نشریه علمی علوم و فنون نقشه برداری، دوره نهم، شماره ۴، خرداد ماه ۱۳۹۹ - مقاله پژوهشی

مقایسه تحلیلی روش‌های محاسبه کامل بودن اطلاعات مکانی داوطلبانه

مسعود یزدان‌نیک[۱]، رحیم علی‌عباسپور*[۲]، علیرضا چهرقان[۳]

[۱]دانشجوی کارشناسی ارشد سیستم‌های اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران
yazdannik.masoud@ut.ac.ir

[۲]استادیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی - پردیس دانشکده‌های فنی - دانشگاه تهران
abaspour@ut.ac.ir

[۳]استادیار دانشکده مهندسی معدن - دانشگاه صنعتی سهند تبریز
chehreghan@sut.ac.ir

(تاریخ دریافت فروردین ۱۳۹۸، تاریخ تصویب بهمن ۱۳۹۸)

چکیده

محدودیت و فاکتور هزینه در دسترسی به داده‌های مکانی، در سال‌های پیش باعث شد تا کاربران و محققان این داده‌ها محدود به افراد خاص باشد. ازاین‌رو معرفی داده‌های مکانی با دسترسی آزاد و رایگان برای تمامی کاربران، منجر به افزایش محبوبیت اطلاعات مکانی داوطلبانه و به‌صورت خاص OSM به‌عنوان مجموعه داده تکمیلی و یا منبع جایگزین داده‌های رسمی شد. از زمان ایجاد این پروژه تاکنون، یکی از مهم‌ترین چالش‌ها در استفاده ازاین‌گونه داده‌ها بحث تضمین کیفیت آن بوده است. در میان پارامترهای ارزیابی کیفیت داده‌های مکانی بحث کامل‌بودن مجموعه داده مورد بررسی به‌واسطه تأثیر آن بر پارامترهای دیگر ارزیابی کیفیت، همیشه مورد توجه ویژه بوده است. تاکنون در این راستا رویکردهای گوناگونی در تحقیقات مختلف جهت بررسی کامل بودن داده‌های مکانی داوطلبانه ارائه شده است که هر یک از این روش‌ها دارای نقاط قوت و محدودیت‌های خاص خود هستند. در این مقاله تلاش می‌گردد پارامتر کامل بودن اطلاعات مکانی داوطلبانه از طریق روش‌های عارضه‌مبنا و منطقه‌مبنا ارزیابی شده و نتایج آن مورد ارزیابی و بررسی قرار گرفته شود و نقاط ضعف و قوت هر روش در شرایط گوناگون بررسی گردد. نتایج نشان می‌دهد در صورتی بهره‌گیری از روش‌های منطقه‌مبنا، منجر به ارزیابی صحیحی از کامل‌بودن خواهد شد، که مجموعه داده OSM دارای کاربری‌های یکسان با مجموعه داده رسمی و عاری از خطا توپولوژیک باشد. درحالی‌که مزیت روش‌های عارضه‌مبنا، در عین وجود محاسبات بالا و زمان‌بر، رسیدن به دقت بالاتر به عنوان کیفیت داده‌ها است، البته در صورتی که صحت مکانی داده‌ها تضمین شود. با جمع‌بندی بر نتایج حاصل‌شده، توصیه این مقاله در مطالعات ارزیابی کیفیت، استفاده از روش‌های عارضه‌مبنا است.

واژگان کلیدی: اطلاعات مکانی داوطلبانه، ارزیابی کیفیت، کامل‌بودن، روش‌های عارضه مبنا، روش‌های منطقه مبنا

* نویسنده رابط

۱- مقدمه

با راه‌اندازی OSM[۱] در سال ۲۰۰۴ میلادی رویکرد جدیدی در جمع‌آوری و استفاده از داده‌های مکانی ایجاد شد [۱]. در این نوع از فرآیند جمع‌آوری داده، به کاربران علاوه بر استفاده از این داده‌ها قابلیت‌های جدیدی از جمله تولید و ویرایش نیز داده شد [۲]. پیشرفت‌های جدید در حوزه تکنولوژی نظیر فراگیر شدن وسایل هوشمند مجهز به سیستم موقعیت‌یابی جهانی (GPS[۲]) در میان کاربران عادی و دسترسی این افراد به سرویس‌های ارائه تصاویر ماهواره‌ای با قدرت تفکیک مکانی بالا تحت وب نیز باعث قدرتمند شدن این پروژه و محبوبیت آن شد. تا این میزان که در زمان نگارش این مقاله تعداد ۵۱۹۸۷۷۱ کاربر در این پروژه ثبت‌نام کرده و بالغ بر ۵۶۳۱۹۷۵۹۵ عارضه توسط این افراد در این پروژه ثبت شده است.

برای اولین بار در سال ۲۰۰۷ میلادی، گودچایلد این نوع جدید از اطلاعات مکانی را اطلاعات مکانی داوطلبانه (VGI[۳]) نامید. از نظر وی، هر شخص در محیط پیرامون خود می‌تواند مانند یک حسگر جهت جمع‌آوری داده‌های مکانی عمل نماید [۳]. ازاین‌رو می‌توان این انتظار را داشت که برای مناطق با جمعیت زیاد مانند مناطق شهری، داده‌های مکانی با پوشش بالا و سطح جزئیات مناسب وجود داشته باشد. همچنین با توجه به این‌که جامعه تولیدکننده اطلاعات مکانی داوطلبانه مردم عادی هستند، حجم انبوهی از اطلاعات هرروزه در حال تولید است که باعث می‌شود مجموعه داده موردنظر همیشه خاصیت به‌روز بودن خود را حفظ نماید. از سوی دیگر، ویژگی متمایزکننده این مجموعه داده نسبت به داده‌های رسمی مسئله دسترسی رایگان به این داده‌ها به دلیل مردمی بودن و عدم وابستگی به سازمان تجاری خاص است [۴].

تا این قسمت از مقاله تنها به نکات مثبت این پروژه اشاره گردید ولی در مورد داده‌های ثبت شده در این پایگاه داده باید به این نکته توجه داشت که داده‌های این سرویس از منابع مختلف داده نظیر داده‌های جمع‌آوری شده از طریق GPS، رقومی سازی تصاویر ماهواره‌ای و یا عوارض وارد شده از منابع رسمی تشکیل شده‌اند [۱].

در این صورت پارامترهای کیفیت برای کل منطقه برخلاف داده‌های رسمی یکسان نبوده و ناهمگونی میان این پارامترها در سراسر منطقه مشاهده می‌گردد [۵]. با در نظر گرفتن این ناهمگونی در کیفیت و عدم وجود استاندارد و چارچوب‌های پیشگیرانه در جلوگیری از ورود داده‌های باکیفیت پایین در این مجموعه داده و همچنین نیاز به اطلاعات مکانی با صحت بالا در برخی از کاربردها، لزوم تعیین کیفیت این داده‌ها قبل از به‌کارگیری در هر کاربرد احساس می‌شود [۵]. درزمینه کیفیت داده‌های مکانی سازمان بین‌المللی استاندارد ISO/TC211، استاندارد ISO 19157 را تعریف کرده است. در این استاندارد عناصر مختلفی نظیر صحت مکانی، کامل‌بودن، همگونی منطقی، قدرت تفکیک مکانی به عنوان عناصر تعیین کیفیت داده‌های مکانی شناخته شده است [۱].

هر یک از المان‌های استاندارد ISO درارزیابی کیفیت، موضوع‌های گسترده‌ای هستند که با توجه به محدودیت مطالعه، پرداختن به تمامی این عناصر ممکن نخواهد بود. از این‌رو پارامتر کامل‌بودن، که یکی از مهمترین عناصر کیفیت ذکر شده به شمار می‌رود به واسطه اهمیت آن و تاثیر بر معیارهای دیگر [۶]، هدف بررسی این بخش در نظر گرفته شده است. نمونه‌ای از اهمیت معیار ذکر شده را می‌توان در یکی از معمول‌ترین کاربردهای اطلاعات مکانی، مسیریابی دید. چنانچه مجموعه داده ورودی به الگوریتم مسیریابی دارای اطلاعات بدون نظیر در دنیا واقعی و یا نقص اطلاعات باشد، در این صورت به احتمال بالا مسیریابی انجام شده مستعد خطا خواهد بود.

