



ارزیابی تأثیر مقدار ضریب بتا در عملکرد خوشه‌بندی بتای انعطاف‌پذیر در طبقه‌بندی پوشش گیاهی

نغمه پاک‌گهر^۱، جواد اسحاقی‌راد^{۲*}، غلامحسین غلامی^۳، احمد علیجانپور^۴ و دیوید رابرتز^۴

^۱ دکتری جنگلداری، گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه
^۲ استاد، گروه جنگلداری، دانشکده منابع طبیعی، دانشگاه ارومیه، ارومیه
^۳ استادیار، گروه ریاضی، دانشکده علوم، دانشگاه ارومیه، ارومیه
^۴ استاد، گروه اکولوژی، دانشگاه مونتانا، مونتانا

(تاریخ دریافت: ۱۴۰۰/۶/۹؛ تاریخ پذیرش: ۱۴۰۰/۹/۲۱)

چکیده

خوشه‌بندی از پرکاربردترین روش‌های مختلف طبقه‌بندی و خوشه‌بندی بتای انعطاف‌پذیر از روش‌های موفقی سلسله‌مراتبی تجمعی در طبقه‌بندی جوامع گیاهی است. هدف این بررسی، تعیین مقدار بتای مناسب در روش خوشه‌بندی بتای انعطاف‌پذیر است. برای این پژوهش داده‌های پوشش گیاهی از جنگل‌های هیرکانی و جنگل‌های بلوط زاگرس انتخاب شدند و مقدار مختلف بتا در نتایج خوشه‌بندی بتای انعطاف‌پذیر (۰/۱، ۰/۲۵، ۰/۴، ۰/۶، ۰/۸) با چهار معیار ارزیابی‌کننده PARATNA، MRPP، Silhouette و همبستگی فی ارزیابی شد و نتایج هر معیار ارزیابی‌کننده از بهترین به بدترین رتبه‌بندی شدند. سپس با برآورد میانگین کل ارزیابی‌کننده‌ها، عملکرد خوشه‌بندی‌ها مشخص شد. نتایج این پژوهش نشان داد که در داده‌های ناحیه ریشی هیرکانی خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱ - بهترین عملکرد را دارد، اما خوشه‌بندی با مقدار بتای ۰/۲۵ و ۰/۴ - نیز عملکرد مناسبی دارد. در داده‌های ناحیه ریشی زاگرس خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۲۵ - بهترین عملکرد را دارد و خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱ - در رتبه دوم قرار دارد. بنابراین با توجه به تأثیر اهمیت انتخاب درست روش طبقه‌بندی در تفسیر اکولوژیکی نتایج حاصل، این بررسی با در نظر گرفتن همه نتایج، استفاده از ضریب بتای ۰/۱ و ۰/۲۵ - را برای طبقه‌بندی پوشش گیاهی پیشنهاد می‌کند.

واژه‌های کلیدی: جنگل‌های زاگرس، جنگل‌های هیرکانی، سلسله‌مراتبی.

مقدمه

توانایی در فراهم ساختن همزمان آثار اقلیم، خاک و فیزیوگرافی رویشگاه همواره در درجه اول اهمیت قرار دارد (Peet & Roberts, 2013)، اگرچه با وجود هزاران گونه که در جوامع طبیعی وجود دارند و مجموعه داده‌های چندمتغیره و پیچیده را تشکیل می‌دهند، این کار مشکل است (Lengyel et al., 2018).

ایده اصلی طبقه‌بندی، جدا کردن نمونه‌ها از یکدیگر و قرار دادن آنها در گروه‌های شبیه به هم است. به این معنا که نمونه‌های مشابه باید در یک گروه قرار

یافتن الگوها در داده‌های مربوط به گونه‌ها موضوعی بحث‌برانگیز در علم پوشش گیاهی است. در میان روش‌های مختلف تحلیل پوشش گیاهی، روش‌های طبقه‌بندی ابزاری مهم به‌شمار می‌روند (Schmidtlein et al., 2010).

بنابراین روش‌های طبقه‌بندی پوشش گیاهی به‌سرعت توسعه یافته‌اند (Lötter et al., 2013). طبقه‌بندی رویشگاه‌های طبیعی بر مبنای پوشش گیاهی به‌دلیل

تفکیک می‌شوند (Hämäläinen et al., 2017). روش‌های سلسله‌مراتبی تجمعی از یک قطعه نمونه به‌عنوان اولین خوشه شروع شده و سپس در روند خوشه‌بندی دیگر قطعه نمونه‌ها به خوشه‌ها اضافه می‌شوند تا زمانی که همه قطعات نمونه در یک خوشه قرار گیرند. از طرف دیگر در روش‌های مقسمی، در شروع همه قطعات نمونه در یک خوشه قرار دارند و در ادامه در خوشه‌های مختلف قرار می‌گیرند تا زمانی که هر قطعه نمونه در یک خوشه منحصر به فرد قرار گیرد (Roleček et al., 2009; Fernández & Gómez, 2020). تا کنون به‌طور گسترده از انواع روش‌های طبقه‌بندی خوشه‌ای برای تشخیص گروه‌های اکولوژیکی در پژوهش‌های اکولوژی پوشش گیاهی استفاده شده است (EshghiRad et al., 2011; Esmailzadeh & Asadi, 2014; Lechner et al., 2016; Esmailzadeh & Nourmohammadi, 2006; Xianping et al., 2016). اما وجود روش‌های بسیار متفاوت خوشه‌بندی و نبود معیار واحدی که به‌تنهایی بهترین روش خوشه‌بندی را معرفی کند، در تشخیص درستی و نادرستی نتایج توسط پژوهشگر بسیار مشکل و گمراه‌کننده خواهد بود (Mufti, 2006; Bertrand & Bertrand, 2006). با وجود محبوبیت بعضی از روش‌ها نسبت به دیگر روش‌های طبقه‌بندی، تاکنون هیچ روشی به‌طور گسترده از دیدگاه نظری یا عملی پذیرفته نشده است (Belbin & McDonald, 1993). خوشه‌بندی بتای انعطاف‌پذیر (Flexible beta clustering) یکی از روش‌های موفق سلسله‌مراتبی تجمعی در خوشه‌بندی جوامع گیاهی است که در سال‌های اخیر توجه بسیاری از محققان در زمینه اکولوژی را جلب کرده است (Pakgothar et al., 2021). این روش توسط Lance & Williams (1967) ابداع شد و گفته شده است که عملکرد آن حد واسط روش ادغام میانگین (Average linkage method) و روش ادغام دورترین همسایه (Complete linkage method) است (Legendre & Legendre, 1998). این الگوریتم خوشه‌بندی دارای پارامتری به‌نام

گیرند و با نمونه‌های گروه‌های دیگر بیشترین تفاوت را داشته باشند (Jain et al., 1988). بنابراین طبقه‌بندی پوشش گیاهی برای تفکیک داده‌های ناهمگن به گروه‌های همگن‌تر انجام می‌گیرد تا بررسی تغییرات پوشش گیاهی آسان‌تر شود (EshghiRad et al., 2009). طبقه‌بندی پوشش گیاهی از دیرباز در پژوهش‌های جامعه‌شناسی گیاهی کاربرد داشته است و امروزه به‌منزله ابزار تحلیل داده‌های چندمتغیره بخشی جدایی‌ناپذیر از علوم گیاهی محسوب می‌شود (Lengyel & Podani, 2015; Khabazi et al., 2019). گرچه هنوز روش‌های طبقه‌بندی پوشش گیاهی به تصمیمات ذهنی‌ای وابسته‌اند که بر نتایج تجزیه و تحلیل تأثیرات معنی‌داری دارند (De Cáceres et al., 2015). به‌عبارت دیگر، روش‌های عددی برای طبقه‌بندی پوشش گیاهی به کار گرفته می‌شوند تا با کاهش دادن ابعاد داده‌ها، الگوی موجود در جوامع گیاهی را پیدا کنند و تجزیه و تحلیل داده‌ها را از روش ذهنی به روش‌های عینی سوق دهند (Lengyel et al., 2018; Saberi et al., 2021).

