قابلیت پیش‌بینی مدل‌های آماری در ارزیابی توان تولید رویشگاه راش شرقی

سمیه دهقانی نژاد[۱]، سید جلیل علوی[۲]، سید محسن حسینی[۳]

تاریخ دریافت: ۹۶/۴/۲۵ تاریخ پذیرش: ۹۷/۶/۱۹

چکیده

در مطالعه حاضر قابلیت پیش بینی مدل‌های خطی و جمعی تعمیم یافته برای ارتفاع غالب گونه راش به عنوان شاخصی عالی از کیفیت رویشگاه مورد بررسی قرار گرفته است. ارتفاع غالب در مطالعه حاضر به صورت میانگین ارتفاع سه اصله از مرتفع ترین درخت در هر قطعه نمونه تعریف می‌شود. به این منظور، در تیپ‌هایی که در آنها گونهٔ راش غالب بود، ۱۲۷ قطعهٔ نمونه دایره‌ای به مساحت ۱۰۰۰ مترمربع پیاده و در هر یک از آنها ارتفاع و قطر تمام درختان گونهٔ راش قطورتر از ۷/۵ سانتی‌متر علاوه بر ارتفاع از سطح دریا و درصد شیب و آزیموت اندازه‌گیری و ثبت شد. همچنین در مرکز هر قطعهٔ نمونه، از عمق ۱۰-۰ سانتی‌متری، نمونه‌برداری خاک صورت گرفت و متغیرهای فیزیکی و شیمیایی متعددی اندازه‌گیری شد. در تحقیق حاضر، عملکرد پنج روش گزینش متغیر به‌طور جداگانه برای هریک از مدل‌های خطی و جمعی تعمیم‌یافته مورد بررسی قرار گرفته است. به منظور مقایسه کارآیی روش‌های گزینش متغیر در مدل خطی تعمیم‌یافته، از اعتبارسنجی متقابل با ۲۵۰۰ تکرار و در مدل جمعی تعمیم‌یافته از اعتبارسنجی ۱۰-fold استفاده شده است. پس از انتخاب بهترین روش گزینش متغیر در هریک از دو مدل خطی و جمعی تعمیم‌یافته، اهمیت نسبی هر یک از متغیرهای مهم را محاسبه نموده که درنهایت متغیر ارتفاع از سطح دریا به‌عنوان مهمترین متغیر اثرگذار بر ارتفاع غالب گونه راش شناسایی شد. سپس با استفاده از معیارهای ارزیابی حاصل از داده‌های مدل‌سازی مشاهده گردید مدل جمعی تعمیم‌یافته از نظر معیارهای ارزیابی، عملکرد بهتری نسبت به مدل خطی تعمیم-یافته دارد.

واژه‌های کلیدی: توان تولید رویشگاه، عملکرد گونه، اعتبارسنجی، ضریب تبیین، مجذور میانگین مربعات خطا، مدل خطی تعمیم یافته، مدل جمعی تعمیم یافته.

۱- نویسنده مسئول: کارشناسی ارشد جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور، ایران.
Email: somaye8030@gmail.com

[۲] - استادیار گروه جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور، ایران

[۳] - استاد گروه جنگلداری، دانشکده منابع طبیعی و علوم دریایی، دانشگاه تربیت مدرس، نور، ایران

مقدمه

مدل‌سازی متغیر پاسخ با استفاده از متغیرهای تبیینی متعدد یکی از زمینه‌های مطرح در اکولوژی پوشش گیاهی است. این مدل‌ها برای اهداف مختلفی همانند پیش‌بینی پاسخ‌ها برای مشاهدات جدید (۱۹)، برآورد خطر (۸)، ایجاد فرضیه در مورد روابط علت و معلولی (۱۶) و ساخت مدل‌های همبستگی مکانی یا زمانی مقادیر پاسخ (۱۳،۱۷) بکار گرفته می‌شوند.

درسال‌های اخیر روش‌های مدل‌سازی متعددی با استفاده از متغیرهای محیطی برای پیش-بینی عملکرد گونه‌های گیاهی توسعه یافته و با موفقیت‌های مختلفی روبرو شده است که از آن جمله می‌توان به مدل خطی تعمیم‌یافته[1]، مدل جمعی تعمیم‌یافته[2] و همچنین تکنیک-های مختلف داده‌کاوی[3] همانند درخت طبقه-بندی و رگرسیون[4]، شبکه‌های عصبی[5] و درخت رگرسیون تقویت شده[6] اشاره داشت (۳).

مدل‌های خطی تعمیم‌یافته که بسط مدل رگرسیون عام می‌باشد، برای مواقعی که مشاهدات به طور نرمال توزیع نیافته‌اند یا جایی که سایر جنبه‌های مدل رگرسیون عام مناسب نمی‌باشد، ابداع شد. این مدل‌ها که در سال ۱۹۷۲ توسط Nelder و Wedderburn معرفی شده است به طور موفقیت‌آمیز در پژوهش‌های اکولوژیک مورد استفاده قرار گرفته است (۴۰).

استفاده از مدل خطی تعمیم‌یافته در اکولوژی پوشش گیاهی در مقایسه با رویکرد مدل‌سازی جمعی تعمیم‌یافته دارای نقاط ضعفی است که مهمترین آن این است که در مدل‌سازی رابطه عملکرد گونه و محیط، شکل منحنی پاسخ ممکن است از نظر زیستی، غیر واقع بینانه باشد (۱۱)، علاوه بر این مدل‌های خطی تعمیم‌یافته نیاز به پیش‌فرض روابط تابعی بین متغیر پاسخ و متغیرهای پیشگو دارند.

انتخاب نادرست فرض تابعی بودن شکل رابطه، روابط صحیح را می‌پوشاند و برآوردهای اریب-دار ارائه می‌دهد و باعث اشتباه در نتیجه‌گیری می‌شود. برای رفع این مسئله، مدل‌های جمعی تعمیم‌یافته، به عنوان یکی از روش‌های آماری مهم در چهل سال گذشته توسعه پیدا کرده است. این مدل، یک مدل ناپارامتری بوده و بسط مدل‌های خطی تعمیم‌یافته (که خود نیز بسط مدل‌های خطی عام می‌باشند) است. مدل جمعی تعمیم‌یافته روش بسیار مناسبی را برای تجزیه و تحلیل داده‌ها و بررسی رابطه بین متغیرهای وابسته و مستقل ارائه می‌دهد. مدل‌های جمعی تعمیم‌یافته (GAM) نسبت به مدل‌های خطی تعمیم‌یافته (GLM) از چند نظر برتری دارند و هدف از کاربرد این مدل‌ها به حداکثر رساندن کیفیت پیش‌بینی متغیر وابسته، کشف روابط غیرخطی بین متغیر وابسته و مجموعه متغیرهای تبیینی است (۱۰،۹). از آن‌جایی که مدل جمعی تعمیم‌یافته داده محور می‌باشد، سعی می‌نماید رابطه واقعی بین متغیر پاسخ و متغیرهای پیشگو را آشکارسازی نماید.

[1] Generalized Linear Models (GLM)
[2] Generalized Additive Models (GAM)
[3] Data Mining
[4] Classification and Regression Trees (CART)
[5] Artificial Neural Networks (ANN)
[6] Boosted Regression Trees (BRT)

انتخاب متغیر و برآورد ضرایب در مدل رگرسیونی اساسی‌ترین بخش در مدل‌سازی است. از دیدگاه عملی، انتخاب متغیر به تعیین این که کدام متغیر بیشترین تأثیر را بر روی متغیر وابسته دارد، معطوف می‌شود، در صورتی که از دیدگاه آماری روشی را برای دست یافتن به یک توازن بین نکویی برازش و اصل صرفه‌جویی ارائه می‌نماید. به عبارتی، با شناسایی موثر زیر مجموعه‌ای از متغیرهای مهم، انتخاب متغیر می‌تواند هم تفسیرپذیری مدل را تسهیل نماید و هم صحت پیش‌بینی را بهبود بخشد.

