بررسی شایستگی مدل رگرسیون حداقل مربعات در مدلسازی ارتفاع گونهٔ راش نسبت به متغیرهای محیطی در جنگل آموزشی و پژوهشی دانشگاه تربیت مدرس

سید جلیل علوی[۱*]، وریا مردان‌پور[۲] و کارستن اف دورمن[۳]

[۱]عضو هیأت علمی گروه جنگلداری، دانشکدهٔ منابع طبیعی و علوم دریایی نور، دانشگاه تربیت مدرس
[۲]دانشجوی کارشناسی ارشد جنگلداری، دانشکدهٔ منابع طبیعی و علوم دریایی نور، دانشگاه تربیت مدرس
[۳]عضو هیأت علمی گروه بیومتری و تحلیل سیستم‌های محیطی، دانشگاه فرایبورگ، آلمان

(تاریخ دریافت: ۱۳۹۴/۷/۷؛ تاریخ پذیرش: ۱۳۹۴/۱۱/۱۹)

چکیده

در پژوهش حاضر، به‌منظور مدلسازی ارتفاع گونهٔ راش به‌عنوان یک مؤلفهٔ مهم توان تولید رویشگاه نسبت به متغیرهای محیطی در جنگل آموزشی و پژوهشی دانشگاه تربیت مدرس از مدل رگرسیون حداقل مربعات استفاده شد. به این منظور به‌روش منظم – تصادفی ۱۲۳ قطعهٔ نمونه دایره‌ای ۰/۱ هکتاری در منطقه مستقر گردید و در هر یک از قطعات نمونه، ارتفاع کل و قطر برابرسینه تمام درختان راش با قطر بیشتر از ۷/۵ سانتی‌متر اندازه‌گیری شد. در مرکز قطعات نمونه، از عمق ۰تا۱۰ سانتی‌متر نمونه‌های خاک برداشته شده و مجموعه‌ای از خصوصیات فیزیکی و شیمیایی خاک در محل استقرار قطعات نمونه اندازه‌گیری شد. همچنین ارتفاع از سطح دریا، درصد شیب و آزیموت قطعات نمونه نیز ثبت شد. نتایج پژوهش حاضر نشان داد هرچند مدل رگرسیون خطی برحسب ضریب تبیین، کارایی به نسبت خوبی در مدلسازی ارتفاع گونهٔ راش دارد، به‌کارگیری شاخص موران I نشان می‌دهد که مسئلهٔ خودهمبستگی مکانی در مدل رگرسیون وجود دارد. به‌کارگیری مدل‌های خودرگرسیونی با وقفهٔ مکانی و خطای مکانی به‌عنوان رویکردهای جایگزین برای حذف پدیدهٔ خودهمبستگی نشان داد که این مدل‌ها ازنظر معیارهای ضریب تبیین، معیار اطلاعاتی آکائیک و لگاریتم درست‌نمایی کارایی بهتری نسبت به مدل رگرسیون خطی دارند. مقایسهٔ دو مدل خودرگرسیونی توأمان با وقفهٔ مکانی و خطای مکانی با استفاده از آماره‌های فوق نیز نشان می‌دهد که مدل خود رگرسیونی با خطای مکانی بهتر از مدل با وقفه مکانی است. پژوهش حاضر بر اهمیت بررسی و کنترل خودهمبستگی مکانی در مطالعات بوم‌شناسی جنگل تأکید می‌ورزد و دستورالعملی را برای مدلسازی عملکرد گونهٔ راش نسبت به متغیرهای محیطی فراهم می‌کند.

واژه‌های کلیدی: آمارهٔ موران I، خودهمبستگی مکانی، ساختار همسایگی، مدل خطای مکانی، مدل وقفهٔ مکانی.

* نویسندهٔ مسئول شمارهٔ تماس:۰۹۱۱۱۵۸۰۰۹۷ ایمیل: j.alavi@modares.ac.ir

مقدمه و هدف

بررسی روابط عملکرد[1] گونه‌های گیاهی (عملکرد دربرگیرندهٔ اندازه‌گیری بعضی از قسمت‌های گیاه همانند ارتفاع گیاه، طول، شکل و اندازه برگ و غیره است که به شاخص اندازه رشد، قدرت و توان گیاهی مرتبط باشد (Kent, 2011)) و عوامل محیطی، همواره به‌عنوان یک موضوع کلیدی در بوم‌شناسی گیاهی مطرح بوده است و بیش از یک قرن است که بوم‌شناس‌ها برای یافتن عوامل محیطی کنترل‌کننده بر پراکنش، وفور و عملکرد گونه‌های گیاهی تلاش می‌کنند (Comstock and Ehleringer, 1992). روش های کلاسیک بررسی روابط گونه‌های گیاهی و عوامل محیطی، هنگامی می‌تواند به‌عنوان برآوردگری سازگار، مورد اعتماد قرار گیرد که نمونه‌ها همگن و مستقل از یکدیگر باشند؛ اما در عمل، در بسیاری از موارد ممکن است مشاهدات مستقل نباشند و برحسب موقعیت قرارگیری در فضای مورد بررسی به یکدیگر وابسته باشند، ازاین‌رو وجود پدیده‌ای که خودهمبستگی مکانی[2] (SAC) نامیده می‌شود، تحلیل داده‌ها را پیچیده می‌سازد (سوری و منیری جاوید، ۱۳۹۰). خودهمبستگی مکانی الگویی است که در آن مشاهدات با یکدیگر براساس فاصلهٔ جغرافیایی بین مشاهدات در ارتباط‌اند. به‌گونه‌ای که مشاهدات نزدیک به هم وابستگی بیشتر و مشاهدات دورتر از هم وابستگی کمتری دارند. زمانی که SAC مثبت باشد، محل‌ها و مکان‌های مجاور هم نسبت به مکان‌های دورتر مقادیر مشابه‌تری را اتخاذ می‌کنند. استقلال مشاهدات در داده‌هایی که ویژگی مکانی دارند فرض بسیار قوی‌ای است (سوری و منیری جاوید، ۱۳۹۰). جملهٔ مشهوری می‌گوید "هر پدیده‌ای با پدیدهٔ دیگر در ارتباط است، اما ارتباط پدیده‌های نزدیک‌تر قوی‌تر از ارتباط پدیده‌های دور است" (سوری و منیری جاوید، ۱۳۹۰). بدیهی است که تحلیل آماری داده‌های مکانی با روش‌های آماری معمول مقدور نیست، زیرا شرط اساسی استقلال داده‌ها، محقق نیست.

روش‌های رگرسیونی متعددی برای سنجش روابط عملکرد گونه‌های گیاهی و عوامل محیطی به‌کار گرفته شده‌اند که از آن جمله می‌توان به مدل‌های رگرسیون خطی چندگانه، رگرسیون پواسون، رگرسیون لجستیک و مدل‌های رگرسیون جمعی اشاره داشت (Guisan and Zimmermann, 2000). هرچند این تکنیک‌ها پیشرفت‌های زیادی را در مدل سازی روابط گونه‌های گیاهی و عوامل محیط نشان داده‌اند، در شکل استاندارد خود، زمانی که خودهمبستگی مکانی وجود داشته باشد، نمی‌توانند نمایش دقیقی از عوامل محیطی اثرگذار بر وفور گونه را فراهم کنند. ازاین‌رو روش‌های آماری متعددی ابداع شده‌اند که می‌توانند خودهمبستگی مکانی در مدل های وفور گونه_ محیط را لحاظ کنند.