انتخاب روش طبقه‌بندی مناسب تأثیر بسیار مهمی در کارایی نقشه‌های موضوعی ناشی از آن دارد (Bihanta et al., 2020). در میان روش‌های مختلف طبقه‌بندی، روش خوشه‌بندی را می‌توان از پرکاربردترین روش‌ها نام برد (Wang et al., 2019) که براساس آن واحدهای نمونه‌ای مانند قطعات نمونه یا توده‌ها را مبتنی بر تشابه یا نبود تشابه ترکیب گونه‌ای دسته‌بندی می‌کند (Rodriguez et al., 2019). روش‌های خوشه‌بندی به دو دسته خوشه‌بندی سلسله‌مراتبی و خوشه‌بندی غیرسلسله‌مراتبی تقسیم می‌شوند (Peet & Roberts, 2013). خوشه‌بندی سلسله‌مراتبی نوعی ساختار درختی برای نمایش نتایج خوشه‌بندی ایجاد می‌کند. این روش‌ها به ماتریس تشابه/ نبود تشابه یا معیار فاصله نیز وابسته‌اند (Zaki et al., 2014; Hämäläinen et al., 2017) که این روش به روش‌های تجمعی و روش‌های مقسمی

از سطح دریا در اسلام‌آباد غرب (استان کرمانشاه) استفاده شد (جدول ۲).

شیوه اجرای پژوهش

در منطقه مورد مطالعه در جنگلهای هیرکانی، برای تعیین مراکز قطعات نمونه از روش منظم تصادفی با ابعاد شبکه ۲۰۰×۱۰۰ استفاده شد. در هر مرکز قطعه نمونه، قطعات نمونه ۱۰۰ متر مربعی (۱۰ متر × ۱۰ متر) برای بررسی پوشش علفی و ۴۰۰ متر مربعی (۲۰ متر × ۲۰ متر) برای بررسی پوشش درختی و درختچه‌ای انتخاب شد. برآورد فراوانی و پوشش گونه‌های درختی، درختچه‌ای و علفی براساس مقیاس براون بلانکه انجام گرفت (Eshaghi Rad et al., 2009). میانگین هر طبقه (از جدول براون بلانکه) میانگین درصد پوشش هر گونه در نظر گرفته شد. در منطقه مورد مطالعه در جنگلهای بلوط زاگرس، در هر قطعه با استفاده از سه ترانسکت که در فاصله‌های ۲۰۰ متری از یکدیگر قرار گرفته بودند و در جهت شیب پیاده شدند، از پوشش گیاهی قطعات نمونه برداری شد (اولین ترانسکت به صورت تصادفی و ترانسکت‌های بعدی به صورت منظم - تصادفی پیاده شدند). نقاط برداشت پوشش گیاهی در هر ترانسکت در فاصله‌های صفر، ۲۵، ۵۰، ۱۰۰ و ۱۵۰ متری بود (Eshaghi Rad et al., 2017). زیرا این مقدار بیش از دوبرابر فاصله‌ای است که اثر حاشیه‌ای بر ترکیب پوشش زیراشکوب مؤثر است. البته باید یادآور شد که پوشش گیاهی منطقه بر این موضوع تأثیرگذار است (Matlack, 1994). برای برداشت پوشش علفی، در هر نقطه پنج قطعه نمونه ۰/۲۵ متر مربعی (۰/۵×۰/۵ متر) عمود بر ترانسکت در سمت چپ و راست پیاده و در هر کوادرات ضمن شناسایی گونه‌ها با استفاده از جدول براون بلانکه، فراوانی و چیرگی هر گونه ثبت شد. میانگین هر طبقه (از جدول براون بلانکه) به‌عنوان میانگین درصد پوشش هر گونه در نظر گرفته شد.

بتا است که بین ۱ تا ۱- توسط کاربر تعیین می‌شود (Milligan, 1989). تعیین عدد مشخصی برای مقدار بتا برای تضمین بهترین عملکرد این خوشه‌بندی موضوعی چالش‌برانگیز است (Lance & Williams, 1967) و انتخاب هر مقدار بتا در محدوده مذکور، نتایج خوشه‌بندی بتای انعطاف‌پذیر را تحت تأثیر قرار خواهد داد (Everitt et al., 2011). اعتقاد بر این است که مقدار بتای منفی‌تر تأثیر بهتری بر عملکرد خوشه‌بندی دارد (Belbin et al., 1992). Dufrene & Legendre (1997) نیز مقدار بتای ۰/۲۵- را توصیه کردند و بسیاری از محققان نیز از این دستورالعمل پیروی می‌کنند؛ اما دیگر مقادیر مقدار بتا کمتر استفاده شده و تحقیقات کمی به بررسی این موضوع پرداخته‌اند (Lengyel et al., 2020). از این‌رو هدف این پژوهش، ارزیابی تأثیر مقدار بتای مختلف بر خروجی اجرای روش خوشه‌بندی بتای انعطاف‌پذیر در طبقه‌بندی پوشش گیاهی است.

مواد و روش‌ها

منطقه پژوهش

در این بررسی از داده‌های پوشش گیاهی برداشت‌شده از پارسل‌های ۲۹ (مدیریت‌شده) و ۲۶ (شاهد) از سری ۱۰ لالیس، پارسل‌های ۱۳۴ (مدیریت‌شده) و ۱۳۵ (شاهد) از سری ۱ شیراکنس، پارسل‌های ۵۳۰ (مدیریت‌شده) و ۵۴۱ (شاهد) از سری ۵ لاکوبن، پارسل‌های ۷۲۵ (مدیریت‌شده) و ۷۲۴ (شاهد) از سری ۷ واشمرد از سری ۳ جمند، پارسل‌های ۳۱۷ (مدیریت‌شده) و ۳۱۸ (مدیریت‌نشده) واقع در حوزه استحفاظی اداره کل منابع طبیعی نوشهر استان مازندران استفاده شد (Khanalizadeh et al., 2020). مشخصات عمومی پارسل‌های انتخاب‌شده در جدول ۱ ارائه شده است (جدول ۱). همچنین از داده‌های پوشش گیاهی برداشت‌شده از جنگل‌های بلوط زاگرس مربوط به سه قطعه نمونه جنگلی با دامنه جنوبی از جنگل‌های بلوط با شرایط مشابه از نظر ارتفاع

