصد رس و شن و جرم مخصوص ظاهری) به‌عنوان تخمین‌گر برای تخمین رطوبت خاک از طریق توابع انتقالی شبه‌پیوسته استفاده شد. استفاده از رطوبت در ظرفیت مزرعه‌ای در هر دو مرحله آموزش و معتبرسازی هر دو روش رگرسیون خطی و RF موجب بهبود درستی تخمین‌ها شد که این بهبود به‌جز مرحله معتبرسازی روش RF در بقیه موارد معنادار بود. علت این نتیجه این است که رطوبت در ظرفیت مزرعه‌ای یک نقطه از SWRC بوده و اطلاعات مستقیمی از آن را وارد مدل‌ها می‌کند. محققان زیادی از جمله (Schaap et al., 2001) و (Ahuja et al., 1989) گزارش کردند که استفاده از یک یا دو نقطه از SWRC می‌تواند باعث بهبود قابل‌توجه درستی تخمین رطوبت خاک شود.

در تابع ششم استفاده از PWP به‌عنوان تخمین‌گر به‌همراه متغیرهای ورودی تابع پنجم بهبود معناداری در درستی تخمین‌ها در هیچ مرحله از هیچ کدام از دو روش ایجاد نکرد. در نتایج گزارش شده توسط (Schaap et al., 2001) نیز استفاده از PWP بهبود قابل‌توجهی ایجاد نکرده و فقط مجذور میانگین مربعات خطا را از  $0.47$  به  $0.44 \text{ cm}^3 \text{ cm}^{-3}$  کاهش داده است. نتیجه مشابهی توسط (Rastgou et al., 2022) گزارش شده است. رطوبت‌های FC و PWP در تحقیقات زیادی به‌عنوان تخمین‌گر برای تخمین‌های نقطه‌ای و پارامتریک SWRC استفاده شده‌اند (Børgesen & Schaap, 2005; Touil et al., 2016; Twarakavi et al., 2009). شاید علت این نتیجه را بتوان در جداول ۲ و ۳ یافت. ضریب تغییرات بالای رطوبت PWP (جدول ۲) باعث ورود عدم قطعیت بالا به توابع گردیده و از طرف دیگر همبستگی پایین آن با دیگر متغیرهای خاک (جدول ۳) می‌تواند نشانه‌ای از تأثیرگذاری کمتر این متغیر در بهبود تخمین‌ها باشد.

توابع شبه‌پیوسته ۷ تا ۱۰ به‌ترتیب معادل با توابع شبه‌پیوسته ۳ تا ۶ می‌باشند، با این تفاوت که به‌جای درصد رس و شن در توابع شبه‌پیوسته ۳ تا ۶، از  $d_g$  و  $S_g$  به‌عنوان تخمین‌گر در توابع شبه‌پیوسته ۷ تا ۱۰ برای پیش‌بینی SWRC، استفاده شده است. هدف از ایجاد توابع شبه‌پیوسته ۷ تا ۱۰ این بود که بررسی شود آیا استفاده از اطلاعات آماری

