

## کارایی روش رگرسیون کریجینگ در تهیه نقشه توان تولید رویشگاه راش در جنگل پژوهشی دانشگاه تربیت مدرس

زهرا احدی<sup>۱</sup>، سیدجلیل علوی<sup>۲\*</sup> و سید محسن حسینی<sup>۳</sup>

<sup>۱</sup> دانشجوی کارشناسی ارشد جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور  
<sup>۲</sup> استادیار گروه جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور  
<sup>۳</sup> استاد گروه جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور

(تاریخ دریافت: ۱۳۹۶/۴/۱۲؛ تاریخ پذیرش: ۱۳۹۶/۹/۱۲)

### چکیده

تهیه نقشه توان تولید جنگل که توانایی رویشگاه را در ایفای نقش‌هایی همانند تولید چوب و ترسیب کربن نشان می‌دهد، امری اجتناب‌ناپذیر در مدیریت و بهره‌برداری پایدار اکوسیستم جنگلی است. وسعت فراوان جنگل‌های شمال کشور و کوهستانی بودن آنها، سبب شده است که یافتن روش مناسب برای تهیه نقشه ویژگی‌های کمی جنگل ضروری باشد. از طرفی اهمیت تجاری و زیست‌محیطی جنگل‌های خزر و به‌خصوص گونه راش که از اقتصادی‌ترین و فراوان‌ترین گونه‌های پهن‌برگ در این جنگل‌هاست، سبب می‌شود که همواره تهیه اطلاعات کمی و کیفی دقیق از آنها و بررسی تغییرات مربوط به آنها ضروری باشد. تحقیق حاضر در زمینه ارزیابی کارایی روش رگرسیون کریجینگ در تهیه نقشه توان تولید رویشگاه راش در جنگل آمیخته و ناهمسال راش دانشگاه تربیت مدرس انجام گرفت. به‌منظور جمع‌آوری اطلاعات لازم، ۱۲۳ قطعه نمونه دایره‌ای به مساحت ۱۰۰۰ متر مربع در توده‌هایی که گونه راش وضعیت چیره یا چیره‌نما داشت، به روش آماربرداری منظم تصادفی در شبکه‌ای به ابعاد ۱۰۰ × ۱۰۰ متر پیاده شد. پس از محاسبه شاخص توان تولید رویشگاه راش در محل هر یک از قطعات نمونه و استخراج متغیرهای اولیه و ثانویه، کارایی دو رویکرد مدل‌سازی رگرسیون خطی چندگانه و جنگل تصادفی به‌عنوان مدل‌های مبنا در روش رگرسیون کریجینگ در تهیه نقشه توان تولید رویشگاه راش ارزیابی شد. نتایج اعتبارسنجی متقابل با توجه به مقدار میانگین خطای نسبی و مقدار مجذور میانگین مربعات خطای نسبی نشان داد که الگوریتم جنگل تصادفی عملکرد بسیار بهتری از روش رگرسیون کریجینگ با مدل مبنا رگرسیون خطی و روش کریجینگ داشت، به‌طوری که مجذور میانگین مربعات خطا را در حدود ۷۰ درصد کاهش داد. از این‌رو روش رگرسیون کریجینگ با مدل مبنا جنگل تصادفی توانست با دقت بیشتری نقشه توان تولید رویشگاه راش را تهیه کند.

**واژه‌های کلیدی:** جنگل تصادفی، درون‌یابی، رگرسیون، زمین‌آمار، کریجینگ، یادگیری ماشین.

### مقدمه

رویشگاه‌های جنگلی نیز مقوله بسیار مهمی در پژوهش‌های جنگلداری در جهان است که متأسفانه در ایران کمتر به آن توجه شده است و آن عبارت است از ارائه شاخصی که پتانسیل رویشگاه را در ایفای نقش‌های مختلف جنگل همانند تولید چوب،

جنگل‌ها حالت نهایی یک اکوسیستم طبیعی‌اند و از این‌رو به‌منظور درک و مدیریت بهتر باید از جنبه‌های مختلف ارزیابی شوند (Meng et al., 2009). ارزیابی توان تولید

همچنین متغیرهای محیطی را به حساب می‌آورد. بوم‌شناسان جنگل اغلب، متغیرهای پاسخ نظیر شاخص توان تولید را بدون توجه زیاد به محل، پیش‌بینی می‌کنند. این پیش‌بینی‌ها مبتنی بر این فرض است که توزیع توان تولید رویشگاه در هر محل مستقل از دیگر محل‌هاست (Guisan & Zimmermann, 2000). گسترش سامانه‌های اطلاعات مکانی در ارتباط با آمار مکانی نیز موجب شد رویکردهای مبتنی بر محل که با عنوان روش‌های زمین‌آمار طبقه‌بندی می‌شوند و از روش‌های درون‌یابی همانند کریجینگ معمولی در تهیه انواع نقشه‌ها استفاده می‌کنند، رواج یابد (Aertsen et al., 2012). امروزه روش‌های زمین‌آمار و درون‌یابی، زمینه کاربرد خود را در منابع طبیعی به‌ویژه در علوم جنگل پیدا کرده‌اند، به‌طوری که با به‌کارگیری آنها می‌توان متغیرهای جنگل را برآورد و نقشه‌سازی کرد (Akhavan & Kleinn, 2009). به‌تازگی تهیه نقشه‌های پیش‌بینی در مدیریت جنگل به‌سمت استفاده از رویکردهای ترکیبی سوق پیدا کرده است که می‌تواند ترکیبی از رویکردهای مبتنی بر محل و رویکرد مبتنی بر خصوصیت قلمداد شود که در آن هم وابستگی مکانی و هم متغیرهای پیشگو لحاظ می‌شوند. در رویکرد ترکیبی دو زیرگروه اصلی از روش‌ها قرار می‌گیرند (Hengl et al., 2007). روش‌های مبتنی بر رگرسیون کریجینگ.

در خصوص بررسی ویژگی‌های جنگل‌ها با استفاده از فنون زمین‌آمار پژوهش‌های متعددی در جهان انجام گرفته است. برخی از این تحقیقات بیانگر کارایی زیاد و برخی نشان‌دهنده ضعف‌هایی برای این

ترسیب کربن، حفظ خاک، کنترل سیلاب، حفظ تنوع زیستی و تأمین آب نشان می‌دهد (Skovsgaard & Vanclay, 2008).

در بسیاری از موارد به‌علت محدودیت‌های زمانی و مالی، تهیه اطلاعات از تمامی سطح جنگل به‌صورت آماربرداری صددرصد با هدف تهیه نقشه توان تولید جنگل امکان‌پذیر نیست. گاهی امکان اندازه‌گیری شاخص توان تولید رویشگاه به‌طور مستقیم براساس ارتفاع غالب و سن (قطر) توده وجود ندارد، از این‌رو براساس ویژگی‌های رویشگاه همانند اقلیم، توپوگرافی یا خاک، توان تولیدی رویشگاه برآورد می‌شود (Aertsen et al., 2011).