روش‌های آماری متعددی برای پرداختن به خودهمبستگی مکانی وجود دارند (Keitt *et al.*, 2002; Miller *et al.*, 2007) که می‌توان به مدل‌های خودرگرسیونی توأمان[3]، مدل‌های خودرگرسیونی شرطی[4]، مدل‌های حداقل مربعات تعمیم‌یافته[5]، مدل‌های آمیختهٔ خطی تعمیم‌یافته[6] و معادلات برآورد تعمیم‌یافته[7] اشاره داشت، اما تنها تعداد اندکی از آنها در متون بوم‌شناسی به چشم می خورند، زیرا اجرای این مدل‌ها ازنظر ریاضی پیچیده است و ازنظر محاسباتی نیز به زمان زیادی نیاز دارند (Rangel *et al.*, 2006) و از طرف دیگر، نرم‌افزار آماری R که اجرای این مدل‌ها در آن امکان‌پذیر است، نیز چند سالی است که در دسترس همگان قرارگرفته است (Dormann *et al.*, 2007).

[1] Performance
[2] Spatial Autocorrelation (SAC)
[3] Simultaneous Autoregressive Models (SAR)
[4] Conditional Autoregressive Models (CAR)
[5] Generalized Least Squares (GLS)
[6] Generalized Linear Mixed Models (GLMM)
[7] Generalized Estimation Equations (GEE)

روش‌های مختلف مدلسازی که خودهمبستگی مکانی را در تحلیل روابط وفور گونه و متغیرهای محیطی به‌حساب می‌آورند توسط Keitt *et al.* (2002) مقایسه شد. نتایج نشان داد زمانی که خودهمبستگی مکانی در فرایند مدلسازی لحاظ شود، نتیجه‌گیری در مورد متغیرهای محیطی که وفور گونه را کنترل می کنند به‌شدت تغییر می‌کند. Segurado *et al.*(2006) بیان می‌دارند خودهمبستگی مکانی، مسئله‌ای جدی در مدلسازی آشیان بوم‌شناسی گونه‌هاست و معنی‌داری پارامترهای مدل را به‌شدت تغییر می‌دهد. آنها اشاره دارند که مدل‌های خطی و جمعی تعمیم‌یافته به اثرهای خودهمبستگی مکانی حساس‌اند. تکنیک‌های مختلف مدلسازی مکانی نظیر مدل‌های خودرگرسیونی توأمان، مدل‌های خود رگرسیونی شرطی، مدل‌های حداقل مربعات تعمیم‌یافته، مدل‌های آمیخته خطی تعمیم‌یافته و معادلات برآورد تعمیم‌یافته به‌طور جامع توسط Dormann *et al.* (2007) در یک مقالهٔ مروری بررسی شده است. عملکرد سه مدل خودرگرسیونی توأمان (خطای مکانی[1]، وقفهٔ مکانی[2] و مدل مکانی[3] Durbin) و مدل رگرسیون حداقل مربعات[4] (OLS) در مدلسازی پراکنش گونه‌ها توسط Carl (2007) و Kissling با استفاده از چهار مجموعه دادهٔ شبیه‌سازی شده ارزیابی شد و این نتیجه به‌دست آمد که مدل خطای مکانی از مدل‌های دیگر کارایی بهتری دارد. نتایج تحقیق Miller *et al.*(2007) در بررسی رویکردهای مختلفی که وابستگی مکانی را در مدل‌های پیش‌بینی پوشش گیاهی به‌حساب می‌آورند نشان داد که مدل‌های خودرگرسیونی توانایی بیشتری در توصیف وابستگی مکانی کوچک‌مقیاس دارند، درصورتی‌که روش‌های زمین‌آماری برای مدلسازی وابستگی مکانی بزرگ‌مقیاس مناسب‌ترند. مدل‌های رگرسیون حداقل مربعات و خودرگرسیونی توأمان با خطای مکانی با مدل‌های خود رگرسیونی توأمان واریوگرامی و حداقل مربعات تعمیم‌یافته توسط Beguería and Pueyo (2009) مقایسه شدند. نتایج پژوهش آنها نشان داد که مدل واریوگرام، کارایی بهتری از مدل خطای مکانی داشته است، اما مدل حداقل مربعات تعمیم‌یافته نسبت به مدل‌های مبتنی بر خودرگرسیونی توأمان برتر بوده است. روش‌های مختلف مدلسازی شامل مدل حداقل مربعات تعمیم‌یافته، فیلترهای مکانی[4]، مدل موجک[5]، مدل‌های خود رگرسیونی و مدل‌های آمیختهٔ جمعی تعمیم‌یافته برای برآورد ضرایب رگرسیون از مجموعه داده شبیه‌سازی‌شده توسط Beale *et al.* (2010) بررسی شد. نتایج نشان داد که مدل‌های حداقل مربعات تعمیم‌یافته و مدل‌های خود رگرسیونی عملکرد خوبی داشته‌اند. به‌منظور بررسی تأثیر متغیرهای محیطی و انسانی بر پوشش گیاهی در کالیفرنیای آمریکا، Dahlin *et al.* (2014) از مدل‌های خودرگرسیونی استفاده کردند. نتایج پژوهش آنان، بر اهمیت فرایندهای پویا همانند محدودیت پراکنش و تاریخچهٔ آشفتگی در تعیین الگوهای فعلی پوشش گیاهی تأکید می‌کند.

همان‌طور که اشاره شد، وجود SAC در مدل‌ها فرض مستقل بودن داده‌های مکانی را دچار تخطی می‌کند. مدل‌های روابط عملکرد گونه _ محیط که خودهمبستگی مکانی را نادیده می‌گیرند، ممکن است تأکید زیاد و بی‌موردی بر فاکتورهایی داشته باشند که تأثیر اندکی بر عملکرد گونه دارند یا هیچ تأثیری ندارند یا ممکن است منجر به حذف متغیرهای مهم در طول انتخاب مدل شوند (Dormann *et al.*, 2007). جنگل‌های خالص و آمیختهٔ راش از مهم‌ترین و غنی‌ترین جنگل‌های ایران به‌شمار می‌رود که در آن گونهٔ راش حدود ۳۰ درصد از کل حجم سرپا و حدود

[1] Spatial lag model
[2] Spatial Durbin model
[3] Ordinary Least Square
[4] Spatial filters
[5] Wavelet models

۲۳ درصد از کل تعداد درختان این جنگل‌ها را به خود اختصاص داده است (احمدی و همکاران، ۱۳۹۴). ازآنجا که ارتفاع گونهٔ درختی به‌عنوان مؤلفه‌ای مهم در محاسبهٔ توان تولیدی جنگل در توده‌های آمیخته و ناهمسال مطرح است (احمدی و همکاران، ۱۳۹۴)، تحقیق حاضر به‌منظور بررسی کارایی مدل رگرسیون حداقل مربعات در مدلسازی ارتفاع گونهٔ راش (*Fagus orientalis* Lipsky) به‌عنوان یک معیار عملکرد نسبت به متغیرهای خاکی و فیزیوگرافی در جنگل آموزشی و پژوهشی دانشگاه تربیت مدرس صورت می‌گیرد. چنانچه شایستگی مدل رگرسیون حداقل مربعات تأیید نشود، رویکردهای جایگزین معرفی و کارایی آنها ارزیابی می‌شود.