معیارهای بسیاری برای مقایسه آماری مدل های چند متغیره وجود دارد. اخیرا، معیارهای تئوری اطلاعات نظیر معیار اطلاعات آکائیک (AIC) و معیار اطلاعات بیزی شوارتز (BIC) مورد توجه اکولوژیست‌ها قرار گرفته است (۴، ۱۲، ۳۸). به عنوان مثال Mazerolle (۲۰۰۶) اشاره می‌نماید که آماره AIC در انتخاب مدل (مثلا انتخاب متغیر) نسبت به رویکردهای آزمون فرضیه به طور قابل توجهی ارجحیت دارد (۲۱) و Lukacs و همکاران ۲۰۰۷ بیان می‌دارند که تحلیل تبیینی داده‌ها براساس روش‌های آزمون فرضیه صفر همانند انتخاب گام به گام، اندیشه و استدلال را از تحلیل داده حذف می‌نماید (۲۰).

قدیمی‌ترین الگوریتم برای انتخاب متغیرهای تبیینی مراحل گام به گام هستند، که در آن متغیرهای پیشگو داوطلب برای وارد شدن به مدل غربال می‌گردند و متغیرهایی که از قبل در مدل هستند برای حذف‌های بعدی مورد بررسی قرار می گیرند. الگوریتم‌های مدرن‌تر شامل جستجوی کامل زیر مجموعه‌های بیشتری از متغیرهای پیشگو نسبت به رویکرد گام به گام می باشد که همین امر موجب تردید اکولوژیست‌ها در سودمندی روش انتخاب گام به گام متغیر گردیده است. Whittinghamو همکاران (۲۰۰۶)، به عنوان مثال، نسبت به استفاده گسترده از رویکردهای گام به گام در مجلات بوم‌شناختی، با توجه به اریب‌ها و نواقص مسلم رگرسیون گام به گام چندگانه، اظهار تأسف نموده است (۳۹). Mundry & Nunn (۲۰۰۹) نیز همانند دیگران توصیه می‌نمایند زیست‌شناسان در به کارگیری این روش‌ها اجتناب می‌نمایند (۲۵). محققین بسیاری روش‌های انتخاب متغیر را مورد بررسی و ارزیابی قرار دادند (۲۸، ۳۲، ۳۸، ۱۷). Raffalovich و همکاران (۲۰۰۸) مرور جامعی در این زمینه انجام دادند (۲۹). شبیه‌سازی تنها روشی است که در آن مدل واقعی شناخته می‌شود، اما نتایج مطالعات شبیه‌سازی به روش ایجاد داده‌های شبیه سازی و معیارهای انتخاب شده جهت مقایسه تکنیک‌های مختلف مدل‌سازی، وابسته است. بعنوان مثال Murtaugh (۱۹۹۸) دریافت که رویکردهای متکی بر آزمون F و Cp و BIC فراوانی یکسانی از تصمیمات درست در خصوص ورود یا حذف متغیرهای تبیینی ارائه می‌دهد (۲۶). رویکردهای مبتنی بر درخت رگرسیون، نسبت به روش‌های رگرسیونی، به طور قابل توجهی عملکرد ضعیفی دارند. اما Murtaugh (۱۹۹۸) با توجه به اینکه داده‌ها براساس یک مدل رگرسیون خطی شبیه‌سازی

شده بودند، اعتبار مقایسه را مورد سوال قرار داد (۲۶).

از دیدگاه جنگلداری، توان تولید، پتانسیل رویشگاه را برای تولید چوب برای یک گونه مشخص بازگو می‌کند (۳۶)؛ از این‌رو یکی از وظایف خطیر مدیران جنگل کمی‌سازی و سنجش اختلافات و تفاوت‌ها در رویشگاه‌ها است (۳۷). از آن جایی که رویشگاه در تعیین میزان رشد (که یک فاکتور بسیار مهم و کلیدی در برنامه‌های مدیریت جنگل می‌باشد) نقش بسیار مهم ایفا می‌نماید، کیفیت آن نیز میزان رشد و توان تولید را تحت تأثیر قرار می‌دهد (۲۲). یکی از معیارهای مهمی که تحت تأثیر کیفیت رویشگاه می‌باشند، ارتفاع غالب بوده که به صورت ارتفاع ۱۰۰ اصله از قطورترین درختان در هکتار بیان می‌شود(۲). این معیار برخلاف قطر برابر سینه، کم‌تر تحت تأثیر عملیات پرورشی در جنگل قرار می‌گیرد و رابطه نزدیکی با حجم درختان دارد، به همین دلیل معیار مناسبی برای بررسی کیفیت رویشگاه می‌باشد (۲۶).

در تحقیق حاضر، عملکرد پنج روش گزینش متغیر به‌طور جداگانه برای هریک از مدل‌های خطی و جمعی تعمیم‌یافته مورد بررسی قرار گرفته است. از آنجایی که مدل صحیح شناخته شده‌ای از متغیر پاسخ برحسب متغیرهای تبیینی وجود ندارد، اعتبارسنجی متقابل برای به دست آوردن برآورد درست از توانایی پیش بینی استفاده می‌گردد که می‌تواند مقایسه منصفانه‌ای از روش‌های انتخاب متغیر باشد.

مواد و روش‌ها

منطقه مورد بررسی

پژوهش حاضر در جنگل آموزشی و پژوهشی صلاح‌الدین کلا متعلق به دانشکده منابع طبیعی دانشگاه تربیت مدرس صورت گرفته است. منطقه مورد بررسی در سری ۳ آغوزچال از طرح جنگلداری حوزه ۴۶ کجور واقع شده است. این سری در قسمت شمال شرقی طرح کجور و درحوزه استحفاظی اداره کل منابع طبیعی و آبخیزداری استان مازندران- نوشهر، اداره منابع طبیعی و آبخیزداری شهرستان نوشهر و منابع طبیعی توسکاتک قرار دارد. مساحت منطقه ۱۸۱۷ هکتار و ارتفاع از سطح دریا از ۱۵۰ تا ۱۷۵۰ متر می‌باشد. منطقه مورد مطالعه بین عرض جغرافیایی "۳۶°۲۹'۲۳ تا "۳۶°۳۲'۵۶ و طول جغرافیایی "۵۱°۴۳'۲۰ تا "۵۱°۴۷'۳۹ قرار گرفته است. میزان بارندگی در این ناحیه بر اساس داده‌های ایستگاه هواشناسی نوشهر ۱۳۰۸ میلی متر در سال می‌باشد. گرم‌ترین ماه سال تیر و مرداد با میانگین دمای ۲۹/۲ و سردترین ماه سال، بهمن‌ماه با میانگین دمای ۲/۶ ° است. همچنین میانگین دمای سالانه برابر با ۱۵/۹ ° ثبت شده است. تیپ جنگل در منطقه مورد مطالعه آمیخته‌ای از راش به همراه گونه ممرز، انجیلی، افرا، شیردار، نمدار، ون و بلوط می باشد. جنگل دارای کلاسه سنی میانسال، مسن و جوان بوده و دارای تاج پوشش حدود ۷۰ تا ۸۵ ٪ و ۲ تا ۳ اشکوبه می‌باشند. جنگل دارای تراکم انبوه تا نیمه انبوه با ساختار دانه زاد ناهمسال نامنظم است. به طور میانگین گونه

راش حدود ۵۰ درصد از کل حجم سرپا و حدود ۴۰ درصد از کل تعداد درختان را در منطقه مورد مطالعه به خود اختصاص داده است. در سری ۳ آغوزچال هیچ واحد دامداری در داخل جنگل وجود ندارد و در حال حاضر تنها ، دامداران جهت ییلاق قشلاق از جنگل عبور می‌نمایند. روستائیان منطقه نیز از جنگل استفاده تفریحی و توریستی می‌نمایند (۷).