جدول ۱- مشخصات عمومی پارسل‌های انتخاب‌شده در جنگل‌های هیرکانی

سری	پارسل	مساحت (هکتار)	تعداد قطعات نمونه	جهت عمومی- محدوده ارتفاعی (متر)	طول جغرافیایی	عرض جغرافیایی
سری ۴ دهگا	۴۱۴ (مدیریت‌شده)	۷۱	۲۰	شرقی	۵۰°، ۵۲'، ۳۴"	۱۵°، ۴۰'، ۲۵"
	۴۱۲ (شاهد)	۵۹		شمالی	تا ۱۵"	تا ۳۶°، ۴۴'
سری ۱۰ لالیس	۲۹ (مدیریت‌شده)	۶۷	۲۰	شمال شرقی	۵۱°، ۲۳'، ۳۰"	۳۶°، ۴۴'، ۳۲"
	۲۶ (شاهد)	۵۹		شمالی	تا ۵۱°، ۲۸'	تا ۳۶°، ۴۴'، ۳۶"
سری ۱ شیراکنس	۱۳۴ (مدیریت‌شده)	۱۱۳	۲۰	جنوب شرقی	۵۱°، ۳۰'، ۰۰"	۳۶°، ۳۵'، ۲۷"
	۱۳۵ (شاهد)	۱۲۶		جنوب شرقی	تا ۲۸"	تا ۳۶°، ۳۰'، ۰۰"
سری ۵ لاکوبن	۵۳۰ (مدیریت‌شده)	۲۴/۷	۲۰	شمال غربی	۵۱°، ۱۶'، ۱۵"	۳۶°، ۵۰'، ۳۶"
	۵۴۱ (شاهد)	۹۲		شمال غربی	تا ۱۸'، ۵۰"	تا ۳۶°، ۴۰'، ۳۶"
سری ۷ واشمرد	۷۲۵ (مدیریت‌شده)	۳۴	۲۰	جنوب شرقی	۵۱°، ۱۸'، ۱۵"	۳۶°، ۳۷'، ۳۶"
	۷۲۴ (شاهد)	۵۸		شرقی	تا ۵۱°، ۱۸'	تا ۳۶°، ۳۹'، ۳۶"
سری ۳ جمند	۳۱۷ (مدیریت‌شده)	۶۴	۲۰	شمال شرقی	۳۰° و ۰۰' و ۰۰"	۳۶° و ۳۵' و ۳۰"
	۳۱۸ (شاهد)	۴۷		شمال شرقی	۵۱° تا ۲۸"	۳۶° تا ۲۷" و ۳۵' و ۳۶"

جدول ۲- مشخصات عمومی جنگل انتخاب‌شده در جنگل‌های زاگرس

جنگل چهار زبر	مساحت (هکتار)	تعداد قطعه نمونه	جهت عمومی- ارتفاع از سطح دریا (متر)	طول جغرافیایی	عرض جغرافیایی
قطعه ۱	۴/۵	۱۵	شمال شرقی- ۱۷۰۰	۴۶°، ۴۹' تا ۴۶°، ۳۹'	۳۴°، ۱۴' تا ۳۹°، ۹'
قطعه ۲	۵	۱۵	شمال شرقی- ۱۶۸۰	۴۶°، ۴۹' تا ۴۶°، ۳۹'	۳۴°، ۱۴' تا ۳۹°، ۹'
قطعه ۳	۵/۵	۱۵	شمال شرقی- ۱۶۵۰	۴۶°، ۲۹'، ۲"	۳۴°، ۵'، ۲"

روش تحلیل

برای آماده‌سازی داده‌ها از تبدیل هلینگر (Hellinger) استفاده شد (Legendre & Gallagher, 2001) تا پراکنش داده‌ها دارای تقارن بیشتری باشد و تعداد صفرها کم شود. برحسب همگنی داده‌ها، برای داده‌های ترکیب پوشش گیاهی ناحیه هیرکانی دو تا ۵ پنج خوشه و برای داده‌های ترکیب پوشش گیاهی ناحیه زاگرس دو تا شش خوشه انتخاب شد. مراحل

رتبه‌بندی و انتخاب روش خوشه‌بندی مناسب برای هر

یک از داده‌ها به شرح زیر است:

ابتدا خوشه‌بندی‌های مختلف با مقدار بتای متفاوت انجام گرفت. سپس نتایج خوشه‌بندی با هر معیار ارزیابی شده و نتایج به دست آمده رتبه‌بندی شد (از بهترین به بدترین رتبه). در مرحله بعد خوشه‌بندی‌ها با تعداد خوشه مختلف (دو تا شش خوشه برای داده‌های زاگرس و دو تا پنج خوشه برای

(1967) معرفی شده است. آنها به این منظور روش اتصال معمول خوشه‌ها را تغییر دادند و با اتصال دو خوشه i و j یک فاصله بین خوشه k و خوشه ij ایجاد کردند (رابطه ۳)

$$BC_{1,2} = \frac{2 \sum_{i=1}^m \min(Y_{1K}, Y_{2K})}{\sum_{k=1}^m (Y_{1K} + Y_{2K})} \quad \text{رابطه ۲}$$

برای ماتریس داده‌ای با m گونه، مقدار k گونه در i ام قطعه نمونه برابر است با Y_{ik}

$$d_{k(ij)} = \alpha_i d_{ki} + \alpha_j d_{kj} + \beta d_{ij} + \gamma |d_{ki} - d_{kj}| \quad \text{رابطه ۳}$$

جایی که d_{ij} فاصله بین خوشه‌های i و j است. از این فرمول به عنوان طرح جدید انعطاف‌پذیر استفاده کردند با مقدار پارامتر $\alpha_i + \alpha_j + \beta = 1$ ، $\alpha_i = \alpha_j$ ، $\beta < 1$ ، $\gamma = 0$. با تغییر مقدار بتا، طرح خوشه‌بندی با ویژگی‌های مختلف فراهم می‌شود. محدوده تعریف شده مقدار بتای -1 تا $+1$ است این خوشه‌بندی حد واسط بین خوشه‌بندی نزدیک‌ترین همسایه و دورترین همسایه است. مقدار بتای کوچک در محدوده منفی همانند -0.25 را توصیه کردند، اگرچه Schneider & Schneider (1985) مقدار بتای -0.5 را پیشنهاد دادند. با توجه به اهمیت انتخاب مقدار بتا در خوشه‌بندی بتای انعطاف‌پذیر، در این بررسی مقدار بتا -0.1 ، -0.25 ، -0.4 ، -0.6 و -0.8 انتخاب شده است تا محدوده مؤثر بتا ارزیابی شود.

معیارهای ارزیابی کیفیت خوشه‌بندی

انتخاب چند شاخص برای گزینش روش خوشه‌بندی مناسب بهتر از استفاده از فقط یک شاخص است، زیرا هر شاخص معایب خاصی دارد (Gordon, 1999) و پیچیدگی داده‌ها نتایج را تحت تأثیر قرار می‌دهد (Olkin & Sampson, 2001). بنابراین در این بررسی همزمان از چند روش ارزیابی خوشه‌بندی استفاده شد تا خوشه‌بندی مناسب با دقت بیشتری انتخاب شود.