روش‌های بسیاری برای برآورد توان تولید یا کیفیت رویشگاه وجود دارد که از آن جمله می‌توان به استفاده از شاخص‌های تنوع زیستی، گیاهان شاخص، بررسی خاک، تجزیه برگ و لاشبرگ، وضعیت ظاهری توده و شاخص رویشگاه<sup>۱</sup> اشاره داشت (Herrera et al., 1999). روش معمول برای ارزیابی کیفیت رویشگاه برای توده‌های همسال، شاخص رویشگاه (ارتفاع درخت در رابطه با سن مرجع) است. با توجه به اینکه در توده‌های ناهمسال و آمیخته، شاخص رویشگاه نمی‌تواند معرف خوبی برای بیان کیفیت رویشگاه باشد، فرم رویشگاه<sup>۲</sup> که ارتفاع درختان در یک توده در یک قطر مرجع است، به‌عنوان معیار مطمئن‌تر برای ارزیابی کیفیت رویشگاه معرفی شد (Vanclay & Henry, 1988). به‌طور کلی از سه رویکرد مختلف می‌توان برای تهیه نقشه توان تولید رویشگاه استفاده کرد: رویکرد مبتنی بر محل<sup>۳</sup> که از پدیده خودهمبستگی مکانی بهره می‌گیرد و متغیر هدف را در فضای جغرافیایی برآورد می‌کند و نه در فضای متغیر پیشگو (Lehmann et al., 2002)؛ رویکرد مبتنی بر ویژگی<sup>۴</sup> که با مرتبط کردن متغیر هدف مورد نظر به یک یا چند متغیر محیطی بدون در نظر گرفتن محل‌های مکانی، برآوردها را انجام می‌دهد؛ و رویکرد ترکیبی<sup>۵</sup> که وابستگی مکانی و

<sup>1</sup> Site index

<sup>2</sup> Site form

<sup>3</sup> Location-based approach

<sup>4</sup> attribute-based approach

<sup>5</sup> hybrid approach

امروزه برآورد مشخصه‌های جنگل با استفاده از روش‌های زمین‌آمار و به‌کمک تصاویر ماهواره‌ای می‌تواند دقت مناسبی داشته باشد، برای تهیه نقشه ارتفاع تاج درختان از روش رگرسیون کریجینگ استفاده کردند. آنها برای بهبود دقت از روش رگرسیون کریجینگ استفاده کردند و توانستند دقت را ارتقا دهند.

اهمیت جنگل‌های خزری از نظر تجاری و زیست‌محیطی ایجاب می‌کند که همواره اطلاعات کمی و کیفی دقیقی از آنها تهیه و تغییرات مربوط به آنها بررسی شود. با توجه به گستردگی جنگل‌ها و کوهستانی بودن عرصه‌ها، امکان صرف زمان و هزینه زیاد و آماربرداری صددرصد برای تهیه نقشه‌های دقیق و صحیح وجود ندارد، از این رو یافتن روشی مناسب برای تهیه نقشه‌های ویژگی‌های کمی جنگل‌ها و بهبود دقت برآورد با صرف هزینه و زمان اندک به‌منظور مدیریت پایدار و بهره‌برداری مستمر از این جنگل‌ها ضروری به‌نظر می‌رسد. با توجه به پیشینه پژوهش‌های صورت گرفته و تأیید کارایی این روش، تحقیق حاضر با هدف بررسی پتانسیل روش زمین‌آمار رگرسیون کریجینگ در مقایسه با روش کریجینگ معمولی برای تهیه نقشه توان تولید رویشگاه‌های راش، به‌عنوان یکی از فراوان‌ترین و اقتصادی‌ترین گونه‌های پهن‌برگ در جنگل‌های آمیخته و ناهمسال شمال کشور، انجام پذیرفت.

## مواد و روش‌ها

### منطقه پژوهش

پژوهش حاضر در جنگل آموزشی و پژوهشی صلاح‌الدین کلا در استان مازندران متعلق به دانشکده منابع طبیعی دانشگاه تربیت مدرس صورت گرفت. منطقه تحقیق در سری ۳ آغوزچال از طرح جنگلداری حوزه ۴۶ کجور واقع شده است. مساحت منطقه در حدود ۲۰۰ هکتار است (شکل ۱). تیپ جنگل در منطقه، آمیخته‌ای از راش به‌همراه گونه‌های ممرز،

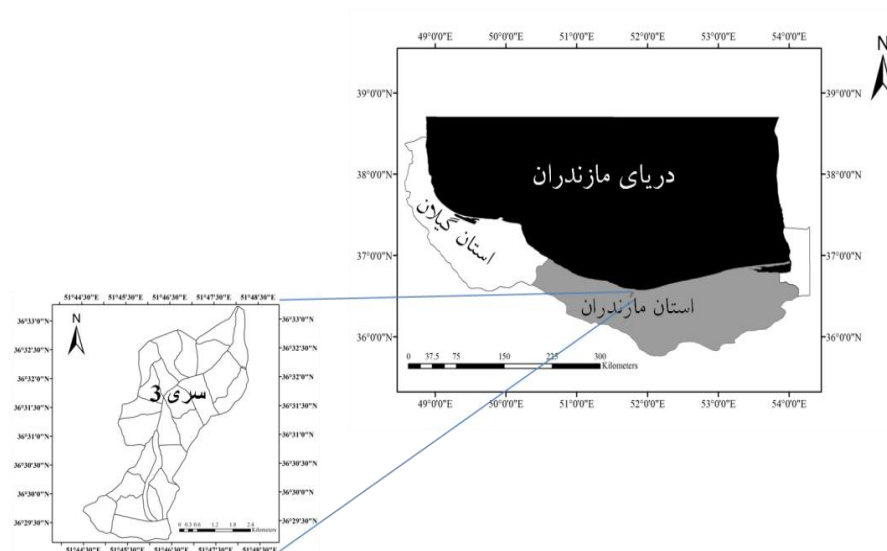
فنون هستند. شایان ذکر است که بیشتر تحقیقات در خصوص استفاده از رگرسیون کریجینگ به تهیه نقشه ویژگی‌های خاک مربوط است. سوابق تحقیقات داخلی در زمینه کاربرد روش‌های برآورد مکانی در تهیه نقشه ویژگی‌های جنگل‌ها همگی درباره استفاده از روش‌های کریجینگ، کوکریجینگ و معکوس فاصله وزنی (IDW<sup>۱</sup>) است.

در تحقیقی (Meng et al. (2009 روش‌های کریجینگ معمولی، کوکریجینگ و رگرسیون کریجینگ را برای پیش‌بینی پارامترهای جنگل همانند رویه زمینی، ارتفاع، اوضاع سلامت و زی‌توده در غرب جورجیا در آمریکا به‌کار گرفتند. نتایج نشان داد که رگرسیون کریجینگ با کمترین خطا و بیشترین ضریب تبیین، روش برتر است. در نیوزیلند با استفاده از روش‌های رگرسیون کریجینگ، کریجینگ معمولی و معکوس فاصله وزنی، روش‌های پیش‌بینی مکانی توان تولید و ارتفاع غالب گونه *Pinus radiata* توسط Palmer et al. (2009) مقایسه شد. نتایج تحقیق او نشان داد که در مقیاس محلی روش کریجینگ بیشترین کارایی را داشت و در مقیاس منطقه‌ای نتایج روش‌ها به‌نسبت مشابه بود. در پژوهشی دیگر (Palmer et al. (2010 در بررسی توان تولید گونه *Pinus radiata* اشاره کردند که رگرسیون کریجینگ توانست مقادیر ضریب همبستگی را از ۰/۶ به ۰/۷ افزایش دهد. همچنین (Aertsen et al. (2012 در منطقه معتدل فلاندر در بلژیک، پنج روش تهیه نقشه شاخص رویشگاه شامل کریجینگ معمولی، رگرسیون و سه روش هیبرید (geomatching، کوکریجینگ و رگرسیون کریجینگ) را برای دو گونه درختی مهم (بلوط و راش) مقایسه کردند. نتایج مطالعه نشان داد که نمی‌توان گفت در هر شرایطی یک روش از بقیه روش‌ها بهتر است. در منطقه گویان فرانسه (Fayad et al. (2016 برای تأکید بر این موضوع که

<sup>1</sup> Inverse Distance Weighted

درصد از کل تعداد درختان را در منطقه به خود اختصاص داده است (Ahmadi et al., 2015).