روش بررسی

به منظور جمع‌آوری اطلاعات مورد نیاز از ۱۲۷ قطعات نمونه دایره‌ای شکل با مساحت ۱۰۰۰ مترمربع در توده‌هایی که در آن‌ها گونه راش غالب بوده است استفاده شد. پس از پیاده نمودن قطعات نمونه، در داخل هر یک از آن‌ها مشخصه‌های هر درخت از جمله نوع گونه، قطر برابر سینه و همچنین ارتفاع کامل تمام درختان راش با قطر بیشتر از ۷/۵ سانتی‌متر مورد اندازه‌گیری قرار گرفت. اندازه‌گیری ارتفاع کامل درختان راش در هر قطعه نمونه با استفاده از دستگاه Vertex 3 انجام گرفت. ارتفاع غالب که در اینجا به صورت میانگین ارتفاع سه اصله از مرتفع‌ترین درخت در هر قطعه نمونه تعریف می‌شود، معمولاً به عنوان شاخصی عالی از کیفیت رویشگاه شناخته می‌شود. ویژگی‌های عمومی رویشگاه مثل جهت جغرافیایی (آزیموت)، ارتفاع از سطح دریا و درصد شیب نیز یادداشت گردید. جهت جغرافیایی با استفاده از رابطه زیر به شاخص تابش خورشیدی[1] تبدیل گردید که در آن θ مقدار آزیموت جهت بر حسب درجه می‌باشد.

مقدار شاخص تابش خورشیدی بین صفر و یک می‌باشد و جهت شمال شرقی دارای مقدار صفر (خنک‌ترین دامنه) و جهت جنوب غربی دارای مقدار یک (گرمترین دامنه) می‌باشد (۱).

$$TRASP = [1 - \cos((\pi/180)(\theta - 30))]/2 \qquad (۱)$$

در مرحله بعد در هر قطعه نمونه، نمونه‌برداری خاک از عمق ۰ تا ۱۰ سانتی‌متر صورت گرفت. نمونه‌ها خاک همچنین در مرکز هر قطعه نمونه، از لایه اول نمونه‌برداری خاک صورت گرفت. پس از خشک کردن نمونه‌ها در هوای آزاد و جدا کردن ریشه‌ها، سنگ و سایر ناخالصی‌ها از آن، کلوخ‌ها خرد و از الک ۲ میلی‌متری عبور داده شد. در بررسی فاکتورهای خاک، مجموعه‌ای از خصوصیات فیزیکی و شیمیایی خاک از جمله وزن مخصوص ظاهری به روش کلوخ و پارافین، بافت خاک به روش هیدرومتری، pH خاک با دستگاه pH متر الکتریکی، درصد ماده آلی به وسیله روش Walkley & Black، فسفر قابل‌جذب به روش السون، پتاسیم به وسیله دستگاه طیف‌سنج اتمی و نیتروژن به روش کجلدال اندازه‌گیری شد (۱۴).

در مطالعه حاضر به منظور بررسی رابطه بین ارتفاع غالب گونه راش و متغیرهای خاکی و فیزیوگرافی از مدل خطی و جمعی تعمیم‌یافته در نرم‌افزار R استفاده گردید. در این پژوهش با توجه به ماهیت متغیر ارتفاع غالب بعنوان متغیر پاسخ، از توزیع گوسی استفاده شده است.

[1]- Radiation Index

روش‌های انتخاب متغیر

هدف از روش‌های انتخاب متغیر انتخاب زیرمجموعه‌ای از متغیرهای پیشگو است که در مفهوم ارائه شده مهم‌ترین هستند. به‌طورکلی متغیرهای پیشگو زیاد در مدل باعث کم شدن اریبی پیش‌بینی‌ها می‌شوند، اما دارای واریانس بالاتری هستند. وجود متغیرهای پیشگو زیاد در مدل رگرسیونی به‌طور معمول باعث (بیش برازش) Over-fitting می‌شود درحالی‌که مخالف آن Under-fitting نامیده می‌شود (۳۳). در چنین مواردی مجموعه‌ای از شیوه-های انتخاب متغیر پیشنهاد می‌شود. مهم‌ترین روش‌ها شامل موارد زیر می‌باشند:

۱- شیوه انتخاب پیشرو: این شیوه با معادله‌ای که تنها یک جمله ثابت دارد و هیچ متغیر پیشگویی را دربر ندارد، شروع می‌شود. متغیر اولی که در معادله وارد می‌شود متغیری است که بیشترین همبستگی ساده را با متغیر پاسخ Y دارد. اگر ضریب رگرسیون این متغیر اختلاف معنی‌داری با صفر داشته باشد در معادله باقی می‌ماند و برای متغیر دوم تحقیق می‌شود. این روش وقتی خاتمه می‌یابد که آخرین متغیری که وارد معادله می‌شود یک ضریب رگرسیون معنی‌دار نداشته باشد و خطای آزمون به ۵ درصد برسد (۶).

۲- شیوه حذف پسرو: در این روش ابتدا کلیه متغیرهای مستقل وارد معادله شده و اثر کلیه متغیرها بر روی متغیر وابسته سنجیده می-شود. اما در این روش به مرور متغیرهای ضعیف‌تر، یکی پس از دیگری از معادله خارج شده و نهایتا این مراحل تا زمانی ادامه می‌یابد که خطای آزمون معنی‌داری به ۱۰ درصد برسد (۱۵).

۳- شیوه گام‌به‌گام: این روش در اصل یک شیوه‌ی انتخاب پیشرو است ولی با این شرط اضافی که در هر مرحله امکان حذف یک متغیر را مانند حذف پسرو در نظر می‌گیرد. در این شیوه، متغیری که در مراحل قبلی انتخاب وارد معادله شده است ممکن است در مراحل بعدی حذف شود (۶). این روش زمانی پایان می‌یابد که خطای آزمون معنی‌داری به ۵ درصد برسد (۱۵).

۴- ارزیابی تمام معادلات ممکن: لازمه‌ی این روش این است که تمام معادلات رگرسیون دارای یک متغیر مستقل، دو متغیر مستقل، سه متغیر مستقل والی آخر را برازش داد. این معادلات با استفاده از معیار مناسب ارزیابی شده و بهترین مدل رگرسیون گزینش می-گردد. اگر فرض شود که جمله β_0 در تمام معادلات وجود داشته باشد و اگر k متغیر مستقل وجود داشته باشد، 2^k معادله باید بررسی شوند (۳۱).

معیارهای انتخاب متغیر

از جمله معیارهایی که برای ارزیابی پتانسیل زیرمجموعه‌های متغیرهای پیش‌بینی کننده استفاده می‌شود شامل موارد زیر است (۳۳).

ضریب تبیین (R^2)

ضریب تبیین مدل رگرسیون، به‌عنوان نسبتی از تغییرپذیری متغیر پاسخ توصیف شده توسط

مدل رگرسیون است که به‌صورت ذیل تعریف شده است:

$$R^2 = \frac{SSreg}{SST} = 1 - \frac{RSS}{SST} \quad (۲)$$

در عمل معمول است متغیری که دارای مقدار R^2 بیشتری است به‌عنوان متغیر مهم‌تر انتخاب شود. که این معیار بر پایه حداقل مربعات است و سایر معیارهای ارزیابی بر اساس تئوری حداکثر احتمال می‌باشند (۳۴).

