فصلنامه مهندسی حمل و نقل، دوره ۱۲، شماره ۳، بهار ۱۴۰۰، ص. ۶۶۹-۶۹۶
مقاله پژوهشی

ارزیابی تحلیلی روش‌های محاسبه صحت اطلاعات مکانی داوطلبانه برای شبکه معابر درون شهری

مسعود یزدان نیک، دانشجو کارشناسی ارشد سیستم‌های اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران

رحیم علی‌عباسپور (مسئول مکاتبات)، استادیار، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران

Email: abaspour@ut.ac.ir

علیرضا چهرقان، استادیار دانشکده مهندسی معدن، دانشگاه صنعتی سهند تبریز

چکیده

انجام مطالعات و برنامه‌ریزی‌های شهری یکی از ضروریاتی است که با گسترش شهرها و جمعیت مناطق شهری لازم است به آن پرداخته شود. در میان داده‌های موردنیاز جهت انجام چنین مطالعاتی، شبکه عوارض خطی شهری نظیر شبکه معابر دارای اهمیت ویژه‌ای است. ویژگی‌های برجسته داده‌های مکانی داوطلبانه[1] نظیر پوشش جامع و بروز بودن آن باعث می‌گردد مجموعه داده داوطلبانه به‌عنوان مجموعه داده مناسب برای انجام مطالعات شناسایی گردد ولی با توجه به کیفیت ناهمگون این مجموعه داده لازم است ابتدا پارامترهای کیفیت آن ارزیابی و کیفیت مجموعه داده تائید گردد. ازاین‌رو در این مقاله با توجه به لزوم ارزیابی پارامترهای کیفیت برای داده‌های مکانی داوطلبانه و اهمیت بسیار بالای معیار صحت داده‌های مکانی و عدم وجود مطالعه جامع با هدف مقایسه معیارهای ارزیابی در مطالعات پیشین، لازم به نظر رسید که مطالعه جامعی بر روش‌های ارزیابی صحت مکانی انجام شود تا امکان بررسی نقاط قوت و چالش‌های موجود هر روش مشخص گردد. تحلیل‌های انجام‌شده بر معیارهای ارزیابی صحت‌مکانی عارضه‌مبنا، نامناسب بودن معیار فاصله میانگین در عدم تمایز میان صحت‌مکانی عوارض در حالات خاص و مزیت فاصله هاسدورف نسبت به فاصله حریم افزایشی با سرعت پردازش بالاتر را نتیجه داد. براساس این معیار صحت‌مکانی مجموعه داده داوطلبانه شهر تهران در مناطق ۱۱ تا ۱۴ تهران به میزان ۵/۲ متر ارزیابی شد که حاکی از صحت‌مکانی مطلوب داده‌های داوطلبانه این مناطق است. همچنین در دسته دوم روش‌های ارزیابی به سبب وجود همبستگی زیاد (۰/۸۷) پارامتر کامل‌بودن و صحت مکانی، توصیه این مطالعه ارزیابی کامل‌بودن دو مجموعه داده داوطلبانه و رسمی نسبت به هم، به منظور جلوگیری از بروز خطاهای احتمالی است. در این گروه از روش‌ها، معیار فاصله حریم به واسطه تاثیر ناپذیری از حالات خاص به عنوان روش مناسب انتخاب و صحت مکانی مجموعه داده داوطلبانه شهر تهران با شعاع حریم ۵ متر، ۰/۷۸۱ ارزیابی شد

واژه‌های کلیدی: اطلاعات مکانی داوطلبانه، روش‌های سلول‌مبنا، روش‌های عارضه‌مبنا، صحت مکانی، شبکه معابر.

۱. مقدمه

یکی از اصـــلی‌ترین مؤلفه‌های زیرســـاخت اطلاعات مکانی[۲]، اطلاعات مرتبط با شبکه راه‌های شهری است که کاربرد وسیعی در مســـائـل مختلف مربوط بـه حوزه مـدیریت شـــهری دارد [چهرقـان و علی‌عباس پور, ۲۰۱۸]. مطالعاتی نظیر مســـیریابی بهینـه شـــبکه حمل‌ونقل عمومی در یک شـــهر [افندی زاده و همکاران, ۲۰۱۹; تیموریان و همکاران, ۲۰۱۹]، شـناسایی نقاط حادثه خیر در معابر [Rassafi et all., 2012] و مســـیریابی بهینـه شـــهری [ذوالفقاری و کرکه آبادی, ۲۰۱۹; شـــاکری و همکاران, ۲۰۱۹; خمر و همکاران, ۲۰۱۷] حوزه‌های مرتبط با مدیریت شـــهری هســـتند که با توجه به گســـتردگی نیازهای آن ایجـاب می‌کنـد، مجموعه داده انتخابی جهت انجام مطالعات تا حـد قـابـل‌قبول دقیق، کامل و بروز باشـــد. تمامی ویژگی‌های مطرح شـــده برای مجموعـه داده‌هـای مکـانی عمـدتـاً در یک مجموعـه داده موجود نبوده و هر مجموعه داده دارای نواقص و محدودیت‌هایی اســـت. به‌عنوان‌مثال مجموعه داده رســـمی که عمـدتـاً در کاربردهای متنوع از آن اســـتفاده می‌گردد علی‌رغم کیفیت مشـخص و همگن در سراسر مجموعه داده ممکن است برای نوع خـاصـــی از عوارض و کـاربری‌ها و یا اهمیت پایین منطقه فاقد اطلاعات باشد. همچنین به‌واسطه هزینه بالا عملیات به‌روزرســانی عمدتاً به‌روزرســانی مجموعه داده رســمی در حد روزانه و یا ماهانه مقدور نیست [Haklay, 2010].

رسـیدن به اهدافی نظیر به‌روزرسانی سریع و پوشش جامع منطقـه اهدافی هســـتند که لازمه رســـیدن به آن تغییر رویکرد فرآیند جمع‌آوری داده اســـت [Lauriault and Mooney, 2014] رویکرد سـنتی جمع‌آوری داده مکانی نیازمند تجهیزات پرهزینه و نیروهای متخصـص بوده که تولید اطلاعات مکانی را محـدود بـه مراکز خـاص می‌کنـد [چهرقـان و علی‌عباس پور, ۲۰۱۸]. ازاین‌رو درصـــورتی‌که افراد اســـتفاده‌کننده از داده‌های مکانی خود در فرآیند ایجاد آن مشـارکت داشـته باشند به دلیل جامعه بســـیار زیاد افراد دخیل رســـیدن به اهداف ذکرشـــده دسـترس‌پذیر تر گردد [Bruns, 2008]. معرفی نسل جدیدی از وب تحـت عنوان وب ۲ و همه‌گیر شـــدن فنّاوری‌های نظیر ســـیســـتم موقعیت‌یابی جهانی[۳] میان عموم افراد منجر به تولید حجم انبوهی از اطلاعات مکانی در سـال‌های اخیر شـده است [Arsanjani et al., 2015]. ســـرویـس‌هـایـی نـظـیـر Wikimapia، Google Map Maker و OpenStreetMap ازجملـه ســـرویس‌های اطلاعات مکانی مشارکتی هستند که در این میان OSM به‌واسطه تعداد کاربران زیاد و گستردگی داده‌ها به‌عنوان مشهورترین سرویس ثبت و به اشـتراک‌گذاری داده‌های مکانی داوطلبانه شـناسـایی می‌شـود [چهرقان و علی عباس پور, ۲۰۱۸].

باوجود این‌که عملیات سـاخت نقشـه جهانی توسـط این پروژه‌ها به‌ســرعت در حال انجام اســت و داده‌هـــای مناطق مختلف به‌آسـانی قابل دریافت اسـت، ولی به‌موجب استفاده از فعالیت‌هـــای جمع سپاری در مقیاس وسیع به‌خصوص مقیاس جهانی، ســـاخت منابع قابل‌اطمینان داده همواره با مشـــکلات خاص خود همراه اسـت. دلیل این‌گونه از مشکلات را این‌طور می‌توان توجیه نمود که: «فعالیت‌های جمع ســـپاری که توســـط گروه عظیمی از مشـارکت‌کنندگان به‌صـورت مسـتقل از هم و بدون نظارت سـرپرست انجام می‌شود [Meier, 2015] و هر یک از مشـــارکت‌کنندگان طبق علایق فردی خود عملیات تولید داده را انجـام می‌دهند [Bégin et al., 2013]، باعث ایجاد خطاهای متنوع در مجموعه داده نهایی می‌شـوند». این موضـوع هـمـچنین بـا توجـه بـه این واقعیـت کـه طیف عظیمی از مشـــارکـت‌کننـدگان این مجموعه داده‌ها نه‌تنها در این زمینه از اطلاعات متخصص نبوده بلکه کاملاً مبتدی و ناآشنا به مسائل و اســتانداردهـــای علم جمع‌آوری این نوع از داده هســتند نیز تشـدید می‌یابد [Haklay, 2010]. به‌صـورت کلی منابع خطا

در این داده‌ها با پدیده‌های اجتماعی مرتبط بوده و این مسئله در داده‌های داوطلبانه با آنچه در داده‌های رسمی دیده می‌شود کاملاً متفاوت است [Mashhadi et al.,2015].

VGI به سبب برخورداری از میلیون‌ها حس‌گر آشنا به محیط اطراف خود می‌تواند تغییرات محیطی را بسیار سریع در مجموعه داده خود به‌روزرسانی نماید [Goodchild, 2007]. درست است که ویژگی فوق تأثیر مشهودی بر نتایج حاصل از آنالیزهای مختلف دارد ولی موضوع مهم‌تر در مورد این داده‌ها مسئله کیفیت است و نیازمند این است داده‌های مکانی داوطلبانه از منظر دقت مورد ارزیابی قرار گیرد.

بنابراین با توجه به اهمیت و کاربردهای وسیع داده‌های مکانی داوطلبانه ضروری است قبل از به‌کارگیری این داده‌ها، جنبه‌های مختلف کیفیت این داده‌ها (ارائه‌شده توسط سازمان استانداردسازی ISO[4]) تعیین گردد. در استاندارد ISO مؤلفه‌های کامل‌بودن[5]، سازگاری منطقی[6]، صحت مکانی[7]، صحت زمانی[8] و صحت توصیفی[9] به‌عنوان مؤلفه‌های تعیین‌کننده کیفیت داده‌های مکانی معرفی‌شده‌اند [Neis and Zielstra, 2014].

برای ارزیابی پارامترهای مختلف کیفیت داده‌های مکانی داوطلبانه، روش‌های متعددی در مقالات مختلف معرفی و پیاده‌سازی شده است. به‌صورت کلی روش‌های مورد استفاده را به دودسته کلی می‌توان تقسیم‌بندی نمود. در دسته اول روش‌هایی طبقه‌بندی شده‌اند که کیفیت داده‌های داوطلبانه را با مقایسه با یک مجموعه رسمی متناظر محاسبه می‌نمایند [Antoniou and Skopeliti, 2015]. درحالی‌که در دسته دیگر روش‌هایی قرار می‌گیرد که کیفیت داده‌های داوطلبانه را بدون نیاز به داده رسمی و با استفاده از ماهیت داده‌ها و ویژگی‌های مختلف درون آن محاسبه می‌نمایند [H. Zhang and Malczewski, 2018].

از آن‌جاکه تنوع روش‌های به‌کار برده شده در مقالات ارزیابی پارامتر صحت‌مکانی بسیار متنوع است، لازم دیده شد تا مقاله پیش‌رو با هدف مقایسه روش‌های ارزیابی صحت‌مکانی مبتنی بر مقایسه با مجموعه داده معتبر نگارش شود. انتخاب صحت‌مکانی از میان پارامترهای مختلف کیفیت با توجه به اهمیت این معیار و تاثیر آن بر توانایی پاسخگویی مجموعه داده به کاربردهای مختلف بوده است [Chrisman, 1991]. طبیعی است که در هریک از این روش‌ها پیش‌فرض‌هایی لحاظ شده است که بر نتایج ارزیابی موثر هستند. شناخت این موارد و تاثیر آن می‌تواند هنگام ارزیابی کیفیت در تحقیق‌های آینده مفید واقع شود. علاوه بر ارزیابی نتایج حاصل شده محدودیت‌های مرتبط به داده در زمان ارزیابی و حجم محاسبات هر روش از دیگر جنبه‌های درنظر گرفته شده در این مطالعه است. بنابراین در این مقاله سعی شده است علاوه بر مقایسه روش‌های منطقه‌مبنا و عارضه‌مبنا کیفیت به مقایسه روش‌های متعلق به هر رویکرد نیز پرداخته شود.

در نهایت با مقایسه روش‌های ارزیابی موجود و انتخاب معیار مناسب، صحت‌مکانی مجموعه داده داوطلبانه مناطق ۱۱ تا ۱۴ تهران مورد ارزیابی قرار گرفت. باتوجه به این موضوع که تمامی روش‌های بررسی شده در این مطالعه مبتنی‌بر مقایسه با مجموعه داده رسمی هستند، از داده‌های رسمی تهیه شده توسط سازمان شهرداری تهران در مقیاس ۱:۲۰۰۰ استفاده شده است.

این مقاله در چهار بخش ارائه‌شده است. پس از بیان مقدمه در بخش اول، در بخش دوم به توضیح روش‌های ارزیابی صحت مکانی ارائه‌شده در پژوهش‌های مختلف به همراه روابط آن‌ها پرداخته‌شده است. سپس در بخش سوم به محاسبه معیارهای ارائه‌شده و مقایسه‌ی نتایج به‌دست‌آمده و نقاط ضعف و قوت هر یک از این روش‌ها پرداخته‌شده است. در قسمت پایانی با مقایسه جامع بر روی تمام نتایج حاصل از قسمت قبل به جمع‌بندی و نتایج حاصل از تحقیق پرداخته می‌شود.

۲. پیشینه تحقیق

در بسیاری از تحقیقات انجام‌شده درزمینه بررسی کیفیت داده‌های داوطلبانه که پارامترهای مختلف کیفیت را با مقایسه با مجموعه داده رسمی محاسبه می‌نمایند، شهرهای اروپایی به‌عنوان منطقه موردمطالعه استفاده‌شده است. دلیل این امر را می‌توان این‌گونه توجیه نمود که به‌واسطه معرفی OSM در اروپا و استقبال زیاد از این پروژه، استفاده از شهرهای اروپایی به دلیل میزان مشارکت بالاتر و همچنین مجموعه داده‌های کامل‌تر امری منطقی است.