انجیلی، افرا، شیردار، نمدار، ون و بلوط با ساختار دانه‌زاد ناهمسال نامنظم است. به‌طور میانگین، گونه راش حدود ۵۰ درصد از کل حجم سرپا و حدود ۴۰



شکل ۱- موقعیت منطقه پژوهش

تربیت مدرس، مدل Chapman-Richards بهترین مدل در توصیف رابطه قطر و ارتفاع گونه راش است. با استفاده از این مدل و روش معادله اختلاف، شاخص توان تولید رویشگاه راش بر اساس میانگین ارتفاع و قطر برابرسینه گونه راش در هر قطعه نمونه محاسبه شد (Ahmadi et al., 2015).

### روش درون‌یابی رگرسیون کریجینگ

در زمین‌آمار فرض بر این است که نمونه‌های انتخاب‌شده از جامعه مستقل نیستند، بلکه تا فاصله معینی به‌صورت مکانی به هم وابستگی دارند (Goovaerts, 1997). به‌طور کلی زمین‌آمار شامل دو بخش اصلی واریوگرافی<sup>۱</sup> و کریجینگ<sup>۲</sup> است (Akhavan et al., 2012). رگرسیون کریجینگ نوعی روش درون‌یابی هیبرید است که ترکیبی از یک مدل رگرسیونی با کریجینگ ساده یا معمولی است. شرط

### شیوه اجرای پژوهش

اطلاعات لازم از ۱۲۳ قطعه نمونه دایره‌ای با مساحت ۱۰۰۰ متر مربع در توده‌هایی که گونه راش در آنها غالب بود، به روش آماربرداری منظم تصادفی در شبکه آماربرداری به ابعاد ۱۰۰ × ۱۰۰ متر در سطحی حدود ۲۰۰ هکتار با دامنه ارتفاعی ۱۰۰۰ تا ۱۵۰۰ متر از سطح دریا جمع‌آوری شد. پس از پیاده کردن قطعات نمونه در جنگل، قطر برابرسینه و همچنین ارتفاع کامل همه درختان راش با قطر بیشتر از ۷/۵ سانتی‌متر اندازه‌گیری شد.

برای برآورد معیار فرم رویشگاه ابتدا باید رابطه قطر برابرسینه و ارتفاع درخت مدلسازی شود. پس از انتخاب مناسب‌ترین مدل براساس معیارهای ارزیابی عملکرد مدل، به روش منحنی راهنما یا معادله اختلاف، معادله یا منحنی‌های شاخص توان تولید برای گونه مورد نظر در قطر مرجع تهیه می‌شود. براساس مطالعات صورت‌گرفته در جنگل دانشگاه

<sup>۱</sup> Variography

<sup>۲</sup> Kriging

منطقه تحقیق از نرم‌افزار<sup>۱</sup> SAGA 2.0.8 تهیه شد. به‌منظور بررسی روابط پیچیده محیط و جنگل، مدل رگرسیون خطی را می‌توان با الگوریتم‌های یادگیری ماشین جایگزین کرد. الگوریتم‌های متداول یادگیری ماشین عبارت‌اند از شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان، درختان طبقه‌بندی و رگرسیون و جنگل تصادفی (Hengl et al, 2015). در تحقیق حاضر به دو دلیل از روش جنگل تصادفی به‌عنوان مدل مبنای رگرسیون کریجینگ در تهیه نقشه توان تولید رویشگاه راش شرقی استفاده شد: ۱. در مطالعات متعدد ثابت شده است که الگوریتم جنگل تصادفی عملکرد بهتری از رگرسیون دارد؛ ۲. این روش، برخلاف مدل رگرسیون خطی، نیازمند توزیع احتمال برای متغیر هدف نیست و می‌تواند روابط غیرخطی پیچیده را نیز برازش دهد (Alavi et al., 2017). جنگل تصادفی روشی جدید و قدرتمند در داده‌کاوی است؛ با این حال، این روش در مطالعات اکولوژیک به‌نسبت ناشناخته است. رویکرد جنگل تصادفی مبتنی بر ایجاد تعداد زیاد درخت تصمیم و ترکیب آنها برای پیش‌بینی است (Cutler et al., 2007). جنگل‌های تصادفی برای ایجاد هر درخت از نمونه bootstrap استفاده می‌کند که تنها ۶۶ درصد از داده اولیه را شامل می‌شود. سپس یک متغیر پیشگو به‌طور تصادفی در خلال فرایند ایجاد درخت معرفی می‌شود. متغیر مورد استفاده در ساخت درخت به‌صورت تصادفی انتخاب می‌شود. پارامترهای کلیدی برای مدل جنگل تصادفی، تعداد درختان و تعداد متغیرهای پیشگو هستند. براساس این دو پارامتر، درخت تصمیم به بزرگ‌ترین اندازه ممکن رویانده و بدون هرس رها می‌شود (Breiman, 2001).

مشخصه‌های مشتق از مدل رقومی ارتفاع (DEM<sup>۲</sup>) در بسیاری از مطالعات مربوط به طبقه‌بندی

استفاده از رگرسیون کریجینگ آن است که یک یا چند متغیر کمکی مستقل به تعداد زیاد در منطقه موجود بوده و بین متغیر وابسته و متغیرهای کمکی، همبستگی خوبی وجود داشته باشد (Pebesma, 2006). در این روش ابتدا با استفاده از مدل رگرسیون خطی چندگانه یا هر روش مدلسازی دیگر، رابطه بین متغیر اصلی (معیار توان تولید رویشگاه) و متغیرهای کمکی، برآورد می‌شود و سپس باقی‌مانده مدل مورد استفاده با روش کریجینگ معمولی درون‌یابی شده و به مقادیر برآوردشده از رابطه رگرسیونی اضافه می‌شود تا مقادیر نهایی درون‌یابی به‌دست آید. معادله نهایی روش رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه به‌صورت رابطه ۱ است (Minasny & McBratney, 2007).

$$\hat{z}(s_0) = \sum_{k=0}^p \hat{\beta}_k + q_k(s_0) + \sum_{i=1}^n \lambda_i \cdot e(s_i) \quad \text{رابطه ۱}$$

که در آن ضرایب مدل رگرسیون برآوردشده؛  $q_k(s_0)$  مقادیر متغیرهای کمکی در محل  $s_0$ ؛  $p$  تعداد متغیرهای کمکی یا پیشگو؛ و  $n$  تعداد مشاهدات است.  $\lambda_i$  وزن کریجینگ است که توسط ساختار مکانی باقی‌مانده‌ها تعیین می‌شود.  $e(s_i)$  مقدار باقی‌مانده در محل  $s_i$  است. قبل از درون‌یابی به روش کریجینگ باید یک مدل نظری را بر واریوگرام تجربی برازش داد، زیرا زمین‌آمار روشی وابسته به مدل است. مدل‌های متفاوتی وجود دارد؛ از جمله مدل‌های خطی، کروی، نمایی و گوسی که می‌توان از آنها استفاده کرد (Akhavan & Kleinn, 2009). مراحل تهیه نقشه با استفاده از روش رگرسیون کریجینگ در شکل ۲ ارائه شده است.