کامل‌بودن یک مجموعه داده مکانی ارتباط میان اشیا درون پایگاه داده را با نمونه واقعی اشیا در دنیا واقعی بررسی می‌کند [۷, ۸]. علاوه‌بر ماهیت هندسی اشیا، کامل‌بودن دامنه اطلاعات توصیفی عوارض و ارتباطات میان عوارض را نیز پوشش می‌دهد که در استاندارد ISO این معیار به دو زیر شاخه omission (اطلاعات رسمی بدون نظیر در مجموعه داده داوطلبانه) و commission (اطلاعات داوطلبانه بدون نظیر در مجموعه داده رسمی) تقسیم شده است [۹]. رویکردهای متنوعی در مطالعات پیشین که در بخش پیش‌رو به آن پرداخته شده است برای ارزیابی کامل‌بودن داده‌ها به کار گرفته شده‌اند. در حالت کلی این روش‌ها براساس واحد محاسباتی قابل طبقه‌بندی به دو دسته روش‌های منطقه‌مبنا و عارضه‌مبنا هستند.

---

۱ Open Street Map
۲ Global Positioning System
۳ Volunteered Geographic Information

طیف گسترده روش‌های ارزیابی کامل‌بودن و حساسیت این روش‌ها به برخی از خطاها در مجموعه داده و نیاز به بکارگیری پیش پردازش‌های خاص به منظور حذف این خطاها هدف اصلی پژوهش انجام شده است، چراکه در صورت نادیده‌گیری موارد مطرح شده در هر معیار، باتوجه به شرایط مجموعه داده، ممکن است نتایج دارای اعتبار نباشد. علاوه‌بر مقایسه روش‌های ارزیابی کامل‌بودن، نتایج این مطالعه می‌تواند دید روشنی در خصوص کامل‌بودن داده‌های OSM مناطق مطالعه شده تهران، برای خوانندگان ایجاد نماید.

قسمت مشترک میان تمامی روش‌های ارزیابی کیفیت در این مطالعه، نیاز به مجموعه داده رسمی در عملیات مقایسه است. بر همین اساس انتخاب مجموعه داده رسمی مناسب، گامی مهم در فرآیند ارزیابی است. مناطق ۱۱ تا ۱۴ تهران به سبب در دسترس بودن اطلاعات داوطلبانه و رسمی، جهت تحلیل‌های بعدی انتخاب شده‌اند. در خصوص مجموعه داده رسمی لازم به اشاره است، مقیاس این مجموعه داده ۱:۲۰۰۰ و گردآوری آن از سازمان شهرداری تهران انجام شده است.

در ادامه، این مقاله در سه قسمت اصلی ارائه شده است. در بخش دوم این مقاله به توضیح روش‌های ارائه‌شده در پژوهش‌های مختلف به همراه روابط آن‌ها پرداخته شده است. سپس بخش سوم این مقاله به محاسبه معیارهای ارائه شده قسمت قبل و مقایسه نتایج به دست آمده و نقاط ضعف و قوت هر یک از این روش‌ها می‌پردازد. در قسمت پایانی این مقاله با مقایسه جامع بر روی تمام نتایج حاصل از قسمت قبل به جمع‌بندی و نتایج حاصل از تحقیق پرداخته می‌شود.

۲- پیشینه تحقیق

در مطالعه ژانگ و همکاران در بررسی پارامترهای ارزیابی کیفیت در مناطق مختلف جهان نتیجه‌گیری شد که از میان ۶۰ مطالعه ارزیابی کیفیت داده‌های OSM در حدود ۶۰ درصد به موضوع کامل‌بودن داده‌های داوطلبانه پرداخته بودند [۱۰]. دلیل توجه زیاد به این معیار ارزیابی کیفیت، اهمیت و تاثیر آن بر سایر معیارهای ارزیابی کیفیت است. مطالعاتی که در ادامه این بخش ارائه شده‌اند بخشی از مهم‌ترین مطالعات با رویکرد ارزیابی کامل‌بودن مجموعه داده داوطلبانه مناطق مختلف جهان هستند.

زیلسترا و زیپف در ابتدای سال ۲۰۱۰ تفاوتی ۷ درصدی میان مجموع طول شبکه راه‌ها در پایگاه داده OSM و مجموع طول این عوارض در داده‌های رسمی TeleAtlas محاسبه کردند. همچنین در شهرهای بزرگ به دلیل تراکم جمعیتی بالاتر برخلاف مناطق پراکنده روستایی روند رشد در معیار کامل‌بودن کاملاً محسوس است. حتی برخی از نقاط شهری شامل اطلاعات بیشتری نسبت به مجموعه داده رسمی است [۱۱].

جهت ارزیابی پارامتر کامل‌بودن داده‌های مکانی داوطلبانه کشور فرانسه، گیرس و تویا در سال ۲۰۱۰ علاوه بر بررسی کامل‌بودن شبکه عوارض مرتبط به معابر شهری، کامل‌بودن نوع عوارض دیگر (نقطه - چندضلعی) را مورد بررسی قرار دادند. نتایج این بررسی نشان داد، تعداد عوارض و میزان مشارکت عوارض داوطلبانه در داخل یک منطقه به وضوح با میزان تعداد کاربران فعال در آن ناحیه مرتبط است [۱۲].

نیس و همکاران (۲۰۱۱) تأکید کردند که ارتباط قوی میان تراکم جمعیت و میزان کامل‌بودن وجود دارد. با مقایسه دو مجموعه داده رسمی و داوطلبانه در سال‌های ۲۰۰۷ تا ۲۰۱۰ نشان داده شد، شکاف اطلاعاتی میان داده-های داوطلبانه و رسمی نهایتا در سال ۲۰۱۲ به صفر خواهد رسید [۱۳] ولی در مورد اطلاعات مربوط به قوانین و محدودیت‌های حرکتی (کاربردهای مسیریابی) حداقل بازه زمانی ۵ ساله جهت حذف شکاف اطلاعاتی مورد نیاز است.

لودویگ و همکاران (۲۰۱۱) پارامتر کامل‌بودن را برای مجموعه داده داوطلبانه آلمان، در نواحی مختلف شهری نسبت به داده رسمی مورد بررسی و ارزیابی قرار دادند. ارزیابی کامل‌بودن مجموعه داده داوطلبانه آلمان سطح بالایی از کامل‌بودن را نشان می‌دهد ولی در سطح منطقه-ای رفتار این کمیت متغیر است به طوری که در مناطق پرجمعیت شهری، این کمیت به میزان ۹۷ درصد و در مناطق خالی از سکنه ۱۷ درصد گزارش شده است [۱۴].

مقاله کوکلتسوس و همکاران در سال ۲۰۱۲ نیز به‌نوعی توسعه مقاله هاکلای و همکاران (۲۰۱۰) است. با توجه به این‌که در مقاله هاکلای و همکاران (۲۰۱۰) جهت تعیین معیارهای کامل‌بودن مجموعه داده موردنظر، عوارض نظیر به‌صورت دستی انتخاب‌شده‌اند، با افزایش وسعت منطقه موردمطالعه و به‌روزرسانی‌های زیاد عملاً استفاده از این روش غیرممکن می‌گردد. به همین دلیل در این مقاله

الگوریتمی جهت تناظر یابی عوارض معرفی‌شده است و پارامترهای کامل‌بودن برای منطقه پیشنهادی شهر لیورپول با مساحت ۱۸۷۰ کیلومترمربع محاسبه‌شده است [۸].

تا پیش از سال ۲۰۱۳ هدف مطالعات VGI تنها ارزیابی و مقایسه کیفیت در مناطق مختلف جهان بود. شناسایی کامل‌بودن عوارض چندضلعی به عنوان نوعی از عوارض برداری و اهمیت این عوارض در برنامه‌ریزی‌های شهری و مدیریت بحران، انگیزه اصلی هچ و همکاران در انجام مطالعه جدید بود. مشابه با رویکرد بهره‌گرفته شده در مطالعه پیش‌رو، کامل‌بودن عوارض چندضلعی توسط دو دسته اصلی روش‌ها (منطقه‌مبنا و عارضه‌مبنا) ارزیابی شد که بر پایه نتیجه‌گیری حاصل شده روش‌های عارضه‌مبنا هرچند نیاز به محاسبات کمتری داشته ولی در نمایش جزئیات با محدودیت روبه‌رو هستند. یکی از محدودیت-های اساسی روش منطقه‌مبنا وابستگی بسیار زیاد به شیوه مدل‌سازی داده‌های داوطلبانه بوده و گاها منجر به تولید نتایج غیرواقعی به‌عنوان مقدار کامل‌بودن می‌گردد [۱]. تفاوت مطالعه هچ و مطالعه کنونی در نوع عوارض بررسی شده و مقایسه معیارها نسبت به یکدیگر است. در مطالعه هچ تنها روش‌های منطقه‌مبنا و عارضه‌مبنا نسبت به هم مقایسه شده است در حالی که در این مطالعه علاوه بر بررسی این دو رویکرد، روش‌های ارزیابی هر دسته نیز نسبت به یکدیگر ارزیابی شده‌اند.