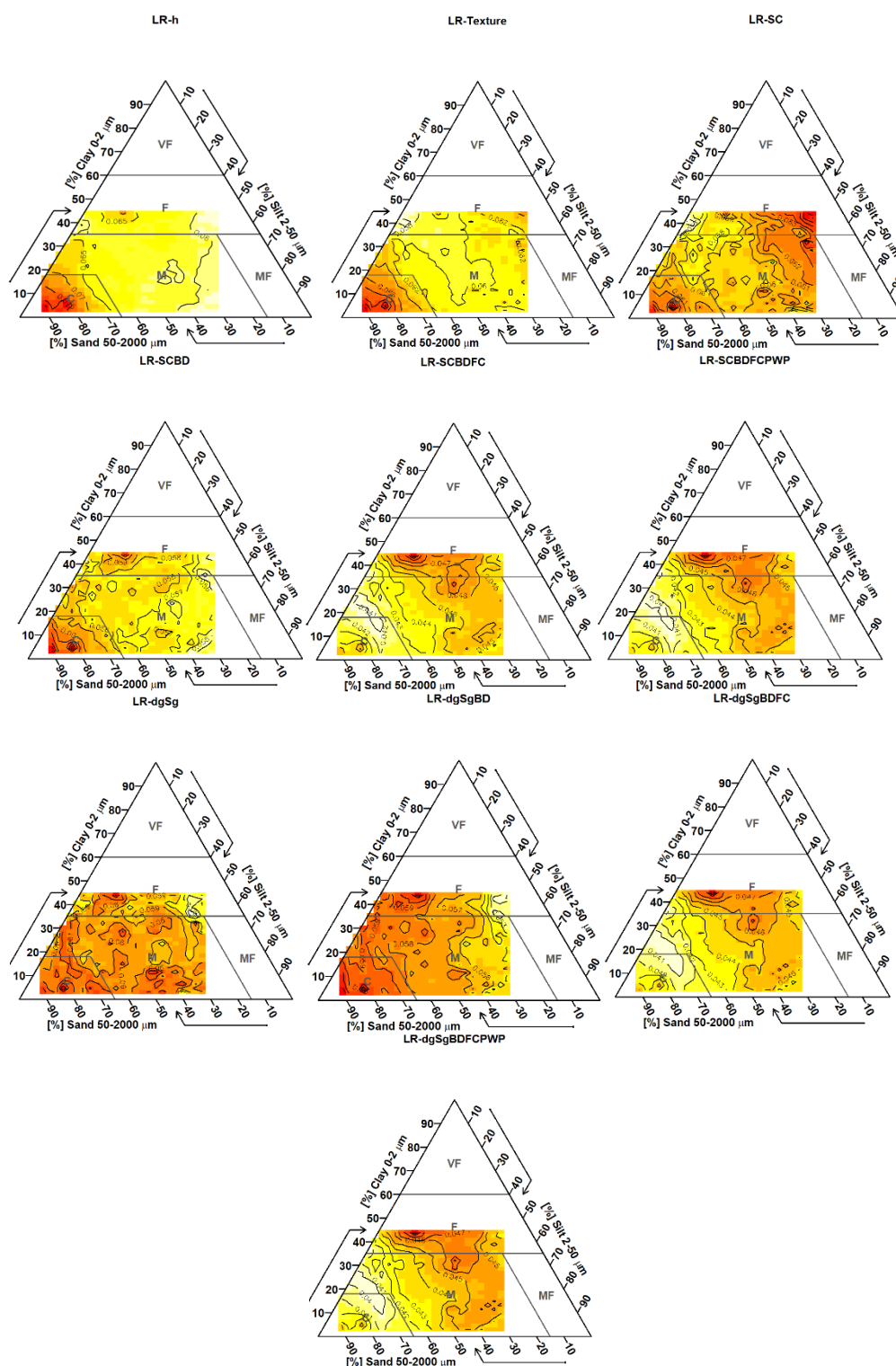
توزیع اندازه ذرات به جای اجزای بافت خاک می‌تواند بهبودی در درستی تخمین SWRC از طریق توابع شبه‌پیوسته ایجاد کند یا خیر؟

در تابع شبه‌پیوسته هفتم استفاده از  $d_g$  و  $S_g$  به‌عنوان تخمین‌گر به‌جای درصد رس و شن در تابع شبه‌پیوسته سوم، باعث بهبود تخمین‌ها در هر دو مرحله آموزش و معتبرسازی هر دو روش رگرسیون خطی و RF گردید که این بهبود به‌جز مرحله آموزش روش رگرسیون خطی در بقیه موارد معنادار بود (جداول ۴ و ۵)؛ ولی در توابع شبه‌پیوسته ۸ تا ۱۰ به‌ترتیب در مقایسه با توابع شبه‌پیوسته ۴ تا ۶ بهبود معناداری در درستی تخمین‌ها در هیچ مرحله از هیچ کدام از دو روش مشاهده نشد. یعنی استفاده از  $d_g$  و  $S_g$  به‌عنوان تخمین‌گر به‌جای درصد رس و شن در توابع شبه‌پیوسته مزیت خاصی نداشت. نتایج گزارش شده توسط Rastgou et al. (2022) بر خلاف نتایج این تحقیق است، هر چند بهبود ایجاد شده در اثر استفاده از  $d_g$  و  $S_g$  به‌عنوان تخمین‌گر به‌جای درصد رس و شن در توابع پارامتریک آن‌ها هم ناچیز بود. شاید علت تفاوت در نتایج، در نوع توابع ایجاد شده در دو تحقیق باشد. چراکه Rastgou et al. (2022) از طریق توابع پارامتریک SWRC را تخمین زدند، ولی در این تحقیق تخمین‌ها از طریق توابع شبه‌پیوسته انجام شد. بر اساس مطالعات برخی محققین مانند Tomasella et al. (2003) و Rawls et al. (1991) ذرات شن بر نگر داشت آب خاک در مکش‌های پایین و نزدیک اشباع و ذرات رس بر نگر داشت آب خاک در مکش‌های بالا و نزدیک PWP تأثیر بیشتری دارند. بنابراین، احتمالاً با توجه به اینکه در توابع انتقالی شبه‌پیوسته همبستگی متغیرهای ورودی به‌طور مستقیم با رطوبت خاک بر درستی تخمین‌ها موثر است، و با در نظر گرفتن اثر مستقیم اجزای بافت خاک بر SWRC، لذا استفاده از  $d_g$  و  $S_g$  به‌عنوان تخمین‌گر به‌جای درصد رس و شن نتوانست بهبود معناداری در درستی تخمین‌ها ایجاد کند.