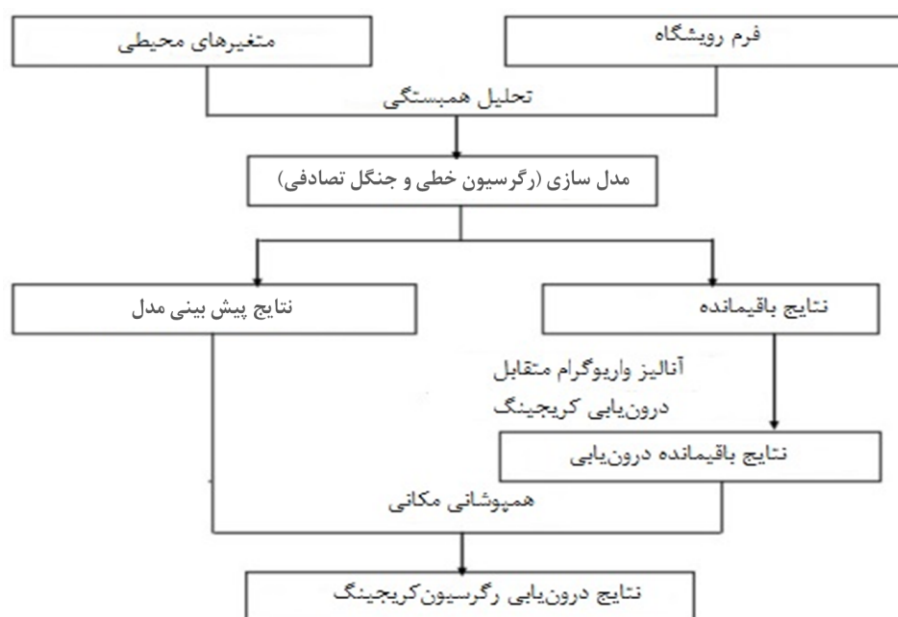
از آنجا که نقشه متغیرهای خاک و اقلیم برای منطقه تحقیق موجود نیست، متغیرهای اولیه و ثانویه توپوگرافی با استفاده از مدل رقومی ارتفاع برای

<sup>۱</sup> System for Automated Geoscientific Analyses

<sup>۲</sup> Digital Elevation Model

خصوصیات ترکیبی، مشخصه‌هایی دارند که مرکب از ویژگی‌های اولیه به‌همراه شاخص‌های فیزیکی یا تجربی هستند که تغییرپذیری مکانی فرایندهای خاص در سیمای اراضی را بیان می‌کنند (Ghanbari et al., 2009).

خصوصیات و برای پیش‌بینی خاک، پوشش گیاهی و پراکنش گونه‌ها در مقیاس‌های متفاوت استفاده شده‌اند. به‌طور معمول، خصوصیات اولیه توپوگرافی، شامل ویژگی‌هایی است که به‌طور مستقیم از DEM محاسبه می‌شوند. از سوی دیگر خصوصیات ثانویه یا



شکل ۲- مراحل تهیه نقشه با استفاده از رگرسیون کریجینگ

رابطه‌های زیر  $Z(x_i)$  مقدار متغیر ناحیه‌ای  $x$  در نقطه  $i$ ،  $\bar{Z}(x_i)$  برآورد مقدار متغیر ناحیه‌ای  $x$  در نقطه  $i$  و  $\bar{Z}$  مقدار میانگین متغیر ناحیه‌ای  $x$  است.

$$ME = \frac{1}{N} \sum_{i=1}^N [z(x_i) - \hat{z}(x_i)] \quad \text{رابطه ۲}$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |z(x_i) - \hat{z}(x_i)| \quad \text{رابطه ۳}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(x_i) - \hat{z}(x_i)]^2} \quad \text{رابطه ۴}$$

در تحقیق حاضر به‌منظور ارزیابی صحت رگرسیون کریجینگ و مقایسه آن با روش کریجینگ از روش اعتبارسنجی متقابل 10-fold استفاده شده است. در این روش کل داده‌ها به ۱۰ قسمت تقسیم شده و هر بار یکی از این قسمت‌ها به‌منظور اعتبارسنجی کنار گذاشته می‌شود و با ۹ قسمت دیگر مدل‌سازی صورت می‌گیرد. روش‌های مورد استفاده در تحقیق حاضر با استفاده از آماره‌های میانگین خطا<sup>۱</sup> (ME)، میانگین قدرمطلق خطا<sup>۲</sup> (MAE)، جذر میانگین مربعات خطا<sup>۳</sup> (RMSE)، مقادیر نسبی میانگین خطا<sup>۴</sup> (ME<sub>r</sub>) و مقادیر نسبی مجذور میانگین مربعات خطا<sup>۵</sup> (RMSE<sub>r</sub>) ارزیابی شدند (Ahadi et al., 2017). نحوه محاسبه هر یک از آماره‌های مورد بررسی در رابطه‌های ۲ تا ۶ ارائه شده است. در

<sup>1</sup> Mean Error

<sup>2</sup> Mean Absolute Error

<sup>3</sup> Root Mean Square Error

<sup>4</sup> Relative Mean Error

<sup>5</sup> Relative Root Mean Square Error

## روش تحلیل

تجزیه و تحلیل‌های مرتبط با مدلسازی در نرم‌افزار آماری R نسخه 3.4.1 (R Core R Development Team, 2017) و با استفاده از بسته‌های gstat، randomForest و GSIF.nlme انجام گرفت.

## نتایج

اطلاعات اولیه آماری شاخص توان تولید رویشگاه راش همانند حداقل، حداکثر، میانگین و ضریب تغییرات در جدول ۱ ارائه شده است.

جدول ۱- اطلاعات اولیه آماری شاخص توان تولید رویشگاه راش شرقی

ویژگی زیست‌سنجی	حداقل	حداکثر	میانگین	ضریب تغییرات
فرم رویشگاه (متر)	۳۰/۲۱	۳۹/۳۱	۳۴/۷۸	۶/۲۸

دارد یا خیر. به‌کارگیری معیار اطلاعات آکائیک (AIC) و آزمون نسبت درست‌نمایی ( $P < 0.01$ ) نشان داد که ساختار وابستگی مکانی در باقی‌مانده‌های مدل رگرسیونی و جنگل تصادفی وجود دارد. افزودن ساختار مکانی به مدل رگرسیونی سبب شد مقدار AIC به‌طور معنی‌دار کاهش یابد؛ اما بین مدل‌های مختلف با ساختار مکانی اختلاف بسیار اندکی وجود دارد. از آنجا که متداول‌ترین مدلی که بیشترین کاربرد را در مطالعات زیست‌محیطی دارد، مدل کروی است، در تحقیق حاضر نقشه‌سازی توان تولید رویشگاه راش با بهره‌گیری از این مدل صورت گرفت. درصد ساختار مکانی واریوگرام برای باقی‌مانده مدل رگرسیونی در حدود ۵۰ درصد و برای روش جنگل تصادفی در حدود ۶۵ درصد محاسبه شد که نشان‌دهنده ساختار مکانی متوسط برای باقی‌مانده‌ها در هر دو مدل است. مشخصه‌های واریوگرام برازش‌شده با مدل کروی با استفاده از باقی‌مانده مدل رگرسیونی خطی چندگانه و جنگل تصادفی در جدول ۲ ارائه شده است. در شکل

$$MEr = \frac{ME}{\bar{z}(x_i)} \times 100$$

رابطه ۵

$$RMSEr = \frac{RMSE}{\bar{z}(x_i)} \times 100$$

رابطه ۶

این معیارها نشان می‌دهند که بین مقدار برآوردشده و مشاهده‌شده چقدر اختلاف وجود دارد. چنانچه مقدار این معیارها برابر با صفر باشد، نشان‌دهنده آن است که مقادیر نمونه خوب برآورد شده است.