داده‌های هیرکانی) انجام گرفت و در هر بار نتایج هر معیار رتبه‌بندی شد. سپس میانگین هر معیار برای خوشه‌بندی با بتای مختلف و برای تعداد خوشه‌های مختلف محاسبه شد و در نهایت از اعداد محاسبه‌شده در مرحله قبل میانگین گرفته شد و رتبه‌بندی نهایی انجام گرفت. به این ترتیب خوشه‌بندی مناسب معرفی شد. برای بررسی بهتر، نتایج هر ارزیاب به صورت نمودار جعبه‌ای رسم شد (Aho et al., 2008; Schmidtlein et al., 2010; Roberts, 2015). همه محاسبات در نرم‌افزار R 3.6.1 در بسته‌های آماری (Oksanen et al., 2013) vegan، (Maechler et al., 2013) cluster، (R Core Team, 2019). مراحل یادشده در زیر به صورت اجمالی تشریح شده است.

تبدیل داده هلینگر

تبدیل داده هلینگر برای داده‌های درصد تاج‌پوشش گونه‌ها بسیار مناسب است که وزن کمی به گونه‌های نادر می‌دهد. با توجه به رابطه ۱ تبدیل داده‌ها شامل تقسیم هر مقدار به مجموع هر ردیف در ماتریس داده می‌شود و سپس مجذور هر مقدار محاسبه می‌شود (Legendre & Gallagher, 2001).

$$y'_{ij} = \sqrt{\frac{y_{ij}}{y_{i+}}} \quad \text{رابطه ۱}$$

j شاخص گونه، i شاخص قطعات نمونه یا رویشگاه و i^+ نشان‌دهنده مجموع ردیف‌ها برای i ام نمونه است

خوشه‌بندی بتای انعطاف‌پذیر

در این بررسی از روش اندازه‌گیری فاصله بری-کورتیس (Bray-Curtis) استفاده شد که براساس پژوهش‌های متعدد، روش اندازه‌گیری مناسب برای داده‌های جوامع اکولوژیکی گیاهی است (Bray & Curtis, 1957; Roberts, 2015) (رابطه ۲) و برای اتصال خوشه‌ها روش اتصال بتای انعطاف‌پذیر انتخاب شد (رابطه ۳) این روش توسط Lance & Williams

شاخص میانگین سیلوئت

این معیار برای یافتن تعداد بهینه خوشه‌ها و ارزیابی کیفیت خوشه‌ها استفاده می‌شود. این معیار از فشردگی (میانگین فاصله داخل خوشه‌ها) و جدایش (میانگین فاصله بین خوشه‌ها) استفاده می‌کند (Rousseeuw, 1987). میانگین سیلوئت $s(i)$ به صورت رابطه ۴ تعریف می‌شود:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad \text{رابطه ۴}$$

که در آن $a(i)$ عدم تشابه بین نمونه‌هاست و $b(i)$ عدم تشابه با نزدیک‌ترین خوشه است. محدوده این شاخص بین ۱- تا ۱ است. مقدار $s(i)$ نزدیک ۱ نشان‌دهنده طبقه‌بندی خوب است. اگر $s(i)$ صفر باشد طبقه‌بندی منطقی به نظر نمی‌رسد و اگر $s(i)$ ۱- باشد نشان‌دهنده طبقه‌بندی اشتباه است.

آنالیز پارتیشین (PARTANA)

این معیار با استفاده از رابطه ۵ مقدار نسبت تشابه درون گروهی به بین گروهی را محاسبه می‌کند (Aho et al., 2008).

$$P = \frac{\sum_{z=1}^c \sum_{i \in z} \sum_{j=i+1}^N S_{ij} / \sum_{z=1}^c (n_z^2 - n_z) / 2}{\sum_{i=1}^{N-1} \sum_{j=i+1}^N S_{ij} / \sum_{z=1}^{c-1} \sum_{k=z+1}^c n_z \cdot n_k} \quad \text{رابطه ۵}$$

P نسبت PARTANA است، N تعداد کل نمونه‌ها، C تعداد خوشه‌ها، n_z تعداد نمونه در خوشه z ام ($z = 1, 2, \dots, C$) و S_{ij} تشابه بین نمونه‌های i و j است؛ بنابراین نشان می‌دهد که i و j نمونه‌هایی از خوشه‌های متفاوت هستند. مقادیر بیشتر این شاخص نشان‌دهنده خوشه‌بندی بهتر است.

ضریب همبستگی فی

این ضریب مقدار تعلقه گونه‌ها را براساس داده‌های حضور و غیاب محاسبه می‌کند. این شاخص می‌تواند مقدار تعلقه گونه‌ها را در داده‌های با تعداد رویشگاه‌های مختلف محاسبه کند (Tichy & Chytry, 2006). ضریب همبستگی فی (Φ) محدوده بین ۱- تا

+۱ است. مقدار فی مثبت برای شناسایی گونه‌های معرف مهم است و مقدار فی منفی برای یافتن مقدار اختلاف بین جوامع اهمیت دارد. گونه‌های با مقدار مثبت به‌عنوان گونه‌های تعلقه شناخته شده و فقط گونه‌هایی با مقدار $p\text{-value} < 0.001$ در آزمون فیشر بررسی شدند و براساس مجموع مقادیر فی‌های مثبت معنی‌دار رتبه‌بندی انجام می‌گیرد.

روش پاسخ چندگانه جایگشت

پاسخ چندگانه جایگشت (Multi Response Permutation Procedure) روش نوعی روش غیرپارامتریک برای آزمون اختلاف میان دو یا چند گروه است (Mielke & Berry, 2007). این روش آماری برای بررسی اختلاف گروه‌ها با در نظر گرفتن همه متغیرها کاربرد دارد (Esmailzadeh & Nourmohammadi, 2016). تحلیل MRPP با ایجاد دو آماره T و A امکان بررسی کیفیت گروه‌ها را فراهم می‌کند. آماره T برای تعیین مقدار اختلاف بین گروه‌های طبقه‌بندی شده استفاده می‌شود و آماره A همگنی داخل گروه‌ها را تعیین می‌کند.

نتایج

پنج روش خوشه‌بندی با چهار معیار ارزیابی کننده در دو سری داده پوشش جنگلی بررسی شد. برای یافتن بتای مناسب در الگوریتم خوشه‌بندی بتای انعطاف‌پذیر، مقادیر مختلف بتا ارزیابی شد. برای هر الگوریتم، میانگین رتبه در بین معیارهای ارزیابی کننده مختلف و تعداد مختلف خوشه‌ها محاسبه شد و با استفاده از رتبه میانگین در بین الگوریتم‌های مختلف روش مناسب معرفی شد. از آنجا که درج تک‌تک نتایج خوشه‌بندی برای تعداد مختلف خوشه‌ها و بررسی آنها توسط معیارهای ارزیابی کننده امکان‌پذیر نبود، نتایج میانگین رتبه‌بندی شده هر معیار ارزیابی کننده و میانگین کل ارزیابی کننده به صورت جدول ۳ برای داده‌های هیرکانی و جدول ۴ برای داده‌های زاگرس ارائه شده است.