### توزیع RMSE روی مثلث بافت خاک

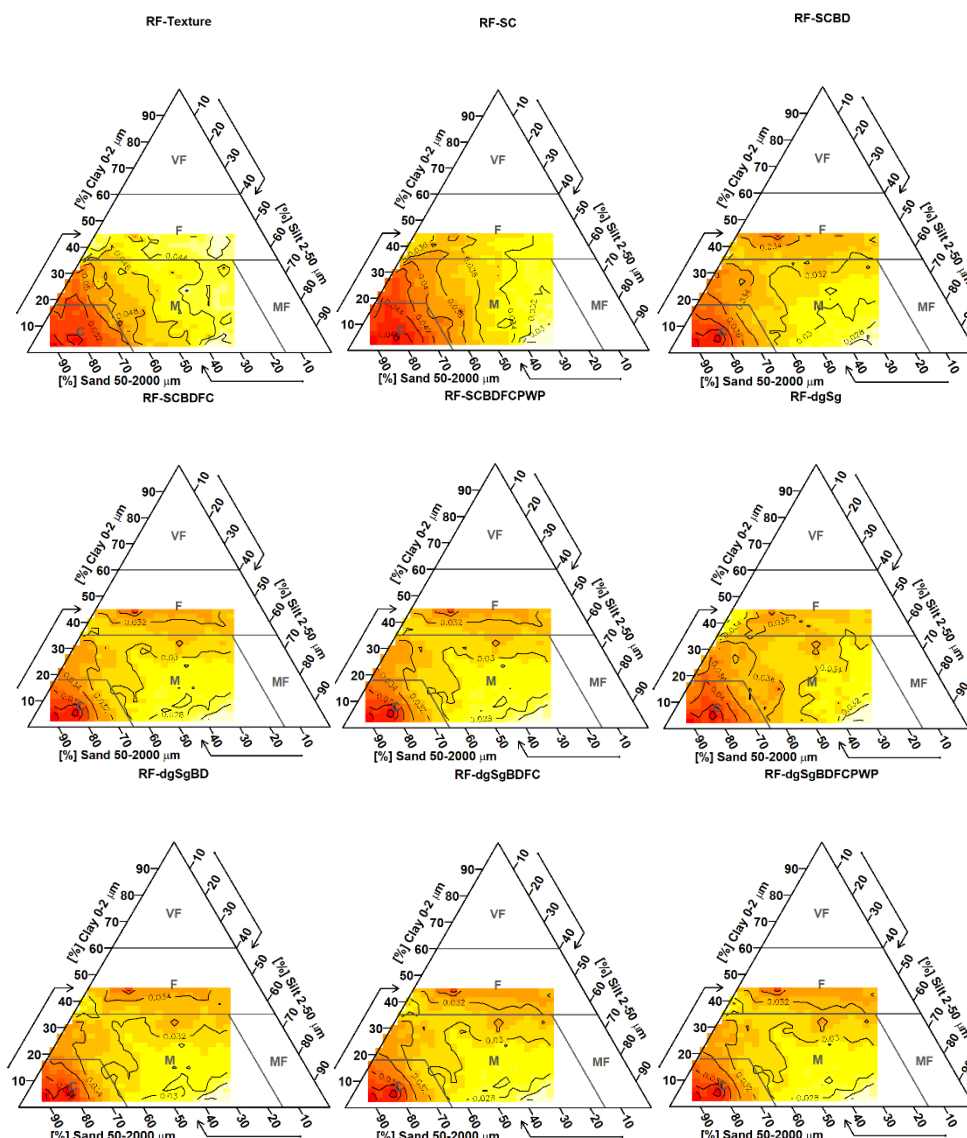
در شکل‌های ۲ و ۳ توزیع RMSE در مرحله معتبر سازی به‌ترتیب برای روش‌های رگرسیون خطی و RF بر روی مثلث بافت خاک نشان داده شده است. هدف از آوردن این آنالیزها این است که بررسی شود آیا درستی تخمین دو روش و در توابع متفاوت وابسته به کلاس بافت خاک است یا خیر؟ و اینکه آیا توزیع خطای تخمین در روش‌ها و توابع مختلف از الگوی مشابهی تبعیت می‌کند یا خیر؟

نتایج نشان داد که توزیع خطا بر روی مثلث بافت خاک در روش رگرسیون خطی هیچ وابستگی به بافت خاک ندارد. چراکه در توابع ۱، ۲، ۴، ۷ و ۸ بیشترین مقدار خطا در بافت درشت به‌دست آمد (شکل ۲)، ولی در توابع ۵، ۶، ۹ و ۱۰ کمترین مقدار خطا مربوط به کلاس بافتی درشت بود. در هر دو روش توابعی که متغیرهای ورودی مشابه دارند، دارای الگوی توزیع خطای مشابه بر روی مثلث بافت خاک می‌باشند. مثلاً در روش رگرسیون، توابعی که در آن‌ها از مقدار رطوبت در ظرفیت مزرعه‌ای یا نقطه پژمردگی دائم به‌عنوان تخمین‌گر استفاده شده است، دارای الگوی توزیع خطای مشابه بر روی مثلث بافت خاک می‌باشند (شکل ۲) و در آن‌ها، کمترین خطا در بافت درشت بوده، و با ریزتر شدن بافت خاک میزان خطا افزایش می‌یابد. در روش RF در توابعی که از نقاط رطوبتی به‌عنوان تخمین‌گر استفاده کرده‌اند، بیشترین خطا در بافت درشت بوده، کمترین خطا مربوط به بافت متوسط و در بافت‌های ریزتر میزان خطا متوسط است (شکل ۳). بنابراین، توزیع خطا بر روی مثلث بافت خاک وابسته به نوع متغیرهای ورودی و روش ایجاد توابع است. در واقع توانایی دو روش RF، LR در استخراج الگوی روابط بین متغیرهای ورودی و خروجی، با تغییر متغیرهای ورودی در بافت‌های مختلف تغییر، می‌کند. (Pachepsky & Rawls (2004) نیز گزارش کردند که، به‌علت ناهمگن بودن توزیع خطا بر روی مثلث بافت خاک نمی‌توان وابستگی خطا به بافت خاک در تخمین خصوصیات هیدرولیکی را تأیید نمود.