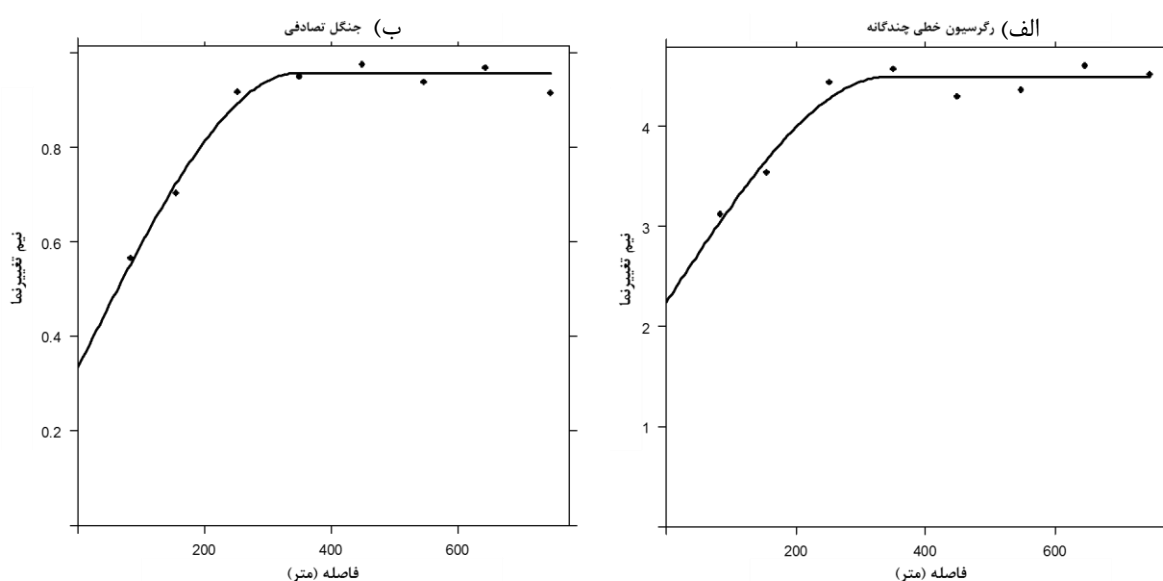
قبل از ورود متغیرهای مشتق از مدل رقومی زمین در مدل رگرسیون خطی چندگانه و جنگل تصادفی، به‌منظور بررسی هم‌خطی متغیرهای محیطی از روش تحلیل خوشه‌ای سلسله‌مراتبی با استفاده از مربع همبستگی اسپیرمن در بسته Hmisc در نرم‌افزار آماری R (Harrell, 2016) بهره گرفته شد. بررسی هم‌خطی متغیرهای محیطی با استفاده از تحلیل خوشه‌ای و همبستگی اسپیرمن نشان داد که همبستگی قوی بین مدل رقومی زمین و شبکه کانال و درصد شیب دامنه و طول شیب وجود دارد. بر این اساس تصمیم گرفته شد متغیرهای شبکه کانال و طول شیب حذف و متغیرها حفظ شوند.

از آنجا که در روش رگرسیون کریجینگ، ابتدا روش‌های مدلسازی بر روی متغیرهای کمکی که برای کل منطقه در دسترس است، اعمال می‌شود و سپس از کریجینگ معمولی برای درون‌یابی باقی‌مانده‌ها از مدل رگرسیونی استفاده می‌شود، باید بررسی شود که آیا وابستگی مکانی در مدل‌های مورد استفاده وجود

۳ نیز تغییرنمای باقی‌مانده‌های رگرسیون خطی و روش جنگل تصادفی ارائه شده است.

جدول ۲- مشخصه‌های واریوگرام برازش شده با مدل کروی و با استفاده از باقی‌مانده مدل رگرسیون خطی چندگانه و جنگل تصادفی

مدل	طول گام (متر)	اثر قطعه‌ای	حد آستانه	دامنه تأثیر (متر)	ساختار مکانی
مدل رگرسیون خطی چندگانه	۱۰۰	۲/۲۵	۴/۵۰	۳۴۳	٪۵۰
مدل جنگل تصادفی	۱۰۰	۰/۳۳	۰/۹۶	۳۴۸	٪۶۵



شکل ۳- تغییر نمای باقیمانده مدل رگرسیون خطی چندگانه (الف) و جنگل تصادفی (ب)

هرچه این شاخص از صفر دور شود نشان می‌دهد که اختلاف بین مقادیر مشاهده شده و برآورد شده بیشتر است. شاخص میانگین قدرمطلق خطا برای سنجش دقت روش قابل استفاده است. هرچه مقدار به صفر نزدیک‌تر باشد، دقت مکانی روش بیشتر شده و هرچه از صفر دور شود، از دقت آن کاسته می‌شود (Ahadi et al., 2017). براساس یافته‌های این تحقیق کمترین خطای مشاهدات مربوط به روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی است.

نتایج محاسبه معیارهای مختلف خطا، برای دو روش کریجینگ معمولی و روش رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه و جنگل تصادفی در جدول ۳ ارائه شده است. مقادیر میانگین خطا و میانگین قدر مطلق خطا مقدار اربیبی را نشان می‌دهند که در حالت ایده‌آل باید مساوی صفر باشند. مقادیر مثبت یا منفی شایان توجه آنها به ترتیب نشان‌دهنده برآورد بیشتر یا کمتر از مقدار واقعی‌اند. در بیشتر موارد بین مقدار برآورد شده و مشاهده شده اختلاف‌هایی وجود دارد. بنابراین

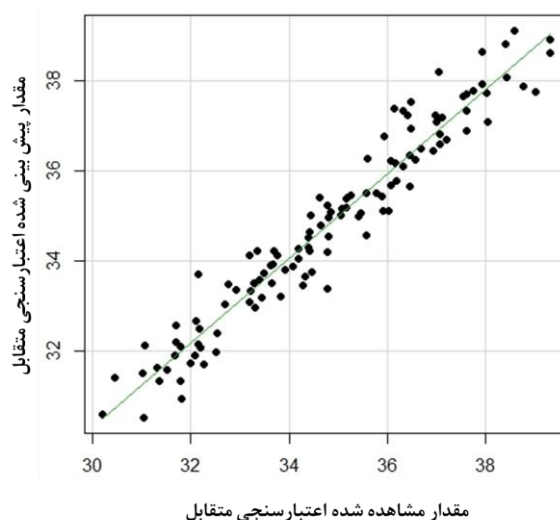


جدول ۳- نتایج مقایسه روش‌های مختلف درون‌یابی مکانی برای تخمین توان تولید رویشگاه

R <sup>2</sup> %	RMSEr	RMSE	MEr	MAE	ME	روش درون‌یابی
۱۵	۵/۸۶	۲/۰۴	-۰/۱۳	۱/۵۹	-۰/۰۵	کریجینگ
۲۴	۵/۵۵	۱/۹۳	-۰/۰۹	۱/۵۴	-۰/۰۳	رگرسیون کریجینگ
۸۳	۱/۶۱۵	۰/۵۶۲	-۰/۰۱۳	۰/۴۳۹	-۰/۰۰۵	رگرسیون کریجینگ جنگل تصادفی

رگرسیون کریجینگ با مدل مبنای جنگل تصادفی کمتر از ۲۰ درصد بود (جدول ۳). از این‌رو نتیجه‌گیری می‌شود که استفاده از روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی در مطالعه حاضر سبب بهبود بسیار معنی‌دار آماری در صحت تهیه نقشه نسبت به روش کریجینگ و روش رگرسیون با مدل مبنای رگرسیون خطی چندگانه شد. در شکل ۴ نمودار اعتبارسنجی متقابل برای توان تولید رویشگاه راش با استفاده از روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی به‌همراه بهترین خط برازش‌یافته ارائه شده است. همان‌طور که مشاهده می‌شود پراکندگی نقاط حاصل از اعتبارسنجی متقابل نسبت به خط برازش‌یافته کم است که بر کارایی بسیار خوب الگوریتم جنگل تصادفی دلالت دارد.