فرقانی و دلاور (۲۰۱۴) از رویکرد منطقه‌مبنا برای ارزیابی میزان کامل‌بودن داده‌های داوطلبانه شهر تهران بهره گرفته-اند. پس از تقسیم ناحیه مورد بررسی به نواحی کوچک‌تر از پارامتر نسبت مجموع طول عوارض در هر دو ناحیه جهت تعیین میزان کامل‌بودن داده‌های مکانی داوطلبانه هر سلول به داده رسمی متناظر استفاده شده است [۱۵].

مشهدی و همکاران نیز مشابه با سایر محققان پس از بررسی مجموعه داده داوطلبانه شهر لندن در سال ۲۰۱۵، توزیع غیر یکنواخت کامل‌بودن در سراسر منطقه بررسی شده را تایید کردند. براساس نظر نویسنده، رفتار غیر یکنواخت کامل‌بودن با فاکتورهای اجتماعی جامعه نظیر تراکم جمعیت، فاصله از مراکز اصلی شهر و اوضاع اقتصادی متناسب است، به‌طوری‌که افزایش ۵۰ نفر در هکتار متناظر با ۲۵ درصد رشد کامل‌بودن و کاهش فاصله ۵ کیلومترنسبت به مرکز شهر، معادل ۲۸ درصد افزایش کامل‌بودن است [۱۶].

مشابه با یافته‌های حاصل شده در مطالعه مشهدی و همکاران، کامبویم نیز به نتایج یکسانی در مورد تاثیر جمعیت و فاصله از مراکز شهری، بر کامل‌بودن داده‌های مکانی داوطلبانه کشور برزیل دست یافت. دیدگاه نویسنده پس از تحلیل کامل‌بودن شهر مختلف بر این است که شهرهای پرجمعیت در نزدیکی مراکز شهرهای مهم، بهتر مورد مشارکت واقع شده‌اند [۱۷].

الگو ناهمگون پارامترهای کیفیت در شهرهای مختلف محدود به نوع خاصی از عوارض داوطلبانه نیست. بررسی ۷۸ منطقه شهر در آمریکا با تمرکز بر عوارض داوطلبانه دارای کاربری مختص دوچرخه نشان داد کامل‌بودن این عوارض نسبت به داده‌های رسمی از ۰ تا ۹۴ درصد متغیر است [۱۸].

با توجه به بررسی‌ها و آنالیزهای کم انجام شده بر روی مجموعه داده داوطلبانه کانادا توسط تنی (۲۰۱۴) [۱۹] و میر (۲۰۱۵) [۲۰]، این‌بار ژانگ و همکاران در سال ۲۰۱۸ به بررسی پارامترهای کیفیت در این منطقه پرداختند. یافته‌های این مطالعه نشان می‌دهد بر خلاف مناطق شهری تراکم عوارض موجود در مجموعه داده رسمی در مناطق روستایی و حاشیه شهر بیشتر از مناطق شهری است. الگو تراکم متفاوت برخی از مناطق روستایی براساس نظر نویسندگان این مقاله به واسطه داده‌های رسمی است که در سال ۲۰۰۸ به مجموعه داده داوطلبانه افزوده شده است [۲۱].

در مقاله چهرقان و علی‌عباسپور (۲۰۱۸) پس از تناظریابی عوارض ۲۲ گانه تهران کامل بودن مجموعه داده داوطلبانه نسبت به مجموعه رسمی محاسبه شده است. بررسی روند تغییرات مشارکت‌های انجام‌شده در سال‌های ۲۰۱۳ تا ۲۰۱۷ رشد ۸۷/۲ درصدی میزان مشارکت کاربران را نمایش می‌دهد که موجب محاسبه نرخ بالای کامل‌بودن عوارض داوطلبانه به میزان ۸۷ درصد شده است. [۲۲].

دیدگاه بیشتر مطالعات ارزیابی کیفیت در سراسر جهان، پایین بودن پارامترهای مشارکتی و میزان کامل-بودن داده‌های داوطلبانه کشورهای درحال توسعه به توسعه یافته است. ژو در سال ۲۰۱۹ کامل‌بودن عوارض داوطلبانه چین را مورد بررسی قرار داد. نتایج تجزیه و تحلیل نشان داد، مجموعه داده داوطلبانه چین به حد کافی کامل نیست ولی در پاسخگویی به برخی از کاربرد ها مناسب است. همچنین در مورد کامل‌بودن عوارض به تفکیک نوع آن‌ها نظر نویسنده بر این است: پایین‌ترین نرخ میزان مشارکت متعلق به خیابان های تک‌باند، بدون نام و دارای کاربری خصوصی است [۲۳].

در تمامی مطالعات اشاره شده در این بخش، کلیه نتایج، قابل خلاصه‌سازی در مفاهیمی همچون بهبود کامل‌بودن در اثر گذشت زمان و الگوی مناسب‌تر کامل‌بودن مناطق شهری نسبت به مناطق حاشیه شهر قابل خلاصه‌سازی است. ولی در هیچ یک از مطالعات، وابستگی معیارهای معرفی شده به حالات خاص و پیش‌پردازش‌های ضروری پیش از بکارگیری این معیارها پرداخته نشده است. اهمیت پیش‌پردازش‌های ذکر شده در برخی حالات به میزانی است که ممکن است نتایج حاصل‌شده از معیار ارزیابی صحیح نباشد. از این‌رو در نگارش مطالعه کنونی سعی شده است به مسائلی از این دست پرداخته شده باشد.

۳- کامل‌بودن

روش‌های معرفی‌شده در مقالات پیشین با توجه به واحد محاسباتی آن‌ها قابل دسته‌بندی در دو گروه هستند. گروه اول تحت عنوان روش‌های منطقه‌مبنا و گروه دوم تحت عنوان روش‌های عارضه‌مبنا شناخته می‌شوند.

۳-۱- روش‌های منطقه‌مبنا

همان‌گونه که در قسمت مقدمه اشـاره شد برخلاف داده‌های رسمی که پارامترهای کیفیت در آن‌ها در سراسر مجموعه داده ثابت و مشخص است، در مجموعه داده داوطلبانه پارامترهای کیفیت به صورت ناهمگون در سراسر منطقه تغییر می‌نماید. ازاین‌رو ارزیابی پارامتر کیفیت برای کل ناحیه مورد بررسی نمی‌تواند میزان قابل قبولی باشد. رویکرد منطقه‌مبنا نیز در اصل برای حل این مشکل معرفی شده است. در داده‌های OSM با در نظر گرفتن این موضوع که بیشتر کاربران داده‌های مناطق محلی و تحت شناخت خود را ویرایش می‌کنند، ازاین‌رو می‌توان انتظار داشت که کیفیت این داده‌ها در هر منطقه خاصیت به نسبت همگنی داشته باشد. یعنی انتظار می‌رود که معیارهای کیفیت ارزیابی شده برای هر یک از زیر نواحی به مقدار واقعی این کمیت‌ها نزدیک باشد.

از مزایای به‌کارگیری این دسته از روش‌ها می‌توان به محاسبات کمتر و سرعت پردازش بالاتر آن‌ها اشاره نمود. با این وجود این روش‌ها در نمایش جزئیات محدود بوده و میان میزان سطح نمایش جزئیات و اندازه سلول انتخاب شده رابطه مستقیم دیده می‌شود [۱].

در انتخاب اندازه زیر نواحی توصیه می‌گردد همیشه جانب تعادل رعایت شود چراکه در صورت انتخاب این مقدار به اندازه زیاد برآورد دقیقی از پارامترهای کیفیت امکان‌پذیر نخواهد بود و در حالت مقابل سادگی محاسباتی این دسته از روش‌ها در مقابل روش‌های عارضه‌مبنا از بین خواهد رفت. در تمامی موارد انجام شده تعیین این میزان نیز به‌صورت تجربی انجام گرفته و رابطه ریاضی مشخصی برای آن وجود ندارد. در مقالات مختلف نیز این میزان با در نظر گرفتن تراکم شبکه راه‌ها و وسعت منطقه، در بازه عددی ۲۰۰ تا ۲۰۰۰ متر انتخاب شده است [۱۵, ۲۱, ۲۴-۲۶].