مواد و روش اجرای تحقیق

مشخصات منطقهٔ تحقیق

تحقیق حاضر در جنگل آموزشی و پژوهشی صلاح‌الدین‌کلا متعلق به دانشکدهٔ منابع طبیعی دانشگاه تربیت مدرس در سطحی حدود ۲۰۰ هکتار با دامنهٔ ارتفاعی ۱۰۰۰ تا ۱۵۰۰ متر از سطح دریا، صورت گرفت. جهت جغرافیایی غالب در منطقه، شمالی است و پس از آن جهت‌های غربی و جنوبی بیشترین سهم را دارند. شیب غالب منطقه بین ۵۰-۲۵ و ۸۰-۵۰ درصد است. منطقه بین عرض جغرافیایی "۳۶°۲۹'۲۴ تا "۳۶°۳۰'۴۲ و طول جغرافیایی "۵۱°۴۵'۵۹ تا "۵۱°۴۶'۵۱ قرار گرفته است. مقدار بارندگی در منطقه ۱۳۰۸ میلی‌متر در سال است. گرم‌ترین ماه سال تیر و مرداد با میانگین دمای ۲۹/۲ و سردترین ماه سال، بهمن با میانگین دمای ۲/۶ درجهٔ سانتی‌گراد است. همچنین میانگین دمای سالانه برابر با ۱۵/۹ درجهٔ سانتی‌گراد ثبت شده است (بی‌نام، ۱۳۹۲). تیپ جنگل در منطقه آمیخته‌ای از راش به‌همراه گونهٔ ممرز (*Carpinus betulus*)، انجیلی (*Parrotia persica*)، افرا (*Acer velutinum*)، شیردار (*Acer leatum*)، نمدار (*Tilia begonifolia*)، ون (*Fraxinus excelsior*) و بلوط (*Quercus castaneifolia*) است. به‌طور میانگین گونهٔ راش حدود ۵۰ درصد از کل حجم سرپا و حدود ۴۰ درصد از کل تعداد درختان منطقه را به خود اختصاص داده است. تیپ‌های خاک منطقه، تکامل نیافته راندزین تا راندزین شسته‌شده، قهوه‌ای جنگلی با pH قلیایی و قهوه‌ای شسته‌نشده تا پسدوگلی است (بی‌نام، ۱۳۹۲).

جمع‌آوری داده‌ها

به‌منظور جمع‌آوری اطلاعات، لازم از ۱۲۳ قطعه‌نمونه دایره‌ای با مساحت ۱۰۰۰ متر مربع که به‌روش آماربرداری منظم تصادفی در شبکه‌ای به ابعاد ۱۰۰ × ۱۰۰ متر در منطقه مستقر شد، استفاده شد. پس از پیاده کردن قطعات نمونه در جنگل، قطر برابرسینه و همچنین ارتفاع کامل تمام درختان راش با قطر بیشتر از ۷/۵ سانتی‌متر اندازه‌گیری شد. در داخل هر قطعه‌نمونه ارتفاع از سطح دریا، درصد شیب و جهت جغرافیایی نیز ثبت شد. جهت جغرافیایی با استفاده از رابطهٔ ۱ به شاخص تابش خورشیدی[1] تبدیل شد که در آن θ مقدار آزیموت جهت برحسب درجه است. مقدار شاخص بار گرمایی بین صفر و یک است؛ جهت شمال شرقی دارای مقدار صفر (خنک‌ترین دامنه) و جهت جنوب غربی دارای مقدار یک (گرم‌ترین دامنه) است (Aertsen *et al.*, 2010).

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

در مرکز هر قطعه‌نمونه، نمونه‌های خاک تا عمق ۱۰ سانتی‌متری (علوی و همکاران، ۱۳۹۰) از چهار جهت جغرافیایی و مرکز قطعه‌نمونه برداشت و سپس برای آزمایش مشخصه‌های فیزیکوشیمیایی با یکدیگر ترکیب شدند. پس از خشک‌کردن نمونه‌ها در هوای آزاد و جدا کردن ریشه‌ها، سنگ و سایر ناخالصی‌ها از

[1] Radiation Index

آن، کلوخ‌ها خرد شده و از الک ۲ میلی‌متری عبور داده شد. در بررسی فاکتورهای خاک، مجموعه‌ای از خصوصیات فیزیکی و شیمیایی خاک ازجمله وزن مخصوص ظاهری به‌روش کلوخ و پارافین، بافت خاک به‌روش هیدرومتری، pH خاک با دستگاه pH متر الکتریکی، درصد مادهٔ آلی به‌روش والکلی و بلک، فسفر قابل‌جذب به روش السون، پتاسیم با دستگاه طیف‌سنج اتمی و نیتروژن به‌روش کجلدال اندازه‌گیری شد (جعفری حقیقی، ۱۳۸۲).

تجزیه‌وتحلیل آماری داده‌ها

به‌منظور مدلسازی ارتفاع گونهٔ راش و متغیرهای محیطی از مدل رگرسیون حداقل مربعات استفاده شد. مدل رگرسیون حداقل مربعات که به‌دلیل اجرا و تفسیر آسان کاربردی گسترده در بوم‌شناسی جنگل دارد، رویکردی است که در آن مدل رگرسیون تنها با متغیرهای تبیینی محیطی و بدون درنظر گرفتن همبستگی خودکار مکانی برازش می‌یابد؛ بنابراین تغییر در عملکرد و وفور گونه تنها به تغییر در متغیرهای محیطی مرتبط است. تجزیه‌وتحلیل رگرسیونی در صورتی معتبر خواهد بود که فرضیه‌های رگرسیون همانند نرمال بودن، همگنی واریانس و مستقل بودن برقرار باشد. نقص این فرضیه‌ها، ممکن است به استدلال‌های آماری نادرست منجر شود. Zuur *et al.* (2007) اظهار کرده‌اند که تخطی از مستقل بودن داده‌های مکانی جدی‌ترین مسئله است، ازاین‌رو قبل از پرداختن به روش‌های مدلسازی که خودهمبستگی مکانی را به‌حساب می‌آوردند، ابتدا باید بررسی شود که آیا خودهمبستگی مکانی ممکن است تجزیه‌وتحلیل داده‌ها را تحت تأثیر قرار دهد؛ بدین معنا که آیا باقی‌مانده‌های مدل، خودهمبستگی مکانی را نشان می‌دهند یا خیر. برای بررسی و آزمون ضریب خودهمبستگی مکانی و معنی‌داری آن می‌توان از یکی از روش‌های آمارهٔ موران I[1]، همبستگی نگار موران I[2]، همبستگی نگار Geary[3] و سمی‌واریوگرام[4] استفاده کرد. در مطالعات تجربی، اغلب از آمارهٔ موران I در تحلیل خودهمبستگی مکانی استفاده می‌شود (Dormann *et al.*, 2007). آمارهٔ موران I برای متغیر X در خصوص مناطق مختلف از طریق رابطه ۲ محاسبه می‌گردد (Lee and Wong, 2001):

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}c_{ij}}{s^2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{s^2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}$$

رابطه ۲

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$$

رابطه ۳

در رابطهٔ فوق، x_i و x_j مقادیر x در مورد مناطق (نمونه‌های) مختلف، و s^2 واریانس نمونه (رابطهٔ ۳) است. W_{ij} موقعیت مجاورت i و j نسبت به یکدیگر و به عبارتی نوع ارتباط مکانی آنهاست که ماتریس وزن نیز نام برده می‌شود. این ماتریس به‌روش‌های مختلف به‌دست می‌آید که از آن جمله می‌توان به ماتریس مجاورت خطی[5]، مجاورت رخمانند[6]، مجاورت ملکه‌مانند[7]، و مجاورت فیل‌مانند[8] اشاره کرد. از انواع دیگر ارتباط مکانی می‌توان به فاصلهٔ معکوس[9]، مربع فاصلهٔ معکوس[10]، باند فاصلهٔ ثابت[11]، ناحیهٔ بی‌تفاوتی[12] و

[1] Moran's I
[2] Moran's I correlogram
[3] Geary's c correlogram
[4] Semivariogram
[5] linear contiguity
[6] Rook contiguity
[7] Queen contiguity
[8] Bishop contiguity
[9] Inverse Distance
[10] Inverse Distance Squared
[11] Fixed Distance Band
[12] Zone of Indifference