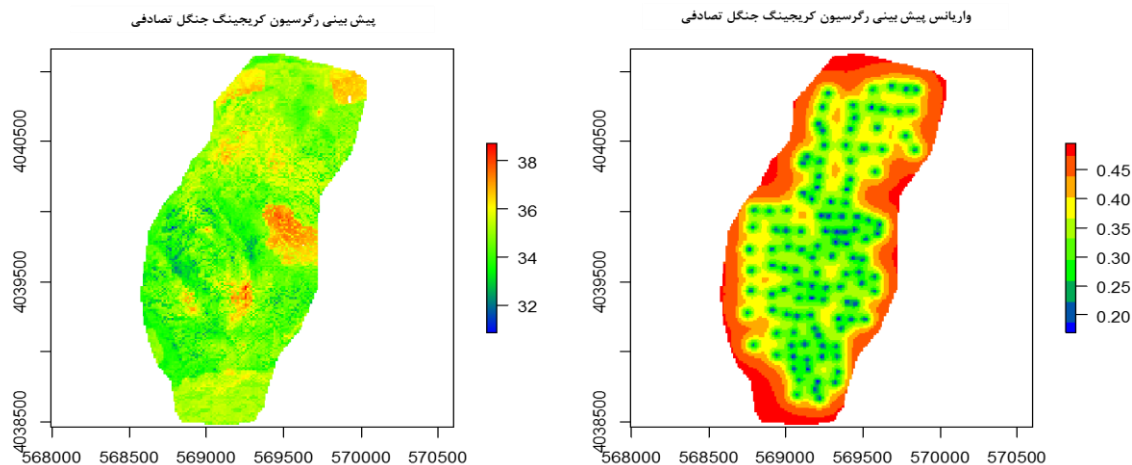
بر اساس معیارهای ارزیابی روش‌های مورد استفاده در تحقیق حاضر، عملکرد روش رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه، اندکی بهتر از روش کریجینگ معمولی بود، اما روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی عملکرد بسیاری خوبی داشت؛ به‌طوری که میزان RMSE را نسبت به دو روش یادشده بیشتر از ۷۰ درصد کاهش داد. نکته شایان توجه در تحقیق حاضر این بوده است که هرچند روش رگرسیون با مدل مبنای رگرسیون خطی چندگانه توانست نسبت به روش کریجینگ، ضریب تبیین را در حدود ۱۰ درصد بهبود بخشد، بخش بسیار زیاد تغییرپذیری در توان تولید توسط مدل رگرسیون خطی چندگانه توجیه نشد (حدود ۷۵ درصد)، در صورتی که تغییرپذیری توجیه‌نشده در توان تولید رویشگاه راش برای روش



شکل ۴- نمودار اعتبارسنجی متقابل روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی

میزان خطای آن در شکل ۵ نشان داده شده است.

نقشه حاصل از برآورد به روش رگرسیون کریجینگ با مبنای جنگل تصادفی به همراه



شکل ۵- نقشه توان تولید رویشگاه راش شرقی (الف) و میزان خطای آن (ب) به روش رگرسیون کریجینگ جنگل تصادفی

جدول ۴ نشان داده شده است.

نتایج مربوط به آماره‌های توصیفی پیش‌بینی روش رگرسیون کریجینگ با مبنای جنگل تصادفی در

جدول ۴- نتایج مربوط به آماره‌های توصیفی پیش‌بینی روش رگرسیون کریجینگ با مبنای جنگل تصادفی

ویژگی زیست‌سنجی	حداقل	حداکثر	میانگین	ضریب تغییرات
فرم رویشگاه (متر)	۳۰/۹۲	۳۸/۵۷	۳۴/۷۹	۲/۶۰

## بحث

رگرسیون کریجینگ با مدل مبنای رگرسیون خطی و روش کریجینگ معمولی در تهیه نقشه توان تولید رویشگاه راش دارد. هرچند عملکرد دو روش اخیر اختلاف چندانی با یکدیگر ندارد. زمانی که همبستگی قوی بین متغیر پاسخ و متغیرهای کمکی وجود نداشته یا ناچیز باشد، رگرسیون کریجینگ به کریجینگ خالص تبدیل می‌شود؛ اما چنانچه باقی‌مانده‌ها هیچ‌گونه خودهمبستگی مکانی را نشان ندهند، رگرسیون کریجینگ همان مدل رگرسیون خطی چندگانه است. با توجه به اینکه ضریب تبیین

هدف تحقیق حاضر، ارزیابی کارایی روش رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه و جنگل تصادفی در تهیه نقشه توان تولید رویشگاه راش شرقی بود. بررسی عملکرد روش‌های مذکور با استفاده از رایج‌ترین معیارهای ارزیابی همانند میانگین خطا، میانگین قدر مطلق خطا، ضریب تبیین، مجذور میانگین مربعات خطا نشان داد که روش رگرسیون کریجینگ با مدل مبنای جنگل تصادفی کارایی بسیار خوبی نسبت به روش

کاربردهای نیازمند روش‌های پیش‌بینی مکانی در مقیاس محلی، استفاده از روش رگرسیون کریجینگ مفید است، همخوانی دارد. از طرفی Palmer et al. (2009) بیان کردند که در مقیاس محلی، روش کریجینگ بهترین کارایی را در تهیه نقشه توان تولید *Pinus radiata* نسبت به رگرسیون کریجینگ و IDW داشته است. در مقیاس منطقه‌ای نتایج روش‌ها به نسبت مشابه بوده است.

شایان ذکر است که هرچند مدل رگرسیون خطی، در صورت استفاده مناسب، رویکرد ارزشمندی است، برای بسیاری از وضعیت‌های واقعی خیلی ساده است و توسط فرض‌های اساسی محدود می‌شود. به همین دلیل از چند دهه گذشته، پیشرفت‌های فزاینده‌ای در مدل‌های آماری گوناگون که در اکولوژی استفاده می‌شوند، صورت گرفته است. دقت پیش‌بینی بسیار زیاد، ارائه روش جدید در تعیین اهمیت نسبی متغیر، توانایی مدلسازی اثرهای متقابل پیچیده میان متغیرهای پیش‌بینی‌کننده، انعطاف‌پذیری برای انواع مختلف تجزیه و تحلیل داده‌های آماری از جمله رگرسیون، طبقه‌بندی و یک الگوریتم برای محاسبه داده‌های از دست‌رفته، از عمده ویژگی‌های الگوریتم جنگل تصادفی است که آن را از دیگر روش‌ها متمایز می‌کند (Cutler et al., 2007). همچنین زمانی که اثرهای متقابل پیچیده بین متغیرهای پیشگو و متغیر پاسخ وجود دارد و نیز زمانی که متغیرهای پیشگو هم‌خطی زیادی با هم دارند به کارگیری مدل جنگل تصادفی برای پیش‌بینی پراکنش (وقوع و وفور) یک گونه بسیار مفید است. مزیت دیگر مدل جنگل تصادفی این است که نیازمند فرض نرمال‌یته نیست و می‌تواند به روابط غیرخطی نیز بپردازد.

نکته حائز اهمیت در تحقیق حاضر، اندازه تفکیک DEM به کاررفته است. در تحقیقی Mitsuda et al. (2007) با استفاده از پنج نوع DEM با اندازه تفکیک ۱۲/۵ تا ۱۰۰ متر شاخص رویشگاه را برای جنگلکاری‌ها در ژاپن پیش‌بینی کردند. نتایج

در تحقیق حاضر برای مدل رگرسیون خطی چندگانه کم بوده (جدول ۳) و از طرف دیگر، ساختار مکانی باقی‌مانده‌ها هم در حد متوسط بوده است، روش رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه نتایجی مشابه مدل رگرسیون خطی چندگانه و کریجینگ دارد که از این نظر با پژوهش (Ahadi et al. 2017) درباره مقایسه دو روش کریجینگ و IDW در تهیه نقشه توان تولید رویشگاه راش شرقی مطابقت دارد.