هاکلای (۲۰۱۰) را می‌توان از اولین پژوهش‌های انجام‌شده درزمینه ارزیابی کیفیت داده‌های مکانی داوطلبانه به شمار آورد. در این مقاله به‌واسطه میزان مشارکت بالای مردم شهر لندن نسبت به سایر شهرهای اروپایی، از این شهر به‌عنوان منطقه موردمطالعه استفاده شده است. در این مقاله به‌منظور ارزیابی صحت مکانی از معیار میزان قرارگیری طول عارضه در ناحیه حریم عارضه رسمی متناظر استفاده‌شده است [Haklay, 2010]. این معیار به‌نوعی حالت ساده شده روش فاصله حریم افزایشی است که در مقاله گودچایلد و هانتر (۱۹۹۷) معرفی شده است [Goodchild and Hunter, 1997].

کونادی (۲۰۰۹) در پژوهش خود پارامترهای کیفیت نظیر صحت مکانی و کامل‌بودن را به‌صورت منطقه‌مبنا و در شهر آتن پیاده‌سازی می‌کند. در این پژوهش صحت مکانی عوارض خطی داوطلبانه در اکثر مناطق به میزان بسیار بالایی (بیش از ۹۰ درصد) محاسبه‌شده است [Kounadi, 2009].

گیرس و تویا (۲۰۱۰) را می‌توان ازجمله مقالات شاخص درزمینه بررسی پارامترهای کیفیت داده‌های مکانی داوطلبانه به شمار آورد. در این مقاله که به‌نوعی توسعه پژوهش هاکلای (۲۰۱۰) است، پارامترهای کیفیت داده‌های مکانی برای یکی از شهرهای فرانسه بررسی‌شده است. به‌صورت نمونه صحت مکانی عوارض مختلف نظیر عوارض نقطه‌ای، خطی و چندضلعی در این پژوهش با استفاده از روش‌های مختلف تعیین‌شده است [Girres and Touya, 2010].

یکی از نقاط ضعف پژوهش انجام‌شده توسط هاکلای در سال ۲۰۱۰ استفاده از روش‌های دستی در تعیین عوارض نظیر است [Haklay et al., 2010]، به‌طوری‌که به‌کارگیری این روش در تکرار ارزیابی برای مناطق بزرگ‌تر و یا به‌روزرسانی داده‌های داوطلبانه، مشکل است. بر همین اساس کوکولتسوس در سال ۲۰۱۱ روش تناظریابی چند مرحله‌ای برای تناظریابی دو مجموعه داده داوطلبانه و رسمی را ارائه کرد. الگوریتم معرفی‌شده در این مقاله بر روی مجموعه داده شهر لیورپول مورد بررسی قرار گرفت و پس از تعیین عوارض متناظر، پارامترهای ارزیابی کیفیت داده‌های مکانی نظیر کامل بودن و صحت مکانی بررسی شد [Koukoletsos et al., 2011].

در مقاله انجام‌شده توسط جاکسون و همکاران (۲۰۱۳) با توجه به تحقیقات کم انجام شده درزمینه بررسی کیفیت داده‌های داوطلبانه و تمرکز این مقالات بر روی داده‌های خطی، این بار در این پژوهش به ارزیابی پارامترهای کیفیت برای داده‌های نقطه‌ای پرداخته‌شده است. در این مقاله برای ارزیابی پارامتر صحت مکانی یک زیرمجموعه از نقاط زیرساختی یک منطقه شهری آمریکا انتخاب‌شده است. در این پژوهش پس از انجام تناظریابی برای شناسایی نقاط نظیر از فاصله اقلیدسی به‌منظور تعیین پارامتر صحت مکانی استفاده‌شده است [Jackson et al., 2013].

به‌صورت مشابه فان و همکاران (۲۰۱٤) پارامترهای کیفیت داده‌های مکانی داوطلبانه نظیر کامل‌بودن و صحت مکانی را مورد ارزیابی قرار داده‌اند. به‌منظور تعیین معیار صحت مکانی عوارض چندضلعی در این مقاله پس از تناظریابی دو مجموعه داده از میانگین فاصله اقلیدسی نقاط نظیر استفاده‌شده است. همچنین از معیار تشابه شکل جهت تعیین صحت

چندضلعی‌های مجموعه داوطلبانه نسبت به مجموعه رسمی استفاده‌شده است [Fan et al., 2014].

وانگ و همکاران (۲۰۱۳) صحت مکانی داده‌های داوطلبانه شهر ووهان در کشور چین را مورد بررسی قرار دادند. در این پژوهش صحت مکانی برای کل شهر به‌صورت یکجا و با استفاده از روش ارائه‌شده در مقاله هاکلای (۲۰۱۰) مورد محاسبه قرار گرفت [Wang et al., 2013]. به دلیل پرداخته نشدن به پارامترهای کیفیت به‌صورت جزئی این بار ژانگ و همکاران (۲۰۱٤) به بررسی صحت مکانی عوارض از نوع خطی و نقطه‌ای برای کل کشور چین پرداختند. به‌واسطه وسعت بسیار بالای مناطق مورد مطالعه در این پژوهش ابتدا منطقه به سلول‌های مربعی با طول ۱۰۰ کیلومتر شکسته و مطابق با روش استفاده‌شده در مقاله وانگ و همکاران (۲۰۱۳) صحت مکانی عوارض برای هر سلول تعیین شد. نتایج نشان می‌دهد که پارامتر صحت مکانی برای شهرهای مختلف چین از لحاظ وسعت و جمعیت رفتار متفاوتی از خود نشان می‌دهد. برای مثال در شهرهای پرجمعیت و دارای مشارکت بالاتر نظیر پکن و شانگهای میزان صحت مکانی عوارض بالاتر از شهر ارومچی با میزان مشارکت کمتر است [Zheng and Zheng, 2014].

ژانگ و همکاران در سال ۲۰۱۷ صحت مکانی عوارض داوطلبانه خطی را به تفکیک نوع عوارض در کشور کانادا مورد ارزیابی قرار دادند. در این مقاله نتایج نشان می‌دهد که حدود ۷۷/۵ درصد از عوارض داوطلبانه در حریم ۵ متری عوارض رسمی قرار می‌گیرند که این میزان به‌اندازه ٤/۵ درصد از میزان محاسبه شده برای مجموعه داده داوطلبانه آلمان توسط لودویگ و همکاران [Ludwig et al., 2011] بیشتر است [Zhang and Malczewski, 2018].

از آن‌جاکه روش فاصله حریم یک روش پرکاربرد در ارزیابی صحت‌مکانی عوارض در مقالات ارزیابی کیفیت است. ژانگ و همکاران در سال ۲۰۱۹ روش مبتنی بر انتشار خطا نقاط تشکیل دهنده خطوط برای ایجاد روش واحد در انتخاب ابعاد حریم طراحی کردند. رویکرد ارائه شده در این مطالعه با توجه به مشکل اصلی روش ارزیابی ناحیه حریم، عدم وجود مبنا و رابطه مشخص ریاضی در تعیین شعاع حریم طراحی شده است. نتایج ارزیابی صحت مکانی هنگ‌کنگ نشان می‌دهد عوارض OSM مرتبط به این مناطق دارای صحت‌مکانی ٦۷/۰۹ درصد است [Zhang et al., 2019].

کاسترو و همکاران در سال ۲۰۱۹ باتوجه به این واقعیت که مجموعه داده OSM، شرایط یه مجموعه کلان داده را داراست، با رویکرد مبتنی بر کلان داده به ارزیابی صحت مکانی عوارض در سه منطقه شهری در کشور اکوادور پرداخت. خلاصه نتایج این مقاله را می‌توان عدم وجود رابطه میان تعداد نسخه و میانگین خطا هندسی به شمار آورد. به عبارتی کاسترو پس از ارزیابی‌های خود معتقد است: استدلال تعداد نسخه بالاتر منجر به صحت مکانی بالاتر خواهد شد همیشه فرضیه صحیحی نمی‌تواند تلقی شود [Castro et al., 2019].

ژو در سال ۲۰۱۷، به تحلیل روش مرسوم ارزیابی صحت مکانی، آنالیز بافر پرداخت. بررسی‌ها در این مطالعه هاکی از آن است که محدودیت اصلی این روش، مرتبط با حالتی است که مجموعه داده معتبر و OSM مورد بررسی مطابقت یک به یک نداشته باشند. در چنین حالتی ارزیابی صحت‌مکانی منجر به ارزیابی نادرستی خواهد شد [Zhou 2017].

بر روی مجموعه داده‌های داوطلبانه شهر تهران نیز تاکنون مطالعاتی صورت گرفته است. ازجمله مقاله فرقانی و دلاور (۲۰۱٤) که به معرفی شاخص جدیدی برای ارزیابی کیفیت داده‌های مکانی داوطلبانه پرداخته است. در این مقاله پس از تقسیم دو مجموعه داده داوطلبانه و رسمی به نواحی کوچک‌تر، کیفیت این داده‌ها از طریق معیارهای اختلاف مساحت پوش محدب[10] ، فاصله میان مراکز هندسی و شاخص توزیع جهت[11] دو مجموعه داده تعیین‌شده است. درنهایت میزان سازگاری

شاخص‌ها در داده‌های داوطلبانه نسبت به داده رسمی متناظر به کمک منطق فازی محاسبه‌شده است. نتایج به‌دست‌آمده در این مقاله نشـــان می‌دهد که در داده‌های داوطلبانه این مناطق میزان کیفیت بسیار بالا یا بسیار کم نسبت به مجموعه رسمی مشاهده نمی‌شـود و اغلب سـلول‌ها (حدود ۸۰ درصـد) دارای کیفیت متوسط هستند [Forghani and Delavar, 2014].

در پژوهش انجام‌شده توسط محمدی و ملک (۲۰۱۵) برای ارزیـابی صـحت مکانی عوارض داوطلبانه مکانی از داده‌های داوطلبانه شهر تهران و داده رسمی متناظر آن استفاده‌شده است. در این مقـالـه پس از معرفی برخی از معیـارهـای تأثیرگذار در کیفیت و طبقه‌بندی این معیارها در چهار دسـته مکانی، زمانی، فردی و ویژگی‌های ذاتی عارضـه، ارتباط این کمیت‌ها با مؤلفه نهایی صحت مکانی، به‌واسطه ارتباط غیرخطی و پیچیده موجود از طریق روش‌های شـبکه عصـبی تعیین‌شده است. درنهایت با آموزش شبکه عصبی ایجادشده این امکان وجود خواهد داشت کـه کیفیـت داده‌های داوطلبانه برای مناطق بدون داده رسـمی متناظر محاسبه گردد [Mohammadi and Malek, 2015].

نصـیری و همکاران نیز در سـال ۲۰۱۸ رویکردی جدید بر مبنا دیاگرام ورونی برای بهبود کیفیت داده‌های مکانی داوطلبانه با اسـتفاده از سـابقه داده‌ها ارائه کردند. نتایج به‌کارگیری این رویکرد بر نسـخه تاریخچه عوارض بهبود صـحت مکانی و کامل‌بودن داده‌های داوطلبانه به میزان ۱۲/۸ و ۱۰/۹ درصـد را نشان داد [Nasiri et al., 2018].

۳. صحت مکانی عوارض خطی

روش‌های اسـتفاده‌شده در مقالات مختلف به‌منظور محاسبه صـحت مکانی داده‌هـای داوطلبانه را می‌توان در دودسـته تقسـیم‌بنـدی نمود. در دسـتـه اول روش‌هـایی تحت عنوان منطقه‌مبنا[12] قرار می‌گیرد که صـحت مکانی را پس از تقسـیم منطقه موردمطالعه به نواحی کوچک‌تر و به‌صورت کلی محاسبه می‌نمایند. در گروه دوم روش‌هایی جای گرفت اسـت که واحد محاسـباتی آن‌ها به‌صورت عارضه‌مبنا[13] بوده و صحت مکانی را برای هر یـک از عوارض محاسـبه می‌نمایند [Zhang and Malczewski, 2017].

۳-۱ روش‌های منطقه‌مبنا

یکی از روش‌های سـاده در محاسـبه کیفیت داده‌های مکانی محاسـبه این پارامترها برای کل منطقه به‌صـورت یکجا اسـت. علی‌رغم پیاده‌سازی ساده این روش‌ها، از این روش‌ها فقط برای مناطق با وسعت کم (در حدود یک محله) می‌توان استفاده کرد. برای منـاطق شـهری بـا وسـعت زیاد به دلیل ناهمگونی‌های ذکرشده در کیفیت و پوشش این داده‌ها عملاً نتایج به‌دست‌آمده قـابـل اعتماد نخواهد بود. راهکار ارائه شـده برخی از مقالات برای مقابله با این مشکل و همچنین مشکلات خاص استفاده از روش‌های تناظریابی و پیش‌پردازش‌های آن، تقسـیم ناحیه مورد مطالعه به بخش‌های کوچک‌تر (برای مثال سـلول‌های مربعی با طول ۲۰۰ متر) بوده اسـت [Forghani and Delavar, 2014; Haklay et al., 2010; Koukoletsos et al., 2011; Kounadi, 2009; Wang et al., 2013].

یکی از مسـائـل کلیدی در این این‌گونه از روش‌ها تعیین مقدار مسـاحت و وسعت زیر مناطق ایجادشده است. در تعیین این مقدار باید به این نکته توجه داشت که همان‌طور که زیاد در نظر گرفتن این مقـدار بـاعث تولید نتایج غیرقابل اطمینان و در مواردی نادرسـت می‌شود، کوچک در نظر گرفتن این مقدار نیز بـاعـث افزایش بـار محاسـباتی الگوریتم می‌گردد. بنابراین در این‌گونه از مسـائل سـعی می‌گردد همیشه حالت تعادلی در نظر گرفته شود و مقدار این کمیت از طریق آزمون و خطا و با توجه به تراکم راه‌ها مشخص می‌شود. در اغلب مقالات این مقدار در بـازه عـددی ۲۰۰ تـا ۱۰۰۰ مـتـر در نـظر گرفتـه شـده

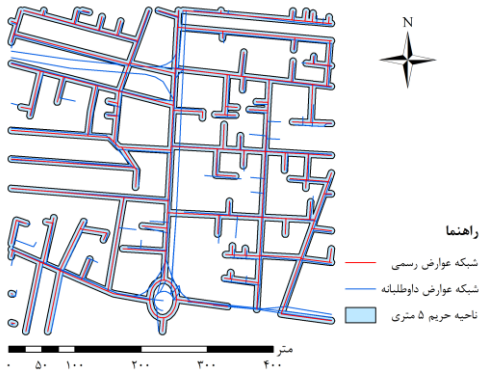
است[Forghani and Delavar, 2014; Koukoletsos et al., 2011; Kounadi, 2009].