مجاورت چندضلعی[1] اشاره کرد که ضریب خودهمبستگی موران I از طریق فواصل اقلیدسی[2] و منهتن[3]، با توجه به شکل استانداردشدهٔ ماتریس به‌دست می‌آید (رفیعی دارانی و قربانی، ۱۳۹۳). مقدار شاخص موران I بین ۱ و ۱- متغیر است. مقادیر نزدیک به ۱ نشان‌دهندهٔ الگوی خوشه‌ای و به‌عبارتی خودهمبستگی مثبت است و ارزش نزدیک به ۱- نشان می‌دهد که الگوی همسان یا متفرق و به‌عبارتی خودهمبستگی منفی وجود دارد. در صورت نبود خودهمبستگی مکانی، مقدار عددی شاخص موران صفر است. معنی‌داری آمارهٔ موران از طریق آزمون Z بررسی می‌شود که مقدار Z استاندارد آمارهٔ موران از طریق رابطهٔ ۴ محاسبه می‌شود. مقادیر به‌دست‌آمده با مقادیر Z جدول قابل مقایسه است (رفیعی دارانی و قربانی، ۱۳۹۳):

$$Z_I = \frac{I - E(I)}{VAR(I)}$$ رابطهٔ ۴

که در آن، I مقدار آمارهٔ موران I، E(I) مقدار مورد انتظار برای آمارهٔ موران و VAR(I) واریانس شاخص موران I است. مقدار آماره‌ها و آزمون‌های یادشده، مربوط به وضعیتی است که در آن هدف پژوهش، بررسی خودهمبستگی مکانی یک متغیر در نقاط یا مکان‌های مختلف است. چنانچه در پژوهشی، مدل رگرسیونی مدنظر باشد، با توجه به وجود داده‌های مختصات‌دار، فرض مستقل بودن باقی‌مانده‌ها در مدل رگرسیون حداقل مربعات، با اعمال آماره موران I به باقی‌مانده‌ها بررسی می‌شود. در چارچوب مدل رگرسیونی، آماره موران I از طریق رابطهٔ ۵ محاسبه می‌شود (Zuur *et al.*, 2007):

$$I = (n / S_0)e'We / e'e$$ رابطهٔ ۵

در رابطهٔ بالا، W ماتریس وزنی است که همانند ماتریس وزنی آمارهٔ موران I در قسمت قبل به‌عنوان نوع همسایگی تعریف می‌شود، S_0، مجموع عناصر ماتریس W و e باقی‌ماندهٔ حاصل از مدل رگرسیون معمولی است. معنی‌داری آمارهٔ موران I همانند بالا از طریق آزمون Z صورت می‌گیرد. در پژوهش حاضر آمارهٔ موران I با استفاده از بستهٔ spdep (Bivand and Piras, 2015) در نرم‌افزار آماری R نسخه 3.0.2 محاسبه و معنی‌داری آن بررسی شد.

معنی‌دار شدن آزمون موران I در خصوص همبستگی مکانی، بدین معناست که از فرض مستقل بودن در مدل رگرسیون خطی معمولی تخطی‌شده است، ازاین‌رو مدل‌های جایگزین که وابستگی مکانی را منظور می‌کنند، باید به کار گرفته شوند. با به‌کارگیری این مدل‌ها، اریبی کاهش و دقت افزایش می‌یابد (Dormann *et al.*, 2007). در مدل‌های خطی با داده‌های دارای توزیع نرمال، خودهمبستگی مکانی می‌تواند توسط رویکردهای مرتبط با مدل‌های خودرگرسیونی توأمان در نظر گرفته شود. در مدل SAR که تعمیم مدل رگرسیون حداقل مربعات معمولی است، رابطهٔ همسایگی برحسب یک ماتریس n×n از وزن‌های مکانی (w) بیان می‌شود. مشخص کردن ماتریس وزن‌های مکانی با شناسایی ساختار همسایگی هر سلول آغاز می‌شود. پس‌ازآن، همسایه‌ها وزن‌دهی می‌شوند. همسایه‌های نزدیک‌تر، وزن‌های بیشتر؛ و همسایه‌های دورتر وزن‌های کمتری دریافت می‌کنند. برای کددهی ماتریس وزن‌های مکانی روش‌های متعددی وجود دارد برای مثال ۱- کددهی دودویی (محل‌ها به‌صورت همسایه یا غیرهمسایه فهرست می‌شوند)؛ ۲- استانداردسازی ردیف‌ها؛ ۳- ثابت کردن واریانس (Begueria and Pueyo, 2009).

در خصوص مدل‌های خودرگرسیونی توأمان، دو نوع وابستگی وجود دارد: ۱- وقفهٔ مکانی، نوعی از وابستگی مکانی است که فرض می‌کند مقدار متغیر پاسخ در یک محل مشخص نه‌تنها تابعی از متغیرهای مستقل در

[1] Polygon contiguity
[2] Euclidean
[3] Manhattan

محل i است، اما به مقدار متغیر پاسخ در محل‌های مجاور j نیز بستگی دارد. مدل خودرگرسیونی توأمان با وقفهٔ مکانی به‌صورت رابطهٔ ۶ است:

$$\mathbf{Y} = \rho\mathbf{WY} + \mathbf{X\beta} + \varepsilon$$ رابطهٔ ۶

که در آن ρ ضریب خود رگرسیونی، W ماتریس وزن‌های مکانی، و β برداری است که شیب‌های مرتبط با متغیرهای تبیینی X و ε خطا را نشان می‌دهد.

۲- خطای مکانی، نوعی از وابستگی مکانی است که فرض می‌کند فرایند خودهمبستگی تنها در جملهٔ خطا مشاهده می‌شود، نه در متغیر پاسخ و متغیرهای تبیینی. برای مدل خودرگرسیونی توأمان با خطای مکانی مدل رگرسیونی معمولی ($Y=\beta X+\varepsilon$) توسط جمله $\lambda W\mu$ که ساختار مکانی ($\lambda W\mu$) در جمله خطا را که ازنظر مکانی همبسته هستند (μ) را نشان می‌دهد. مدل خطای مکانی به‌صورت رابطهٔ ۷ است:

$$\mathbf{Y} = \mathbf{X\beta} + \lambda\mathbf{W\mu} + \varepsilon$$ رابطهٔ ۷

که λ ضریب خودرگرسیونی مکانی است. تحلیل مدل‌های خودرگرسیونی توأمان در بسته Spdep در نرم‌افزار آماری R نسخه 3.0.2 انجام گرفت. همان‌طور که اشاره شد برای استفاده از توابع خودرگرسیونی توأمان، محقق باید ابتدا فواصل همسایگی را مشخص کند، سپس ماتریس وزن‌های مکانی توسط وزن‌دهی همسایه‌ها با یک طرح کددهی مشخص (برای مثال دودویی، استانداردسازی ردیفی و ثابت نمودن واریانس) محاسبه می‌شود. در پژوهش حاضر به‌منظور تعیین بهترین فاصلهٔ همسایگی، ۲۰ فاصلهٔ همسایگی ۵۰ متری مشخص شد و با برنامه‌نویسی در نرم‌افزار R بهترین فاصلهٔ همسایگی براساس معیار اطلاعاتی آکائیک تعیین شد. در پژوهش حاضر از سیستم کددهی استانداردسازی ردیفی که در مدل‌های خود رگرسیونی ارجح است (Dormann *et al.*, 2007) استفاده شده است. انتخاب هر یک از مدل‌های خودرگرسیونی (وقفهٔ مکانی و خطای مکانی) و مقایسهٔ آنها با مدل رگرسیون حداقل مربعات براساس آماره‌های ضریب تبیین، اشتباه معیار باقی‌مانده، لگاریتم درست‌نمایی[1]، معیار اطلاعاتی آکائیک[2] (AIC)، صورت می‌گیرد. لگاریتم درست‌نمایی، آماره‌ای است که برای تعیین مقادیر ضرایب برآوردی رگرسیون بهینه می‌شد. ازآنجاکه مقادیر لگاریتم درست‌نمایی تابعی از تعداد نمونه‌اند، نمی‌توانند به‌تنهایی شاخصی از نکویی برازش باشد و باید از معیارهای دیگر همانند ضریب تبیین و معیار اطلاعاتی آکائیک استفاده کرد. ازآنجاکه هدف، به حداکثر رساندن لگاریتم درست‌نمایی است، مقادیر بزرگ‌تر، بهتر است (Burnham and Anderson, 2002). AIC معیاری برای سنجش نیکویی برازش است که توسط هیروتوگو آکائیکه در سال ۱۹۷۴پیشنهاد شد. این معیار با برقرار کردن تعادل میان دقت مدل و پیچیدگی آن به انتخاب بهترین مدل آماری کمک می‌کند (Akaike, 1974). در حالت کلی، AIC به‌صورت رابطهٔ ۸ محاسبه می‌شود (Burnham and Anderson, 2002):