AIC (معیار اطلاعات آکائیک)

معیار اطلاعات آکائیک به دو روش ایجاد می-شود. عمومی‌ترین روش مبتنی بر تعادل نیکویی برازش و سادگی مدل است، و این‌چنین تعریف شده است که مدل و متغیر با مقادیر پایین‌تر AIC مناسب‌ترین مدل و با اهمیت‌ترین متغیر می‌باشد. و رابطه آن به‌صورت ذیل می‌باشد (۵):

$$AIC = n\log\left\{\frac{RSS}{n}\right\} + 2P \quad (۳)$$

یا

$$AIC = n\ln(RMSE) + 2P \quad P = K - 2 \quad (۴)$$

K (تعداد پارامترهای تخمین زده شده در مدل برازش داده شده) برابر با حد پیچیدگی و p تعداد متغیرهای مدل است (۱). معیار AIC وقتی‌که N به سمت بی‌نهایت میل می‌کند تمایل به انتخاب مدل‌هایی دارد که خیلی پیچیده هستند.

BIC (معیار اطلاعات Bayesian)

شوارتز[1] (۳۵) معیار اطلاعات Bayesian را پیشنهاد داد و به این صورت تعریف شده است که مدل و متغیر دارای مقدار BIC پایین‌تر مدل مناسب‌تر و متغیر مهم‌تر می‌باشد. این معیار برای نمونه‌های محدود به علت امتیازهای منفی زیادی که برای پیچیدگی در نظر می‌گیرد اغلب مدل‌هایی که خیلی ساده‌اند را انتخاب می‌کند (۳۴). و رابطه آن به‌صورت زیر می‌باشد (۱).

$$BIC = n\ln(RMSE) + P\ln(n) \quad (۵)$$

معیارهای مورد استفاده برای انتخاب متغیر

در این مطالعه به دلیل پایگاه اطلاعات داخلی منطقه، وجود تعداد زیاد متغیرهای مستقل، دقت در مدل‌سازی، و هم‌چنین به دلیل نتایج حاصل از منابع موجود، انتخاب متغیر صورت گرفت. بنابراین به منظور انتخاب متغیرهای تأثیرگذار بر عملکرد گونه راش در مدل خطی تعمیم‌یافته از روش گام‌به‌گام با استفاده از معیارهای AIC و BIC[2] و روش بهترین زیر مجموعه با بهره‌گیری از بسته bestglm استفاده شده است. برای انتخاب متغیرهای مهم در مدل جمعی تعمیم‌یافته در بسته gam از روش گام‌به‌گام و معیارهای AIC و BIC در ترکیب با بسته GRASP (Generalized Regression Analysis and Spatial Prediction) که توسط Lehmann و همکاران (۲۰۰۲) ارائه گردید (۱۸)، استفاده شد. به منظور انتخاب مهمترین متغیرها در بسته

[1] Schwarz
[2] Akaike and Baysian Information Criterion

mgcv از روش پیشرو با استفاده از معیار AIC و یا روش انقباض[1] و روش بهترین زیرمجموعه استفاده گردید. به منظور مقایسه کارآیی روش‌های گزینش متغیر در مدل‌های خطی و جمعی تعمیم‌یافته از اعتبارسنجی متقابل استفاده گردید. به همین دلیل داده‌ها به دو بخش مدل‌سازی و اعتبارسنجی تقسیم شدند. ۷۵ درصد داده‌ها به بخش مدل‌سازی و ۲۵ درصد باقیمانده نیز به بخش اعتبارسنجی تخصیص داده شد. برای مدل خطی تعمیم-یافته از اعتبارسنجی متقابل با ۲۵۰۰ تکرار استفاده شده است. هم‌چنین برای گزینش متغیر در مدل خطی تعمیم‌یافته از اعتبارسنجی ۱۰-fold استفاده گردید.

معیارهای ارزیابی مدل

در این مطالعه به منظور مقایسه کارآیی مدل-های خطی و جمعی تعمیم‌یافته، از معیارهای ارزیابی حاصل از داده‌های مدل‌سازی نظیر ضریب تبیین[2] و مجذور میانگین مربعات خطا[3] و ضریب تبیین تعدیل یافته[4] استفاده شده است، که روابط آنها به ترتیب در زیر آمده است:

$$R^2 = \frac{SSreg}{SST} = 1 - \frac{RSS}{SST} \quad (۶)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - Q^\wedge_i)^2}{n}} \quad (۷)$$

$$R^2_{adj} = 1 - \frac{(n-1)\sum_{i=1}^{n}(h_i - \hat{h_i})^2}{(n-p)\sum_{i=1}^{n}(h_i - h^-)^2} \quad (۸)$$

که Q_i مقدار مشاهده شده، $Q^\wedge_i$ مقدار برازش یافته، n تعداد مشاهده‌ها، p تعداد متغیرهای پیشگو، h_i مقدار مشاهده شده و $\hat{h_i}$ مقدار برآورد شده می‌باشد.

نتایج

در جدول ۱ برای هر یک از روش‌های انتخاب متغیر، متوسط تعداد متغیرهای تبیینی موجود در مدل‌های برازش یافته حاصل از مجموعه داده‌های آموزشی برای متغیر ارتفاع غالب، ارائه شده است. در مدل خطی تعمیم‌یافته برازش داده شده، میانگین تعداد متغیرها در روش بهترین زیرمجموعه BIC در مقایسه با سایر روش‌ها کمتر می‌باشد. همچنین در مدل جمعی تعمیم‌یافته برازش یافته، میانگین تعداد متغیرها در روش بهترین زیرمجموعه BIC، در مقایسه با سایر روش‌ها کمتر می‌باشد.

شکل ۱ نمودار خلاصه گرافیکی از نتایج مدل خطی تعمیم‌یافته ارائه شده در جدول ۱ را نمایش می‌دهد. به‌طور کلی، در مدل خطی تعمیم‌یافته روش‌های مبتنی بر AIC مدل‌های بزرگتری را ایجاد نمودند؛ به این دلیل که AIC جریمه نسبتاً کوچکی را برای پیچیدگی مدل در نظر می‌گیرد. مدل‌های حدواسط با استفاده از روش مبتنی بر آزمون F و BIC ایجاد شدند. از طرف دیگر، RMSE برای مدل‌های مبتنی بر BIC بیشتر از AIC می-باشد.

[1] Shrinkage Method
[2] Coefficient of Determination
[2] Root Mean Square Error (RMSE)
[4] Adjusted Coefficient of Determination

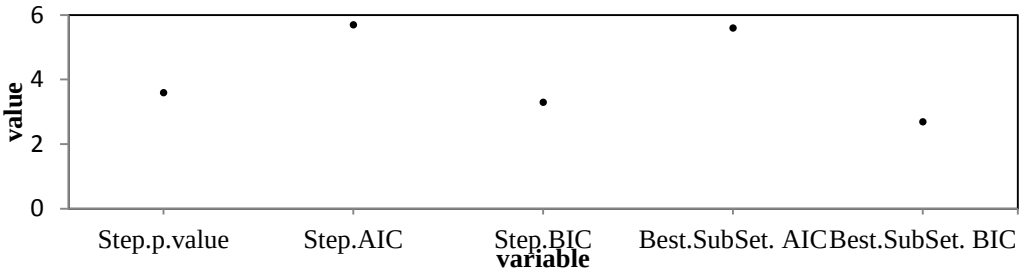
در مدل جمعی تعمیم‌یافته نیز مدل بزرگتر با استفاده از روش مبتنی بر AIC و انقباض ایجاد شده و مدل‌های حدواسط با استفاده از روش‌های پیشرو و روش مبتنی بر BIC و بسته GRASP ایجاد شدند. شکل ۲ نمودار گرافیکی از نتایج مدل جمعی تعمیم‌یافته در جدول ۱ را نمایش می‌دهد.