جدول ۴ میانگین رتبه‌بندی شده الگوریتم‌های مختلف خوشه‌بندی را با تعداد مختلف خوشه برای هر معیار ارزیابی در داده‌های زاگرس نشان می‌دهد. براساس نتایج به‌دست آمده، الگوریتم خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۲۵ - بیشترین رتبه را در بین دیگر رتبه‌بندی‌های دارد. روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱ - در رتبه دوم قرار دارد و ضعیف‌ترین خوشه‌بندی برای این مجموعه داده بتای انعطاف‌پذیر ۰/۸ - معرفی شد.

نتایج هر معیار ارزیابی‌کننده برای تعداد مختلف خوشه در داده‌های زاگرس در شکل ۲ نمایش داده شده است. نتایج این نمودارها برای معیار ارزیابی‌کننده‌های سیلوئت، همبستگی فی و معیار آنالیز پارتیشن روش خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۱ - بهترین عملکرد را دارد، اما ارزیابی‌کننده MRPP عملکرد خوشه‌بندی بتای انعطاف‌پذیر ۰/۶ - را بهتر می‌داند.

جدول ۳ نتایج رتبه‌بندی شده پنج الگوریتم خوشه‌بندی را براساس میانگین معیارهای ارزیابی‌کننده مختلف برای تعداد مختلف خوشه‌بندی نشان می‌دهد. برای داده‌های هیرکانی عملکرد خوشه‌بندی بتای انعطاف‌پذیر ۰/۱ - بهتر از دیگر خوشه‌بندی‌هاست. الگوریتم خوشه‌بندی بتای انعطاف‌پذیر ۰/۲۵ - و ۰/۴ - در رتبه بعدی قرار دارند. اما الگوریتم خوشه‌بندی ۰/۶ - ضعیف‌تر از دیگر خوشه‌بندی‌هاست.

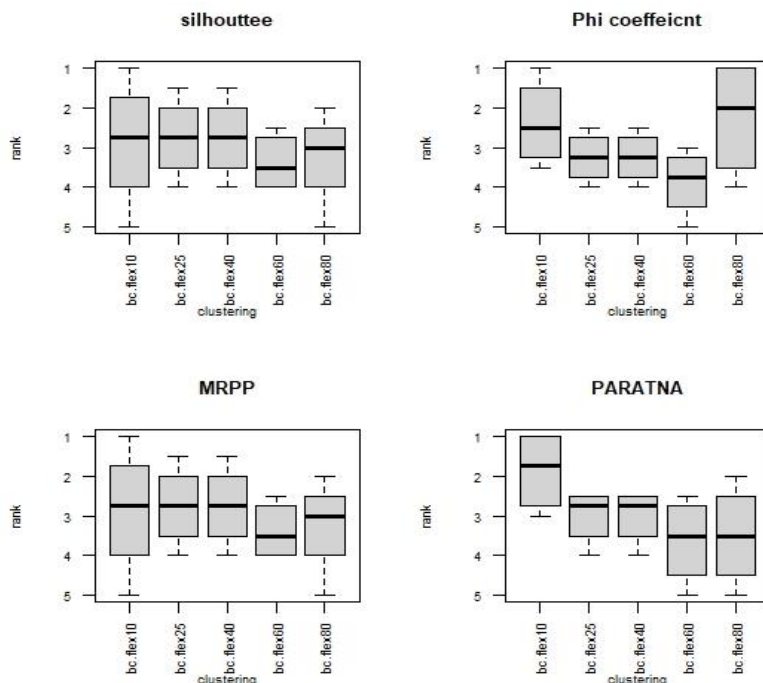
نتایج میانگین هر معیار ارزیابی برای تعداد مختلف خوشه برای داده‌های هیرکانی به‌صورت نمودار جعبه‌ای در شکل ۱ ارائه شده است. براساس معیار سیلوئت و MRPP روش خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۱ -، ۰/۲۵ - و ۰/۴ - رتبه بیشتری دارند، اما معیارهای همبستگی فی و معیار آنالیز پارتیشن فقط روش خوشه‌بندی بتای انعطاف‌پذیر ۰/۱ - را به‌عنوان بهترین خوشه‌بندی معرفی می‌کنند.

جدول ۳- نتایج رتبه‌بندی کلی الگوریتم‌های مختلف خوشه‌بندی برای داده‌های هیرکانی

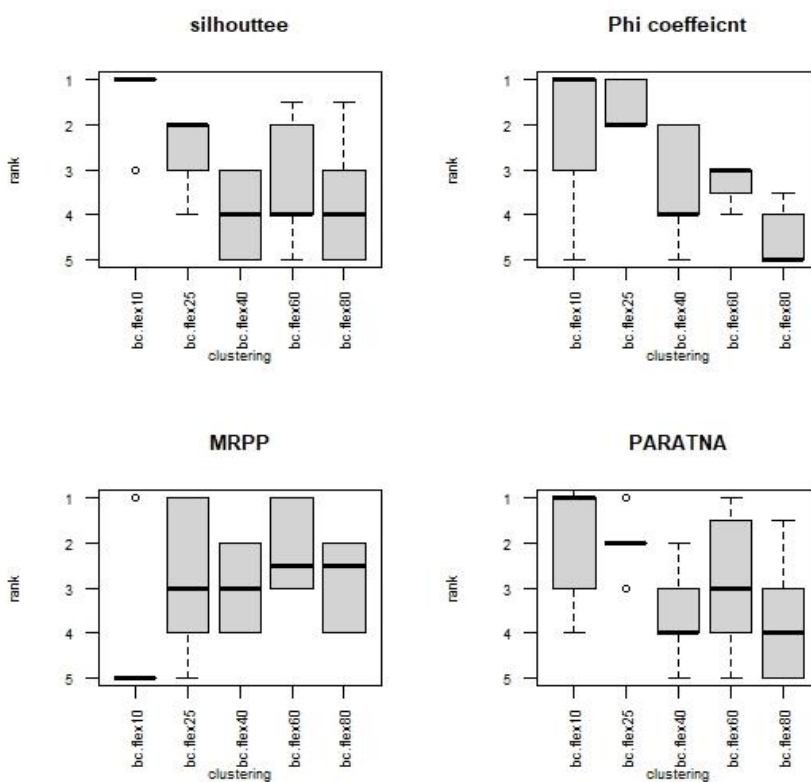
رتبه	میانگین	MRPP	آنالیز پارتیشن	فی	سیلوئت	خوشه‌بندی
۱	۲/۴۳	۲/۷۵	۱/۸۷	۲/۳۷	۲/۷۵	Flexible-β(-0.1)
۲/۵	۲/۹۴	۲/۷۵	۳	۳/۲۵	۲/۷۵	Flexible-β(-0.25)
۲/۵	۲/۹۴	۲/۷۵	۳	۳/۲۵	۲/۷۵	Flexible-β(-0.4)
۵	۳/۵۶	۳/۳۷	۳/۶۲	۳/۸۷	۳/۳۷	Flexible-β (-0.6)
۴	۳/۰۶	۳/۲۵	۳/۵	۲/۲۵	۳/۲۵	Flexible-β (-0.8)