شکل ۲- نمایش توزیع مجذور میانگین مربعات خطای مرحله معبرسازی روش رگرسیون خطی (LR) بر روی مثلث بافت خاک برای ده تابع شبه پیوسته. در داخل مثلثها VF: خیلی ریز، F: ریز، M: متوسط، MF: متوسط ریز و C: درشت. علائم اختصاری مذکور در بالای مثلثها در جدول ۱ معرفی شده است. مقدار RMSE بر روی خطوط هم مقدار در داخل شکلها نوشته شده است. در تمام شکلها کم رنگی و پررنگی به ترتیب نشان دهنده مقادیر کمتر و بیشتر RMSE می باشد.





شکل ۳- نمایش توزیع مجذور میانگین مربعات خطای مرحله معتبرسازی روش جنگل تصادفی (RF) بر روی مثلث بافت خاک برای نه تابع شبه‌پیوسته. در داخل مثلث‌ها VF: خیلی ریز، F: ریز، M: متوسط، MF: متوسط ریز و C: درشت. علائم اختصاری مذکور در بالای مثلث‌ها در جدول ۱ معرفی شده است. مقدار RMSE بر روی خطوط هم‌مقدار در داخل شکل‌ها نوشته شده است. در تمام شکل‌ها کم‌رنگی و پررنگی به ترتیب نشان‌دهنده مقادیر کمتر و بیشتر RMSE می‌باشد.

### مقایسه درستی تخمین‌های دو روش RF و LR

نتایج مقایسه درستی دو روش در دو مرحله آموزش و معتبرسازی بر اساس آماره آکایک در جدول ۶ آمده است. تفاوت‌های معنادار بین دو روش با هایلایت نشان داده شده است. در تمام ۹ تابع شبه‌پیوسته که توسط هر دو روش ایجاد شد، درستی تخمین‌ها در هر دو مرحله آموزش و معتبرسازی در روش RF به‌طور معنادار و قابل‌توجهی بیشتر از رگرسیون خطی بود. تا کنون از روش RF برای ایجاد توابع انتقالی شبه‌پیوسته استفاده نشده است. هر چند به‌طور کلی نیز از این روش در علوم خاک کمتر استفاده شده است. برتری روش RF بر برخی روش‌ها توسط محققان مختلفی گزارش شده است. (Araya & Ghezzehei (2019) از چهار روش SVM، k-NN، RF و درخت رگرسیون تقویت شده (BRT) جهت تخمین هدایت هیدرولیکی اشباع استفاده کرده و گزارش کردند که درستی توابع ایجاد شده توسط روش‌های

BRT و RF نسبت به دیگر روش‌ها بالاتر بود. (Gunarathna et al. (2019a) با مقایسه عملکرد سه روش ANN، k-NN و RF در ایجاد توابع نقطه‌ای تخمین رطوبت خاک، درستی بالاتر روش RF را گزارش کردند. (Rastgou (2020) درستی بالاتر روش RF نسبت به رگرسیون خطی را در تخمین نقطه‌ای و پارامتریک SWRC گزارش کرد. نتیجه مشابهی توسط Jamshidi et al. (2019) در ارزیابی روش‌های نقشه‌برداری رقومی در تهیه نقشه سه‌بعدی کربن آلی خاک گزارش شده است.

برتری RF نسبت به رگرسیون خطی دارای علت‌های متفاوتی می‌باشد. روش RF یک روش گروهی است و برای تخمین مکرر متغیر خروجی، چندین الگوریتم درختی را ترکیب می‌کند. جنگل تصادفی یک روش مدل‌سازی قوی بوده و می‌تواند با برقراری ارتباط غیرخطی بین متغیرهای ورودی و خروجی الگوهای پیچیده بین آن‌ها را یاد بگیرد. زمانیکه تعداد نمونه‌ها در مقایسه با تعداد تخمین‌گرها نسبتاً کمتر باشد، عملکرد RF بهتر خواهد شد (Svetnik et al., 2003). از طرف دیگر درختان RF، بدون هرس کردن رشد کرده و در این روش آموزش بیش‌برازشی بر درستی تخمین مدل تأثیری نمی‌گذارد. RF می‌تواند تخمین غیرمرتبط داخلی از خطای تعمیم را محاسبه کند، بدون این که از زیرمجموعه‌های داده‌های خارجی استفاده کند (Breiman, 2001). بنابراین، به‌طور کلی، علت برتری روش RF نسبت به رگرسیون خطی، یادگیری بهتر روابط غیرخطی و حساسیت کمتر در برابر گیر افتادن در حداقل‌های موضعی می‌باشد.