۳-۱-۱- معیار مجموع طول

برای محاسبه‌ی کامل‌بودن یک مجموعه داده نسبت به مجموعه دیگر، در بیشتر مقالات از شاخص نسبت یا تفاضل مجموع طول شبکه راه‌ها در دو مجموعه داده استفاده شده است [۱۰, ۱۷, ۲۱, ۲۷]. به‌منظور تعیین پارامتر کامل‌بودن داده‌های داوطلبانه نسبت به داده‌های رسمی پس از محاسبه مجموع طول عوارض در هر ناحیه از طریق رابطه (۱) قابل محاسبه است[۲۸].

$$C = \frac{\sum \text{VGI length}}{\sum \text{Refrence length}} \qquad (۱)$$

در این رابطه C مقدار کامل بودن، ∑VGI length مجموع طول شبکه عوارض مجموعه داده داوطلبانه و ∑Refrence length مجموع طول شبکه عوارض مجموعه داده رسمی متناظر است.

۳-۱-۲- معیار تعداد نقاط

در تعیین کامل‌بودن داده‌های نقطه‌ای روش مرسوم، بررسی نسبت تعداد عوارض نقطه‌ای مجموعه داده مورد بررسی به رسمی است. در صورت کاهش بعد عوارض از نوع خط به نقطه‌ای امکـان به کارگیـری این روش برای داده‌های خطی وجود خواهد داشت (شکل ۱).

شکل ۱- استخراج نقاط سازنده شبکه معابر برای قسمتی از مجموعه داده مورد مطالعه

این رویکرد در پژوهش میر (۲۰۱۵) جهت ارزیابی کامل‌بودن داده‌های داوطلبانه به کار گرفته شده است [۲۰]. در این روش پس از تقسیم ناحیه مورد بررسی به زیر نواحی کوچکتر، کلیه نقاط سازنده خطوط (نقاط ابتدائی، انتهائی و نقاط شکستگی) استخراج می‌گردد. سپس کامل‌بودن داده‌های داوطلبانه نسبت به مجموعه رسمی از طریق رابطه (۲) محاسبه می‌گردد.

$$C = \frac{\sum \text{VGI point number}}{\sum \text{Refrence point number}} \qquad (۲)$$

در این رابطه $\sum$VGI point number مجموع تعداد نقاط داوطلبانه و $\sum$Refrence point number مجموع تعداد نقاط رسمی است.

۳-۱-۳- معیار تعداد عوارض

در این روش برای ارزیابی کامل‌بودن داده‌های مورد بررسی به داده رسمی، از نسبت تعداد عوارض موجود در شبکه معابر مجموعه داده مورد بررسی به تعداد عوارض شبکه معابر مجموعه رسمی در ناحیه مورد نظر استفاده می‌گردد (رابطه (۳)) [۱۲].

$$C = \frac{\sum \text{VGI line number}}{\sum \text{Refrence line number}} \qquad (۳)$$

در این رابطه $\sum$VGI line number مجموع تعداد نقاط داوطلبانه و $\sum$Refrence line number مجموع تعداد نقاط رسمی است.

۳-۱-۴- معیار چگالی عوارض

با تقسیم طول شبکه عوارض بر مساحت ناحیه مورد بررسی، چگالی عوارض در ناحیه مد نظر محاسبه می‌شود [۱۸]. بنابر ماهیت این معیار، از این معیار می‌توان برای مقایسه مناطق مختلف با مساحت‌های گوناگون نظیر مناطق مختلف شهری بهره گرفت. در رابطه (۴) D چگالی عوارض مورد بررسی با مجموع طول $\sum$Line length در ناحیه‌ای به مساحت A می‌باشد.

$$D = \frac{\sum \text{Line length}}{A} \qquad (۴)$$

پس از تعیین چگالی عوارض دو مجموعه داده از تفاضل دو مقدار چگالی عوارض داوطلبانه و رسمی میزان کامل‌بودن داده‌های داوطلبانه قابل ارزیابی است [۱۷, ۱۸].

$$C = D_{VGI} - D_{Refrence} \qquad (۵)$$

در رابطه (۵) D_{VGI} و $D_{Refrence}$ چگالی عوارض داوطلبانه و رسمی و مقدار C میزان کامل‌بودن است.

۳-۱-۵- ارزیابی نتایج معیارها

یافته‌های حاصل شده در مطالعات بررسی الگو کامل-بودن عوارض داوطلبانه، وضعیت مطلوب‌تر مناطق شهری نسبت به مناطق روستایی و اطراف شهر را نشان داده است. بر همین اساس با توجه به این‌که مناطق انتخابی از مناطق پرجمعیت مرکزی شهر تهران انتخاب شده‌اند، به احتمال زیاد مجموعه داده داوطلبانه مرتبط با این مناطق، دارای اطلاعات کامل و بروز است. با توجه به این که تراکم بالاتر عوارض داوطلبانه نسبت به مجموعه نظیر رسمی تنها در سطح فرضیه است، نیاز است به کمک آزمون‌های آماری، معنادار بودن فرضیه مورد آزمایش قرار گیرد.

با توجه به این‌که هدف آزمون بررسی تراکم بیشتر یک مجموعه داده نسبت به مجموعه داده دیگر است، آزمون t با دو نمونه مستقل، آزمون آماری مناسب برای آزمایش فرضیه است. فرض صفر آزمون در صورتی که، برتری تراکم عوارض مجموعه داده رسمی نسبت به مجموعه داده

داوطلبانه معرفی شود، چنان‌چه p-value محاسبه شده از آزمون کمتر از سطح معناداری آزمون (۰/۰۵) باشد فرض صفر رد شده و فرض مقابل به صورت برتری تراکم شبکه عوارض داوطلبانه نسبت به رسمی پذیرفته می‌شود. با استناد به نتایج پیشین، انتظار مطالعه رد فرض صفر در نظر گرفته شده آزمون است.

در خصوص استفاده از آزمون t لازم است دو شرط تبعیت از توزیع نرمال و برابری واریانس نمونه‌های اخذ شده از دو جامعه برقرار باشد. البته در صورت عدم تبعیت نمونه‌ها از توزیع نرمال، امکان نادیده‌گرفتن این پیش شرط در ازای انتخاب تعداد نمونه زیاد از جامعه به واسطه قضیه حد مرکزی وجود دارد. در خصوص واریانس نابرابر نمونه‌های دو جامعه نوع دیگر آزمون t، t welch جایگزین مناسب خواهد بود [۲۹].

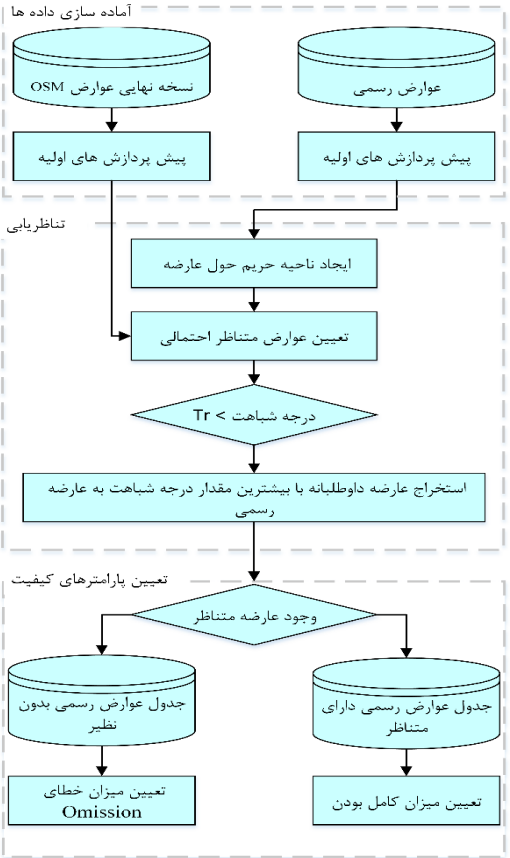
۳-۲- روش‌های عارضه‌مبنا

برخلاف روش‌های منطقه‌مبنا که پس از انجام پیش-پردازش‌های اولیه استفاده از آن‌ها به سهولت انجام می‌گیرد، به واسطه عواملی نظیر استانداردهای متفاوت دو مجموعه داده و چارچوب‌های مختلف جمع‌آوری داده، پیاده‌سازی روش‌های عارضه‌مبنا با معضلات مخصوص به خود همراه خواهد بود. مهم‌ترین گام در استفاده از این روش‌ها شناخت عوارض متناظر در دو مجموعه داده مورد بررسی است.