جدول ۱. میانگین تعداد متغیرهای انتخاب شده، ضریب تبیین و RMSE برای معیار ارتفاع غالب

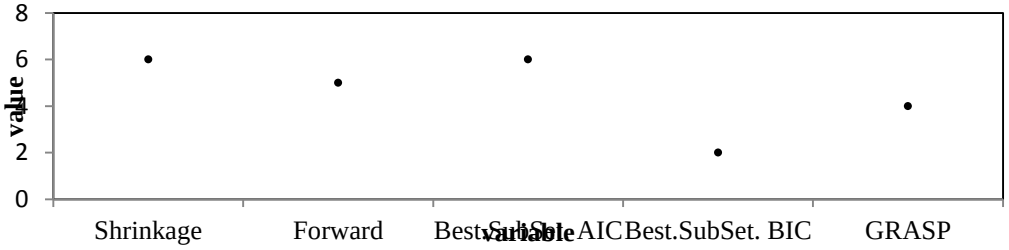
	GLM				GAM					
مدل / معیار	Step. p.value	Step. AIC	Step. BIC	Best. SubSet.AIC	Best. SubSet.BIC	Shrinkage	Forward	Best. SubSet.AIC	Best. SubSet.BIC	Grasp
میانگین تعداد متغیر انتخاب شده	۳/۶	۵/۷	۳/۳	۵/۶	۲/۷	۶	۵	۶	۲	۴
ضریب تبیین برای داده مدل-سازی	۳۲/۸۴	۳۶/۷۷	۳۲/۰۳	۳۶/۲۳	۳۰/۳۳	۳۸	۳۶	۳۸	۲۷	۳۱/۱۱
RMSE برای داده مدل‌سازی	۳/۴۸	۳/۳۸	۳/۵۱	۳/۴۰	۳/۵۶	۳/۳۷	۳/۴۳	۳/۳۶	۳/۶۵	۳/۵۴
ضریب تبیین برای داده اعتبارسنجی	۲۰/۶۸	۲۰/۸۵	۲۰/۹۸	۲۰/۵۶	۲۰/۴۹	۲۹/۵۰	۳۰/۸۳	۳۱/۵۴	۳۲/۲۱	۲۴/۳۸
RMSE برای داده اعتبارسنجی	۳/۹۳	۳/۹۵	۳/۹۲	۳/۹۳	۳/۹۱	۳/۷۴	۳/۷۳	۳/۷۳	۳/۷۶	۳/۷۲

با توجه به ضریب تبیین و RMSE حاصل از داده‌های مدل‌سازی برای مدل خطی تعمیم-یافته در جدول فوق، به دلیل اینکه بیشترین ضریب تبیین و کمترینRMSE بدست آمده مربوط به روش step.AIC است پس عملکرد روش مذکور از بقیه روش‌ها مناسب‌تر است. اما انتخاب روش بهینه گزینش متغیر، براساس داده‌های اعتبارسنجی صورت می‌گیرد. با توجه به معیارهای بدست آمده برای این داده‌ها، عملکرد روش‌ها تقریبا شبیه هم می‌باشد ولی روش step.BIC عملکرد بهتری دارد. از آنجا که بین دقت و پیچیدگی مدل باید توازن و تعادلی برقرار باشد، روش step.BIC به‌عنوان روش گزینش متغیر در مدل خطی تعمیم‌یافته برای ارتفاع غالب انتخاب می‌شود.

در مدل جمعی تعمیم‌یافته برای متغیر ارتفاع غالب نیز روش بهترین زیرمجموعه AIC بهترین عملکرد را دارد. همان‌طور که اشاره شد بین دقت مدل و پیچیدگی مدل باید توازنی برقرار باشد. از این رو روش بهترین زیرمجموعه AIC به‌عنوان روش گزینش متغیر انتخاب می‌گردد.



شکل ۱. میانگین تعداد متغیرهای انتخاب شده برای ارتفاع غالب با استفاده از پنج روش در مدل خطی تعمیم‌یافته



شکل ۲. میانگین تعداد متغیرهای انتخاب شده برای ارتفاع غالب با استفاده از پنج روش در مدل جمعی تعمیم‌یافته

تعیین اهمیت هر کدام از متغیرهای مهم

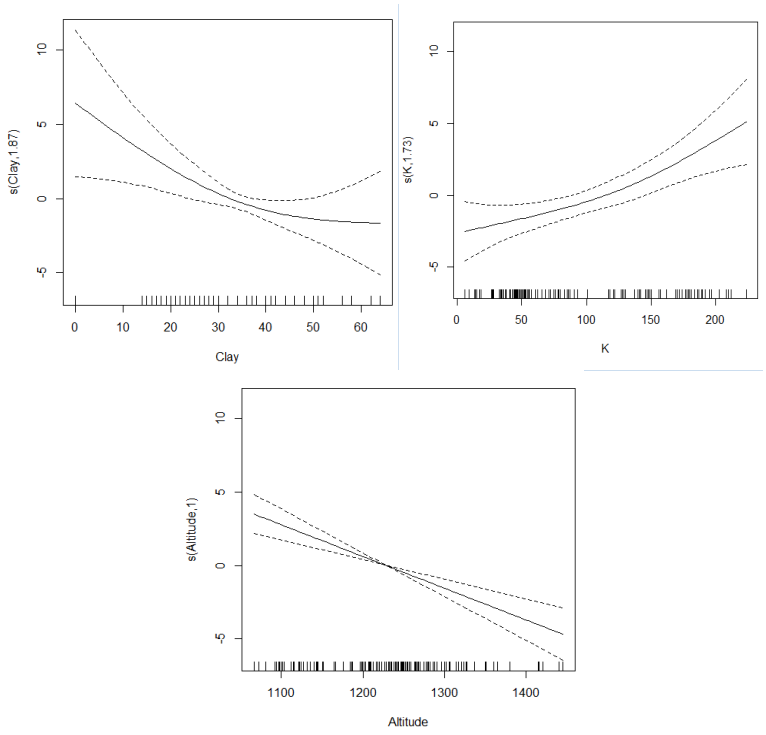
پس از انتخاب روش گزینش متغیر در هریک از مدل‌ها، اهمیت نسبی هر یک از متغیرهای مهم را به دست آورده و مهمترین متغیر اثرگذار را در دو مدل خطی و جمعی تعمیم-یافته برای معیارهای عملکرد گونه راش انتخاب نمودیم. برای به دست آوردن اهمیت نسبی متغیرها از بسته MuMIn در نرم‌افزار R استفاده شده است. که نتایج آن را در جدول ۲ مشاهده می‌کنید.

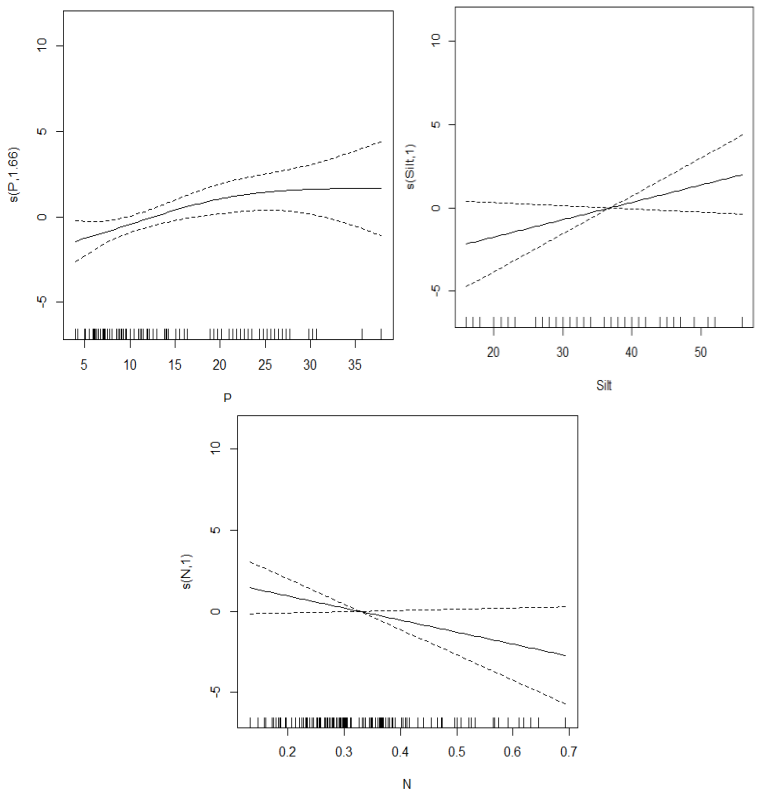
جدول ۲. اهمیت نسبی متغیرهای مهم بر معیارهای عملکرد گونه راش