$$AIC = -2loglike + 2K$$ رابطهٔ ۸

که در آن loglike لگاریتم درست‌نمایی مدل رگرسیون مورد نظر و K تعداد پارامترهای برآوردی مدل رگرسیونی است. ازآنجاکه مدل دارای کمترین AIC بهترین است، می‌توان استنباط کرد که کدام مدل وضعیت بهتری نسبت به بقیهٔ مدل‌ها دارد (Akaike, 1974). به‌عنوان یک قاعدهٔ کلی، اختلاف AIC سه و بیشتر شاهدی آشکار بر بهبود عملکرد مدل‌هاست (Fotheringham *et al.*, 2001).

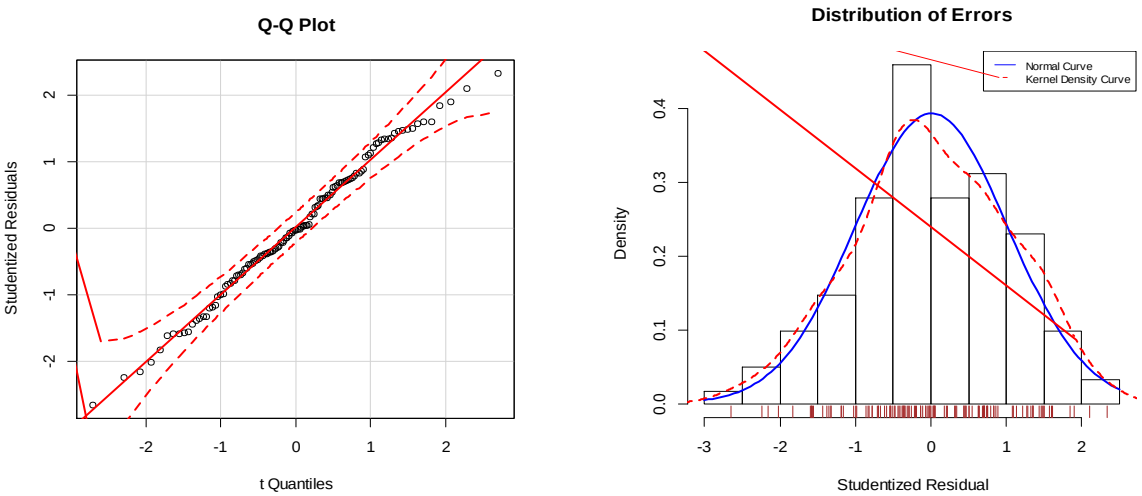
[1] Log-likelihood
[2] Akaike information criterion

نتایج

نتایج برازش مدل رگرسیون خطی (F= ۵/۵۶, P-value = ۰/۰۰۰) نشان می‌دهد که مدل رگرسیون خطی بسیار معنی‌دار است. گزینش متغیرها با استفاده از روش گام‌به‌گام و معیار اطلاعاتی بیزی[1] یا BIC در مدل رگرسیون خطی نشان داد از ۱۱ متغیر تبیینی، اثرهای اصلی متغیرهای ارتفاع از سطح دریا، وزن مخصوص ظاهری خاک، درصد رس، درصد سیلت، نیتروژن، پتاسیم، فسفر، نسبت کربن به ازت و شاخص تابش خورشیدی و اثر متقابل تعدادی از متغیرها بر ارتفاع درخت راش تأثیرگذار بوده‌اند (جدول ۲). با به‌کارگیری مدل رگرسیون خطی ۵۶/۶۳ درصد از تغییرات در متغیر ارتفاع درخت راش توجیه شده است.

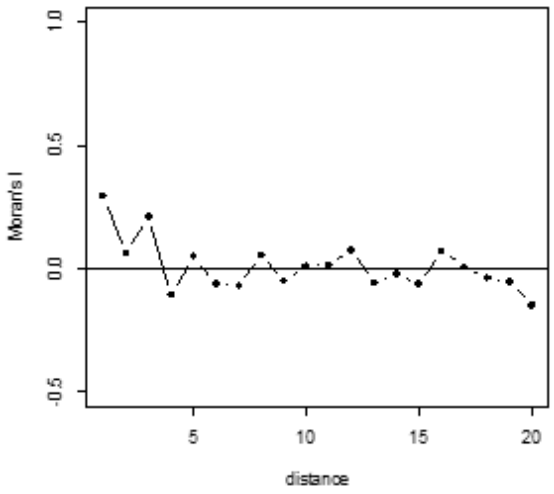
در خلال فرایند مدلسازی، کفایت مدل باید با بررسی فرضیه‌های آن بررسی شود. ابزارهای گرافیکی همانند هیستوگرام، نمودار q-q باقی‌مانده‌های استیودنت‌شده (شکل ۱) نشان می‌دهد که در خصوص تخطی از فرض نرمال بودن مسئله‌ای وجود ندارد. به‌کارگیری آزمون Shapiro-Wilk نیز نشان‌دهنده برقرار بودن فرض نرمال بودن داده‌هاست (W= ۰/۹۹۲ ,P-value= ۰/۷۲۸). بررسی فرض همگنی واریانس با استفاده از تابع ncvTest در بستهٔ car در نرم‌افزار آماری R حکایت دارد که این فرض برقرار بوده (Chisquare = ۲/۶۸ ،P-value = ۰/۱۰۲) و از آن تخطی نشده است. آزمون Breush-Pagan در بستهٔ lmtest در نرم‌افزار آماری R نیز مؤید برقراری فرض ثابت بودن واریانس است (P-value = ۰/۱۴، BP = ۳۰/۳۸).

همبستگی‌نگار موران I برای باقی‌مانده‌های مدل رگرسیون خطی با استفاده از تابع correlog در بستهٔ ncf (Bjornstad, 2013) به‌طور مشخص خودهمبستگی مکانی را نشان می‌دهد که تأیید می‌کند فرض اینکه خطاها به‌طور مستقل توزیع یافته‌اند، دچار تخطی شده است. شکل ۲ نشان می‌دهد که در فواصل کوتاه خودهمبستگی مکانی مثبت و قوی وجود دارد. علاوه‌بر این، مقدار آمارهٔ موران نیز ۰/۱۶۸ بوده که بسیار معنی‌دار است (P-value =۰/۰۰۰) که نشان می‌دهد خودهمبستگی مکانی قوی در مدل وجود دارد. ازاین‌رو استفاده از رویکرد مدلسازی OLS مناسب نیست و مدل‌های خودرگرسیونی توأمان با وقفه مکانی و خطای مکانی باید ارزیابی شوند.