درزمینه بررسی صحت مکانی عوارض داوطلبانه به‌صورت منطقه‌مبنا، فرقانی و دلاور (۲۰۱٤) از سه معیار اختلاف مساحت پوش محدب، فاصله میان مراکز هندسی و شاخص توزیع جهت دو مجموعه داده استفاده کرده‌اند [Forghani and Delavar, 2014]. در مقالات دیگری نظیر کوکولتسوس و همکاران (۲۰۱۲)، هچ و همکاران (۲۰۱۳) و برووللی و زامبونی (۲۰۱۸) از ترکیب روش‌های عارضه‌مبنا و منطقه‌مبنا استفاده شده است [(Brovelli and Zamboni, 2018; Hecht, et al., 2013; Koukoletsos et al., 2012]. هدف از این ترکیب کاهش وسعت منطقه مورد بررسی در روش‌های عارضه‌مبنا و قابلیت نمایش نتایج به‌صورت گرافیکی است.

۳-۱-۱ آنالیز بافر

نحوه محاسبه صحت مکانی برای هر سلول توسط این معیار به این صورت است که ابتدا برای عوارض رسمی موجود درون هر سلول ناحیه حریمی به عرض مشخص ایجاد می‌گردد و سپس طول عوارض مورد بررسی درون ناحیه حریم محاسبه می‌شود. نسبت طول محاسبه‌شده به طول کلی عوارض مورد بررسی در سلول انتخاب‌شده صحت مکانی سلول انتخابی را نمایش می‌دهد. انتخاب عرض حریم در این روش کاملاً تجربی بوده و بر اساس عرض خیابان‌ها و یا دقت مجموعه داده رسمی انتخاب می‌گردد و هیچ‌گونه مبنا تئوری و محاسباتی برای انتخاب عرض حریم وجود ندارد.

$$PA = \frac{\sum l_i}{\sum L_T} \qquad (۱)$$



شکل ۱. ایجاد ناحیه حریم حول عارضه رسمی جهت تعیین طول عارضه داوطلبانه قرارگرفته در این ناحیه

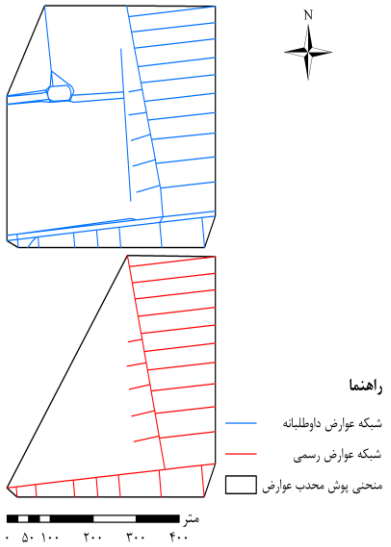
در رابطه (۱) PA صحت‌مکانی سلول انتخابی، $\sum l_i$ مجموع طول عوارض قرارگرفته در ناحیه حریم ایجادشده عوارض رسمی و $\sum L_T$ مجموع طول کل عوارض داوطلبانه در محدوده موردبررسی است. شکل ۱ حریم ۵ متری ایجادشده حول عوارض رسمی جهت تعیین صحت مکانی عوارض داوطلبانه را نمایش می‌دهد.

۳-۱-۲ پوش محدب

پوش محدب نقاط در فضای اقلیدسی، کوچک‌ترین مجموعه محدب دربرگیرنده این نقاط تعریف می‌شود [de Berg et al., 2000]. میزان اختلاف مساحت پوش محدب ایجادشده می‌تواند به‌عنوان پارامتر مرتبط با صحت مکانی عوارض در نظر گرفته شود. بدیهی است که هرچه میزان اختلاف بین دو مقدار محاسبه‌شده بیشتر باشد میزان صحت مکانی کمتری برای عوارض داوطلبانه نتیجه‌گیری می‌شود.

در رابطه (۲)، $Area_{OSM}$ مساحت پوش محدب عوارض داوطلبانه و $Area_{Refrence}$ مساحت پوش محدب عوارض رسمی است.پوش محدب عوارض برای هر یک از دو مجموعه داده را نشان می‌دهد.

$$PA = Area_{OSM} - Area_{Refrence} \qquad (۲)$$

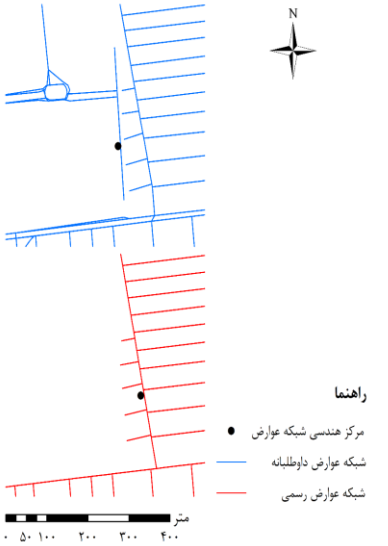


شکل ۲. ایجاد پوش محدب عوارض خطی رسمی و داوطلبانه

۳-۱-۳ مرکز هندسی شبکه راه‌ها

این معیار گرایش مرکزی یک مجموعه داده مکانی را مشخص می‌نماید. خروجی این معیار یک عارضه نقطه‌ای است که فاصله این نقطه تا سایر نقاط مجموعه داده کمترین میزان ممکن خواهد بود (شکل ۳). از مزایای این معیار می‌توان به مقاومت آن در برابر خطاهای فاحش اشاره کرد، به‌صورتی که با ورود نقطه نادرست به مجموعه داده، مرکز هندسی محاسبه‌شده در دو حالت تغییر مکانی جزئی خواهد داشت. این روش نخست در سال ۱۹۶۲ توسط کوهن و کونه به‌صورت یک الگوریتم تکراری با هدف کمینه‌سازی فاصله اقلیدسی نقطه مورد نظر تا سایر نقاط مجموعه داده معرفی شد [Kulin and Kuenne, 1962]. پس از تعیین مرکز هندسی هریک از دو مجموعه داده در هر سلول از فاصله میان این دو نقطه به‌عنوان معیار صحت مکانی عوارض می‌توان استفاده کرد. به این صورت که برای مجموعه داده مورد بررسی هرچه میزان فاصله محاسبه‌شده مقدار بزرگ‌تری باشد صحت مکانی کمتری انتظار می‌رود. رابطه (۳) نحوه محاسبه اختلاف مرکز هندسی در دو مجموعه داده رسمی و داوطلبانه را نشان می‌دهد. در این رابطه، ED فاصله اقلیدسی میان مراکز هندسی دو مجموعه داده (Mc_{OSM} و $Mc_{Refrence}$) است.

$$PA = ED(Mc_{OSM} - Mc_{Refrence}) \quad (۳)$$

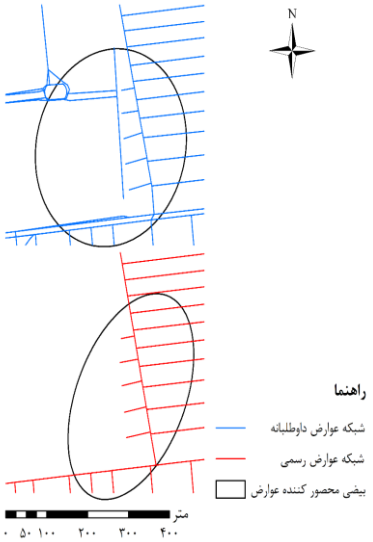


شکل ۳. ایجاد مراکز هندسی عوارض خطی رسمی و داوطلبانه

۳-۱-۴ شاخص توزیع جهت

این معیار ویژگی‌های مکانی عوارض برداری نظیر پراکندگی داده‌ها، گرایش مرکزی و جهت‌گیری عوارض را نشان می‌دهد. خروجی این معیار یک چندضلعی به شکل بیضی است که مرکز این بیضی بر روی مرکز هندسی عوارض واقع شده و ابعاد این بیضی با پارامتر انحراف معیار ورودی رابطه مستقیم دارد (شکل ٤). درصورتی‌که عوارض ورودی این معیار به‌صورت نرمال اطراف مرکز هندسی آن‌ها توزیع‌شده باشند، با انتخاب پارامتر انحراف معیار به میزان یک سیگما بیضی مورد نظر حدود ٦٣ درصد و با انتخاب این معیار برابر با سه سیگما بیضی ایجادشده حدود ۹۹/۹ درصد از عوارض را احاطه خواهد کرد پس از تشکیل این بیضی برای هر یک از دو مجموعه داده از ویژگی‌های مختلف این شکل نظیر جهت، ابعاد و مرکز آن به‌منظور تعیین صحت مکانی دو مجموعه داده می‌توان استفاده کرد.

$$PA = Or_{OSM\ Ellipse} - Or_{Refrence\ Ellipse} \quad (٤)$$



شکل ٤. ایجاد بیضی جهت عوارض خطی رسمی و داوطلبانه

رابطه (۳) نحوه محاسبه اختلاف جهت در دو مجموعه داده رسمی و داوطلبانه را نشان می‌دهد. که در این رابطه، $Or_{OSM\ Ellipse}$ توجیه بیضی مربوط به داده‌های داوطلبانه و $Or_{Refrence\ Ellipse}$ توجیه مربوط به عوارض رسمی است.

۳-۲ روش‌های عارضه‌مبنا

برخلاف روش‌های منطقه‌مبنا که پس از انجام پیش‌پردازش‌های اولیه نظیر یکسان‌سازی سیستم مختصات و حذف داده‌های اضافی استفاده از آن‌ها به سهولت انجام می‌گیرد، پیاده‌سازی و استفاده از دسته دوم روش‌ها (روش‌های عارضه‌مبنا) پیچیدگی‌ها و مشکلات خاص مربوط به خود را دارد. همان‌طور که در قسمت مقدمه اشاره شد، برای به‌کارگیری این روش‌ها نیاز به شناسایی عوارض متناظر در دو مجموعه داده است. از طرفی به این دلیل که دو مجموعه موردبررسی از دو منبع با استانداردهای متفاوت دریافت شده‌اند، هیچ‌گونه داده اضافی برای شناسایی عوارض متناظر در این دو مجموعه داده به‌صورت ضمنی وجود ندارد. ازاین‌رو لازم است عوارض متناظر به‌صورت دستی و یا اتوماتیک تعیین‌شده و به الگوریتم تعیین معیار صحت‌مکانی معرفی گردند. ولی به دلیل تعداد بسیار بالای عوارض موجود در دو مجموعه داده نیاز به الگوریتم تناظریابی سریع و کارا احساس می‌شود. در مقالات کوکولتسوس و همکاران (۲۰۱۱)، کوکولتسوس و همکاران (۲۰۱۲)، فان و همکاران (۲۰۱٦)، چهرقان و علی‌عباسپور (۲۰۱۷) و دهقانی و همکاران (۲۰۱۷) از روش‌های تناظریابی به منظور غنی‌سازی داده‌ها، تعیین پارامترهای کیفیت نسبت به مجموعه داده دوم و یا بهبود قدرت الگوریتم‌های تناظریابی در شناسایی عوارض متناظر استفاده شده است [Chehreghan and Ali Abbaspour, 2017; Dehghani, et al., 2017; Fan, et al., 2016; Koukoletsos et al., 2011; Koukoletsos et al., 2012; Lyu, et al., 2017].

منظور از تناظر یابی عوارض برداری، اتصال این عوارض در سطوح جزییات مختلف به‌منظور تشکیل پایگاه داده چند مقیاسی است[M. Zhang, 2009]. درواقع این عملیات، فرآیند جستجوی عوارض متناظر از میان مجموعه داده‌های مختلف با ارتباط مکانی و مشابهت معنایی قوی است[Chehreghan and Ali Abbaspour, 2017]. در مقالات مختلف از ویژگی‌های گوناگون داده‌های مکانی نظیر اطلاعات توصیفی، اطلاعات هندسی و یا توپولوژیک به‌منظور تناظریابی داده‌های برداری استفاده‌شده است. در این مقاله از ویژگی هندسی داده‌های برداری به‌واسطه موجود بودن این ویژگی برای تمام داده‌های برداری استفاده‌شده است.

شکل ۵ مراحل شناسایی عوارض متناظر بین دو مجموعه داده رسمی و داوطلبانه را نشان می‌دهد. همان‌طور که در شکل نمایان است برای انجام تناظریابی میان دو مجموعه داده، مشابه هر پردازش دیگر نیاز به یک مجموعه از پیش‌پردازش‌های اولیه برای آماده‌سازی داده‌ها احساس می‌شود. در این مجموعه علاوه بر پیش‌پردازش‌های ذکرشده در روش‌های منطقه‌مبنا نیاز به انجام پیش‌پردازش‌های دیگری نظیر عملیات حذف خطای توپولوژیک و ساختار گراف در شبکه معابر ایجاد می‌گردد.

منظور از ساختار گراف این مفهوم است که یال‌های این شبکه در نقاط میانی خود تقاطع نداشته ولی در نقاط پایانی خود می‌توانند هر درجه‌ای داشته باشند [Chehreghan and Ali Abbaspour, 2017].

انجام این پردازش برای مجموعه داده رسمی ضروری نیست زیرا در این مجموعه داده این ساختار به‌صورت ضمنی وجود داشته و در ساخت آن ساختار گراف رعایت گردیده است. ولی در مورد مجموعه داده داوطلبانه به دلیل عدم وجود این استاندارد در ساخت عوارض نیاز به انجام این پیش‌پردازش ضروری است.