معیار عملکرد	ارتفاع غالب	
متغیر	**GLM**	**GAM**
	Step.BIC	Best.SubSet.AIC
Altitude	۰/۹۹۹۸	۱
Bulk_Density	–	–
Clay	۰/۹۹۹۴	۰/۸۹
CN	–	–
K	۰/۷۵۶۵	۰/۹۷
N	۰/۸۰۷۸	۰/۶۹
P	–	۰/۸۹
pH	–	–
Silt	–	۰/۷۵
Slope	–	–
TRASP	–	–

با توجه به جدول (۲) برای معیارهای عملکرد گونه راش در مدل‌های خطی و جمعی تعمیم-یافته ، متغیر ارتفاع از سطح دریا دارای بیش-ترین درجه اهمیت بوده و به عنوان مهم‌ترین متغیر محیطی اثرگذار بر هر سه معیار عملکرد گونه راش شناسایی گردید.

شکل‌های زیر نوع ارتباط معیارهای عملکرد گونه راش را با متغیرهای محیطی را نشان می‌دهند.





شکل ۳. نمودار رابطه بین متغیرهای محیطی اثرگذار و معیار عملکرد ارتفاع غالب گونه راش

قابلیت پیش‌بینی

مقایسه روش‌های مبتنی بر رگرسیون

در مدل خطی تعمیم‌یافته مقایسه مجذور میانگین مربعات خطا و ضریب تبیین برای روش‌های مبتنی بر رگرسیون حکایت از شباهت مقادیر مرتبط با رویکردهای مبتنی بر رگرسیون دارد. با وجود تعداد زیاد متغیر پیشگو در روش‌های مبتنی بر AIC در مقایسه با روش‌های F-test و BIC که منتج به ضریب تبیین بیشتر برای داده‌های مدل‌سازی گردیده است، متوسط قابلیت پیش‌بینی تمام پنج روش تقریباً یکسان بوده است.

مقایسه عملکرد مدل‌های خطی و جمعی تعمیم‌یافته

در این مطالعه به‌منظور ارزیابی عملکرد مدل خطی و جمعی تعمیم‌یافته از معیارهای ضریب تبیین و مجذور میانگین مربعات خطا حاصل از داده‌های مدل‌سازی استفاده گردید. با توجه به جدول شماره ۳ مدل جمعی تعمیم‌یافته بیشترین ضریب تبیین و کمترین مجذور میانگین مربعات خطا را دارد بنابراین مدل مذکور عملکرد بهتری نسبت به مدل خطی تعمیم‌یافته دارد.

جدول ۳. معیارهای ارزیابی مدل‌های خطی و جمعی تعمیم یافته برای ارتفاع غالب گونه راش

معیار عملکرد	ارتفاع غالب

معیار ارزیابی	RMSE	R^2%	R^2_{adj}%
مدل خطی تعمیم‌یافته	۳/۵۵	۳۱	۲۹
مدل جمعی تعمیم‌یافته	۳/۳۶	۳۸	۳۵

بحث و نتیجه‌گیری

بررسی شایستگی و کفایت مدل رگرسیون خطی تعمیم‌یافته

در مطالعه حاضر عملکرد روش‌های مختلف انتخاب متغیر در مدل خطی تعمیم‌یافته مورد بررسی قرار گرفت. قابلیت پیش‌بینی روش‌های مبتنی بر رگرسیون تقریباً مشابه بوده است. هر چند اندازه مدل‌ها متفاوت بوده است اما در خصوص قابلیت پیش‌بینی، تفاوت اندکی بین روش گام‌به‌گام و روش بهترین زیرمجموعه‌ها وجود دارد. این نتایج با اظهارات برخی از بوم-شناسان در مورد ضعف روش گام‌به‌گام مغایرت دارد (۲۵، ۳۹).

روش‌های مبتنی بر معیار اطلاعات آکائیک-هایی با بیشترین تعداد متغیر پیشگو را ایجاد نمودند. از آن‌جا که قابلیت پیش‌بینی این مدل‌ها به‌طور قابل توجهی بهتر از مدل‌های حاصل ازروش آزمون F و BIC نبوده است، از این‌رو انتخاب تکنیک‌های گزینش متغیر باید براساس ملاحظات دیگر صورت گیرد. همان-طور که Murtaugh (۱۹۹۸) اشاره می‌نماید محقق ممکن است یک روش گزینش متغیر محافظه‌کارانه‌تر (F-test یا BIC) و یا روش-های آزادانه‌تر (AIC) را ترجیح دهد.

بسیاری از متخصصان آمار اشاره دارند که تفسیر p-value متغیرهای تبیینی در مدل رگرسیون مشکل یا غیرممکن است، چرا که p-value عدم قطعیت در ساخت مدل را توضیح نمی‌دهد (۳۰، ۲۳). برخی از محققین استفاده از آزمون F در انتخاب متغیر را بد جلوه داده-اند (۴،۲۵). با این وجود آماره F می‌تواند یک رویکرد ارزشمند در بیان توازن بین تغییرات توجیه شده و پیچیدگی مدل باشد. در واقع F-to-enter ، ضریب تبیین تعدیل یافته، و AIC می توانند به عنوان موارد خاص از شکل آماره مالو Cp تلقی شوند (۲۳). بطور مثال در جدول (۴-۱) با اعمال روش انتخاب گام‌به‌گام (روش ۱) با توجه به مقدار p-value (در آستانه ۰/۰۵) میانگین تعداد متغیر ۳/۶ بدست می‌آید. چنانچه از آستانه ۰/۰۱ استفاده گردد، مدل به طور متوسط ۳/۲ متغیری ایجاد می-شود. در نهایت، با استفاده از آستانه ۰/۱۰ مدل ۴/۶ متغیری ایجاد می‌شود. از این رو، هنگامی که استفاده از AIC و یا BIC به تنهایی امکان‌پذیر نیست سطح معنی‌داری آزمون F می‌تواند بعنوان پارامتر تنظیم جریمه پیچیدگی مدل در نظر گرفته شود (۳۲). در این زمینه رد کردن روش انتخاب گام‌به‌گام متغیر با استفاده از آزمونF بعنوان ابزار مناسب برای گزینش، تنها به خاطر استفاده از آزمون فرضیه ناعادلانه به نظر می‌رسد (۲۰،۳۹).

یادآوری این نکته مهم است که مقایسه روش های گزینش در مطالعه حاضر بر اساس قابلیت پیش‌بینی است، چنانچه هدف از مدل‌سازی تبیین یا شناسایی روابط علت و معلولی باشد معیار مقایسه روش‌ها می‌تواند از معیارهایی که در اینجا در نظر گرفته شده متفاوت باشد. در زمان ساخت مدل، بین دو هدف اصلی باید تمایز قائل شد. هدف اول پیش‌بینی است که

در آن ساختار مدل از ملاحظه کمی برخوردار است و هدف دوم، تبیین است که در آن تلاش می‌شود متغیرهای پیشگوی مهم شناسایی شود و بینشی نسبت به رابطه بین متغیر پیشگو و متغیر پاسخ از طریق ساختار مدل حاصل گردد. در خصوص پیش‌بینی، برازش مدل و مجذور میانگین مربعات خطای پیش بینی معیارهای اصلی برای کفایت مدل می باشند (۳۲).