شکل ۱- هیستوگرام و نمودار q-q باقی‌مانده‌های استیودنت‌شده مدل برازش‌داده‌شده

[1] Bayesian Information Criteria



شکل ۲- همبستگی‌نگار باقی‌ماندهٔ مدل رگرسیون خطی

همان‌طور که اشاره شد، گام نخست در به‌کارگیری مدل‌های خودرگرسیونی توأمان، مشخص کردن فواصل همسایگی است. نتایج به‌کارگیری AIC برای تعیین فاصلهٔ همسایگی نشان داد که فاصلهٔ ۱۵۰ متری، کمترین مقدار معیار اطلاعاتی آکائیک را دارد و ازاین‌رو به‌عنوان فاصلهٔ در نظر گرفتن همسایگی در مدل‌های خودرگرسیونی درنظر گرفته می‌شود. نتایج آماری مدل برازش‌یافتهٔ SAR با استفاده از رویکرد درست‌نمایی حداکثر نشان می‌دهد که پارامتر خودهمبستگی مکانی برای مدل با وقفه مکانی تأثیر مثبت دارد (Rho = ۰/۱۶) و معنی‌دار است (P-value = ۰/۰۰۶). از طرف دیگر آزمون نسبت درست‌نمایی برای این پارامتر نیز بسیار معنی‌دار است (P-value = ۰/۰۰۵). اشتباه معیار باقی‌مانده برای مدل SAR با وقفهٔ مکانی کمتر از مدل رگرسیون خطی است (۲/۶۸ در مقابل ۳/۰۹). مقدار AIC برای مدل SAR با وقفهٔ مکانی (۶۳۹/۳۵) کمتر از مدل رگرسیون خطی (۶۴۴/۹۷) است. در مقایسه با مدل رگرسیون خطی مقدار ضریب تبیین به ۵۹/۲۵ درصد افزایش یافته است. آزمون نسبت درست‌نمایی که مدل OLS را به‌عنوان مدل صفر در نظر می‌گیرد نیز نشان می‌دهد که ضریب خودرگرسیونی مکانی معنی دار است (P-value = ۰/۰۰۶, LR = ۷/۶۱) که حاکی از آن است که برازش مدل با افزودن ماتریس وزن مکانی بهبود یافته است و افزودن پارامتر خودرگرسیونی توانسته است عملکرد مدل OLS را کمی بهبود بخشد. برازش مدل SAR با خطای مکانی نیز نشان می‌دهد که پارامتر خودهمبستگی مکانی برای این مدل تأثیر مثبت دارد (Lambda = ۰/۵۳) و بسیار معنی‌دار است (P-value = ۰/۰۰۰). اشتباه معیار باقیمانده برای مدل SAR با خطای مکانی کمتر از مدل رگرسیون خطی (۲/۵۲ در مقابل ۳/۰۹) و مدل SAR با وقفه مکانی است. مقدار AIC برای این مدل (۶۳۴/۶۱) هم کمتر از مدل رگرسیون خطی و هم مدل SAR با وقفهٔ مکانی است. با به‌کارگیری مدل خودرگرسیونی توأمان با خطای مکانی ضریب تبیین بهبود یافته کمتر از (۶۰/۸۱ درصد) و لگاریتم درست‌نمایی نیز افزایش‌یافته است. با مراجعه به جدول ۱ مشاهده می‌شود که با افزودن ماتریس وزن مکانی در مدل‌های خودرگرسیونی برازش مدل با توجه به شاخص‌های آماری بهبود یافته است.

جدول ۱- ارزیابی مدل رگرسیون معمولی و مدل‌های خودرگرسیونی توأمان بر مبنای آماره‌های ارزیابی مدل

مدل	ضریب تبیین (٪)	AIC	لگاریتم درست‌نمایی	اشتباه معیار باقی‌مانده
رگرسیون معمولی	۵۶/۶۳	۶۴۴/۹۷	-۲۹۷/۴۸	۳/۰۹
خودرگرسیونی توأمان با خطای مکانی	۶۰/۸۱	۶۳۴/۶۱	-۲۹۱/۳۰	۲/۵۲
خودرگرسیونی توأمان با وقفهٔ مکانی	۵۹/۲۵	۶۳۹/۳۵	-۲۹۳/۶۸	۲/۶۸

در جدول ۲ مقدار برآوردی ضرایب رگرسیون به همراه P-value تنها برای اثرهای اصلی متغیرهای تبیینی مهم و اثرگذار ارائه شده است. با مراجعه به این جدول مشاهده می‌شود که در برآورد پارامترها و P-value توسط مدل‌های خودرگرسیونی مکانی و غیرمکانی، اختلافاتی وجود دارد. به‌عنوان مثال مقدار P-value برای وزن مخصوص ظاهری که در مدل رگرسیون خطی معمولی در سطح ۰/۰۵ معنی‌دار نبوده است به ۰/۰۴ تغییر یافته است. همچنین در مقدار برآورد پارامترها نیز اختلاف فاحشی به‌خصوص برای مقدار عرض از مبدأ، وزن مخصوص ظاهری و شاخص تابش خورشیدی حاصل شده است. درمجموع، P-value مدل‌های SAR در مقایسه با OLS کاهش بیشتری داشته است که نشان‌دهندهٔ معنی‌دارتر بودن آن پارامتر است.

جدول ۲- نتایج آماری مدل‌های رگرسیونی برای مدلسازی رابطهٔ بین ارتفاع گونهٔ راش و متغیرهای تبیینی

مدل رگرسیونی / متغیر تبیینی	رگرسیون حداقل مربعات		مدل خطای مکانی		مدل وقفهٔ مکانی	
	برآورد پارامتر	P-Value	برآورد پارامتر	P-Value	برآورد پارامتر	P-Value
عرض از مبدأ	۲۴۸/۵۶	۰/۰۰۰	۱۸۵/۵۰۸	۰/۰۰۰	۲۴۳/۷۸	۰/۰۰۰
ارتفاع از سطح دریا	-۰/۱۶	۰/۰۰۰	-۰/۱۳۷	۰/۰۰۰	-۰/۱۷	۰/۰۰۰
وزن مخصوص ظاهری	۲/۶۱	۰/۷۵۰	۱۴/۰۱۸	۰/۰۴۱	۳/۴۴	۰/۶۲۶
درصد رس	-۲/۸۸	۰/۰۰۰	-۲/۵۵۳	۰/۰۰۰	-۳/۰۰	۰/۰۰۰
نسبت کربن به ازت	-۸/۲۴	۰/۰۰۱	-۵/۱۷۱	۰/۰۰۷	-۸/۱۱	۰/۰۰۰
پتاسیم	-۰/۰۶	۰/۵۸۴	-۰/۰۰۲	۰/۹۸۶	-۰/۱۰	۰/۳۰۱
ازت	-۱۶/۷۷	۰/۰۰۰	-۱۴/۳۴۸	۰/۰۰۰	-۱۶/۴۸	۰/۰۰۰
فسفر	۰/۶۸	۰/۰۰۰	۰/۵۳۲	۰/۰۰۰	۰/۵۶	۰/۰۰۰
اسیدیته	-۵/۵۷	۰/۰۳۳	-۲/۸۷۹	۰/۱۸۲	-۴/۱۷	۰/۰۶۷
درصد سیلت	۰/۳۱	۰/۳۷۵	۰/۵۱۴	۰/۰۷۱	۰/۴۰	۰/۱۹۹
شاخص تابش خورشیدی	-۰/۹۹	۰/۹۱۰	-۵/۳۷۱	۰/۴۴۶	-۰/۳۴	۰/۹۴۸

بحث

به‌طور معمول، فرضیه‌های مرتبط با پراکنش، وفور و عملکرد گونه‌های گیاهی نسبت به متغیرهای محیطی با بهره‌گیری از مدل رگرسیون حداقل مربعات ارزیابی می‌شود (Guisan and Zimmermann, 2000). در پژوهش حاضر نیز کفایت و شایستگی مدل رگرسیون خطی معمولی به‌منظور مدلسازی ارتفاع گونهٔ راش به‌عنوان یک مؤلفهٔ مهم در محاسبهٔ شاخص توان تولید رویشگاه نسبت به متغیرهای خاکی و فیزیوگرافی