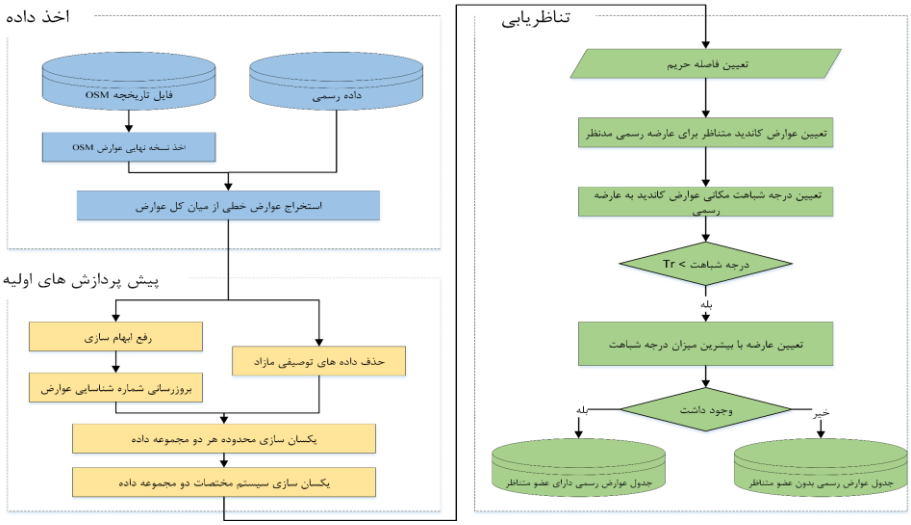
در تناظریابی عوارض برداری با منابع مختلف این امکان وجود دارد که نمایش یک عارضه در یک مجموعه‌ی داده نسبت به مجموعه دیگر متفاوت باشد، در این صورت الگوریتم تناظریابی باید قابلیت شناخت تمامی روابط ممکن میان عوارض دو مجموعه داده نظیر روابط یک به هیچ، یک‌به‌یک، یک به چند و چند به چند را داشته باشد[Dehghani et al., 2017; Sheeren et al., 2004].

پس اخذ داده‌ها و انجام پیش‌پردازش‌های لازم در گام نخست لازم است برای هر یک از عوارض رسمی عوارض متناظر احتمالی (تحت عنوان عوارض کاندید) مشخص گردند. این عملیات از طریق ایجاد ناحیه حریم اطراف هر عارضه رسمی و استخراج عوارض داوطلبانه دارای همپوشانی انجام می‌شود.

در گام دوم هدف تعیین درجه شباهت مکانی هر یک از عوارض کاندید با عارضه رسمی است. برای تعیین درجه شباهت مکانی دو عارضه نسبت به هم معیارهای هندسی و توپولوژیک متعددی در مقالات مختلف اشاره شده است [Yan, 2010; Zhang, 2009].

در این مقاله از معیار هندسی ناحیه مشترک میان حریم عوارض در شناسایی عوارض متناظر استفاده‌شده است [Chehreghan and Ali Abbaspour, 2017; Fan et al., 2016]. دلیل انتخاب معیار فوق از میان معیارهای تعیین شباهت هندسی عوارض به‌واسطه محاسبه بیشترین مقدار recall و precision در شناسایی عوارض متناظر دریکی از مناطق شهری مشابه منطقه انتخاب‌شده این پژوهش است. رابطه (۵) نحوه محاسبه درجه شباهت مکانی عارضه موردبررسی به عارضه رسمی را نمایش می‌دهد.

$$C = \frac{2A_i}{A_{PL_1} + A_{PL_2}} \qquad (۵)$$



شکل ۵. فلوچارت الگوریتم استفاده‌شده به‌منظور تناظریابی عوارض برداری

در رابطه (۵) A_i مساحت ناحیه مشترک بین دو حریم و A_{PL_1}، A_{PL_2} مساحت حریم هریک از خطوط موردبررسی است.

در گام نهایی به‌منظور تعیین عارضه متناظر عارضه رسمی انتخاب‌شده از میان مقادیر درجات شباهت مکانی محاسبه‌شده، عارضه داوطلبانه با بالاترین میزان شباهت انتخاب می‌شود. درصورتی‌که این مقدار از حد آستانه موردنظر کمتر باشد عارضه بدون عضو متناظر شناخته می‌شود. مقدار حد آستانه در بسیاری از مقالات به‌صورت تجربی تعیین‌شده و در این مقاله نیز این مقدار برابر با مقدار ۰/۸ در نظر گرفته‌شده است [Chehreghan & Ali Abbaspour, 2018].

در بسیاری از روش‌های عارضه‌مبنا برای بررسی صحت مکانی مجموعه داده‌های داوطلبانه از معیارهای مبتنی بر محاسبه فاصله استفاده‌شده است. در ادامه پرکاربردترین معیارهای مورداستفاده تشریح می‌گردد.

۳-۲-۱ فاصله حریم افزایشی

در مقالات ارزیابی صحت مکانی عوارض به‌صورت عارضه‌مبنا روش ناحیه حریم به دو شیوه متفاوت پیاده‌سازی شده است:

درروش اولیه مشابه با روش معرفی شده ناحیه حریم در قسمت روش‌های منطقه‌مبنا، عرض حریم به میزان دلخواه با توجه به شرایط مجموعه داده رسمی انتخاب شده و میزان طول قرارگیری عارضه به طول کلی آن به عنوان صحت‌مکانی عارضه درنظرگرفته می‌شود.

در روش دوم، ارزیابی صحت مکانی برخلاف روش اولیه تعیین می‌گردد. در این روش درصد طول قرارگیری عارضه در ناحیه حریم به طول کلی آن برابر با مقدار خاصی (عمدتاً ۹۵ درصد) انتخاب می‌شود و بر اساس مقدار درنظرگرفته شده فاصله حریم طی فرآیند تکراری تا رسیدن درصد طول عارضه درون ناحیه حریم به طول کلی آن به مقدار دلخواه تغییر می‌کند.

در این روش فاصله حریم محاسبه‌شده برای عوارض به‌عنوان صحت مکانی عارضه انتخاب می‌شود. شکل ٦ نحوه تغییر میزان عرض حریم ایجادشده حول عارضه رسمی برای رسیدن نسبت ذکرشده به مقدار ۹۵ درصد را نمایش می‌دهد. در این مطالعه از رویکرد فوق پس انجام تناظریابی داده‌ها به‌منظور تعیین صحت مکانی عوارض استفاده می‌گردد.

شکل ٦. ایجاد ناحیه حریم حول عارضه رسمی جهت ارزیابی صحت مکانی عارضه داوطلبانه

مسئله مهم در به‌کارگیری روش فاصله حریم افزایشی به‌منظور تعیین صحت مکانی عوارض خطی این نکته است که ماهیت عوارض ورودی این تحقیق با ماهیت عوارض موردبررسی گودچایلد و هانتر (۱۹۹۷) متفاوت است و از روش ارائه‌شده این مقاله نمی‌توان برای عوارض خطی موجود در این مقاله استفاده کرد. دلیل این امر را به این صورت می‌توان توضیح داد که در گودچایلد و هانتر (۱۹۹۷) عوارض خطی مورد بررسی پژوهش دارای پیچیدگی بسیار زیاد در طول خود بودند (عوارض ساحلی طویل و با شکستگی‌های زیاد) ولی عمده عوارض موجود در مجموعه داده این مقاله عوارض خطی بدون شکستگی هستند (عوارض خیابان با طول محدود) که به دلیل توازی خطوط مرجع و خطوط مورد بررسی استفاده از روش مقاله گودچایلد و هانتر (۱۹۹۷) به دلیل صفر شدن مخرج کسر عملاً غیرممکن خواهد بود.

$$X_{i+1} = \frac{(P_i - P_d) \times (X_i - X_{i-1})}{(P_i - P_{i-1})} + X_i \qquad (٦)$$

در رابطه (٦) مقادیر X میزان حریم در هر مرحله و مقادیر P میزان نسبت طول قرارگرفته عارضه مورد بررسی در حریم عارضه رسمی به طول کلی عارضه را نشان می‌دهد. در حالتی که عارضه مورد بررسی نسبت به عارضه رسمی حالت توازی داشته باشد پس از چندین تکرار مسئله باحالتی مواجه می‌گردد که نسبت P محاسبه‌شده در مرحله i ام و i-1 ام برابر شده و باعث صفر شدن مخرج کسر می‌گردد.

برای حل مشکل به وجود آمده در این مقاله از راهکار ارائه‌شده در کوکولتسوس و همکاران (۲۰۱۱) به‌منظور تعیین فاصله حریم استفاده‌شده است[Koukoletsos et al., 2011]. اساس کار روش پیشنهادی این مقاله به این صورت است که ابتدا با توجه به‌دقت مجموعه داده رسمی فاصله حریم مقداردهی شده و با محاسبه پارامتر P در هر مرحله مقدار فاصله حریم دو برابر و یا نصف فاصله حریم مرحله قبل انتخاب می‌گردد. این عملیات تا رسیدن مقدار P به میزان دلخواه ادامه می‌یابد. در این حالت نیز به دلیل توازی خط مرجع و خط موردبررسی امکان عدم همگرایی مسئله وجود خواهد داشت. برای رفع این مشکل در این مقاله شرط حداکثر تکرار به الگوریتم مسئله اضافه‌شده است (فلوچارت شکل ۵).

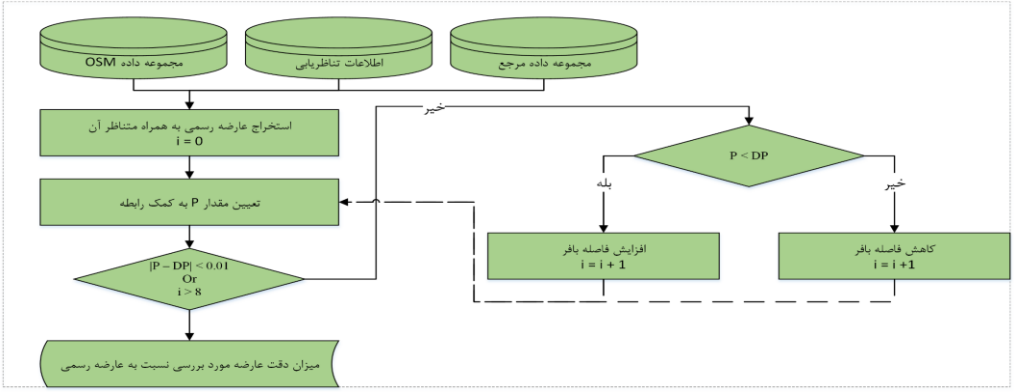
۳-۲-۲ فاصله هاسدورف

فاصله هاسدورف بیشترین مقدار انحراف میان دو چندخطی را محاسبه می‌نماید. برای محاسبه این کمیت کافی است یک‌بار بیشترین فاصله خط دوم تا خط اول d1 و بار دیگر بیشترین فاصله خط اول تا خط دوم d2 محاسبه گردد. فاصله هاسدورف میان این دو خط مقدار بیشینه میان دو مقدار محاسبه‌شده خواهد بود [Girres & Touya, 2010].

$$\text{Hausdurff distance} = \text{Max}(d1 . d2) \qquad (۷)$$

۳-۲-۳ فاصله میانگین

این معیار اولین بار توسط مک‌مستر در سال ۱۹۸٦ به‌منظور بررسی نتیجه تعمیم یک چندخطی نسبت به چندخطی اصلی و بررسی میزان جابجایی آن تعریف شد [McMaster, 1986] . بنا بر تعریف این مقاله فاصله میانگین بین دو چندخطی از نسبت مساحت سطح ایجاد شده میان دو خط به میانگین طول دو چندخطی محاسبه می‌شود. این معیار برخلاف فاصله هاسدورف که حداکثر فاصله میان دو چندخطی را محاسبه می‌نماید، مقدار میانگین فاصله دو چندخطی را محاسبه می‌کند. یکی از چالش‌های موجود در محاسبه این معیار حالات متفاوت قرارگیری عوارض خطی نسبت به همدیگر است که باعث پیچیده شدن الگوریتم محاسباتی می‌گردد. در این مقاله به‌منظور محاسبه مساحت محصور میان دو پاره‌خط از رویکردی که در



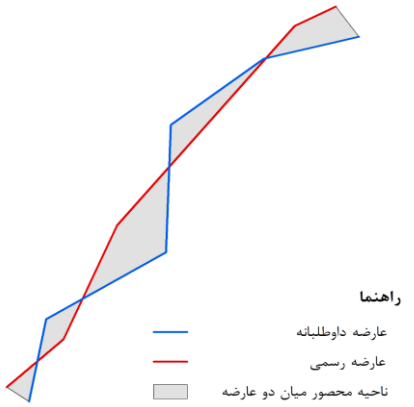
شکل ۷. الگوریتم استفاده‌شده برای محاسبه فاصله حریم افزایشی

ادامه تشریح می‌شود استفاده‌شده است. در گام اول لازم است که وضعیت نقاط ابتدایی و انتهایی دو خط نسبت به هم تعیین گردد. برای این منظور می‌توان از فاصله اقلیدسی میان نقاط ابتدایی و انتهایی خط اول با نقاط ابتدایی و انتهایی خط دوم استفاده کرد. در گام دوم از طریق اتصال نقاط ابتدایی و انتهایی یک‌چند ضلعی ایجاد می‌گردد (شکل ۸). پس از تعیین مساحت این چندضلعی و محاسبه طول خطوط سازنده آن معیار صحت مکانی چندخطی اول نسبت به چندخطی دوم از طریق رابطه (۸) تعیین می‌گردد. در این رابطه $2A_i$ مساحت محصور میان دو پاره مدنظر L_{PL_2} و L_{PL_1} طول این دو پاره خط است.

$$\text{Average distance} = \frac{2\,A_i}{L_{PL_1} + L_{PL_2}} \qquad (۸)$$

٤. پیاده‌سازی

منطقه مورد مطالعه جهت ارزیابی روش‌های معرفی‌شده در بخش قبل مناطق ۱۱ تا ١٤ شهر تهران است (شکل ۹). دلیل انتخاب شهر تهران، میزان مشارکت بالاتر افراد نسبت به سایر مناطق کشور در ساخت مجموعه داده داوطلبانه این شهر است. مناطق انتخاب‌شده در این پژوهش مساحتی در حدود ٥٠/٥٦ کیلومترمربع داشته و تعداد ۲۹۲۳۹ عارضه خطی در مجموعه داده رسمی و تعداد ١٦١٢١ عارضه در مجموعه داده داوطلبانه این مناطق به ثبت رسیده است.