جدول ۶- مقایسه درستی تخمین‌های دو روش رگرسیون خطی (LR) و جنگل تصادفی (RF) بر اساس آماره آکایک.

تابع	آموزش		معتبرسازی	
	LR	RF	LR	RF
۱	۲۱۴۳	-	-۴۳	-
۲	۲۰۱۱	۱۳۶۰	-۵۸	-۱۲۶
۳	۱۹۷۱	۶۷۷	-۶۲	-۱۸۵
۴	۱۸۰۱	۳۴۴	-۸۲	-۲۱۹
۵	۱۱۳۸	۲۴۰	-۱۵۴	-۲۲۸
۶	۱۱۰۹	۲۳۷	-۱۵۶	-۲۲۷
۷	۱۹۱۶	۵۷۴	-۶۸	-۱۹۵
۸	۱۷۷۹	۳۳۱	-۸۳	-۲۱۸
۹	۱۱۳۴	۲۳۷	-۱۵۴	-۲۲۷
۱۰	۱۱۱۰	۲۳۵	-۱۵۷	-۲۲۶

### نتیجه‌گیری کلی

در این تحقیق توابع انتقالی شبه‌پیوسته با دو روش رگرسیون خطی و جنگل تصادفی ایجاد گردید. متغیرهای مختلفی مربوط به توزیع اندازه ذرات خاک و برخی مکش‌های منتخب از منحنی نگهداری آب خاک به‌عنوان تخمین‌گر استفاده شد. روش رگرسیونی تنها با استفاده از مکش خاک به‌عنوان تخمین‌گر، تابع انتقالی شبه‌پیوسته قابل‌قبولی ایجاد کرد. بنابراین، می‌توان با جدیدترین روش‌های تخمین، تابع معتبرتری را ایجاد نمود که جایگزین مدل‌های فراوان منحنی نگهداری آب خاک خواهد شد. در ایجاد توابع انتقالی شبه‌پیوسته استفاده از اجزای بافت خاک (مقدار رس و شن) به‌جای کلاس بافت خاک باعث بهبود معنادار درستی تخمین منحنی نگهداری آب خاک گردید، ولی استفاده از ویژگی‌های آماری

توزیع اندازه ذرات خاک، مانند میانگین و انحراف معیار هندسی، بهبود قابل توجهی ایجاد نکرد. بنابراین، در صورت در دسترس بودن مقادیر رس و شن، استفاده از آن‌ها برای تخمین منحنی نگهداری آب خاک از طریق توابع انتقالی شبه‌پیوسته قابل توصیه می‌باشد. بر اساس نتایج این تحقیق در صورت دسترسی به رطوبت در FC و PWP، استفاده از رطوبت FC در توابع انتقالی شبه‌پیوسته کفایت نموده و نیازی به استفاده از رطوبت PWP به‌عنوان تخمین‌گر نیست. علت این نتیجه همبستگی قوی‌تر رطوبت در مکش‌های میانی با ویژگی‌های خاک و ضریب تغییرات کمتر آن در مقایسه با رطوبت در مکش‌های بالا می‌باشد. هر چند علت این نتیجه نیاز به تحقیقات بیشتری دارد. وابسته بودن توزیع خطای تخمین (RMSE) در روی مثلث بافت خاک به نوع متغیرهای ورودی و روش ایجاد توابع، اهمیت انتخاب ترکیب مناسبی از متغیرهای ورودی و روش ایجاد توابع شبه‌پیوسته در تخمین منحنی نگهداری آب خاک را نشان می‌دهد. با توجه به برتری قابل توجه روش جنگل تصادفی به رگرسیون خطی در ایجاد توابع انتقالی شبه‌پیوسته، علاوه بر توصیه به استفاده از این روش در ایجاد توابع مشابه، بررسی قابلیت دیگر روش‌های پیشرفته در تخمین منحنی نگهداری آب خاک با استفاده از توابع انتقالی شبه‌پیوسته توصیه می‌گردد. استفاده از مکش آب خاک، درصد رس و شن، جرم مخصوص ظاهری و رطوبت در FC به‌عنوان تخمین‌گر در توابع انتقالی شبه‌پیوسته با روش RF بالاترین کارایی را در تخمین منحنی نگهداری آب خاک داشت.

### سپاسگزاری

این تحقیق با پشتیبانی دانشگاه بوعلی سینا همدان انجام شده است. بدینوسیله از حمایت دانشگاه مذکور قدرانی می‌شود.