رگرسیون کریجینگ از جمله روش‌هایی است که برای بررسی تغییرات مکانی متغیرهای پیوسته و دودویی پیشنهاد شده است. با به‌کارگیری روش رگرسیون خطی چندگانه برای مدلسازی توان تولید رویشگاه راش نسبت به متغیرهای اولیه و ثانویه توپوگرافی، درصد تغییرات بسیار زیادی تبیین نشد؛ از این رو می‌توان استنباط کرد که مدل رگرسیون خطی چندگانه برخلاف روش قدرتمند جنگل تصادفی، شایستگی لازم را برای بررسی توان تولید رویشگاه راش ندارد.

برخلاف مطالعه حاضر که در آن رگرسیون کریجینگ با مدل مبنای رگرسیون خطی چندگانه اختلاف فاحشی با روش کریجینگ نداشت، تحقیقات (Meng et al. 2009)، (Viana et al. 2012) و (Nussbaum et al. 2014) همگی تأییدکننده عملکرد بهتر روش رگرسیون کریجینگ نسبت به روش‌های کریجینگ و کوکریجینگ در مطالعات مربوط به جنگل بوده‌اند. تحقیق (Hengle et al. 2015) در خصوص به‌کارگیری الگوریتم جنگل تصادفی به‌عنوان مدل مبنای رگرسیون کریجینگ در تهیه نقشه ویژگی‌های خاک در آفریقا نشان داد که در مقایسه با مدل رگرسیون خطی چندگانه، روش جنگل تصادفی توانست مجذور میانگین مربعات خطا را ۱۵ تا ۷۵ درصد کاهش دهد که با تحقیق حاضر تطابق دارد. همچنین نتایج تحقیق حاضر با پژوهش (Berterretche et al. 2005) که بیان داشتند برای

بیش‌برازش رخ ندهد.

در پایان باید اشاره داشت که رگرسیون کریجینگ روشی انعطاف‌پذیر برای مدلسازی و تهیه نقشه است و با توجه به عملکرد بسیار خوب آن، امید است به‌عنوان بخشی معمولی در جعبه ابزار زمین‌آماری در نرم‌افزارهای GIS لحاظ شود. همان‌طور که پیشتر نیز اشاره شد در سال‌های اخیر روش‌های مدلسازی گوناگونی توسعه یافته است که از آن جمله می‌توان به روش‌های مختلف داده‌کاوی همانند درخت طبقه‌بندی و رگرسیون، ماشین‌های بردار پشتیبان، شبکه‌های عصبی، جنگل تصادفی و درخت رگرسیون تقویت‌شده اشاره داشت. این روش‌ها که به روش‌های یادگیری ماشین<sup>۱</sup> معروف‌اند، اغلب در علوم جنگل از روش‌های دیگر کمتر استفاده شده‌اند؛ از این رو یکی از وظایف برنامه‌نویسان در آینده نزدیک، تلفیق روش‌های یادگیری ماشین، مدلسازی خودکار واریوگرام، الگوریتم فراابتکاری<sup>۲</sup> تبریید شبیه‌سازی شده<sup>۳</sup> (به‌منظور بهینه‌سازی الگوی نمونه‌برداری)، طبقه‌بندی فازی نظارت‌نشده و مشابه آنها در محیط کاربری GIS خواهد بود. از آنجا که بسیاری از روش‌های آماری را می‌توان به‌صورت خودکار انجام داد، یکپارچه‌سازی GIS و الگوریتم‌های آماری، باید امکان درون‌یابی متغیرها را با استفاده از متغیرهای پیشگوی زیاد به‌راحتی و به‌سرعت فراهم کند. با وجود این، تحلیلگران باید نتایج را کنترل کنند و برای این کار دیدی کلی از الگوریتم‌های مورد استفاده داشته باشند.

نشان داد که بهترین مدل با استفاده از DEM با اندازه تفکیک ۱۲/۵ متر به‌دست آمده است. در مطالعه حاضر اندازه تفکیک DEM اولیه در حدود ۱۰۰ متر بود و بعداً به ۲۵ متر تغییر یافت. تحلیل‌های صورت‌گرفته براساس DEM اولیه رضایت‌بخش نبود. با در نظر گرفتن این موضوع که استفاده از DEM با دقت بیشتر سبب بهبود نتایج رگرسیون کریجینگ می‌شود، بلافاصله بعد از دریافت DEM با دقت بیشتر و اندازه تفکیک ۱۲/۵ متر، تمامی تحلیل‌ها از نو انجام گرفت که نتایج آن در تحقیق حاضر ارائه شده است. به‌طور معمول، تحلیل روابط گونه-محیط با بهره‌گیری از مدل رگرسیون حداقل مربعات ارزیابی می‌شود. از طرف دیگر، تجزیه و تحلیل داده‌های پیچیده اکولوژیکی نیازمند روش‌های تحلیلی قوی و انعطاف‌پذیر است. پژوهش حاضر نشان داد که به‌کارگیری روش جنگل تصادفی به‌عنوان مدل مبنای در روش رگرسیون کریجینگ، متدولوژی مؤثری بوده است.

باید اشاره داشت که محدودیت‌هایی برای استفاده رگرسیون کریجینگ وجود دارد. چنانچه هر کدام از این مشکلات رخ دهد، رگرسیون کریجینگ ممکن است نتایجی حتی بدتر از روش‌های درون‌یابی دیگر ارائه دهد؛ به‌عنوان مثال رگرسیون کریجینگ به‌طور کامل متکی بر کیفیت داده‌هاست. چنانچه داده‌ها با استفاده از طرح آزمایش‌های غلط نمونه‌برداری شده باشند، پیش‌بینی‌ها ممکن است حتی بدتر از روش‌های درون‌یابی ساده باشد. حتی یک داده پرت نیز می‌تواند پیش‌بینی رگرسیون کریجینگ را تحت تأثیر قرار دهد. نکته مهم دیگر برای مدلسازی واریوگرام، این است که تعداد کافی جفت نقطه در فاصله‌های مختلف باید در دسترس باشد. Webster and Oliver (2001) دست‌کم ۵۰ و ترجیحاً ۳۰۰ نقطه را برای برآورد واریوگرام توصیه می‌کنند. Hengl et al. (2007) تأکید کردند که رگرسیون کریجینگ فقط برای مجموعه داده‌هایی با بیش از ۵۰ مشاهده کلی استفاده شود تا مسئله