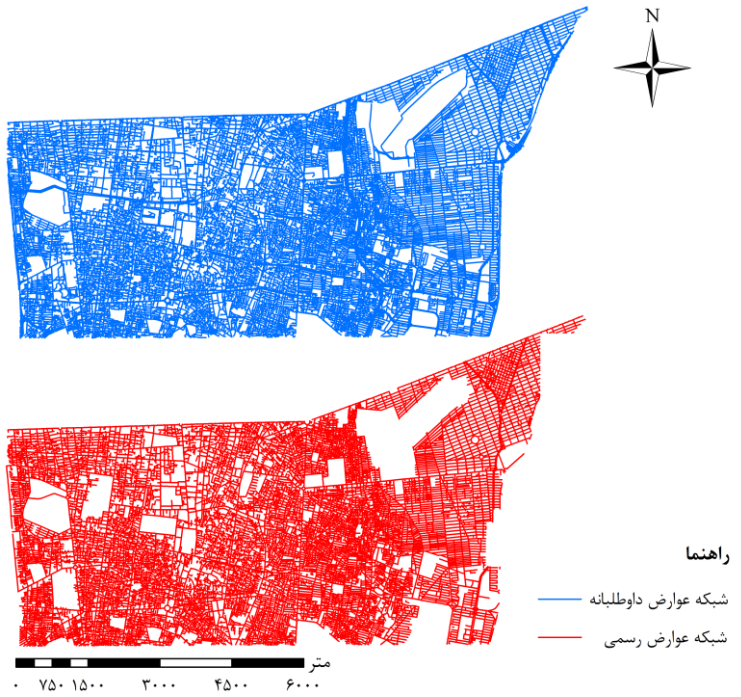
شکل ۸ ناحیه محصور میان دو عارضه موردبررسی

به‌صورت جزئی‌تر میزان طول این عوارض در مجموعه داده رسمی ۱۲۳۵ کیلومتر و میزان طول این عوارض برای مجموعه داده داوطلبانه ۱۸۱۸ کیلومتر محاسبه‌شده است. این ۱۸۱۸ کیلومتر شبکه عوارض خطی توسط ۵۰۵۳۲ مشارکت از ابتدای ساخت این مجموعه داده تاکنون، توسط ۲۸۳ نفر ایجادشده است. در چهار منطقه انتخاب شده، منطقه ۱۱ و ۱۳ دارای بیشترین و کمترین نرخ تراکم عوارض هستند. بررسی پارامترهای آماری طول عوارض مشارکت‌شده نشان می‌دهد میانگین ١٦١٢١ عارضه مشارکت شده برابر با ۱۱۳/۸۵ با انحراف معیار ۳۱۰/۵ متر است.

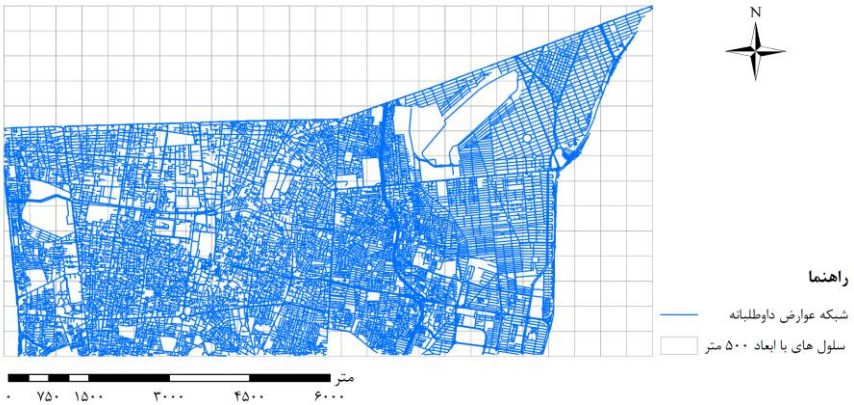
همان‌طور که اشاره شد به‌منظور پیاده‌سازی روش‌های بررسی‌شده در قسمت قبل لازم است تا قالب دو مجموعه‌ی داده یکسان گردد. ازاین‌رو عملیات یکسان‌سازی سیستم مختصات، سیستم تصویر و یکسان‌سازی محدوده بر روی دو مجموعه داده صورت می‌گیرد.

٤-١ ارزیابی صحت مکانی درروش منطقه‌مبنا

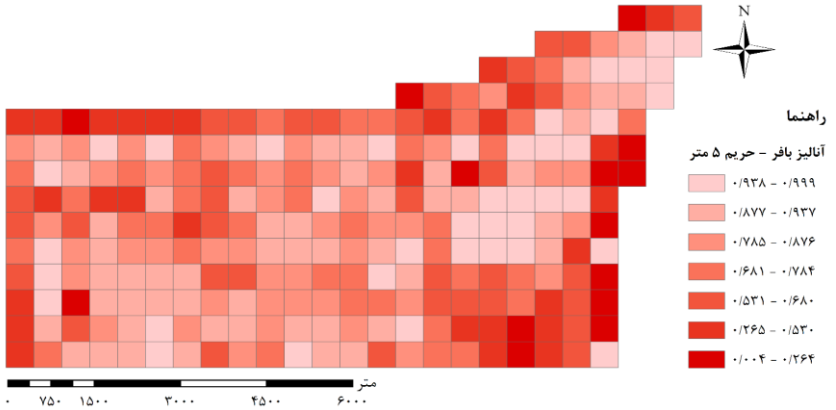
همان‌طور که بیان گردید جهت پیاده‌سازی روش‌های منطقه‌مبنا نیاز به شبکه‌بندی منطقه موردمطالعه است. با انجام بررسی بر روی میزان تراکم شبکه راه‌ها، شبکه‌بندی با ابعاد ۵۰۰ در ۵۰۰ متر متناسب با محدوده دو داده موردبررسی ایجاد و عوارض موجود در دو لایه با استفاده از این شبکه برش داده می‌شود (شکل ۱۰).



شکل ۹. داده‌های VGI (الف) و رسمی (ب) مورداستفاده در مناطق ۱۱ تا ۱٤ شهر تهران

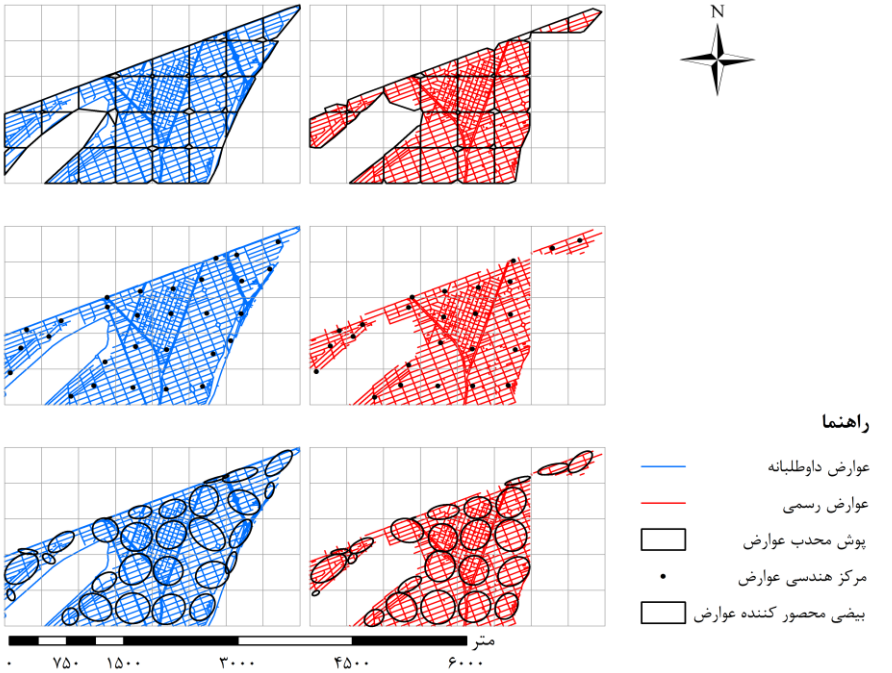


شکل ۱۰. تقسیم منطقه‌ی موردمطالعه به بخش‌های کوچک‌تر به‌منظور بررسی پارامترهای کیفیت

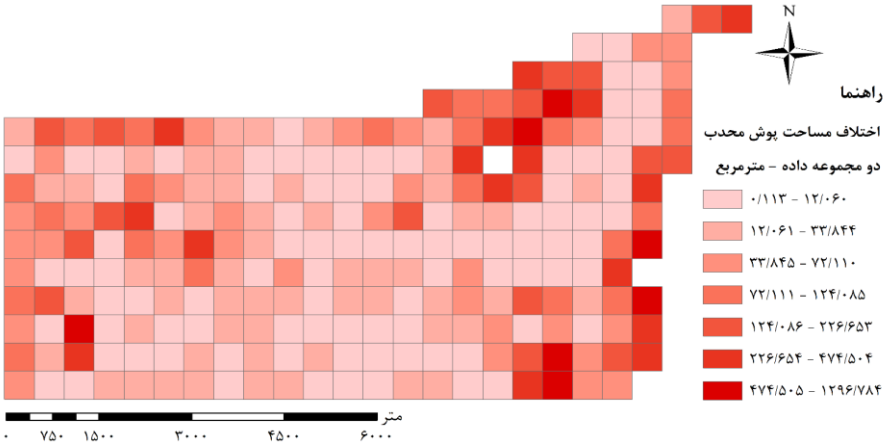


شکل ۱۱. آنالیز بافر در تعیین صحت مکانی عوارض داوطلبانه نسبت به رسمی

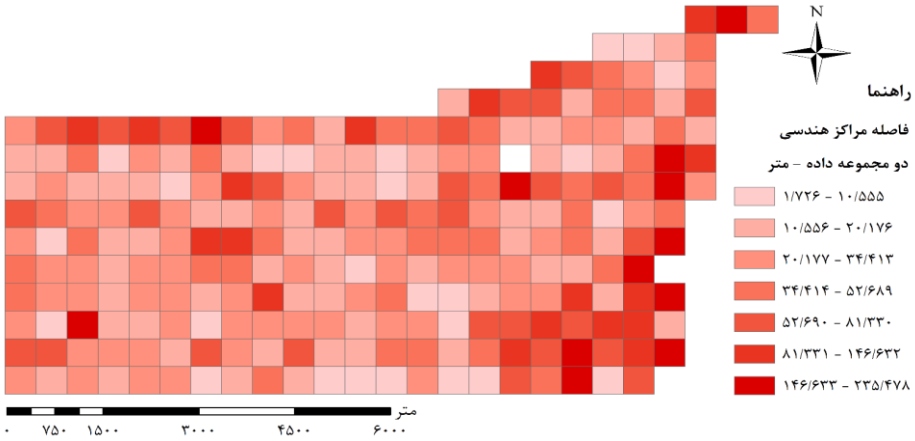
نخستین بخش از روش‌های منطقه‌مبنا ارزیابی صحت‌مکانی به پیاده‌سازی شیوه مرسوم تعیین صحت مکانی، آنالیز ناحیه حریم پرداخته‌شده است. آنالیز ناحیه حریم به‌واسطه پیاده‌سازی ساده و تأثیرپذیری پایین از خطاهای فاحش، آنالیز پرکاربرد در مطالعات پیشین بوده است [Zhou et al., 2014]. به سبب ویژگی‌های اشاره‌شده، این معیار جهت مقایسه نتایج حاصل‌شده از سایر معیارها قرارگرفته است. شکل ۱۱ صحت مکانی عوارض داوطلبانه به ازای ایجاد ناحیه حریم به شعاع ۵ متر را نمایش می‌دهد. شکل ۱۲، شکل ۱۳ و شکل ۱٤ مربوط به پیاده‌سازی معیارهای مساحت پوش محدب، مرکز هندسی و توجیه بیضی محصورکننده عوارض هستند. چندضلعی‌های ایجادشده به‌عنوان پوش محدب عوارض پس از استخراج نقاط عوارض، توسط الگوریتم گراهام اسکن[14] ایجادشده که برای قسمتی از منطقه انتخاب‌شده این چندضلعی‌ها در ترسیم‌شده است (شکل ۱۲).



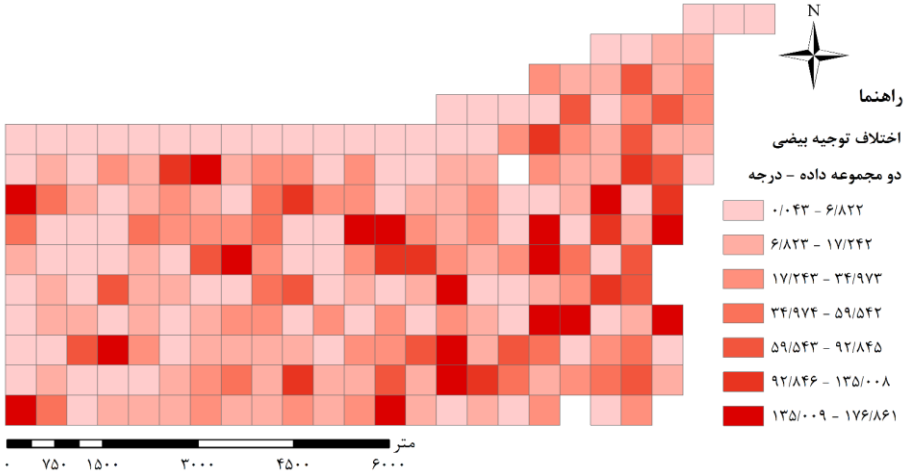
شکل ۱۲. نمایش آنالیزهای پوش‌محدب، مرکز هندسی و بیضی محصورکننده بر قسمتی از منطقه مطالعاتی



شکل ۱۳. اختلاف مساحت پوش محدب دو مجموعه‌ی داده رسمی و داوطلبانه



شکل ١٤. فاصله میان مراکز هندسی دو مجموعه‌ی داده رسمی و داوطلبانه



شکل ١٥. اختلاف جهت شاخص توزیع جهت دو مجموعه‌ی داده رسمی و داوطلبانه

کیفیت یک مجموعه داده نسبت به مجموعه داده دیگر کمیتی است مشخص که روش ارزیابی نباید تأثیری در نتایج حاصل‌شده از آن داشته باشد. ازاین‌رو محاسبه صحت مکانی بالا برای ناحیه خاص توسط یک معیار و محاسبه صحت مکانی پایین برای همان ناحیه توسط معیار دیگر توجیه‌پذیر نیست و نیازمند شناسایی ضعف‌های مربوط به هر معیار است. چنین وضعیتی در نتایج حاصل‌شده از دو معیار آنالیز بافر و توجیه بیضی محصورکننده عوارض وجود دارد. شکل ١٦ نمونه‌ای از ضعف معیار توجیه بیضی محصورکننده داده‌ها را نشان می‌دهد. حذف یکسری عوارض از مجموعه داده موردبررسی طبیعتاً بر صحت مکانی مابقی عوارض نباید تأثیرگذار باشد بلکه بر کامل‌بودن آن تأثیرگذار است.