### منابع مورد استفاده

#### References

- Ahuja, L., Cassel, D., Bruce, R. & Barnes, B. (1989). Evaluation of spatial distribution of hydraulic conductivity using effective porosity data. *Soil Science*, 148, 404–411.
- Akaike, H. (1974). New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Araya, S.N. & Ghezzehei, T.A. (2019). Using Machine Learning for Prediction of Saturated Hydraulic Conductivity and Its Sensitivity to Soil Structural Perturbations. *Water Resources Research*, 55, 5715-5737. <https://doi.org/10.1029/2018WR024357>
- Bayat, H., Neyshabouri, M., Mohammadi, K. & Nariman-Zadeh, N. (2011). Estimating water retention with pedotransfer functions using multi-objective group method of data handling and ANNs. *Pedosphere*, 21, 107–114. [https://doi.org/10.1016/S1002-0160\(10\)60085-9](https://doi.org/10.1016/S1002-0160(10)60085-9)
- Bayat, H., Neyshaburi, M.R., Mohammadi, K., Nariman-Zadeh, N. & Irannejad, M. (2013a). Improving water content estimations using penetration resistance and principal component analysis. *Soil and Tillage Research*, 129, 83–92. <https://doi.org/10.1016/j.still.2013.01.009>
- Bayat, H., Neyshaburi, M.R., Mohammadi, K., Nariman-Zadeh, N., Irannejad, M. & Gregory, A.S. (2013b). Combination of artificial neural networks and fractal theory to predict soil water retention curve. *Computers and Electronics in Agriculture*, 92, 92-103. <https://doi.org/10.1016/j.compag.2013.01.005>
- Børgesen, C.D. & Schaap, M.G. (2005). Point and parameter pedotransfer functions for water retention predictions for Danish soils. *Geoderma*, 127, 154–167. <https://doi.org/10.1016/j.geoderma.2004.11.025>
- Brady, N.C. & Weil, R.R. (2010). *Elements of The Nature and Properties of Soils*. Pearson Educational International Upper Saddle River, NJ, USA.

- Breiman, L. (1984). *Classification and regression trees*. Routledge, New York, USA.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Cueff, S., Coquet, Y., Aubertot, J.-N., Bel, L., Pot, V. & Alletto, L. (2021). Estimation of soil water retention in conservation agriculture using published and new pedotransfer functions. *Soil and Tillage Research*, 209, 104967. <https://doi.org/10.1016/j.still.2021.104967>
- Dane, J. & Hopmans, J.W. (2002). Water retention and storage. Pp.671-717. In: Dane, J. H. & Topp, G. G. (eds.) *Methods of soil analysis. Part 4. Physical Methods*. Soil Science Society of America, Book Series No. 5. Madison, WI.
- Dexter, A.R., Czyz, E.A., Richard, G. & Reszkowska, A. (2008). A user-friendly water retention function that takes account of the textural and structural pore spaces in soil. *Geoderma*, 143, 243-253. <https://doi.org/10.1016/j.geoderma.2007.11.010>
- Ebrahimi, E., Bayat, H., Sadeghi, S., Fallah, M. & Jorreh, M. (2016). Using compression curve characteristics to estimate water content by the van Genuchten model. *Iranian Journal of Soil and Water Research*, 47, 217-228. (in Persian with English abstract) <https://doi.org/10.22059/ijswr.2016.58328>
- Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Monographs on Statistics and Applied Probability. Chapman & Hall, CRC Press, London, UK.
- Gee, G.W. & Or, D. (2002). Particle-size analysis. Pp. 255–293. In: Dane, J. H. & Topp, G. G. (eds.) *Methods of soil analysis. Part 4. Physical Methods*. Soil Science Society of America, Book Series No. 5. Madison, WI, USA.
- Grossman, R. & Reinsch, T. (2002). Bulk density and linear extensibility. Pp. 201-228. In: Dane, J. H. & Topp, G. G. (eds.) *Methods of soil analysis. Part 4. Physical Methods*. Soil Science Society of America, Book Series No. 5. Madison, WI, USA.
- Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K. & Kumari, M.K.N. (2019a). Machine learning approaches to develop pedotransfer functions for tropical Sri Lankan soils. *Water*, 11, 1940. <https://doi.org/10.3390/w11091940>
- Gunarathna, M.H.J.P., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.K.N. & Amarasekara, M.G.T.S. (2019b). Pedotransfer functions to estimate hydraulic properties of tropical Sri Lankan soils. *Soil and Tillage Research*, 190, 109-119. <https://doi.org/10.1016/j.still.2019.02.009>
- Haghverdi, A., Cornelis, W.M. & Ghahraman, B. (2012). A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. *Journal of Hydrology*, 442-443, 46-54. <https://doi.org/10.1016/j.jhydrol.2012.03.036>
- Haghverdi, A., Öztürk, H.S. & Cornelis, W.M. (2014). Revisiting the pseudo continuous pedotransfer function concept: Impact of data quality and data mining method. *Geoderma*, 226-227, 31-38. <https://doi.org/10.1016/j.geoderma.2014.02.026>
- Haghverdi, A., Öztürk, H.S. & Durner, W. (2018). Measurement and estimation of the soil water retention curve using the evaporation method and the pseudo continuous pedotransfer function. *Journal of Hydrology*, 563, 251-259. <https://doi.org/10.1016/j.jhydrol.2018.06.007>
- Hillel, D. (1998). *Environmental Soil Physics: Fundamentals, Applications, and Environmental Considerations*. Elsevier Academic Press, Amsterdam, San Diego.
- Hillel, D. (2004). *Introduction to Environmental Soil Physics*. Elsevier Academic Press, Amsterdam, San Diego.
- Jamshidi, M., Delavar, M.A., Taghizadehe-Mehrjerdi, R. & Brungard, C. (2019). Evaluating digital soil mapping approaches for 3D mapping of soil organic carbon. *Iranian Journal of Soil Research*, 33, 227-239. (in Persian with English abstract) <https://doi.org/10.22092/ijsr.2019.119764>
- Khlosi, M., Alhamdoosh, M., Douaik, A., Gabriels, D. & Cornelis, W. (2016). Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. *European Journal of Soil Science*, 67, 276-284. <https://doi.org/10.1111/ejss.12345>