نکته مهم دیگر این است که استفاده از داده های واقعی در مطالعه حاضر، به جای داده شبیه‌سازی، استنباط و نتیجه‌گیری مربوطه را دشوار می‌سازد. از طرفی دیگر، نمونه‌برداری از مجموعه داده‌ها که نتایج را به جمعیت بزرگتر تعمیم دهد، نیز کار دشواری است (۲۷). در مطالعه حاضر از ۱۲۷ قطعه نمونه استفاده گردیده است. افزایش شدت نمونه‌برداری می تواند تاحدی بر مسائل فوق فایق آید. شبیه سازی تنها روش مشخص برای مقاسیه تکنیک های مدل‌سازی است که در آن مدل درست شناخته می‌شود. پژوهشگران داده‌های شبیه سازی شده بسیار مختلفی را در اختیار دارند و از معیارهای مختلفی برای مقایسه روش‌هایی که ترکیب نتایج آنها دشوار است استفاده می نمایند (۲۷). Raffalovich و همکاران (۲۰۰۸) قابلیت روش‌های متعدد در ملحوظ نمودن متغیرهای مهم و حذف متغیرهای نامربوط را مورد ارزیابی قرار دادند و دریافتند که روش‌های مبتنی بر رگرسیون گام‌به‌گام و BIC بهترین عملکرد را دارند، در صورتی که روش‌های مبتنی بر AIC کارایی نامناسبی داشتند و از آن‌ها باید اجتناب نمود. Murtaugh (۱۹۹۸) از سوی دیگر تفاوت اندکی در قابلیت روش‌های مبتنی بر آزمون F، BIC و Cp مشاهده نمود. Murtaugh در سال ۲۰۰۹ نیز دریافت که عملکرد روش‌های گام‌به‌گام و بهترین زیرمجموعه مشابه می‌باشد و انتخاب معیار برای مقایسه مدل‌ها (AIC، BIC و آماره F) تأثیر اندکی بر قابلیت پیش بینی دارد. نتایج مطالعه حاضر که در آن بهترین روش برای انتخاب مدل آماری وجود ندارد، با مطالعات Murtaugh (۲۰۰۹) مطابقت دارد. لازم به ذکر است که گستره وسیعی از روش‌های گزینش متغیر وجود دارد، که هرکدام از آنها می‌تواند مدل‌های ارزشمندی ایجاد کند که برای کاربرد خاصی مورد توجه قرار گیرد. این که کدام مدل مفیدترین می‌باشد، با استفاده از روش بدست آمدن مدل‌ها تعیین نمی‌گردد، بلکه بیشتر از طریق متناسب بودن آن مدل برای آن هدف تعیین می‌شود (۲۷).

بررسی مدل‌سازی عملکرد گونه راش با استفاده از مدل جمعی تعمیم‌یافته

بهره‌گیری از مدل جمعی تعمیم‌یافته در این مطالعه در پیش‌بینی معیارهای عملکرد راش شرقی نسبت به متغیرهای محیطی نشان داد که مدل جمعی تعمیم‌یافته به دلیل دارا بودن هموارسازهای قوی قادر به کشف روابط غیرخطی بین متغیرها بوده و قادر است پیش‌بینی دقیقی از معیارهای عملکرد ارائه دهد. نتایج این مطالعه نشان داد که کارایی مدل جمعی تعمیم‌یافته در ارزیابی عملکرد گونه راش مناسب می‌باشد. Moisen و همکاران (۲۰۰۶) نیز حضور و غیاب و سطح

مقطع در هکتار گونه‌های مختلف را با استفاده از مدل‌های جمعی تعمیم‌یافته، درخت رگرسیون تقویت شده و روش‌های درخت مبنا پیش‌بینی نمودند و نتیجه گیری نمودند مدل-های جمعی تعمیم‌یافته (GAM) و درخت رگرسیون تقویت شده (BRT) نتایج بهتری را ارائه دادند. هم‌چنین Aertsen و همکاران (۲۰۱۰) نیز مدل جمعی تعمیم‌یافته (GAM) را به عنوان مناسب‌ترین روش مدل‌سازی توان تولیدی رویشگاه در جنگل‌های کوهستانی مدیترانه‌ای در جنوب ترکیه معرفی نمودند.

در این پژوهش از مدل جمعی تعمیم یافته برای بررسی رابطه عملکرد گونه راش نسبت به متغیرهای محیطی استفاده شد و نتایج قابل قبولی نیز به‌دست آمده است. با توجه به نتایج مطالعه حاضر شش متغیر ارتفاع از سطح دریا، درصد رس، مقدار پتاسیم، درصد ازت، مقدار فسفر و درصد سیلت ارتباط معنی‌داری با ارتفاع غالب گونه راش و چهار متغیر ارتفاع از سطح دریا، درصد رس، درصد ازت و مقدار آزیموت نیز ارتباط معنی‌داری با سطح مقطع گونه راش دارند. علاوه بر این هشت متغیر ارتفاع از سطح دریا، جرم مخصوص ظاهری، درصد CN، مقدار فسفر، مقدار اسیدیته خاک، درصد سیلت، درصد شیب و مقدار آزیموت بر معیار عملکرد تراکم گونه راش اثرگذار می-باشند. همان‌طور که اشاره شد با توجه به معیار اهمیت نسبی، ارتفاع از سطح دریا مهم-ترین متغیر اثرگذار برمعیارهای عملکرد گونه راش شناسایی شد.

از مدل‌های جمعی برای مدل‌سازی انواع داده‌ها با مقیاس‌های مختلف در سایر علوم نیز استفاده گردیده است که در آن‌ها برتری مدل‌های جمعی تعمیم یافته بر مدل‌های خطی تعمیم یافته نشان داده شده است. در خاتمه باید خاطر نشان کرد که بیان ریاضی روابط بین متغیرهای محیطی و عکس‌العمل گونه‌ها تنها کمکی برای تفسیر مشاهدات میدانی است و حتی قوی‌ترین همبستگی‌ها را نمی‌توان قطعی فرض کرد.

نتیجه‌گیری

یکی از زمینه‌های مطرح در اکولوژی پوشش گیاهی، تحلیل و درک روابط عملکرد گونه گیاهی و عوامل رویشگاهی است. شناخت رابطه بین عوامل محیطی با پراکنش و عملکرد گونه‌های گیاهی نقش مهمی در طرح‌های زیست محیطی و مدیریتی ایفا می‌نماید. در این مطالعه تعداد ۱۲۷ قطعه نمونه دایره‌ای شکل با مساحت ۱۰۰۰ مترمربع به صورت سیستماتیک تصادفی در جنگل آموزشی و پژوهشی دانشگاه تربیت مدرس پیاده شد و ارتفاع و قطر تمام درختان گونه راش شرقی در هر پلات اندازه‌گیری شد و از عمق صفر تا ۱۰ سانتی‌متری نمونه خاک تهیه گردید. داده‌های جمع‌آوری شده در این مطالعه را به دو دسته داده مدل‌سازی و اعتبارسنجی تقسیم نمودیم. به منظور بررسی رابطه بین عملکرد گونه راش و متغیرهای خاکی و فیزیوگرافی از مدل خطی و جمعی تعمیم‌یافته در نرم‌افزار R استفاده - گردید. در این پژوهش مدل جمعی تعمیم-یافته با استفاده از دو بسته آماری mgcv و

gam در نرم‌افزار R برازش داده شد. در هریک از مدل‌های خطی و جمعی، روش‌های انتخاب گزینش متغیر اجرا گردید و درنهایت نتایج نشان داد که در مدل خطی تعمیم‌یافته برازش داده شده به معیارهای ارتفاع غالب، سطح مقطع و تراکم به ترتیب روش‌های Step. BIC، Best. SubSet. AIC و Step.AIC بعنوان بهترین روش انتخاب متغیر، معرفی شدند. در این مطالعه مدل خطی تعمیم‌یافته توانست ۳۱ درصد تغییرات ارتفاع غالب، ۹/۷۳ درصد تغییرات سطح مقطع و ۱۸/۶۷ درصد تغییرات تراکم را توجیه کند.