جدول ۴- نتایج رتبه‌بندی کلی الگوریتم‌های مختلف خوشه‌بندی برای داده‌های زاگرس

رتبه	میانگین	MRPP	آنالیز پارتیشن	فی	سیلوئت	خوشه‌بندی
۲	۲/۴	۴/۲	۱/۸	۲/۲	۱/۴	Flexible-β(-0.1)
۱	۲/۲	۲/۸	۲	۱/۶	۲/۶	Flexible-β(-0.25)
۴	۳/۵	۳	۳/۶	۳/۴	۴	Flexible-β(-0.4)
۳	۲/۸	۲/۱	۲/۹	۳/۳	۳/۳	Flexible-β (-0.6)
۵	۳/۷	۲/۹	۳/۷	۴/۵	۳/۷	Flexible-β (-0.8)



شکل ۱- نمودار جعبه‌ای نتایج رتبه‌بندی ارزیابی‌کننده‌ها (RARATNA و Silhouette width, Phi coefficient, MRPP) برای خوشه‌بندی‌های مختلف ناحیه‌ی رویشی هیرکانی، محور افقی نشان‌دهنده‌ی خوشه‌بندی‌های مختلف و محور عمودی نشان‌دهنده‌ی رتبه‌ها (رتبه‌ی یک نشان‌دهنده‌ی بهترین خوشه‌بندی و رتبه‌ی پنج نشان‌دهنده‌ی بدترین خوشه‌بندی)



شکل ۲- نمودار جعبه‌ای نتایج رتبه‌بندی ارزیابی‌کننده‌ها (RARATNA و Silhouette width, Phi coefficient, MRPP) برای خوشه‌بندی‌های مختلف ناحیه‌ی رویشی زاگرس، محور افقی نشان‌دهنده‌ی خوشه‌بندی‌های مختلف و محور عمودی نشان‌دهنده‌ی رتبه‌ها (رتبه‌ی یک نشان‌دهنده‌ی بهترین خوشه‌بندی و رتبه‌ی پنج نشان‌دهنده‌ی بدترین خوشه‌بندی)

بحث

در داده‌های پوشش گیاهی منطقه رویشی زاگرس نیز روش خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۲۵- عملکرد بهتری از خوشه‌بندی‌های بررسی شده دارد. بررسی در داده‌های جنوب شرقی استرالیا و داده‌های شبیه‌سازی نشان داد که خوشه‌بندی سلسله‌مراتبی بتای انعطاف‌پذیر روشی منطقی برای افزایش پیشگویی عملکرد مدل است (Lyons et al., 2016). نتایج بررسی در جنگل‌های بومی Mpumalanga در منطقه KwaZulu و Limpopo در آفریقای جنوبی نشان شد که روش خوشه‌بندی اتصال وارد و خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۲۵- اغلب عملکرد خوبی در بین ماتریس‌های مختلف اندازه‌گیری داشتند (Lötter et al., 2013) و در بیشتر مطالعات قبلی تنها استفاده از بتای انعطاف‌پذیر با مقدار ۰/۲۵- رایج بود چرا که عملکرد آن مشابه خوشه‌بندی وارد است که برای خوشه‌های کروی با تعداد خوشه‌ها برابر مناسب است. بنابراین می‌توان بیان کرد که خوشه‌بندی بتای انعطاف‌پذیر نیز برای داده‌هایی با توزیع کروی مناسب است که خوشه‌های کروی و هم‌اندازه ایجاد می‌کنند (Aho et al., 2008; Schmidlein et al., 2010).

نتایج پژوهش در مراتع مجارستان، جنگل‌های وایومینگ و یوتا در ایالات متحده آمریکا و همچنین داده‌های شبیه‌سازی شده نشان داد که با افزایش مقدار بتا، خوشه‌بندی بتای انعطاف‌پذیر که به‌عنوان خوشه‌بندی اولیه برای الگوریتم‌های خوشه‌بندی Optsil، REMOS1 و REMOS2 به کار برده می‌شود، میانگین شاخص سیلوئت کاهش می‌یابد (Lengyel et al., 2019). آنها بر تأثیر گروه‌بندی در مقادیر کمتر بتا تأکید کردند و اظهار داشتند که با افزایش مقدار بتا وزن بیشتری به زنجیرها داده می‌شود. آنها همچنین بیان کردند که مقدار بتای ۱- در عملکرد خوشه‌بندی بتای انعطاف‌پذیر مساوی با عملکرد روش خوشه‌بندی روش اتصال دورترین همسایه است. بتای صفر با روش خوشه‌بندی ادغام میانگین برابر است و بتای ۱+ برابر با

خوشه‌بندی روش بسیار پیچیده‌ای است، زیرا انتخاب روش خوشه‌بندی نیازمند تصمیم‌گیری متعدد راهبردی است (Rodriguez et al., 2019). در این پژوهش خوشه‌بندی با چهار ارزیابی‌کننده مختلف بررسی شد. انتخاب چند شاخص برای انتخاب روش خوشه‌بندی مناسب بهتر از استفاده از تنها یک شاخص است، زیرا هر شاخص معایب خاص خود را دارد (Gordon, 1999) و پیچیدگی داده‌ها نتایج را تحت تأثیر قرار می‌دهد (Olkin & Sampson, 2001). الگوریتم ترکیبی بتای انعطاف‌پذیر با تغییر محدوده پارامتر بتا از ۱- تا ۱+ سبب ایجاد خوشه‌بندی‌های مختلفی می‌شود (Lance & Williams 1967) که در محدوده مثبت مقدار بتای خوشه‌بندی همانند نزدیک‌ترین همسایه که تمایل به ایجاد خوشه‌ها به حالت زنجیره دارد ایجاد می‌کند؛ اما در محدوده صفر تا ۱- که در این پژوهش نیز بررسی شد، خوشه‌بندی‌های بهتری ایجاد می‌شود (Lengyel et al., 2018). در این بررسی خوشه‌بندی بتای انعطاف‌پذیر بررسی شد، زیرا توانایی ارائه بهتری برای نمایش جوامع گیاهی را با حداقل از دست دادن اطلاعات پوشش گیاهی یا بدون از دست دادن این اطلاعات دارد (Milligan, 1989). براساس پژوهش‌های انجام‌گرفته بهترین محدوده بتا بین ۰/۴- تا ۰/۱- است (Milligan, 1989). نتایج این بررسی در داده‌های ناحیه رویشی هیرکانی مشهود است و محدوده بتای ۰/۴- تا ۰/۱- عملکرد بهتری را نشان می‌دهد. اما نتایج کلی این پژوهش نشان داد که روش بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و ۰/۲۵- عملکردی بهتر از دیگر بتای استفاده‌شده در این بررسی دارد (جدول‌های ۳ و ۴). Lance & William (1967) اعتقاد دارند که امکان تعیین مقدار دقیق بتا مقدور نیست. براساس پژوهش‌های تجربی نتایج خوشه‌بندی بتای انعطاف‌پذیر با بتای ۰/۲۵- بهتر از دیگر خوشه‌بندی‌هاست.