ازاین‌رو حذف چندین عارضه نباید منجر به محاسبه اختلاف فاحش در صحت مکانی شود. توجیه ٧١ درجه در حالت ب شکل ١٦ نشان می‌دهد که صحت‌مکانی پس از حذف چند عارضه محدود تغییرات زیادی داشته است.

برخلاف این معیار، فاصله میان مراکز هندسی داده‌ها در دو حالت الف و ب به میزان ٢٤/٧٠ متر جابجا شده است که نسبت به معیار قبل نتایج آن به‌عنوان میزان تغییرات صحت‌مکانی قابل پذیرش‌تر است. بنابراین توصیه این مطالعه

این است حتی‌الامکان از معیار توجیه بیضی محصورکننده جهت ارزیابی صحت مکانی استفاده نگردد.

مقایسه کمیت‌های ارزیابی صحت‌مکانی را می‌توان از زاویه دیگر، شباهت مقادیر صحت‌مکانی سلول‌های مجاور مورد بررسی قرار داد. براساس قانون تابلر [Tobler, 1970]، با توجه به این‌که مناطق نزدیک نسبت به مناطق دور دارای شباهت مکانی بالاتری هستند، انتظار می‌رود صحت مکانی سلول‌های مجاور به یکدیگر شباهت داشته باشند. ارزیابی میزان همبستگی مکانی مقادیر صحت مکانی در ٤ معیار معرفی شده وجود شباهت مکانی در پارامتر کیفیت محاسبه شده توسط آنالیز حریم را نتیجه داد (z-score = ٦/٥٩). بر خلاف آنالیز حریم نتایج بررسی پارامترهای کیفیت مرتبط به شکل‌های ١٣، ١٤ و ١٥ با محاسبه مقدار z-score به میزان ٢/٠٤ ، ١/٠٧ و ١/٩٤ رفتار تصادفی این کمیت‌های در محدوده انتخاب شده را نتیجه‌گیری کرد. براساس این نتایج پیشنهاد این مطالعه استفاده از آنالیز حریم نسبت به سه معیار دیگر (فاصله میان مراکز هندسی، اختلاف مساحت پوش محدب و اختلاف توجیه بیضی توزیع) است.

شکل ١٦. حساسیت معیار توجیه بیضی محصورکننده به حذف یا اضافه شدن عوارض جدید

موضوع دیگری که در طراحی معیارهای ارزیابی کیفیت نیاز است به آن توجه داشت عدم وجود ارتباط میان کمیت‌های ارزیابی کیفیت است، به‌خصوص برای معیار صحت مکانی که کامل‌بودن می‌تواند بر نتایج حاصل‌شده از آن تأثیرگذار باشد. هرچند صحت مکانی عوارض درون یک سلول بالا باشد ضعف در کامل‌بودن عوارض آن سلول می‌تواند منجر به محاسبه صحت مکانی غیرواقعی برای سلول می‌گردد. شکل ١٨ سلولی را نمایش می‌دهد که صحت مکانی عوارض در آن از لحاظ بصری (نسبت به عوارض رسمی موجود) در حد مطلوبی است، ولی معیارهای ارزیابی منطقه‌مبنا نتایج متفاوتی را ارائه می‌دهند. به عنوان مثال فاصله ٣١١ متری میان مرکز هندسی دو مجموعه داده نشان‌دهنده صحت بسیار پایین برای سلول انتخاب‌شده است.

رابطه (٩) کامل‌بودن داده‌های داوطلبانه نسبت به داده‌های رسمی را نمایش می‌دهد. تصویر ترسیم‌شده نشان می‌دهد در سال ٢٠١٨ عمده منطقه دارای مجموعه داده داوطلبانه کامل‌تر نسبت به داده‌های رسمی است (کامل‌بودن ٩٤/٦٥ درصد از سلول‌ها بیشتر از مقدار یک گزارش‌شده است). هدف از بررسی کامل‌بودن داده‌های داوطلبانه در این بخش بررسی ارتباط این کمیت با فاصله میان مراکز هندسی به‌عنوان معیار نشان‌دهنده صحت مکانی است. بر اساس رابطه ٩ به نظر می‌رسد افزایش میزان کامل‌بودن، منجر به افزایش تعداد و پراکندگی عوارض در سلول‌ها نسبت به عوارض رسمی گردد. طبیعی است که افزایش نسبت طول عوارض داوطلبانه نسبت به عوارض رسمی خود باعث ایجاد فاصله زیادتر میان مراکز هندسی دو مجموعه داده گردد.

بررسی آماری ارتباط پارامتر کامل‌بودن و فاصله میان مراکز هندسی دو مجموعه داده به عنوان پارامتر ارزیابی صحت مکانی ارتباط مستقیم و معنادار آماری (ضریب همبستگی پیرسون ٠/٨٧) را نشان می‌دهد (جدول ٢). براساس شکل فلان

همان‌گونه که ملاحظه می‌شـــود در اغلب ســـلول‌های با مقدار کامل‌بودن ۱ تا ۱/۵ میزان فاصله میان مرکز هندسی دو مجموعه داده کمتر از ۳۳ متر محـاســـبه شـــده اســـت درحالی که در ســـلول‌های دارای کامل‌بودن بالاتر و یا پایین‌تر از بازه ۱ تا ۱/۵ میانگین فاصـــله هندســـی میان مراکز هندســـی بیش از ۸۵ متر ارزیابی شده است. این نتایج همچنین با بررسی بصری تغییرات کامل‌بودن و فاصـــله میان مراکز هندســـی دو مجموعه داده در شکل ۱۷ قابل نتیجه‌گیری است.

جدول ۱. ارزیابی صحت‌مکانی عوارض ترسیم شده شکل ۱۸

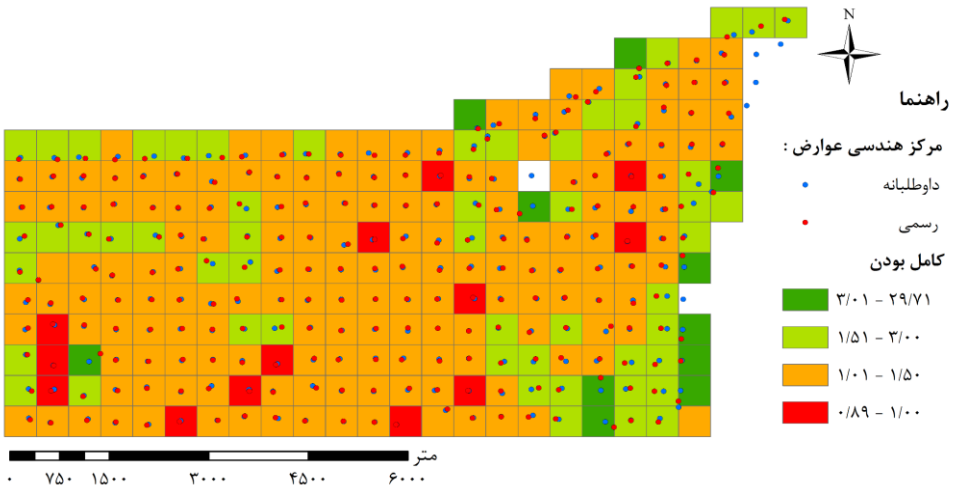
صحت مکانی	معیار انتخابی
۲۰٤۱۷٦ – مترمربع	اختلاف مساحت پوش محدب
۳۱۰/۹٦ – متر	فاصله مراکز هندسی
۷٤/۹۵ – درجه	اختلاف جهت توجیه بیضی محصورکننده
۰/۰٦٦	آنالیز بافر

$$Completness = \frac{\sum osm\ length}{\sum refrence\ length} \qquad (۹)$$

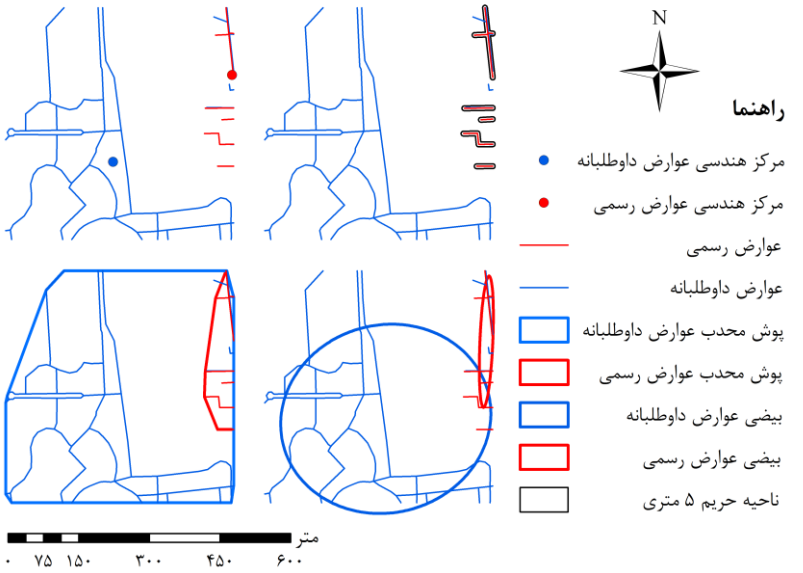
ضـعف معیارهای بررسـی‌شـده را می‌توان به خاطر ماهیت محاسباتی این دسته از روش‌ها به شمار آورد، زیرا در این دسته از روش‌هـا کلیه عوارض بدون درنظرگرفتن وجود رابطه تناظر میـان آن‌هـا وارد عملیات پردازش می‌شـــوند. به‌عنوان مثال در شکل ۱۸، اغلب عوارض داوطلبانه فاقد عضو نظیر در مجموعه داده رسـمی هستند. ازاین‌رو برای تعیین صحت مکانی عوارض داوطلبانه تنها لازم اسـت عوارض رسـمی نظیر آن وارد مرحله پردازش گردنـد، ولی در روش‌هـای منطقـه‌مبنـا چنین امکـانی وجود نـداشـتـه و کلیه عوارض یکجا وارد پروســـه پردازش می‌شـود. در مواجه با داده‌های دارای چنین شرایطی توصیه این مقاله این اســـت در صـــورت وجود اختلاف چشـــمگیر میان کامل‌بودن عوارض داوطلبانه نســـبت به رســـمی، ابتدا برای هر ســـلول عملیات تناظریابی انجام گردد، ســـپس برای عوارض متناظر میان دو مجموعه داده صـــحت‌مکانی مورد ارزیابی قرار گیرد. شـکل ۱۹ صـحت مکانی عوارض پس از انجام عملیات تناظریابی دو مجموعه داده را نمایش می‌دهد.

جدول ۲. نتیجه حاصل از بررسی آزمون پیرسون در تعیین همبستگی دو کمیت کامل‌بودن و فاصله میان مراکز هندسی دو مجموعه داده

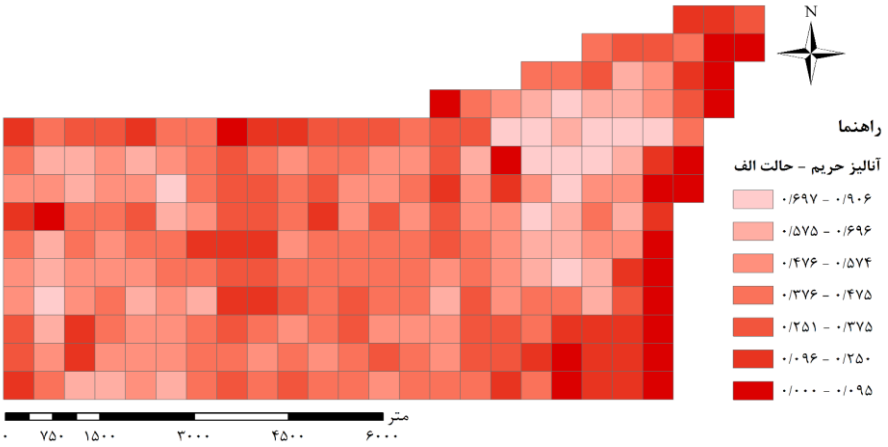
عنوان	میزان همبستگی	آماره کای-اسکور	مقدار P-value	نتیجه آزمون
همبستگی فاصله مراکز هندسی و کامل‌بودن مجموعه داده	۰/۸٦۹	۹/۳۰۳۲	تقریباً صفر	رد فرض صفر



شکل ۱۷. نمایش ارتباط کامل‌بودن و فاصله میان مراکز هندسی عوارض



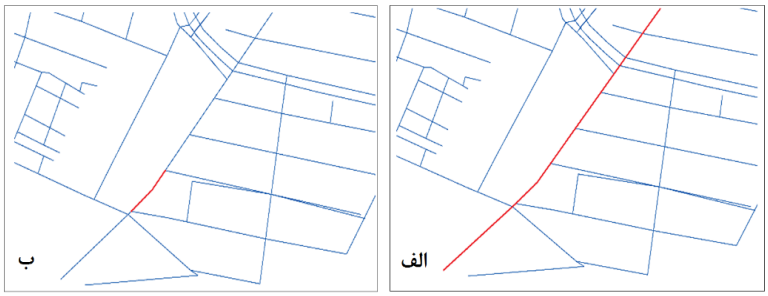
شکل ۱۸. ضعف معیارهای منطقه‌مبنا ارزیابی کیفیت



شکل ۱۹. ترکیب تناظریابی و در رویکرد تعیین کیفیت منطقه‌مبنا

در این قسمت از مقاله به بررسی روش‌های عارضه‌مبنا در تعیین کیفیت داده‌های داوطلبانه پرداخته می‌شود. همان‌طور که بیان شد، در روش‌های تعیین کیفیت به‌صورت عارضه‌مبنا هر عارضه درون یک مجموعه داده با عارضه نظیر خود در مجموعه داده دیگر مقایسه می‌شود. درصورتی‌که دو مجموعه داوطلبانه و رسمی وارد الگوریتم تناظریابی با معیار تعیین درجه شباهت مکانی، ناحیه مشترک میان حریم عوارض شوند این الگوریتم تعداد کمی از عوارض را به‌عنوان عوارض نظیر خروجی می‌دهد. دلیل این امر مربوط به عدم رعایت ساختار گراف در داده‌های داوطلبانه است.