منظور از تناظر یابی عوارض برداری، اتصال این عوارض در سطوح مختلف جزییات به‌منظور تشکیل پایگاه داده چند مقیاسی است [۳۰]. درواقع این عملیات فرآیند جستجوی عوارض متناظر از میان مجموعه داده‌های مختلف با ارتباط مکانی و مشابهت معنایی قوی است. به منظور انجام فرآیند تناظریابی از ویژگی‌های مختلف داده-های برداری نظیر ویژگی‌های هندسی و یا ویژگی‌های توصیفی استفاده می‌گردد. به سبب وجود ویژگی هندسی عوارض برای تمام داده‌های برداری [۳۱] این ویژگی در عملیات تناظریابی این مقاله بهره گرفته شده است.

یکی از استانداردهای داده‌های برداری خطی در هنگام تولید داده رعایت ساختار گراف در آن‌ها است. منظور از ساختار گراف در شبکه معابر این موضوع است که هر یال انتخاب شده در نقاط پایانی خود می‌تواند شامل تقاطع باشد ولی در نقاط میانی، یال‌های گراف نمی‌توانند شامل تقاطع با یال دیگری از گراف باشند [۳۲]. در داده‌های رسمی این ساختار در مراحل تولید داده به صورت ضمنی رعایت می‌شود ولی در داده‌های داوطلبانه به واسطه عدم وجود محدودیت در ورود عوارض، امکان ورود عوارض با تقاطع در نقاط میانی وجود دارد.

اهمیت موضوع فوق در انجام پردازش‌هایی است که در آن‌ها از ویژگی‌های هندسی عوارض نظیر تناظریابی استفاده می‌شود. برای مثال در تناظریابی داده‌های رسمی و داوطلبانه در صورتی که این پیش‌پردازش در نظر گرفته نشود، تعداد کمی از عوارض به عنوان عوارض نظیر تعیین می‌گردند. در عملیات ایجاد ساختار گراف تمامی یال‌های دارای تقاطع در نقاط میانی به یال‌های بدون تقاطع شکسته شده و یال‌های منفصل از هم در صورتی که اتصال آن‌ها منجر به تولید تقاطع در نقاط میانی یال ایجاد شده نگردد به هم متصل می‌شوند [۳۲]. الگوریتم تناظریابی مورد استفاده در این پژوهش از بخش‌های کلی زیر تشکیل شده است (شکل ۲).



شکل ۲- فلوچارت الگوریتم تناظریابی

۱- تعیین عوارض کاندید: برای تعیین عوارض متناظر داده‌های رسمی در مجموعه داده‌های داوطلبانه، ابتدا برای هر یک از عوارض مجموعه داده رسمی حریمی با عرض مشخص (که میزان این حریم به میزان دقت مکانی مجموعه داده رسمی وابسته است) ایجادشده و

عوارض داوطلبانه دارای اشتراک با این ناحیه، به‌عنوان عوارض احتمالی متناظـر استخراج می‌شوند.

۲- تعیین درجه شباهت مکانی عوارض کاندید به عارضه رسمی: شباهت عوارض کاندید استخراج شده در مرحله قبل از طریق معیار هندسی ناحیه مشترک بین حریم عوارض محاسبه می‌گردد (رابطه (۷)).

$$C = \frac{2A_i}{A_{PL_1} + A_{PL_2}} \qquad (۶)$$

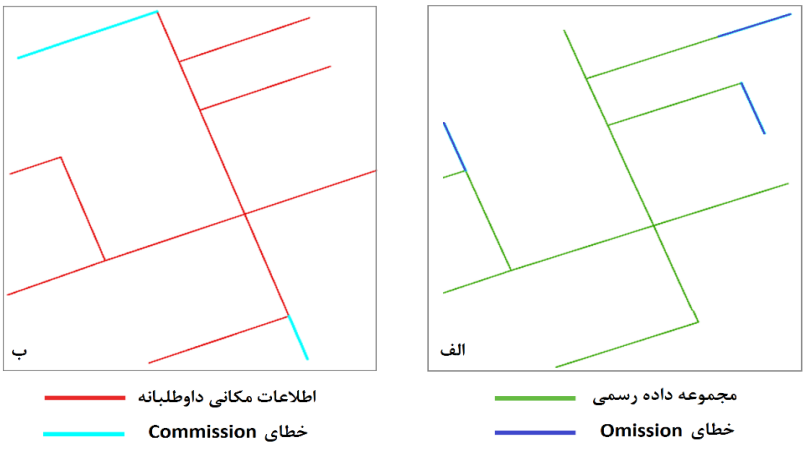
در این رابطه A_i مساحت ناحیه مشترک بین دو حریم و A_{PL_1}، A_{PL_2} مساحت حریم هر یک از خطوط مورد بررسی است.

۳- تعیین عارضه متناظر: از میان عوارض کاندید عارضه‌ای به عنوان عارضه متناظر نهایی شناخته می‌شود که مقدار درجه شباهت مکانی آن بیشترین مقدار میان مقادیر محاسبه‌شده باشد. درصورتی‌که این مقدار از حد آستانه موردنظر کمتر باشد، عارضه بدون عضو متناظر شناختـه می‌شود. مقدار این حـد آستـانه به‌صورت تجربی انتخاب می‌گردد و در بسیاری از تحقیقات این مقدار ۰/۸ در نظر گرفته شده است [۳۱, ۳۳, ۳۴].

۴- محاسبه کامل بودن: پس از انجام فرآیند تناظریابی، عوارض در دو مجموعه داده به دو دسته دارای متناظـر و بدون متناظـر تقسیم می‌شود. از جدول داده‌های متناظر و جدولی که نشان دهنده عوارضی از دو مجموعه است که متناظری برای آن‌ها یافت نشده است، در تعیین پارامترهای کیفیت داده‌های مکانی استفاده می‌گردد [۸]. میزان مجموع طول عوارض دارای نظیر در هر مجموعه داده به مجموع طول کل عوارض آن مجموعه داده، میزان حضور داده‌های یک مجموعه داده در مجموعه داده دوم را نشان می‌دهد. در حقیقت اگر این معیار برای مجموعه داده رسمی محاسبه گردد، میزان حضور داده‌های داوطلبانه را در مجموعه داده رسمی نشان می‌دهد. این مقدار همان میزان کامل‌بودن داده‌های داوطلبانه نسبت به داده‌های رسمی است [۸].

معیار کامل‌بودن با خطاهای Omission و Commission همراه است. خطا Omission به عوارضی اشاره دارد که در دنیای واقعی (مجموعه مرجع) وجود داشته ولی در مجموعه داده مورد بررسی وجود ندارند [۳۵] (حالت الف شکل ۳). در صورتی که خطا Commission نماینده عوارضی از مجموعه داده داوطلبانه است که در دنیا واقعی وجود نداشته و تناظری برای آن‌ها در مجموعه مرجع یافت نمی‌شود (حالت ب شکل ۳). در جدول ۱ نحوه تعیین این دو معیار ذکر شده است. ارزیابی خطا دوم برخلاف خطا اول با پیچیدگی‌های بیشتری همراه است [۸]. یکی از پیچیدگی‌های اساسی در ارزیابی این پارامتر، نیاز به مجموعه داده رسمی با سطح جزئیات برابر با دنیا واقعی است. به همین دلیل به خاطر عدم وجود مجموعه داده رسمی با این شرایط امکان برآورد این پارامتر امکان‌پذیر نیست. به منظور ارزیابی این معیار توصیه می‌شود از شبکه عوارض استخراج شده از تصاویر هوایی به جهت جزئیات بسیار زیاد این اطلاعات برای برآورد این کمیت استفاده گردد.

برای تعیین میزان خطا Commission در مجموعه داده داوطلبانه مورد بررسی کافی است در فلوچارت شکل ۲ در قسمت ورودی داده‌های رسمی، داده داوطلبانه و در قسمت ورودی مجموعه داده داوطلبانه شبکه عوارض مجموعه داده رسمی را قرار داد. مجموع طول عوارض داوطلبانه بدون عضو متناظر در مجموعه دیگر به کل طول عوارض داوطلبانه میزان خطا Commission مجموعه داده داوطلبانه را نشان می‌دهد.