منفی و مثبت، وزن‌دهی غیرواقعی بین قطعات نمونه یا گروه‌ها رخ می‌دهد (Belbin et al., 1992). این یافته منطبق بر یافته‌های این بررسی است. در پژوهش Pakgohar et al. (2021) دریافت کردند که خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- تناسب بیشتری با روش‌های مختلف محاسبه ماتریس فاصله دارد. نتایج این بررسی در هر دو ناحیه ریشی نشان داد که روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و ۰/۲۵- با توجه به معیارهای ارزیابی‌کننده، عملکرد بسیار خوبی دارد، زیرا با افزایش مقدار بتا وزن غیرواقعی بیشتری به فواصل بین گروه‌ها و نمونه‌ها داده می‌شود (Belbin et al., 1992).

نتیجه‌گیری

براساس نتایج این تحقیق می‌توان روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- و ۰/۲۵- را با استفاده از روش اندازه‌گیری فاصله بری-کورتیس به‌منزله نوعی روش خوشه‌بندی مناسب برای طبقه‌بندی پوشش گیاهی و خوشه‌بندی رویشگاه به‌کار گرفت. روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۲۵- برای داده‌های با توزیع کروی پیشنهاد می‌شود، اما برای داده‌های با توزیع بیضی، خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱- توصیه می‌شود که حد واسط روش خوشه‌بندی میانگین و روش خوشه‌بندی وارد است.

روش نزدیک‌ترین همسایه است (Lengyel et al., 2019). با توجه به نتایج این بررسی شاهدیم که با افزایش بتا ($\beta > 0/4$) عملکرد خوشه‌بندی بتای انعطاف‌پذیر کاهش می‌یابد. بتای صفر خوشه‌بندی‌ای همانند روش خوشه‌بندی اتصال میانگین ایجاد می‌کند که برای داده‌های با توزیع کروی و بیضی مناسب است، اما بیش از دیگر روش‌ها حالت زنجیره‌ای در بین خوشه‌ها ایجاد می‌کند. بررسی خوشه‌بندی بتای انعطاف‌پذیر با استفاده از داده‌های شبیه‌سازی‌شده و مجموعه داده‌های جمع‌آوری‌شده از جنگل‌های خزان‌کننده جمهوری چک نشان داد که خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۱-، ۰/۲۵- و ۰/۴- ثبات طبقه‌بندی زیادی دارد (Tichý et al., 2011). در مقایسه روش‌های مختلف خوشه‌بندی سلسله‌مراتبی در داده‌های جمع‌آوری‌شده از پوشش گیاهی شمال کوه‌های راکی و اکوسیستم‌های مرتعی از جنوب شرقی مونتانا، روش خوشه‌بندی بتای انعطاف‌پذیر با مقدار بتای ۰/۲۵- را بهتر از دیگر روش‌های خوشه‌بندی سلسله‌مراتبی معرفی شد (Aho et al., 2008). همچنین در پژوهش دیگر عنوان شد که روش جفت‌گروه بدون وزن با میانگین حسابی (UPGMA)^۱ با مقدار بتای کم و منفی عملکرد بهتری نسبت به روش استاندارد این خوشه‌بندی با بتای صفر دارد و تأکید شد که مقادیر بتای منفی و نزدیک صفر کارکرد بهتری دارد و هرچه مقدار بتا از صفر دور می‌شود در جهت

References

- Aho, K., Roberts, D.W., & Weaver, T. (2008). Using geometric and non-geometric internal evaluators to compare eight vegetation classification results. *Journal Vegetation Science*, 19, 549–562.
- Belbin, L., & McDonald, C. (1993). Comparing three classification strategies for use in ecology. *Journal Vegetation Science*, 4, 341–348.
- Belbin, L., Faith, D.P. & Milligan, G.W. (1992). A comparison of two approaches to beta-flexible clustering. *Multivariate Behavioral Research*, 27(3), 417–433.
- Bray, R.J., & Curtis, J.T. (1957). An ordination of the upl& forest communities of southern Wisconsin. *Ecological Monographs*, 27, 325–349.

1. unweighted pair group method with arithmetic mean

- Bertrand, P., & Bel Mufti, G. (2006). Loevinger's measures of rule quality for assessing cluster stability. *Computational statistics & data analysis*, 50, 992-1015.
- Bihamta, N., Soffianian, A.R., Fakheran, F., & Pourmanafi, S. (2020). Incorporating CART algorithm for mapping Mangrove using Landsat 8 imagery. *Journal of Forest Research and Development*, 5(4), 557-569.
- De Cáceres, M., Chytrý, M., Agrillo, E., & et al (2015). A comparative framework for broad-scale plot-based vegetation classification. *Applied Vegetation Science*, 18(4), 543-60.
- Dufrene, M., & Legendre, P. (1997). Species assemblages & indicator species: the need for a flexible asymmetric approach. *Ecological Monographs*, 67,345-367.
- Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. 5th edition. Chichester: Wiley & Sons, 352p.
- Eshaghi Rad, J., Heidari, M., Mahdavi, A., & Zeinivandzadeh, M. (2011). Impact of recreational activities on vegetation and soil in forest park (case study: Choghasabz Forest Park-Ilam). *Iranian journal of forest*, 3(1),71-80.
- Eshaghi Rad, J., Soleimani, F., & Khodakarami, Y. (2017). Comparison of flora at the edge and within Oak forests in southern slopes of Kermanshah forests. *Applied Biology*, 30(1), 19-36.
- Eshaghi Rad, J., Zahedi, GH., Mohajer, M.R., & Mataji, A. (2009). Relationship between vegetation and physical and chemical properties of soil in Fagetum communities (Case study: Kheiroudkenar forest). *Iranian Journal of Forest and Poplar Research*, 17(2),174-187.
- Esmailzadeh, O., & Nourmohammadi, K. (2016). Introduction of Total Indicator Value Model in Vegetation Classification. *Iranian Journal of Biology*, 30(2), 246-263.
- Esmailzadeh, O., & Asadi, H. (2014). Total Phi Fidelity Index (TPFI) as a new algorithm in plant communities' analysis. *Iranian journal of forest*, 6(2), 215-232.
- Fernández, A., & Gómez, S. (2020). Versatile linkage: a family of space-conserving strategies for agglomerative hierarchical clustering. *Journal of Classification*, 37(3), 584-597.
- Gordon, A. (1999). *Classification, 2nd edition*. New York: Chapman and Hall, 272p
- Hämäläinen, J., Jauhiainen, S., & Kärkkäinen, T. (2017). Comparison of internal clustering validation indices for prototype-based clustering. *Algorithms*, 10(3),105-117.
- Jain, A. K., & Dubes, R.C. (1988). *Algorithms for Clustering Data*. Prentice Hall, New Jersey, 320p.
- Lengyel, A., Landucci, F., Mucina, L., Tsakalos, J.L., & Botta-Dukát, Z. (2018). Joint optimization of cluster number & abundance transformation for obtaining effective vegetation classifications. *Journal of Vegetation Science*, 29(2), 336-347.
- Lengyel, A., Roberts, D.W., & Botta-Dukat, Z. (2020). Comparison of silhouette-based reallocation methods for vegetation classification. *Journal Vegetation Science*, 32, 1-10.
- Lengyel, A., & Podani, J. (2015). Assessing the relative importance of methodological decisions in classifications of vegetation data. *Journal of Vegetation Science*, 26(4), 804-815.
- Lechner, A., McCaffrey, N., McCaffrey, P., Venables, N., & Hunter, J. (2016). Ecoregionalization classification of wetlands based on a cluster analysis of environmental data. *Applied Vegetation Science*, 16(4), 724-735.
- Lyons, M.B., Keith, D.A., Warton, D.I., Somerville, M., & Kingsford, R.T. (2016). Model-based assessment of ecological community classifications. *Journal of Vegetation Science*, 27(4), 704-715.
- Lötter, M.C., Mucina, L., & Witkowski, E.T.F. (2013). The classification conundrum: species fidelity as leading criterion in search of a rigorous method to classify a complex forest data set. *Community Ecology*, 14, 121-132.