- Kotlar, A.M., de Jong van Lier, Q. & de Souza Brito, E. (2020). Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. *Geoderma*, 361, 114098. <https://doi.org/10.1016/j.geoderma.2019.114098>
- Ließ, M., Glaser, B. & Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture: comparison of regression tree and Random Forest models. *Geoderma*, 170, 70-79. <https://doi.org/10.1016/j.geoderma.2011.10.010>
- Merdun, H., Çınar, Ö., Meral, R. & 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. <https://doi.org/10.1016/j.still.2005.08.011>
- Nemes, A., Rawls, W., Pachepsky, Y.A. & van Genuchten, M.T. (2006). Sensitivity analysis of the nonparametric nearest neighbor technique to estimate soil water retention. *Vadose Zone Journal*, 5, 1222-1235. <http://dx.doi.org/10.2136/vzj2006.0017>
- Neyshaburi, M.R., Bayat, H., Mohammadi, K., Nariman-Zadeh, N. & Irannejad, M. (2015). Improvement in estimation of soil water retention using fractal parameters and multiobjective group method of data handling. *Archives of Agronomy and Soil Science*, 61, 257-273. <https://doi.org/10.1080/03650340.2014.924107>
- Nguyen, P.M., Haghverdi, A., De Pue, J., Botula, Y.-D., Le, K.V., Waegeman, W. & Cornelis, W.M. (2017). Comparison of statistical regression and data-mining techniques in estimating soil water retention of tropical delta soils. *Biosystems Engineering*, 153, 12-27. <https://doi.org/10.1016/j.biosystemseng.2016.10.013>
- Pachepsky, Y., Rawls, W., Gimenez, D. & Watt, J. (1998). Use of soil penetration resistance and group method of data handling to improve soil water retention estimates. *Soil and Tillage Research*, 49, 117-126. [https://doi.org/10.1016/S0167-1987\(98\)00168-8](https://doi.org/10.1016/S0167-1987(98)00168-8)
- Pachepsky, Y.A. & Rawls, W. (1999). Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Science Society of America Journal*, 63, 1748-1757. <https://doi.org/10.2136/sssaj1999.6361748x>
- Pachepsky, Y.A., Rawls, W. & Lin, H. (2006). Hydropedology and pedotransfer functions. *Geoderma*, 131, 308-316. <https://doi.org/10.1016/j.geoderma.2005.03.012>
- Pachepsky, Y.A. & Rawls, W.J. (2004). *Development of Pedotransfer Functions in Soil Hydrology*. Elsevier Science.
- Paul, R.K. (2006). *Multicollinearity: Causes, Effects and Remedies*. IASRI, New Delhi.
- Rastgou, M. (2020). Comprehensive comparison of the methods of developing pedotransfer functions (PTFs) and development of new algorithms to predict soil water retention curve (SWRC) and soil hydraulic conductivity curve (SHCC). Ph. D. Thesis, Department of Soil Science and Engineering, Faculty of Agriculture, Bu-Ali Sina University, Hamedan, Iran. (in Persian with English abstract)
- Rastgou, M., Bayat, H., Mansoorizadeh, M. & Gregory, A.S. (2020). Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique. *Computers and Electronics in Agriculture*, 174, 105502. <https://doi.org/10.1016/j.compag.2020.105502>
- Rastgou, M., Bayat, H., Mansoorizadeh, M. & Gregory, A.S. (2022). Estimating soil water retention curve by extreme learning machine, radial basis function, M5 tree and modified group method of data handling approaches. *Water Resources Research*, 58, e2021WR031059. <https://doi.org/10.1029/2021WR031059>
- Rawls, W.J., Gish, T.J. & Brakensiek, D.L. (1991). Estimating Soil Water Retention from Soil Physical Properties and Characteristics. Pp. 213-234. In: Stewart, B.A. (Ed.), *Advances in Soil Science: Volume 16*. Springer New York, New York, NY. [https://doi.org/10.1007/978-1-4612-3144-8\\_5](https://doi.org/10.1007/978-1-4612-3144-8_5)
- Rawls, W., Pachepsky, Y.A., Ritchie, J., Sobecki, T. & Bloodworth, H. (2003). Effect of soil organic carbon on soil water retention. *Geoderma*, 116, 61-76. [https://doi.org/10.1016/S0016-7061\(03\)00094-6](https://doi.org/10.1016/S0016-7061(03)00094-6)
- Romano, N. & Palladino, M. (2002). Prediction of soil water retention using soil physical data and terrain attributes. *Journal of Hydrology*, 265, 56-75. [https://doi.org/10.1016/S0022-1694\(02\)00094-X](https://doi.org/10.1016/S0022-1694(02)00094-X)