در مدل جمعی تعمیم‌یافته نیز روش Best.SubSet.AIC بعنوان بهترین روش گزینش متغیر برای معیارهای عملکرد گونه راش شناسایی شد. در این مطالعه مدل جمعی تعمیم‌یافته توانست بیش از ۳۸ درصد تغییرات ارتفاع غالب، ۲۳ درصد تغییرات سطح مقطع و ۳۸/۶ درصد تغییرات تراکم را به کمک متغیرهای محیطی مورد بررسی توجیه نماید.

با مقایسه مدل‌های خطی و جمعی تعمیم-یافته با استفاده از معیارهای ضریب تبیین و مجذور میانگین مربعات خطا حاصل از داده-های مدل‌سازی، مشاهده می‌شود بیشترین ضریب تبیین و کمترین مجذور میانگین مربعات خطا مربوط به مدل جمعی تعمیم‌یافته است بنابراین می‌توان گفت مدل جمعی تعمیم یافته از نظر معیارهای ارزیابی عملکرد بهتری نسبت به مدل خطی تعمیم‌یافته دارد.

Refrences

1- Aertsen, W., V. Kint, B. De Vos, J. Deckers, J. Van Orshoven, & B. Muys, 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecological Modeling, 221(8): 1119-1130.

2- Alder, D, 1980. Forest volume estimation and yield prediction, F.A.O. Publication, Rome.

3- Berges, L., R. Chevalier, Y. Dumas, A. Franc, J.M. Gilbert, 2005. Sessile oak (*Quercus petraea Liebl*.) site index variations in relation to climate, topography and soil in even-aged high-forest stands in northern France. Annals of Forest Science, 62(5): 391-402.

4- Burnham, K.P., D.R. Anderson, 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd edition, Springer, New York.

5- Burnham, K.P., D.R. Anderson, K.P. Huyvaert, 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral Ecology and Sociobiology 65:23-35.

6- Chatterjee, S., S.A. Hadi, & B. Price, 2004. Regression Analysis by Example. Niroumand, H.A., Publishers University of Mashhad, 384p.

7- Fakoor, E., 2014. Evaluating the site productivity of Oriental Beech using data mining techniques (decision tree). Master's thesis, University of Tarbiat Modares, 79p.

8- Gerritsen, J., J.M. Dietz, & H.T. Wilson, 1996. Episodic acidification of coastal plain streams: an estimation of risk to fish. Ecological Applications, 6(2): 438–448.

9- Hastie, T., R. Tibshirani, 1986. Generalized Additive Model. Statistical sincence, 1(3):297-318.

10- Hastie, T., R. Tibshirani, 1990. Non-parametric logistic and proportional odds regression. Applied statistics, 260-276.
11- Heegaard, E., 2002. The outer border and central border for species–environmental relationships estimated by non-parametric generalised additive models. Ecological Modelling, 157(2): 131-139.
12- Hobbs, N.T., & R. Hilborn, 2006. Alternatives to statistical hypothesis testing in ecology: A guide to self-teaching. Ecological Applications, 16(1): 5–19.
13- Hoeting, J.A., R.A. Davis, A.A. Merton, & S.E. Thompson, 2006. Model selection for geostatistical models. Ecological Applications, 16: 87–98.
14- Jafarihaghighi, M., 2003. Methods of soil analysis, Publication voice of Zoha, 236p, Sari.
15- Kalantari, KH., 2010. Processing and analyzing data in social - economic research, Saba Cultural Publications, 388 p.

16- Knick, S.T., & J.T. Rotenberry, 1995. Landscape characteristics of fragmented shrubsteppe habitats and breeding passerine birds. Conservation Biology, 9(5): 1059–1071.
17- Lee, H., & S. Ghosh, 2009. Performance of information criteria for spatial models. Journal of Statistical Computation and Simulation, 79(1): 93–106.
18- Lehmann, A., J.M. Overton, J.R. Leathwick, 2002. GRASP: generalized regression analysis and spatial prediction. Ecological modelling, 157(2): 189-207.
19- Leigh, G.T., A.J. Read, & P. Halpin, 2008. Fine-scale habitat modeling of a top marine predator: do prey data improve predictive capacity. Ecological Applications, 18(7): 1702–1717.
20- Lukacs, P.M., W.L. Thompson, W.L. Kendall, W.R. Gould, Jr P.F. Doherty, K.P. Burnham, & D.R. Anderson, 2007. Concerns regarding a call for pluralism of information theory and hypothesis testing. Journal of Applied Ecology, 44(2): 456–460.
21- Mazerolle, M.J., 2006. Improving data analysis in herpetology: using Akaike's information criterion (AIC) to assess the strength of biological hypotheses. Amphibia-Reptilia, 27(2): 169–180.
22- McLintock, T.F., & C.A. Bickford, 1957. A proposed site index for red spruce in the Northeast, Publication of Northeastern Forest Experiment Station, U.S. Upper Darby.
23- Miller, A., 2002. Subset Selection in Regression, 2nd edition, Chapman & Hall/CRC, Boca Raton, Florida.
24- Moisen, G.G., E.A. Freeman, J.A. Blackard, T.S. Frescino, N.E. Zimmermann, & Jr T.C. Edwards, 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological modelling, 199(2): 176-187.
25- Mundry, R., & C.L. Nunn, 2009. Stepwise model fitting and statistical inference: turning noise into signal pollution. The American Naturalist, 173(1): 119–123.
26- Murtaugh, P.A., 1998. Methods of variable selection in regression modeling. Communications in Statistics: Simulation and Computation, 27(3): 711–734.
27- Murtaugh, P.A., 2009. Performance of several variable-selection methods applied to real ecological data. Ecology Letters, 12(10): 1061-1068.
28- Olden, J.D., & D.A. Jackson, 2000. Torturing data for the sake of generality: How valid are our regression models? Ecoscience, 7(4): 501–510.

29- Raffalovich, L.E., G.D. Deane, D. Armstrong, & H.S. Tsao, 2008. Model selection procedures in social research: Monte-Carlo simulation results. Journal of Applied Statistics, 35(10): 1093–1114.
30- Ramsey, F., & D. Schafer, 2002. The Statistical Sleuth: A Course in Methods of Data Analysis, 2nd edition, Duxbury Press, Belmont, California.
31- Rezaei, A., & A. Soltani, 2003. Introduction to Applied Regression Analysis, Isfahan University Press, 294 p.
32- Sauerbrei, W., P. Royston, & H. Binder, 2007. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. Statistics in Medicine, 26(30): 5512–5528.
33- Sharifi, A.M., 2014. Evaluate the response species *Festuca ovina* L to some environmental variables using HOF function in catchment's Galandrud, Master's thesis Department of Natural Resources Tarbiat Modares University.
34-Sheather, S., 2009. A Modern Approach to Regression with R, Springer Texts in Statistics, Springer, 393 pp.
35- Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.
36- Skovsgaard, J.P., & J.K. Vanclay, 2008. Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands. Journal of Forestry, 81(1): 13–31.
37- Vanclay, J.K., 1992. Assessing site productivity in tropical moist forests: a review. Forest Ecology and Management, 54(1): 257-287.
38- Ward, E.J., 2008. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. Ecological Modelling, 211(1): 1–10.
39- Whittingham, M.J., P.A. Stephens, R.B. Bradbury, & R.P. Freckleton, 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology, 75(5): 1182–1189.
40- Yee, T.W., N.D. Mitchell, 1991. Generalized additive models in plant ecology, Journal of Vegetation Science, 2(5): 587-602.