شکل ۳- دو خطا omission (**الف**) و Commission (**ب**)

جدول ۱- نحوه محاسبه دو خطا Omission و Commission در روش‌های عارضه‌مبنا

روش‌های عارضه-مبنا	Omission	۱- تناظریابی داده‌های رسمی با داده‌های داوطلبانه ۲- استخراج عوارض مجموعه داده رسمی بدون متناظر ۳- محاسبه طول یا تعداد این عوارض به طول یا تعداد کلی عوارض مجموعه داده رسمی
	Commission	۱- تناظریابی داده‌های داوطلبانه و داده‌های رسمی ۲- استخراج عوارض داوطلبانه بدون متناظر ۳- محاسبه طول یا تعداد این عوارض به طول یا تعداد کلی عوارض داوطلبانه

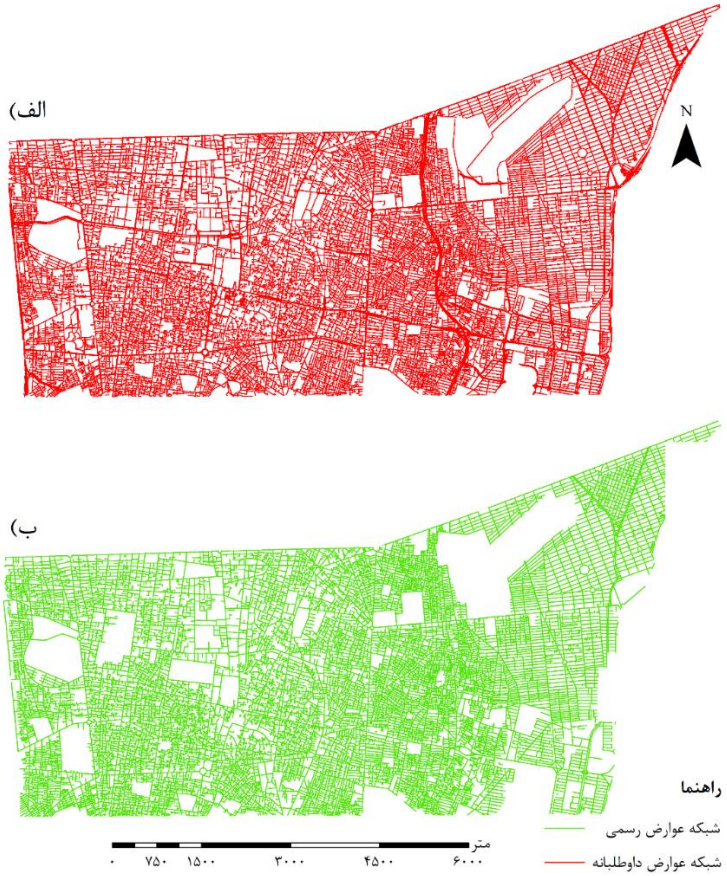
۴- پیاده‌سازی

به منظور پیاده‌سازی روش‌های اشاره شده برای بررسی مولفه کامل بودن مجموعه داده داوطلبانه کشور ایران نیاز به استخراج مناطق با میزان مشارکت بالا است. همان‌گونه که اشاره شد، به دلیل ارتباط این معیار و میزان تراکم جمعیت، این انتظار می‌رود که شهرهای پرجمعیت دارا مجموعه داده‌های به نسبت کامل‌تری باشند. ازاین‌رو داده‌های داوطلبانه مناطق ۱۱ تا ۱۴ کلان‌شهر تهران در مقیاس ۱:۲۰۰۰ به عنوان مناطق مورد مطالعه انتخاب شده است. داده‌های داوطلبانه از وب‌سایت Geofabric در ۱۷ ماه اکتبر سال ۲۰۱۸ دریافت شده و با استفاده از نرم-افزار ArcGis منطقه مورد مطالعه از آن جدا شده است. شکل ۴ داده‌های مورد مطالعه را نشان می‌دهد.

به منظور ایجاد آشنایی بیشتر با مجموعه داده داوطلبانه و رسمی و ایجاد دید کلی نسبت به حجم محاسبات، برخی از ویژگی‌های کمی این دو مجموعه داده نظیر وسعت منطقه، تعداد و طول عوارض موجود در ناحیه و میزان مشارکت‌ها در جدول ۲ ذکر شده است.

جدول ۲- مشخصات کمی دو مجموعه داده انتخاب شده

پارامترها	رسمی	داوطلبانه
مساحت ناحیه (کیلومترمربع)	۵۰/۵۶	۵۰/۵۶
تعداد عوارض	۲۹۲۳۹	۱۶۱۲۱
طول عوارض (کیلومتر)	۱۲۳۵	۱۸۱۸
میزان مشارکت	-	۵۰۵۳۲
تعداد افراد مشارکت کننده	-	۲۸۳



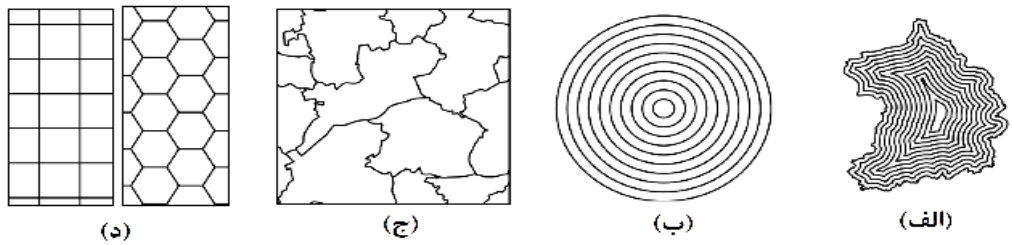
شکل ۴- داده‌های VGI (الف) و رسمی (ب) مورداستفاده در مناطق ۱۱ تا ۱۴ شهر تهران

۴-۱- ارزیابی روش منطقه‌مبنا

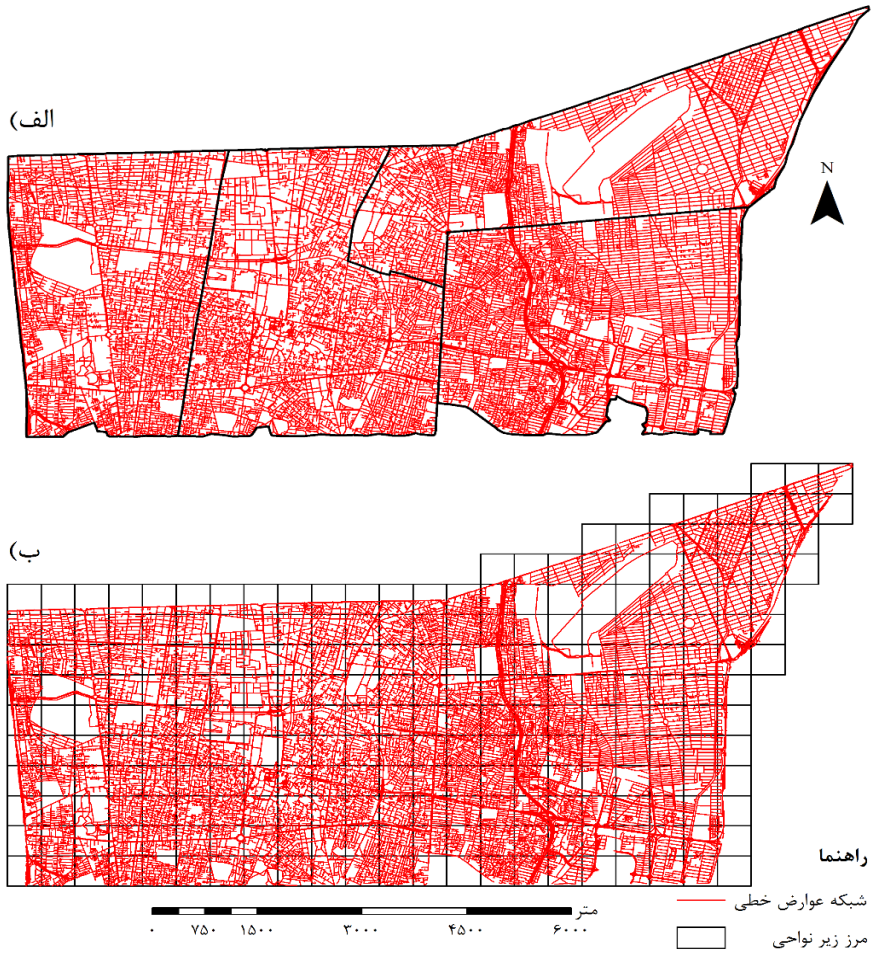
اولین گام در انجام محاسبات مربوط به تعیین پارامترهای کیفیت در روش‌های منطقه‌مبنا، شکستن فضا مسئله به یکسری نواحی با مساحت کمتر است. الگوهای متنوعی برای شکستن ناحیه مورد بررسی به زیر بخش‌های کوچک‌تر وجود دارد. برخی از این الگوها به صورت نمونه در شکل ۵شکل ۵ ذکر شده است [۳۶].