ارزیابی شده است. نکتهٔ اساسی در فرایند مدلسازی، ارزیابی کارایی مدل‌ها با به‌کارگیری معیارهای تشخیص مدل است (Anselin, 2005). تجزیه‌وتحلیل رگرسیونی در صورتی معتبر خواهد بود که فرضیات رگرسیون برقرار باشد. نقص این فرضیه‌های ممکن است به استدلال‌های آماری نادرست بینجامد. نتایج پژوهش حاضر نشان داد که فرضیه‌های نرمال بودن و همگنی واریانس با استفاده از روش‌های گرافیکی و آزمون‌های آماری برای مدل رگرسیون معمولی برقرار بوده است. دربارهٔ با فرض استقلال، پژوهش حاضر نشان می‌دهد که با توجه به مقدار آمارهٔ موران I و معنی‌داری آن، تخطی از این فرض صورت گرفته است. Zuur *et al.* (2007) اشاره دارند که تخطی از فرض مستقل بودن جدی‌ترین مسئله در تجزیه‌وتحلیل رگرسیون خطی است، چراکه آزمون‌های مهم نظیر آزمون F و آزمون t را نامعتبر می‌سازد. استفاده از مدل نامناسب یا ساختار وابستگی در داده‌ها از دلایل اصلی تخطی از فرض استقلال است (Zuur *et al.*, 2007). در پژوهش حاضر به‌منظور اجتناب از مدل نامناسب از مدل کامل (مدل با اثرهای اصلی درجه‌یک و درجه‌دو و اثر متقابل دوطرفه) استفاده شده است، بر این اساس می‌توان اظهار داشت پدیدهٔ خودهمبستگی مکانی حادث‌شده به‌دلیل وجود وابستگی مکانی در داده است. ازاین‌رو مدل رگرسیون خطی کفایت لازم را در خصوص بررسی رابطهٔ ارتفاع گونهٔ راش و متغیرهای محیطی ندارد. Zuur *et al.* (2007) با بررسی ۳۵ مجموعه داده، به این نتیجه رسیدند که در هیچ‌کدام از این داده‌های واقعی یک مدل رگرسیون خطی که تمام فرضیه‌های در آن صادق باشد، یافت نشد که به‌طور واضح، محدودیت رگرسیون خطی را برای تحلیل داده‌های بوم‌شناسی نشان می‌دهد؛ بنابراین استفاده از مدل‌های جایگزین نظیر مدل‌های خودرگرسیونی توأمان که فرض کنند باقی‌مانده ازنظر مکانی خودهمبسته‌اند و همچنین پدیدهٔ خودهمبستگی مکانی را تصحیح کنند، ضروری به نظر می‌رسد.

پژوهش حاضر نشان می‌دهد زمانی که ماتریس وزن‌های مکانی به مدل‌های خودرگرسیونی مکانی افزوده شود، ضریب تبیین مدل از ۵۶/۶۳ درصد به ۵۹/۲۵ در مدل وقفهٔ مکانی و ۶۰/۸۱ درصد در مدل با خطای مکانی افزایش می‌یابد. شایان‌ذکر است که مقایسهٔ این مدل‌ها براساس تنها ضریب تبیین مناسب نیست؛ ازاین‌رو، معیارهای دیگر برازش همانند لگاریتم درست‌نمایی و معیار اطلاعاتی آکائیک باید برای ارزیابی عملکرد مدل‌ها به کار گرفته شود (Anselin, 2005). با اضافه شدن ماتریس وزن مکانی، مقدار لگاریتم درست‌نمایی از ۲۹۷/۴۸- در مدل OLS به ۲۹۳/۶۸- در مدل وقفهٔ مکانی و ۲۹۱/۳۰- در مدل با خطای مکانی افزایش می‌یافت و مقدار AIC از ۶۴۴/۹۷ در مدل OLS به ۶۳۹/۳۵ در مدل وقفهٔ مکانی و ۶۳۴/۶۱ در مدل با خطای مکانی کاهش پیدا کرد. اختلاف در مقادیر AIC بین مدل رگرسیون خطی و دو مدل خود رگرسیونی حداقل ۵/۶ و حداکثر ۱۰/۴ بوده است. با توجه به اینکه مدل‌های خودرگرسیونی اختلاف AIC بیشتر از سه را نشان می‌دهند، می‌توان گفت این مدل‌ها کارایی بهتری نسبت به مدل رگرسیون حداقل مربعات دارند.

بر اساس این مقایسه‌ها می‌توان گفت با به‌کارگیری مدل‌های خودرگرسیونی توأمان بهبود برازش حاصل شده است. مقایسهٔ مستقیم بین دو مدل SAR با وقفهٔ مکانی و خطای مکانی می‌تواند با استفاده از لگاریتم حداکثر درست‌نمایی صورت گیرد که برای مدل وقفهٔ مکانی این مقدار (۲۹۳/۶۸-) کمتر از مدل خطای مکانی (۲۹۱/۳۰-) است که نشان می‌دهد مدل خودرگرسیونی با خطای مکانی عملکرد بهتری نسبت به مدل با وقفهٔ مکانی دارد که حکایت از آن دارد که فرایند خودرگرسیونی تنها در جملهٔ خطا مشاهده شده است و نه در متغیر پاسخ و متغیرهای محیطی. در مقایسه با مدل OLS و SAR با وقفهٔ مکانی ضریب تبیین در مدل SAR با خطای مکانی ۶۰/۸۱ درصد است که نسبت به مدل رگرسیون خطی و مدل وقفهٔ مکانی به‌ترتیب ۷ درصد

و ۳ درصد بهبود یافته است. همچنین مقدار AIC این مدل نیز کمترین بوده است.

همان‌طور که اشاره شد، در مقادیر پارامترهای برآوردی توسط مدل‌های رگرسیونی حداقل مربعات و مدل‌های خود رگرسیونی مکانی تفاوت‌هایی حاصل شده است. همچنین P-value مدل‌های SAR در مقایسه با OLS کاهش بیشتری داشته است، ازاین‌رو استنباط می‌شود وجود خودهمبستگی مکانی در مدل رگرسیون خطی حداقل مربعات، می‌تواند سبب برآورد اریب‌دار در فرایند مدلسازی شود و آزمون‌های معنی‌داری را گمراه سازد.

نکتهٔ مهم در پژوهش حاضر این است که نتایج تحقیق حاضر به داده‌هایی با توزیع نرمال یا گوسی محدود می‌شود. برای داده‌های غیرگوسی (همانند داده‌های حضور – غیاب، داده‌های شمارشی) استفاده از دیگر مدل‌ها همانند مدل‌های autocovariate، معادلات برآورد تعمیم‌یافته و مدل‌های آمیختهٔ خطی تعمیم‌یافته توصیه می‌شود (Dormann *et al.*, 2007). شایان‌ذکر است که ویژگی مدل‌های خودرگرسیونی این است که رابطهٔ همسایگی (فاصلهٔ همسایگی و ماتریس وزن‌های مکانی) باید به‌عنوان یک ورودی برای این مدل‌ها، مشخص شود، که این امر ضعف این روش‌ها محسوب می‌شود. به‌عنوان یک رویکرد جایگزین، می‌توان از مدل حداقل مربعات تعمیم‌یافته استفاده کرد. مدل حداقل مربعات تعمیم‌یافته به‌جای اینکه به ماتریس وزن که قدرت تأثیر محل‌های مجاور را مشخص می‌کند، متکی باشد، از مدل‌های سمی واریوگرام برای مدلسازی استفاده می‌کند. هرچند مدل‌های مبتنی بر حداقل مربعات تعمیم‌یافته در نوشته‌های آماری برای دهه‌ها شناخته‌شده است، کاربرد آنها در بوم‌شناسی و علوم جنگل تاکنون بسیار محدود بوده است، ازاین‌رو توصیه می‌شود در مطالعات آتی عملکرد این روش در مدلسازی ویژگی‌های ساختاری گونهٔ راش نیز ارزیابی شود.