- Legendre, P., & Legendre, L. (1998). Numerical ecology, 2nd Edition. Amsterdam: *Elsevier*, 939p.
- Lance, G.N., & Williams, W.T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The Computer Journal*, 9, 373–380.
- Legendre, P., Gallagher, E.D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2), 271-280.
- Khanalizadeh, A., EshaghiRad, J., Amiri, G.Z., Zare, H., Rammer, W., & Lexer, M.J. (2020). Assessing selected microhabitat types on living trees in Oriental beech (*Fagus orientalis* L.) dominated forests in Iran. *Annals of Forest Science*, 77(3), 1-13.
- Khabazi, F., Esmailzadeh, O., & Najafi, A. (2019). Supervised classification of *Buxus hyrcana* plant communities using artificial neural network. *Iranian Journal of Forest*, 11(3), 387-402
- Mielke, P.W., & Berry, K.J. (2007). Permutation methods: a distance function approach. 2nd edition. New York: *Springer*, 375p
- Matlack, G.R. (1994). Vegetation dynamics at the forest edges, trends in space & successional patterns in the Mediterranean natural of central Chile. *Journal of Arid Environments*, 62, 93-108.
- Milligan, G.W. (1989). A study of the beta-flexible clustering method. *Multivariate behavioral research*, 24(2), 163-176.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., & Studer, M. (2013). Package "Cluster": Cluster Analysis Basics and Extensions. R package version 1.14.4.
- Olkin, I., & Sampson, A.R. (2001). Multivariate analysis: overview. International Encyclopedia of the Social & Behavioral Sciences. New York: *Elsevier*, 239p.
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara R.B., Simpson, G.L., Solymos, P., Stevens, M.H., Wagner, H., & Oksanen, M.J. (2013). Package 'vegan': Community ecology package, 12(9), 1-295.
- Peet, R.K., & Roberts, D.W. (2013). Classification of natural & seminatural vegetation. In: van der Maarel E, Franklin J(eds) *Vegetation ecology*, 2nd edition. Oxford: Oxford University Press, 70p.
- Pakgohar, N., Eshaghi Rad, J., Gholami, G., Alijanpour, A., & Roberts, D.W. (2021). A comparative study of hard clustering algorithms for vegetation data. *Journal of Vegetation Science*, 32(3), 1-11.
- R Core Team (2019). R: a language & environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria
- Rodriguez, M.Z, Comin, C.H., Casanova, D., Bruno, O.M., Amancio, D.R., Costa, L.D.F., & Rodrigues, F.A. (2019). Clustering algorithms: A comparative approach. *PloS one*, 14(1), e0210236.
- Roberts, D.W. (2016). Package 'labdsv: Ordination and Multivariate Analysis for Ecology, R Package Version 1.8-0.
- Roberts, D.W. (2015). Vegetation classification by two new iterative reallocation optimization algorithms. *Plant Ecology*, 216(5), 741-758.
- Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation & validation of cluster analysis. *Journal of computational & applied mathematics*, 20, 53–65.
- Roleček, J., Tichý, L., Zelený, D., & Chytrý, M. (2009). Modified TWINSpan classification in which the hierarchy respects cluster heterogeneity. *Journal Vegetation Science*, 20(4), 596-602.
- Saberi, B.G., Esmailzadeh, O., & Asadi, H. (2021). Evaluating the different indicator species analysis in the classification of plant communities. *Iranian Journal of Forest*, 12(4), 541-555.
- Schmidtlein, S., Tichy, L. Feilhauer, H., & Faude, U. (2010). A brute-force approach to vegetation classification. *Journal Vegetation Science*, 21, 1162–1171.

- Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical & nonhierarchical methods. *Multivariate Behavioral Research*, 20(3), 283-304.
- Tichy, L., & Chytry, M. (2006). Statistical determination of diagnostic species for site groups of unequal size. *Journal of Vegetation Science*, 17(6), 809-818.
- Tichý, L., Chytrý, M., & Smarda, P. (2011). Evaluating the stability of the classification of community data. *Ecography*, 34(5), 807-813.
- Wang, P., Liu, Q., Xu, G., & Wang, K. (2019). A Three-Way Clustering Method Based on Ensemble Strategy & Three-Way Decision. *Information*, 10(2), 59-69.
- Xianping, Z., Mengben, W., Bo. Sh., & Yang, X. (2006). Quantitative classification & ordination of forest communities in Pangquangou National Nature Reserve. *Acta Ecologica Sinica*, 26(3), 754-761.
- Zaki, M.J., Meira, W.J.r. (2014). *Data Mining & Analysis: Fundamental Concepts & Algorithms*. New York: Cambridge University Press. 497p.



Research Article

Evaluating the effect of beta coefficient on the performance of flexible beta clustering in vegetation classification

N. Pakgohar¹, J. Eshaghi Rad^{2*}, GH. Gholami³, A. Alijanpour², and D.W. Roberts⁴

¹ Ph.D. of forestry, Dept. of Forestry, Faculty of Natural Resources, University of Urmia, Urmia, Iran

² Prof., Dept. of Forestry, Faculty of Natural Resources, University of Urmia, Urmia, Iran

³ Assistant Prof., Dept. of mathematics, Faculty of Science, University of Urmia, Urmia, Iran

⁴ Prof., Dept. of Ecology, Montana State University, Bozeman MT, US.

(Received: 10 September 2020; Accepted: 12 December 2021)

Abstract

Among different methods for classification, clustering is commonly used methods. Flexible-Beta clustering is successful hierarchical agglomerative clustering which is employed by ecologists as effective clustering method. The aim of the research was to detect the suitable value of beta for flexible-clustering methods. For this purpose, two different forest regions from Hyrcanian and Zagros Oak regions were selected. The clustering algorithms included Flexible-beta algorithms with five value of beta (-0.1, -0.25-, -0.4, -0.6 and -0.8). Five evaluators (Silhouette, MRPP, PARATNA, Phi coefficient) were employed on each cluster solution to evaluate different clustering algorithms. Algorithms were ranked from best to worst on each clustering evaluator for each data set. The results showed that Flexible-beta clustering with beta value -0.1 had best performance and Flexible-beta clustering with beta value -0.25 and -0.4 were proper performance in Hyrcanian regions. Flexible-beta clustering with beta value -0.25 was superior to others and Flexible-beta clustering with beta value -0.1 had the second rank. Since, choosing the most suitable clustering method is critical for achieving maximally ecological interpretable results, therefore, we suggested using flexible beta clustering with beta value equal to -0.1 and -0.25 in the studies area.

Keyword: Hyrcanian forest, Hierarchical clustering, Zagros forest.