از میان این الگوها، بیشتر از الگوهای شبکه مربعی و یا مرز مربوط به مناطق شهری در کاهش ناحیه محدوده بررسی در روش‌های منطقه‌مبنا استفاده شده است. نمایشی از نحوه تقسیم منطقه به زیر مناطق کوچکتر با استفاده از این دو الگو در حالات الف و ب شکل ۶ آورده شده است.

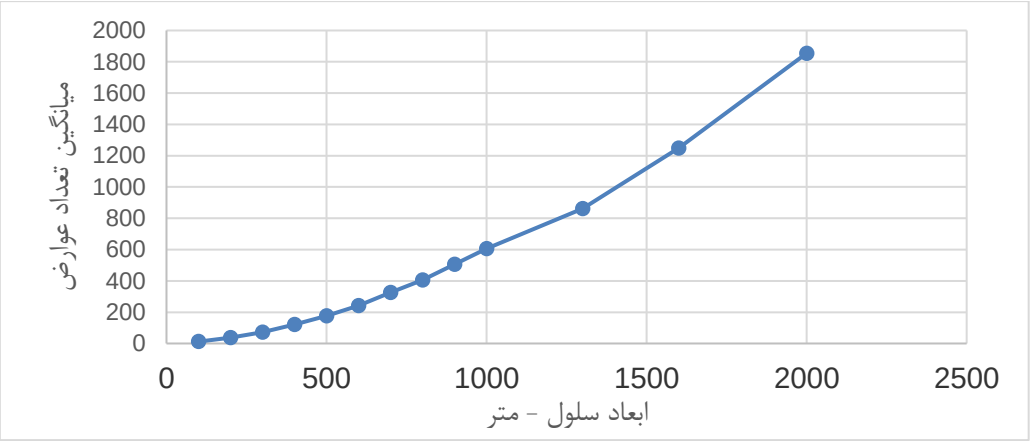
همان‌طور که در قسمت قبل اشاره شد، تعیین ابعاد سلول‌ها با توجه به میزان تراکم عوارض در منطقه به شیوه تجربی تعیین می‌شود. ترسیم نمودارهای نظیر نمودار سرعت پردازشی و میانگین تعداد عوارض در سلول‌ها بر حسب ابعاد سلول می‌تواند در تعیین ابعاد سلول مفید باشد (شکل ۷ و شکل ۸). با مقایسه دو نمودار ترسیم شده در دو شکل ۷ و شکل ۸ با توجه به این که در انتخاب این کمیت لازم است حالت میانه در نظر گرفته شود، اندازه ۵۰۰ متری برای طول اضلاع شبکه در نظر گرفته شده است.



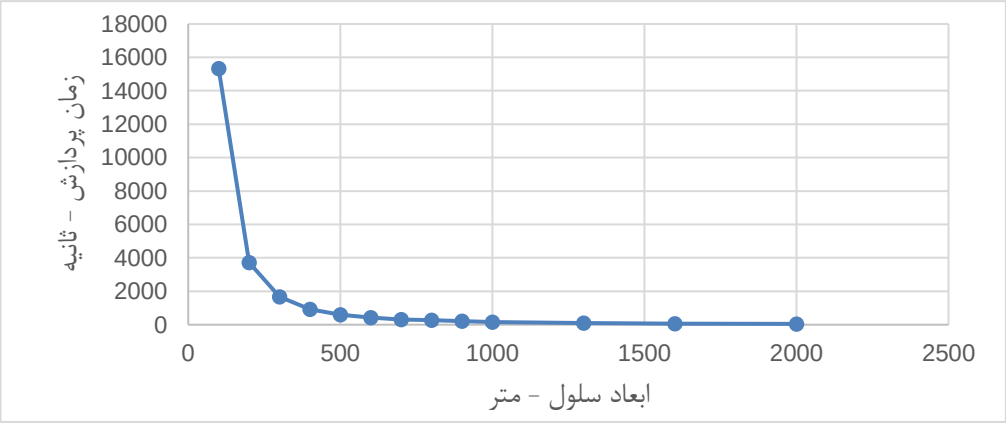
شکل ۵- (الف) حریم ناحیه شهری با عرض‌های مختلف (ب) دوایر هم‌مرکز (ج) مرزهای محلی و شهری (د) شبکه منظم (مربعی - شش‌ضلعی)



شکل ۶- تقسیم منطقه‌ی موردمطالعه به بخش‌های کوچک‌تر به‌منظور بررسی پارامترهای کیفیت براساس (الف) مرزهای شهری (ب) شبکه مربعی



شکل ۷- نمودار میانگین تعداد عوارض بر حسب ابعاد سلول

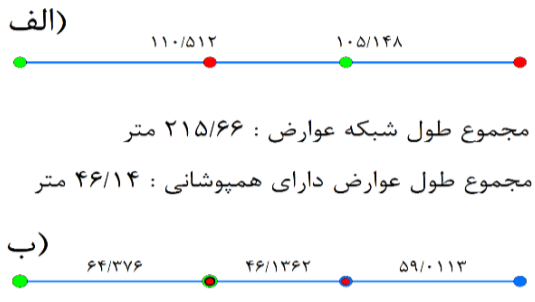


شکل ۸- نمودار زمان پردازش بر حسب ابعاد سلول

شکل ۱۱ نتایج حاصل از محاسبه کامل‌بودن عوارض در هر یک از سلول‌ها، توسط رابطه (۱) را نشان می‌دهد. همان‌گونه که در این شکل مشاهده می‌شود، کامل‌بودن داده‌های مکانی داوطلبانه نسبت به مجموعه رسمی متناظر در بیشتر سلول‌ها دارای وضعیت مطلوب است (سلول‌های سبزرنگ). این تصویر برخی از سلول‌ها مقادیر غیرعادی کامل‌بودن را نشان می‌دهند که با رنگ خاکستری مشخص شده‌اند. استدلال این مطالعه بر این است این نتایج در اثر نقص در عوارض مجموعه داده رسمی حاصل شده است و بهتر است در ارزیابی‌ها این سلول‌ها از پردازش حذف شوند.

علاوه بر این نتایج، ارزیابی‌های انجام‌شده در مقیاس مناطق شهری، هم‌راستا با نتایج حاصل شده از تصویر ۱۱ است (شکل ۱۰). در ۴ منطقه انتخابی منطقه ۱۴ و ۱۳ دارای بیشترین و کمترین میزان تراکم میان مناطق است. منطقه ۱۳ با کمترین میزان تراکم، باز نسبت به منطقه متناظر خود در مجموعه داده رسمی وضعیت مطلوب‌تری داراست.

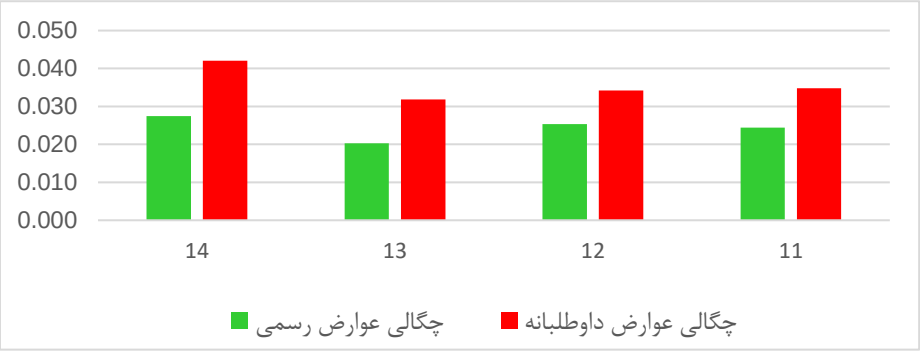
علی‌الرغم کاربرد وسیع مقایسه طول دو مجموعه داده در ارزیابی کامل‌بودن، برخی خطاها در مدل‌سازی می‌تواند تاثیر نادرست بر نتایج ارزیابی داشته باشد. شکل ۹ نمونه‌ای از خطا تاثیرگذار بر معیار طول را نشان می‌دهد. در دو حالت الف و ب این شکل ظاهرا طول عوارض برابر است، در حالی‌که بررسی روابط توپولوژیک دو عارضه، تلاقی نقاط درونی دو پاره‌خط را نمایش می‌دهد. وجود خطا توپولوژیک از این جنس که در داده‌های مکانی داوطلبانه نیز معمول است، منجر به محاسبه سطح بالاتری از کامل‌بودن نسبت به حالت بدون خطا می‌شود.