منابع

احمدی، کوروش، سید جلیل علوی و مسعود طبری کوچکسرایی، ۱۳۹۴. ارزیابی توان تولید رویشگاه راش شرقی (*Fagus orientalis* L.) با استفاده از مدل جمعی تعمیم‌یافته، مجلهٔ جنگل ایران،۷ (۱):۲۲-۲۷.

بی‌نام، ۱۳۹۲. طرح جنگلداری سری ۳ آغوزچال، حوضهٔ آبخیز ۴۶ کجور، تجدیدنظر اول، سازمان جنگل‌ها و مراتع و آبخیزداری کشور، ۱۵۱ ص.

جعفری حقیقی، مجتبی، ۱۳۸۲. روش‌های تجزیهٔ خاک، انتشارات ندای ضحی، ۲۳۶ ص.

رفیعی دارانی، هادی و محمد قربانی، ۱۳۹۳. مشارکت نیروی کار در اقتصاد ملی: تحلیلی در چارچوب رگرسیون فضایی، مجلهٔ تحقیقات مدلسازی اقتصادی، ۱۹ (۱۸): ۱۱۹-۱۴۰.

سوری، داوود و سلیمه منیری جاوید، ۱۳۹۰. مدل تعیین قیمت مسکن، کاربردی از روش رگرسیون موزون جغرافیایی، مجلهٔ مدیریت شهری، ۳۲ (۲۷): ۷-۲۸.

علوی، سید جلیل، قوام‌الدین زاهدی امیری، رامین رحمانی، محمدرضا مروی مهاجر، بارت موریس و جعفر فتحی،۱۳۹۰. تعیین مقدار بهینه و دامنهٔ بوم‌شناختی درخت راش (*Fagus orientalis*) با استفاده از تابع گوسی در جنگل آموزشی و پژوهشی خیرود نوشهر، مجلهٔ محیط‌زیست طبیعی (منابع طبیعی ایران)، ۶۴ (۴): ۳۹۹ -۴۱۵.

Aertsen, W., V. Kint, J. Van Orshoven, K. Ozkan, and 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.

Akaike, H., 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6): 716-723.

Anselin, L., 2005. Spatial regression analysis in R–A workbook, University of Illinois at Urbana,USA press., Arizona, 134 pp.

Beale, C.M., J.J. Lennon, J.M. Yearsley, M.J. Brewer, and D.A. Elston, 2010. Regression analysis of spatial data, *Ecology Letters*, 13(2):246-264.

Beguería, S., and Y. Pueyo, 2009. A comparison of simultaneous autoregressive and generalized least squares models for dealing with spatial autocorrelation, *Global Ecology and Biogeography*, 18(3): 273-279.

Bivand, R., and G. Piras, 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics, *Journal of Statistical Software*, 63(18): 1-36.

Bjornstad, O.N., 2013. ncf: spatial nonparametric covariance functions, R package version 1.1-5. http://CRAN.R-project.org/package=ncf.

Burnham, K.P., and D.R. Anderson, 2002. Model Selection and Inference: A practical information-theoretic approach. Springer-Verlag, New York, 488 pp.

Comstock, J.P., and J.R. Ehleringer, 1992. Plant adaptation in the great basin and Colorado plateau, *The Great Basin Naturalist*, 195-215.

Dahlin, K.M., G.P. Asner, and C.B. Field, 2014. Linking vegetation patterns to environmental gradients and human impacts in a Mediterranean-type island ecosystem, *Landscape Ecology*, 29(9): 1571-1585.

Dormann, C.F., J.M. McPherson, M.B. Araújo, R. Bivand, J. Bolliger, G. Carl, and R. Wilson, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review, *Ecography*, 30(5): 609-628.

Fotheringham, A.S, M.E. Charlton, and C. Brunsdon, 2001. Spatial variations in school performance: a local analysis using geographically weighted regression, *Geographical and Environmental Modelling*, *5*(1): 43-66.

Guisan, A., and N.E. Zimmermann, 2000. Predictive habitat distribution models in ecology, *Ecological Modelling*, 135(2): 147-186.

Keitt, T.H., O.N. Bjørnstad, P.M. Dixon, and S. Citron-Pousty, 2002. Accounting for spatial pattern when modeling organism-environment interactions, *Ecography*, 25(5): 616-625.

Kent, M., 2011. Vegetation description and data analysis: a practical approach. John Wiley and Sons, 414 pp.

Kissling, W.D., and G. Carl, 2007. Spatial autocorrelation and the selection of simultaneous autoregressive models, *Global Ecology and Biogeography*, 17(1): 59-71.

Lee, J., and D. W. Wong, 2001. Statistical analysis with ArcView GIS, John Wiley and Sons inc., Canada, 205pp.

Miller, J., J. Franklin, and R.Aspinall, 2007. Incorporating spatial dependence in predictive vegetation models, *Ecological Modelling*, 202(3): 225-242.

Rangel, T.F., J.A.F. Diniz-Filho, and L.M. Bini, 2010. SAM: a comprehensive application for spatial analysis in macroecology, *Ecography*, 33(1): 46-50.

Segurado, P.A.G.E., M.B. Aráujo, and W.E. Kunin, 2006. Consequences of spatial autocorrelation for niche-based models, *Journal of Applied Ecology*, 43(3): 433-444.

Zuur, A., E.N. Ieno, and G.M. Smith, 2007. Analyzing ecological data, Springer Science and Business Media, 666 pp.

Investigation on the adequacy of ordinary least square regression in modeling the oriental beech tree height in relation to environmental predictors in experimental and educational forest of Tarbiat Modares University

S.J. Alavi[1*], V. Mardanpour[2], and C.F. Dormann[3]

[1]Assistant Prof., Faculty of Natural Resources and Marine Sciences, University of Tarbiat Modares, I. R. Iran
[2]M.Sc.Student, Faculty of Natural Resources and Marine Sciences, University of Tarbiat Modares, I. R. Iran
[3]Prof., Department of Biometry and Environmental System Analysis, University of Freiburg, Germany

(Received: 29 September 2015; Accepted: 8 February 2016)

Abstract

In the current research, the performance of ordinary least square regression model was studied for the task of predicting a key forest structural parameter – tree height – across a study area in Tarbiat Modares University forest research station using a series of edaphic and topographic variables. For this purpose, 123 0.1 ha circular sample plots were established and total height and diameter of *Fagus orientalis* Lipsky trees with DBH ≥ 7.5 cm within each plot was recorded along with elevation, azimuth and slope of the ground. Also, at the center and four geographical aspects of sample plot, soil samples from first layer (0-10 cm) were taken and mixed for analyzing several soil variables. The results showed the OLS model performed moderately well based on R-squared, but exhibited clear signs of spatial autocorrelation (Moran's I =0.168). Adding spatial weighted matrix in spatial simultaneous autoregressive models resulted in removing autocorrelation and statistically significant improvement in model fit. Comparison of spatial lag and error SAR models using AIC, log-likelihood and R squared indicated that error SAR model performs better than lag SAR model. These analyses underscore the importance of controlling for spatial autocorrelation in forest ecology studies and furnish guidelines for future modeling of species performance in relation to environmental predictors.

Keywords: Moran's I statistic, Neighborhoods' structure, Spatial error model, Spatial lag model, Spatial Autocorrelation.

* Corresponding author Tel: 09111580097 Email: j.alavi@modares.ac.ir