<sup>1</sup> Machine Learning

<sup>2</sup> Metaheuristic Algorithm

<sup>3</sup> Simulated Annealing

## References

- Aertsen, W., Kint, V., Van Orshoven, J., & Muys, B. (2011). Evaluation of modelling techniques for forest site productivity prediction in contrasting ecoregions using stochastic multicriteria acceptability analysis (SMAA). *Environmental modelling & software*, 26(7), 929-937.
- Aertsen, W., Kint, V., Von Wilpert, K., Zirlwagen, D., Muys, B., & Van Orshoven, J. (2012). Comparison of location-based, attribute-based and hybrid regionalization techniques for mapping forest site productivity. *Forestry: An International Journal of Forest Research*, 85(4), 539-550.
- Ahadi, Z., Alavi, S.J., & Hoseini, S.M. (2017). Beech forest site productivity mapping using ordinary kriging and IDW (Case study: research forest of Tarbiat Modares University). *Iranian Journal of Forest and Wood Product*, 70(1), 93-102.
- Ahmadi, K., Alavi, S.J., & Tabari Kouchaksaraei, M. (2015). Evaluation of oriental beech (*Fagus orientalis* L.) site productivity using generalized additive model (Case study: Tarbiat Modares University Forest Research Station). *Iranian Journal of Forest*, 7(1), 17-32.
- Akhavan, R., Karami Khorramabadi, M., & Soosani, J. (2012). Application of Kriging and IDW methods in mapping of crown cover and density of coppice oak forests (case study: Kakareza region, Khorramabad). *Iranian Journal of Forest*, 3(4): 305-316.
- Akhavan, R., & Kleinn, C. (2009). On the potential of kriging for estimation and mapping of forest plantation stock (Case study: Beneshki plantation). *Iranian Journal of Forest and Poplar Research*, 17(2), 303-318.
- Alavi, S.J., Nouri, Z., & Zahedi Amiri, Gh. (2017). Determining the most important environmental variables affecting on oriental beech (*Fagus orientalis* Lipsky.) site productivity using random forest technique in Khayroud forest, Nowshar. *Iranian Journal of Forest*, 8(4): 477-492.
- Berterretche, M., Hudak, A.T., Cohen, W.B., Maier sperger, T.K., Gower, S.T., & Dungan, J. (2005). Comparison of regression and geostatistical methods for mapping Leaf Area Index (LAI) with Landsat ETM+ data over a boreal forest. *Remote Sensing of Environment*, 96(1), 49-61.
- Breiman, L. (2001). Random forests, *Machine learning*, 45(1), 5-32.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., & Hess, K.T. (2007). Random Forests for Classification in Ecology. *Ecology*, 88 (11), 2783-2792.
- Fayad, I., Baghdadi, N., Bailly, J.S., Barbier, N., Gond, V., Héroult, B., & Perrin, J. (2016). Regional scale rain-forest height mapping using regression-kriging of spaceborne and airborne LiDAR data: application on French Guiana. *Remote Sensing*, 8(3), 240.
- Goovaerts, P. (1997). Geostatistics for natural resources evaluation. Oxford university press, New York, p. 496.
- Guisan, A., & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling*, 135(2), 147-186.
- Ghanbari, F., Shataee, Sh., Dehghani, A.A., & Ayoubi, Sh. (2009). Tree Density Estimation of Forests by Terrain Analysis and Artificial Neural Network, *Remote Sensing*, 8(3), 240.
- Harrell F.E., 2016. Hmisc: Harrell Miscellaneous. R package version 3.17-2. <https://CRAN.R-project.org/package=Hmisc>
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., & Tondoh, J.E. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PloS One*, 10(6), e0125814.

- Herrera, B., Campos, J.J., Finegan, B., & Alvarado, A. (1999). Factors affecting site productivity of a Costa Rican secondary rain forest in relation to *Vochysia ferruginea*, a commercially valuable canopy tree species. *Forest Ecology and Management*, 118(1), 73-81.
- Hengl, T., Heuvelink, G.B., & Rossiter, D.G. (2007). About regression-kriging: from equations to case studies. *Computers & geosciences*, 33(10), 1301-1315.
- Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., & Tondoh, J.E. (2015). Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLoS one*, 10(6), 1-26.
- Lehmann, A., Overton, J.M., & Leathwick, J.R. (2002). GRASP: generalized regression analysis and spatial prediction. *Ecological modelling*, 157(2), 189-207.
- Mitsuda, Y., Ito, S., & Sakamoto, S. (2007). Predicting the site index of sugi plantations from GIS-derived environmental factors in Miyazaki Prefecture. *Journal of Forest Research*, 12(3), 177-186.
- Minasny, B., & McBratney, A.B. (2007). Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma*, 140(4), 324-336.
- Meng, Q., Cieszewski, C., & Madden, M. (2009). Large area forest inventory using Landsat ETM+: a geostatistical approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(1), 27-36.
- Nussbaum, M., Papritz, A., Baltensweiler, A., & Walthert, L. (2014). Estimating soil organic carbon stocks of Swiss forest soils by robust external-drift kriging. *Geoscientific Model Development*, 7(3), 1197-1210.
- Pebesma, E.J. (2006). The role of external variables and GIS databases in geostatistical analysis. *Transactions in GIS*, 10(4), 615-632.
- Palmer, D.J., Höck, B.K., Kimberley, M.O., Watt, M.S., Lowe, D.J., & Payn, T.W. (2009). Comparison of spatial prediction techniques for developing *Pinus radiata* productivity surfaces across New Zealand. *Forest Ecology and Management*, 258(9), 2046-2055.
- Palmer, J.G., Watt, M.S., Kimberley, M.O., Hock, B.K., Payn, T.W., & Lowe, D.J. (2010). Mapping and explaining the productivity of *Pinus radiata* in New Zealand. *NZ JOURNAL OF FORESTRY*, 55(1), 15- 21.
- R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org/>.
- Skovsgaard, J.P., & Vanclay, J.K. (2008). Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. *Forestry: An International Journal of Forest Research*, 81(1), 13-31.
- Vanclay, J.K., & Henry, N.B. (1988). Assessing site productivity of indigenous cypress pine forest in southern Queensland. *The Commonwealth Forestry Review*, 53-64.
- Viana, H., Aranha, J., Lopes, D., & Cohen, W.B. (2012). Estimation of crown biomass of *Pinus pinaster* stands and shrubland above-ground biomass using forest inventory data, remotely sensed imagery and spatial prediction models. *Ecological Modelling*, 226, 22-35.
- Webster, R., & Oliver, M.A. (2001). *Geostatistics for environmental scientists (Statistics in Practice)*.



## Investigation on the potential of regression kriging for mapping oriental beech forest site productivity in research forest of Tarbiat Modares University

Z. Ahadi<sup>1</sup>, S.J. Alavi<sup>2\*</sup>, and S.M. Hosseini<sup>3</sup>

<sup>1</sup> M.Sc. Student, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, I. R. Iran

<sup>2</sup> Assistant Prof., Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, I. R. Iran

<sup>3</sup> Professor, Faculty of Natural Resources and Marine Sciences, Tarbiat Modares University, Noor, I. R. Iran

(Received: 3 July 2017, Accepted: 3 December 2017)

### Abstract

Forest resources mapping is a prerequisite for sustainable forest management. Site productivity is a key indicator of forest ecosystem services like wood production, carbon sequestration, etc. Due to the extent of Hyrcanian forests and mountainous areas in these forests that are sometimes difficult to access, it seems necessary to find suitable methods for mapping the quantitative parameters in these forests. In this study, site form index which is the most reliable criterion for evaluating site productivity of mixed and uneven stands was used. This study aims at mapping beech forest site productivity by using regression kriging in research forest of Tarbiat Modares University. For this purpose, 123 0.1 ha circular sample plots were laid out in beech dominated stands. The height and diameter of beech trees with  $DBH \geq 7.5$  cm within each plot was recorded. Some primary and secondary variables were also extracted from DEM in sample plots to be used in regression kriging. We investigated the differences between two predictive approaches: random forests and linear regression as the base model for regression kriging technique. Results of 10-fold cross-validation demonstrate that by using criteria such as mean error, mean absolute error, root mean square error, relative mean error, relative root mean square error, the random forests algorithm outperforms the linear regression and kriging techniques, with average decreases of ca. 70% in Root Mean Squared Error (RMSE). Hence, the regression kriging technique with random forest as the base model is recommended in order to better understand the more complex environment-beech forest site productivity relationships.

**Keywords:** Geostatistics, Interpolation, Kriging, Machine learning, Random forest, Regression.