در عملیات ایجاد ساختار گراف در داده‌های داوطلبانه باید توجه داشت که در این عملیات امکان تبدیل یک عارضه خطی



شکل ۲۰. شبکه‌ی معابر داوطلبانه قبل (الف) و بعد (ب) از رفع ابهام سازی

به دو یا چند عارضه مجزا وجود خواهد داشت. بنابراین با توجه شماره شناسایی عوارض تقسیم شده، به‌منظور جلوگیری از ایجاد خطا لازم است شماره شناسایی عوارض پس از عملیات ایجاد ساختار گراف بروز رسانی گردد. در شکل ۲۰ – الف شبکه معابر مجموعه داده داوطلبانه قبل از ایجاد ساختار گراف و در شکل ۲۰ –ب شبکه معابر مجموعه داده داوطلبانه بعد از ایجاد ساختار گراف نشان داده شده است.

پس از رفع ابهام سازی داده‌های داوطلبانه عملیات تناظریابی بر روی دو مجموعه داده رسمی و داوطلبانه انجام شد و از میان تعداد ۲۹۲۳۹ عارضه رسمی تعداد ۱۷٦۸٤ عارضه به عنوان عوارض دارای متناظر در مجموعه داده داوطلبانه انتخاب شد. پس از انجام تناظریابی، صحت مکانی عوارض داوطلبانه دارای متناظر در مجموعه داده رسمی به کمک سه معیار فاصله میانگین، فاصله هاسدورف و آنالیز بافر ارزیابی شد که هیستوگرام مربوط به کمیت‌های فوق در **Error! Reference source not found.Error! Reference source not found.** شکل ۲٤ به همراه پارامترهای آماری آن در **Error! Not a valid bookmark self-reference.** ارائه‌شده است.

بررسی نمودار هیستوگرام دو معیار آنالیز حریم و فاصله هاسدورف نشان می‌دهد رفتار این دو نمودار تا حدی شبیه هم است. وجود ضریب همبستگی ۰/٤۲٦ میان این دو کمیت نشان می‌دهد که امکان استفاده از معیار فاصله هاسدورف در مقابل آنالیز حریم وجود دارد. اهمیت موضوع، به‌واسطه زمان پردازشی بالای معیار آنالیز حریم نسبت به فاصله هاسدورف به دلیل تکراری بودن الگوریتم محاسباتی آن است. بر اساس این مقاله به دلیل رفتار مشابه دو معیار بهتر است از فاصله هاسدورف برای تعیین صحت مکانی عوارض استفاده شود.

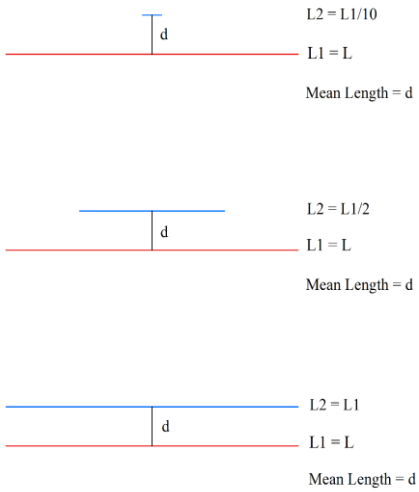
جدول ۳. کمیت‌های آماری روش‌های ارزیابی صحت‌مکانی

معیار	پارامتر	مقدار (متر)
فاصله میانگین	کمترین	۰/۰۰۱
	بیشترین	۲/۰٦
	میانگین	۰/۷۲
	انحراف معیار	۰/٤۳
فاصله هاسدورف	کمترین	۰/۲٤
	بیشترین	۹۱/٦۷
	میانگین	۵/۱٦
	انحراف معیار	۵/۵۷
آنالیز بافر حالت ب	کمترین	۰/۰۱
	بیشترین	**۵۵/٦۸**
	میانگین	۲/۵٦
	انحراف معیار	۲/٦۱
آنالیز بافر حالت الف	کمترین	۰/۳۸۷
	بیشترین	۱
	میانگین	۰/۹۹۲
	انحراف معیار	۰/۰۳۷

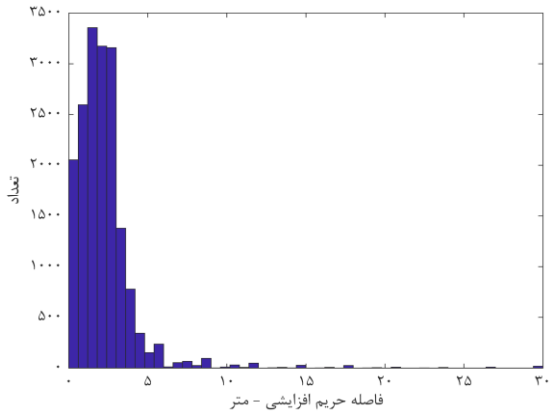
مقایسه هیستوگرام و جدول فاصله میانگین عوارض نشان می‌دهد که کمیت اشاره شده رفتار متفاوتی با دو معیار قبل دارد.

این معیار همان‌گونه که در مطالعات پیشین به آن پرداخته‌شده است بیشتر برای تعیین جابجایی هندسی عارضه به کار گرفته شده است. به‌منظور روشن شده موضوع، یک عارضه رسمی به همراه سه نسخه از عارضه داوطلبانه نظیر آن در نظر گرفته‌شده است (شکل ٢٤). بدیهی است، صحت مکانی سه نسخه در نظر گرفته‌شده متفاوت هستند، ازاین‌رو نتایج حاصل از بهره‌گیری معیارهای ارزیابی کیفیت لازم است چنین موضوعی را تائید نماید. ولی نتیجه جالب این است که فاصله میانگین سه نسخه ترسیم‌شده مقادیر یکسانی (d – فاصله میان دو عارضه) است. این موضوع را می‌توان ضعف فاصله میانگین در ارزیابی کیفیت به شمار آورد و طبق نظر این مطالعه بهتر است از فاصله میانگین در تعیین صحت مکانی عوارض استفاده نگردد.

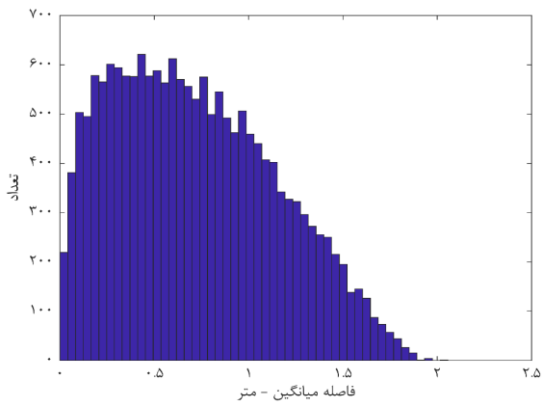
علاوه بر معیارهای اشاره‌شده آنالیز حریم مشابه با آنچه در بخش روش‌ها منطقه‌مبنا ذکر شد (حالت الف) نیز پیاده‌سازی شده است. نتایج این آنالیز نشان می‌دهد در حدود ۸۳/۳ درصد از عوارض صحت مکانی بالاتر از ۹۹ درصد و در حدود ۹/۱۲ درصد از داده‌ها دارای صحت مکانی بین ۹۵ تا ۹۹ درصد هستند. به دلیل تجمع مقادیر صحت مکانی در این بازه‌ها هیستوگرام این معیار ترسیم نشده است.



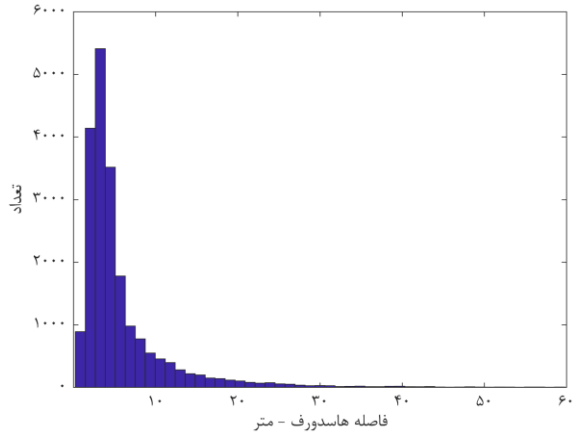
شکل ٢١. فاصله میانگین برابر در سه نسخه عوارض داوطلبانه



شکل ٢٢. هیستوگرام آنالیز بافر (حریم افزایشی) عوارض دارای متناظر



شکل ۲۳. هیستوگرام فاصله میانگین عوارض دارای متناظر



شکل ۲٤. هیستوگرام فاصله هاسدورف عوارض دارای متناظر

۵. نتیجه‌گیری و پیشنهادها

هدف از این مقاله بررسی روش‌های ارائه‌شده در مقالات به‌منظور ارزیابی پارامترهای کیفیت داده‌های مکانی داوطلبانه با تمرکز بر معیار صحت مکانی بود. این پارامتر برای مجموعه داده‌های داوطلبانه مناطق ۱۰ تا ۱٤ شهر تهران با استفاده از دودسته کلی روش‌های ارزیابی پارامترهای کیفیت (عارضه‌مبنا و منطقه‌مبنا) موردبررسی قرار گرفت.

در ارزیابی نخست انجام شده، پارامتر صحت‌مکانی عوارض به شیوه منطقه‌مبنا توسط معیارهای فاصله میان مراکز هندسی دو مجموعه داده، اختلاف مساحت پوش محدب، اختلاف توجیه بیضی توزیع و آنالیز حریم مور ارزیابی قرار گرفت. جنبه مشترک در میان چهار پارامتر ارزیابی صحت‌مکانی، وابستگی بسیار شدید نتایج حاصل شده از این روش‌ها نسبت به پارامتر دیگر کیفیت داده‌ها، کامل‌بودن است (همبستگی ۰/۸۷). بنابراین بررسی کاربری‌های متناظر دو مجموعه داده و حذف اطلاعات مازاد دو مجموعه داده می‌تواند منجر به بهبود ارزیابی صحت‌مکانی شود.

در مقایسه کلی عوارض دو مجموعه متفاوت با استانداردهای متفاوت امکان مقایسه عوارض از نوع متفاوت وجود دارد که همین عامل منجر به ارزیابی غیر واقعی از کیفیت مجموعه داده داوطلبانه خواهد شد. بر اساس بررسی‌های این مقاله بهتر است در مواجه با مجموعه داده‌های دارای استاندارد مختلف، ابتدا طی یک عملیات تناظریابی عوارض غیر مشترک میان دو مجموعه داده حذف گردد، سپس عملیات تعیین کیفیت منطقه‌مبنا برای عوارض دارای متناظر انجام پذیرد. البته این روش برخلاف رویکرد به ظاهر منطقه‌مبنا آن به دلیل استفاده از الگوریتم‌های تناظریابی در شناسایی عوارض متناظر، سرعت اجرایی کند خواهد داشت.

مقایسه ٤ پارمتر ارزیابی منطقه‌مبنا صحت‌مکانی در این مطالعه نشان داد بر خلاف آنالیز حریم، صحت مکانی محاسبه شده توسط معیارهای فاصله میان مراکز هندسی دو مجموعه داده، اختلاف مساحت پوش محدب و اختلاف توجیه بیضی توزیع دارای رفتار تصادفی در سطح منطقه است. این واقعیت همچنین به صورت نمونه با بررسی تغییرات توجیه بیضی توزیع با اضافه شدن یک عارضه به مجموعه داده مورد بررسی، بررسی شد. با جمع‌بندی نتایج حاصل‌شده پیشنهاد این مطالعه استفاده از آنالیز حریم است.

عمده‌ترین ضعف معیار آنالیز حریم در حالت ساده، عدم وجود رابطه دقیق برای تعیین میزان شعاع حریم است. برخلاف

این روش، در حالت دوم پیاده‌سازی این روش با توجه به این‌که عرض حریم ثابت در نظر گرفته نشده و تا رسیدن به درصد مطلوب مقدار آن تغییر می‌کند انتخاب شعاع حریم مسئله‌ساز نخواهد بود. البته لازم است این نکته را در نظر داشت که پیاده-سازی حالت ثانویه آنالیز بافر نسبت به حالت اولیه مشکلات مخصوص به خود را دارد. حجم محاسبات بسیار بالاتر حالت ثانویه نسبت به حالت اولیه یکی از دلایلی است که سبب شده در مطالعات حالت اولیه این معیار نسبت به حالت دیگر ترجیح داده شود. پس از بررسی روش‌های مختلف ارزیابی کیفیت منطقه‌مبنا، پیشنهاد می‌شود در صورت تمایل به استفاده از مدل ساده آنالیز بافر در تعیین صحت مکانی عوارض، بهتر است از مدل انتشار خطا براساس خطا نقاط انتهایی خطوط در مقابل انتخاب شعاع حریم به‌صورت تجربی بهره گرفته شود.

در قسمت پایانی این مقاله پس از تناظریابی داده‌های برداری صحت‌مکانی عوارض نظیر به شیوه عارضه‌مبنا تعیین شد. در میان این روش‌ها، معیار فاصله حریم افزایشی به سبب فرآیند تکراری آن در تعیین صحت مکانی، دارای بیشترین زمان محاسباتی و فاصله هاسدورف در میان این سه معیار دارای کمترین زمان اجرایی بود.

در این مقاله بنابر نتایج به‌دست‌آمده از دو معیار آنالیز بافر و فاصله هاسدورف و همچنین وجود شباهت میان هیستوگرام و همبستگی دو کمیت نتیجه‌گیری شد که فاصله هاسدورف می‌تواند جایگزین مناسبی برای غلبه بر سرعت اجرائی پایین روش آنالیز بافر باشد. عدم توانایی معیار فاصله میانگین در تعیین صحت مکانی برخی از عوارض که نمونه‌ای از آن نیز به آن اشاره شد دلیلی بر ناکارآمدی این معیار است و طبق نظر این مطالعه و مطالعات پیشین بهتر است معیار فوق در کاربردهای دیگری نظیر بررسی جابجایی‌های هندسی یک عارضه به کار گرفته شود.

پس از مقایسه روش‌های تعیین صحت‌مکانی در دو رویکرد منطقه‌مبنا و عارضه‌مبنا و انتخاب معیار مناسب، صحت‌مکانی مجموعه داده داوطلبانه شهر تهران در مناطق ۱۱-۱٤ ارزیابی شد. صحت‌مکانی محاسبه شده برای عوارض در رویکرد منطقه‌مبنا با اتخاذ عرض حریم به میزان ۵ متر ۰/۷۸۱ و صحت‌مکانی عوارض به کمک معیار فاصله هاسدورف ۵/۲ متر محاسبه شد که نشان‌دهنده صحت مکانی بالای این دو مجموعه داده است.