- Schaap, M.G., Leij, F.J. & van Genuchten, M.T. (2001). Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, 251, 163–176. [https://doi.org/10.1016/S0022-1694\(01\)00466-8](https://doi.org/10.1016/S0022-1694(01)00466-8)
- Shirazi, M.A. & Boersma, L. (1984). A unifying quantitative analysis of soil texture. *Soil Science Society of America Journal*, 48, 142-147. <https://doi.org/10.2136/sssaj1984.03615995004800010026x>
- Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P. & Feuston, B.P. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, 43, 1947-1958. <https://doi.org/10.1021/ci034160g>
- Szabó, B., Sztamári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K. & Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics. *Hydrology and Earth System Science*, 23, 2615-2635. <http://dx.doi.org/10.5194/hess-23-2615-2019>
- Tafteh, A., Davatgar, N. & Sedaghat, A. (2022). Estimation of important points on soil water retention curve (SWRC): comparison experimental-physical models and data mining technique. *Arabian Journal of Geosciences*, 15, 968. <https://doi.org/10.1007/s12517-022-10232-0>
- Tomasella, J., Pachepsky, Y., Crestana, S. & Rawls, W. (2003). Comparison of two techniques to develop pedotransfer functions for water retention. *Soil Science Society of America Journal*, 67, 1085-1092. <http://dx.doi.org/10.2136/sssaj2003.1085>
- Tóth, B., Makó, A. & Toth, G. (2014). Role of soil properties in water retention characteristics of main Hungarian soil types. *Journal of Central European Agriculture*, 15, 137-153. <https://doi.org/10.5513/JCEA01/15.2.1465>
- Touil, S., Degre, A. & Chabaca, M.N. (2016). Sensitivity analysis of point and parametric pedotransfer functions for estimating water retention of soils in Algeria. *Soil*, 2, 647. <http://dx.doi.org/10.5194/soil-2-647-2016>
- Twarakavi, N.K., Šimůnek, J. & Schaap, M. (2009). Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Science Society of America Journal*, 73, 1443-1452. <https://doi.org/10.2136/sssaj2008.0021>
- Ungaro, F., Calzolari, C. & Busoni, E. (2005). Development of pedotransfer functions using a group method of data handling for the soil of the Pianura Padano-Veneta region of North Italy: water retention properties. *Geoderma*, 124, 293-317. <https://doi.org/10.1016/j.geoderma.2004.05.007>
- Vereecken, H., Maes, J., Feyen, J. & Darius, P. (1989). Estimating the soil moisture retention characteristic from texture, bulk density, and carbon content. *Soil Science*, 148, 389–403. <http://dx.doi.org/10.1097/00010694-198912000-00001>
- Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M. & Genuchten, M.T. (2010). Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review. *Vadose Zone Journal*, 9, 795–820. <https://doi.org/10.2136/vzj2010.0045>
- Verikas, A., Gelzinis, A. & Bacauskiene, M. (2011). Mining data with random forests: A survey and results of new tests. *Pattern Recognition*, 44, 330–349. <https://doi.org/10.1016/j.patcog.2010.08.011>
- Warkentin, B.P. (1972). Use of the liquid limit in characterizing the clay soils. *Canadian Journal of Soil Science*, 52, 457–464.
- Wiesmeier, M., Barthold, F., Blank, B. & Kögel-Knabner, I. (2011). Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil*, 340, 7-24. <https://doi.org/10.1007/s11104-010-0425-z>
- Williams, J., Prebble, R., Williams, W. & Hignett, C. (1983). The influence of texture, structure and clay mineralogy on the soil moisture characteristic. *Soil Research*, 21, 15–32. <https://doi.org/10.1071/SR9830015>
- Wösten, J., Pachepsky, Y.A. & Rawls, W. (2001). Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *Journal of Hydrology*, 251, 123–150. [https://doi.org/10.1016/S0022-1694\(01\)00464-4](https://doi.org/10.1016/S0022-1694(01)00464-4)
- Zamani, P. (2011). *Statistical Designs in Animal Science*. Bu-Ali Sina University Press, Hamedan, Iran. (In Farsi).