در مقاله پیش‌رو سعی شد یکی از مهم‌ترین جنبه‌های کیفیت داده‌های داوطلبانه، صحت‌مکانی مورد ارزیابی قرار گیرد. با توجه به این واقعیت که رویکردها و روش‌های مختلفی برای این معیار معرفی شده است در این مطالعه به مقایسه تحلیلی این روش‌ها پرداخته شد. حال با توجه به این ارزیابی‌ها توصیه این مطالعه پرداختن به جنبه‌های دیگر کیفیت نظیر کامل‌بودن و روش‌های مختلف ارزیابی آن است. بررسی دقیق‌تر این معیار کیفیت و ارتباط آن با پارامتر صحت‌مکانی به شیوه تحلیلی می‌تواند موضوع مناسبی برای تحقیق‌های آتی باشد. از طرفی با توجه به این واقعیت که در این مقاله تنها صحت‌مکانی نسخه نهایی داده‌های داوطلبانه مورد بررسی واقع شد، بررسی بهبود صحت‌مکانی در نسخه‌های مختلف عوارض مختلف می‌تواند موضوع مهمی برای تحقیق‌های آینده باشد.

٦. پی‌نوشت‌ها

1- Volunteered Geographic Information

2- Spatial Data Infrastructure

3- Global Positioning System

4- International Organization for Standardization

5- Completeness

6- Logical consistency

7- Positional Accuracy

8- Temporal Accuracy

9- Attribute Accuracy

10- Convex Hull

11- Directional Distribution

12- Unit based

13- Object based

14- Graham Scan

۷. مراجع

Antoniou, V. and A. Skopeliti (2015). "Measures and indicators of VGI quality: an overview." ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 2, No. 5, pp. 345.

Arsanjani, J. J., P. Mooney, A. Zipf and A. Schauss (2015). "Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets." OpenStreetMap in GIScience, Springer**:** pp.37-58.

Bégin, D., R. Devillers and S. Roche (2013). "Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours." Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, Vol. 6, No. 2, pp. 149-154.

Brovelli, M. and G. Zamboni (2018). "A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints." ISPRS International Journal of Geo-Information, Vol. 7, No. 8, pp. 289.

Bruns, A. (2008). "The future is user-led: The path towards widespread produsage." Fibreculture journal, Vol 2, No. 11, pp. 25-45.

Castro, R., A. Tierra and M. Luna (2019). "Assessing the Horizontal Positional Accuracy in OpenStreetMap: A Big Data Approach." World Conference on Information Systems and Technologies

Chehreghan, A. and R. Ali Abbaspour (2017). "A new descriptor for improving geometric-based matching of linear objects on multi-scale datasets." GIScience & Remote Sensing, Vol. 54, No. 6, pp. 836-861.

Chehreghan, A. and R. Ali Abbaspour (2018). "A geometric-based approach for road matching on multi-scale datasets using a genetic algorithm." Cartography and Geographic Information Science, Vol. 45, No. 3, pp. 255-269.

Chehreghan, A. R. and R. Ali Abbaspour (2017). "Improvement of Geometric-Based Roads Matching on Multi-Scale Data Structures." Journal of Geomatics Science and Technology, Vol. 6, No. 3, pp. 89-102.

Chrisman, N. R. (1991). "The error component in spatial data." Geographical information systems, Vol. 1, No., pp. 165-174.

de Berg, M., M. van Kreveld, M. Overmars and O. C. Schwarzkopf (2000). "Computational Geometry. Computational Geometry: Algorithms and Applications." Berlin, Heidelberg, Springer Berlin Heidelberg**:** pp. 1-17.

Dehghani, A., A. Chehreghan and R. Ali Abbaspour (2017). "Matching of urban pathways in a multi-scale database using fuzzy reasoning." Geodesy and Cartography, Vol. 43, No. 3, pp. 92-104.

Fan, H., B. Yang, A. Zipf and A. Rousell (2016). "A polygon-based approach for matching OpenStreetMap road networks with regional transit authority data." International Journal of Geographical Information Science, Vol. 30, No. 4, pp. 748-764.

Fan, H., A. Zipf, Q. Fu and P. Neis (2014). "Quality assessment for building footprints data on OpenStreetMap." International Journal of Geographical Information Science, Vol. 28, No. 4, pp. 700-719.

Koukoletsos, T., M. Haklay and C. Ellul (2011). "An automated method to assess data completeness and positional accuracy of OpenStreetMap." GeoComputation

Koukoletsos, T., M. Haklay and C. Ellul (2012). "Assessing data completeness of VGI through an automated matching procedure for linear data." Transactions in GIS, Vol. 16, No. 4, pp. 477-498.

Kounadi, O. (2009). "Assessing the quality of OpenStreetMap data." Msc geographical information science, University College of London Department of Civil, Environmental And Geomatic Engineering.

Kulin, H. W. and R. E. Kuenne (1962). "An efficient algorithm for the numerical solution of the generalized Weber problem in spatial economics." Journal of Regional Science, Vol. 4, No. 2, pp. 21-33.

Lauriault, T. and P. Mooney (2014). "Crowdsourcing: A geographic approach to public engagement." Available at SSRN 2518233, Vol 3. No. 5, pp 23 - 46.

Ludwig, I., A. Voss and M. Krause-Traudes (2011). "A Comparison of the Street Networks of Navteq and OSM in Germany." Advancing geoinformation science for a changing world, Springer**:** pp. 65-84.

Lyu, H., Y. Sheng, N. Guo, B. Huang and S. Zhang (2017). "Geometric quality assessment of trajectory-generated VGI road networks based on the symmetric arc similarity." Transactions in GIS, Vol. 21, No. 5, pp. 984-1009.

Mashhadi, A., G. Quattrone and L. Capra (2015). "The impact of society on volunteered geographic information: The case of OpenStreetMap." OpenStreetMap in GIScience, Springer**:** pp. 125-141.

Forghani, M. and M. R. Delavar (2014). "A quality study of the OpenStreetMap dataset for Tehran." ISPRS International Journal of Geo-Information, Vol. 3, No. 2, pp. 750-763.

Girres, J. F. and G. Touya (2010). "Quality assessment of the French OpenStreetMap dataset." Transactions in GIS, Vol. 14, No. 4, pp. 435-459.

Goodchild, M. F. (2007). "Citizens as sensors: the world of volunteered geography." GeoJournal, Vol. 69, No. 4, pp. 211-221.

Goodchild, M. F. and G. J. Hunter (1997). "A simple positional accuracy measure for linear features." International journal of geographical information science, Vol. 11, No. 3, pp. 299-306.

Haklay, M. (2010). "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets." Environment and planning B: Planning and design, Vol. 37, No. 4, pp. 682-703.

Haklay, M., S. Basiouka, V. Antoniou and A. Ather (2010). "How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information." The Cartographic Journal, Vol. 47, No. 4, pp. 315-322.

Hecht, R., C. Kunze and S. Hahmann (2013). "Measuring completeness of building footprints in OpenStreetMap over space and time." ISPRS International Journal of Geo-Information, Vol. 2, No. 4, pp. 1066-1091.

Jackson, S. P., W. Mullen, P. Agouris, A. Crooks, A. Croitoru and A. Stefanidis (2013). "Assessing completeness and spatial error of features in volunteered geographic information." ISPRS International Journal of Geo-Information, Vol. 2, No. 2, pp. 507-530.

McMaster, R. B. (1986). "A statistical analysis of mathematical measures for linear simplification." The American Cartographer, Vol. 13, No. 2, pp. 103-116.

Meier, J. C. (2015). "An Analysis of Quality for Volunteered Geographic Information." Vol. 4 No. 2, pp 102-125.

Mohammadi, N. and M. Malek (2015). "Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data." Journal of Spatial Science, Vol. 60, No. 1, pp. 119-135.

Nasiri, A., R. Ali Abbaspour, A. Chehreghan and J. Jokar Arsanjani (2018). "Improving the Quality of Citizen Contributed Geodata through Their Historical Contributions: The Case of the Road Network in OpenStreetMap." ISPRS International Journal of Geo-Information, Vol. 7, No. 7, pp. 253.

Neis, P. and D. Zielstra (2014). "Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap." Future Internet, Vol. 6, No. 1, pp. 76-106.

Rassafi, A., F. Momeni, Z. Astaraki and B. Amini (2012). "The application of multi-criteria decision-making analysis to identify hot spots: data envelopment analysis and concordance analysis." Quarterly Journal of Transportation Engineering, Vol. 3, No. 4, pp. 301-314.

Sheeren, D., S. Mustière and J.-D. Zucker (2004). "How to Integrate Heterogeneous Spatial Databases in a Consistent Way?". East European Conference on Advances in Databases and Information Systems, Berlin, Heidelberg.

Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region." Economic geography, Vol. 46, No. 1, pp. 234-240.

Wang, M., Q. Li, Q. Hu and M. Zhou (2013). "Quality analysis of open street map data." ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci, Vol. No., pp. 155-158.

Yan, H. (2010). "Fundamental theories of spatial similarity relations in multi-scale map spaces." Chinese Geographical Science, Vol. 20, No. 1, pp. 18-22.

Zhang, H. and J. Malczewski (2017). "Quality evaluation of volunteered geographic information: The case of OpenStreetMap." Volunteered geographic information and the future of geospatial data, IGI Global: 19-46.

Zhang, H. and J. Malczewski (2018). "Accuracy evaluation of the Canadian OpenStreetMap road networks." International Journal of Geospatial and Environmental Research, Vol. 5, No. 2, pp.

Zhang, M. (2009). "Methods and implementations of road-network matching." doctor engineer, Technische Universität München.

Zhang, W.B., Y. Leung and J.H. Ma (2019). "Analysis of positional uncertainty of road networks in volunteered geographic information with a statistically defined buffer-zone method." International Journal of Geographical Information Science, Vol. 33, No. 9, pp. 1807-1828.

Zheng, S. and J. Zheng (2014). "Assessing the Completeness and Positional Accuracy of OpenStreetMap in China." Thematic Cartography for the Society. Cham, Springer International Publishing: 171-189.

Zhou, P., W. Huang and J. Jiang (2014). "Validation analysis of OpenStreetMap data in some areas of China." The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 40, No. 4, pp. 383.

Zhou, Q. (2017). "Rethinking the buffering approach for assessing openstreetmap positional accuracy." International Cartographic Conference, Springer.

افندی زاده زرگری, ش., بیگدلی, ح., و شاکر, ح. (۲۰۱۹). "استفاده از روش بهینه‌سازی و فرا ابتکاری به‌منظور کاهش سرفاصله زمانی اتوبوس (نمونه موردی: خطوط اتوبوس‌رانی قزوین)." فصلنامه مهندسی حمل و نقل, سال۱۰, دوره ۴, ص, ۸۴۹-۸۳۳.

تیموریان, ف., آل شیخ, ع. ا., علیمحمدی, ع., و صادقی نیارکی, ا. (۲۰۱۹). "توسعه یک سامانه اطلاع‌رسانی و ارزیابی عملکرد اتوبوسرانی شهری با به-کارگیری داده‌های مکانی مردم گستر." فصلنامه مهندسی حمل و نقل, سال ۶, دوره ۲, ص. ۲۲۵-۲۳۶.

چهرقان, ع., و علی عباس پور, ر. (۲۰۱۸). "ارزیابی شباهت هندسی داده های مکانی داوطلبانه در شبکه معابر درون شهری." فصلنامه مهندسی حمل و نقل, سال ۱۰, دوره ۲, ص. ۳۵۷-۳۷۰.

خمر, غ., پاسبان عیسی‌لو, و., و مژگان, ن. (۲۰۱۷). "مطالعه تطبیقی الگوریتم اجتماع مورچگان و ژنتیک در مسیریابی بهینه (مطالعه موردی: شهر پارس آباد و حومه)." فصلنامه مهندسی حمل و نقل, سال ۸, دوره ۳, ص. ۳۸۹-۴۰۴.

ذوالفقاری, ا., و کرکه آبادی, ز. (۲۰۱۹). "مسیریابی هوشمند اکیپ های امدادی با استفاده از الگوریتم تئوری بازی ها نمونه موردی:شهر سمنان." فصلنامه مهندسی حمل و نقل, سال ۵, دوره ۱, ص ۳۲-۱۹.

شاکری, م., علیمحمدی, ع., صادقی نیارکی, ا., و آل شیخ, ع. ا. (۲۰۱۹). "طراحی سیستمی برای حمل و نقل جاده ای با استفاده از تلفیق اطلاعات مکانی داوطلبانه و زیرساخت داده مکانی. فصلنامه مهندسی حمل و نقل".

رحیم علی‌عباسپور، درجه کارشناسی در رشته مهندسی نقشه‌برداری را در سال ۱۳۷۹ از دانشگاه تهران و درجه کارشناسی ارشد در رشته مهندسی نقشه برداری - سیستم‌های اطلاعات مکانی را در سال ۱۳۸۱ از دانشگاه تهران اخذ نمود. در سال ۱۳۸۹ موفق به کسب درجه دکتری سیستم‌های اطلاعات مکانی از دانشگاه تهران گردید. زمینه های پژوهشی مورد علاقه ایشان داده کاوی مکانی-زمانی، بهینه‌سازی مکانی، اطلاعات مکانی داوطلبانه، خدمات مکان‌مبنا (LBS) و محاسبات هندسی بوده و در حال حاضر عضو هیات علمی با مرتبه استادیار در پردیس دانشکده‌های فنی دانشگاه تهران است.

علیرضا چهرقان در سال ۱۳۸۸رشته کارشناسی مهندسی نقشه‌برداری را از دانشگاه تهران اخذ نمود. همچنین کارشناسی ارشد و دکتری خود را به ترتیب در سال های ۱۳۹۰ و ۱۳۹۶ از دانشگاه تهران دریافت کرد. زمینه‌های پژوهشی مورد علاقه ایشان شامل تناطریابی عوارض در پایگاه های داده مکانی، اطلاعات مکانی داوطلبانه، داده کاوی مکانی-زمانی، تصمیم گیری های مکان مبنا، بهینه سازی مکانی، خدمات مکان مبنا و محاسبات هندسی می باشد. در حال حاضر ایشان عضو هیات علمی با مرتبه استادیار در دانشگاه صنعتی سهند است.

مسعود یزدان‌نیک درجه کارشناسی را از مهندسی عمران گرایش نقشه‌برداری را در سال ۱۳۹۶ از دانشگاه اصفهان اخذ نمود و در حال حاضر دانشجو کارشناسی ارشد سیستم‌های اطلاعات مکانی در دانشگاه تهران است. زمینه پژوهشی مورد علاقه ایشان اطلاعات مکانی داوطلبانه، ارزیابی کیفیت داده‌های برداری، محاسبات هندسی و کلان داده است.