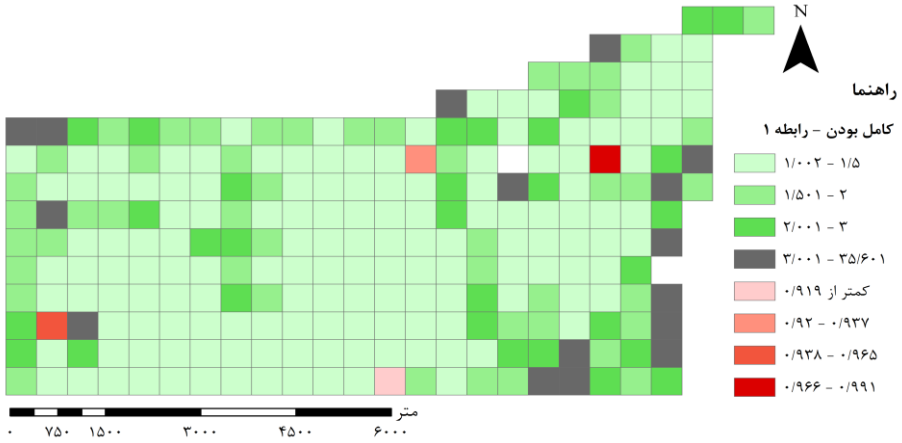
شکل ۹- مجموعه داده دارای عوارض همپوشانی (الف) – مجموعه داده پس از حذف عوارض دارای همپوشانی (ب)

وجود اختلاف ۴۶/۱۴ متری حالت الف نسبت به حالت ب به واسطه وجود خطا توپولوژیک مطرح شده است. از این رو همیشه در بررسی این کمیت برای مجموعه داده‌های داوطلبانه که امکان وجود عوارض دارای همپوشانی در آن امکان‌پذیر است، لازم است طی یک عملیات پیش‌پردازشی، کلیه قسمت‌های تکرار شده از مجموعه داده ورودی حذف گردد.

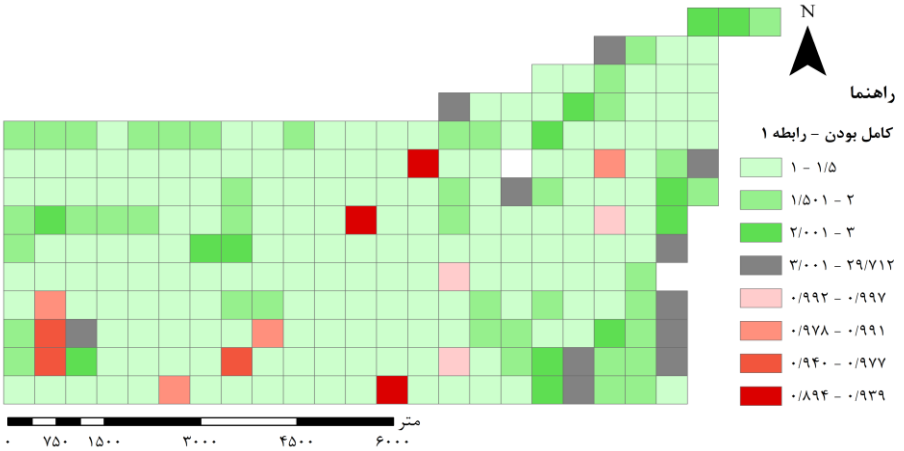
شکل ۱۲ مشابه با شکل ۱۱ میزان کامل‌بودن عوارض پس از حذف عوارض دارای خطا توپولوژیک را نمایش می‌دهد. تاثیر حذف عوارض دارای این خطا بر کل منطقه مطالعاتی کاهش میزان کامل‌بودن از مقدار ۱/۹۳۹ به ۱/۶۵۶ است.



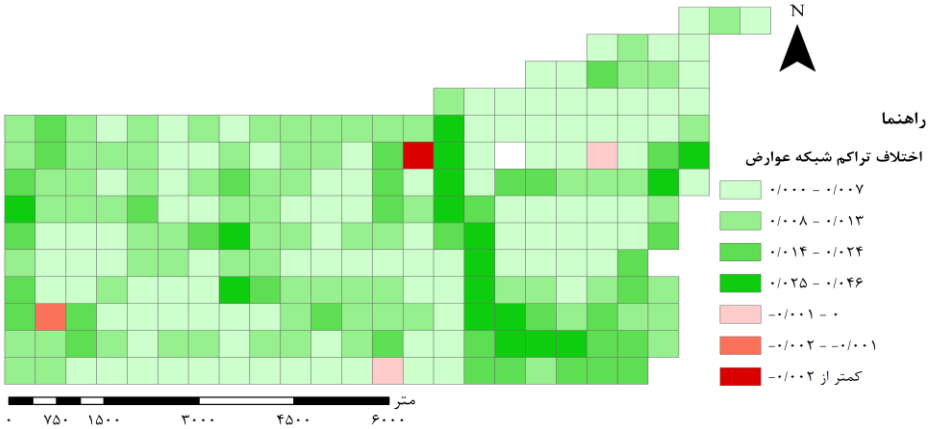
شکل ۱۰- چگالی شبکه عوارض دو مجموعه داده در چهار منطقه مورد بررسی



شکل ۱۱- میزان کامل‌بودن داده‌های داوطلبانه توسط معیار طول – پیش از رفع خطاهای توپولوژیک (عوارض دارای همپوشانی)



شکل ۱۲- میزان کامل‌بودن داده‌های داوطلبانه توسط معیار طول – پس از رفع خطاهای توپولوژیک (عوارض دارای همپوشانی)

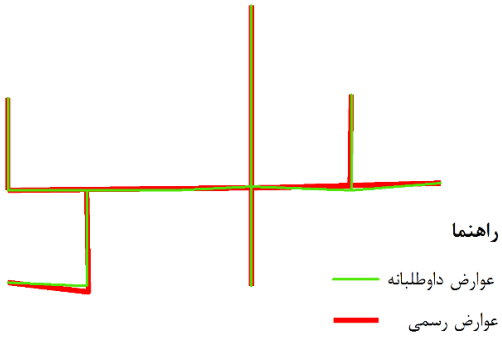


شکل ۱۳- اختلاف تراکم شبکه عوارض داوطلبانه و رسمی

پس از محاسبه کامل‌بودن به کمک معیار طول این بار این پارامتر به کمک معیار تعداد عوارض در هر ناحیه بررسی می‌شود. یکی از ایرادهای این پارامتر نسبت به پارامتر طول وابستگی زیاد این معیار به نحوه مدل‌سازی داده‌های مکانی خطی است [۱]. از شایع‌ترین خطاها در مدل‌سازی داده‌های مکانی برداری خطی موضوع عدم رعایت ساختار گراف است.

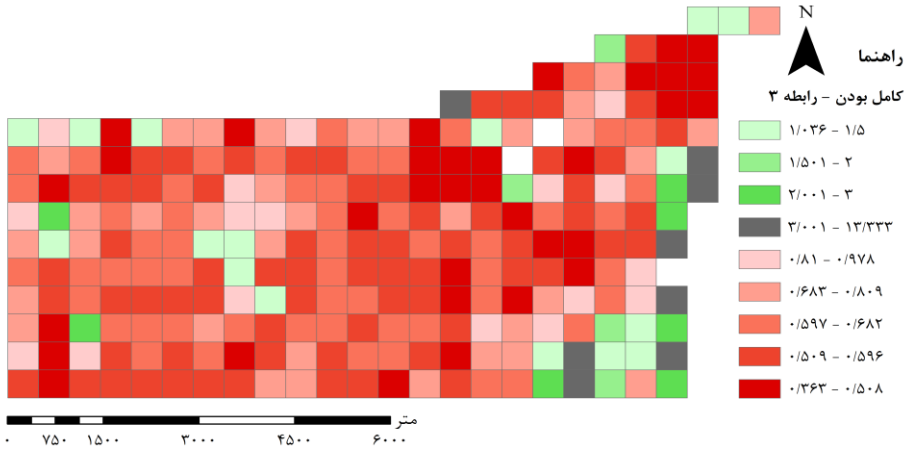
تاثیر عملیات ایجاد ساختار گراف در محاسبه کامل-بودن در شکل ۱۴ ارائه شده است. مجموعه داده داوطلبانه و رسمی در این منطقه شامل ۶۳۵/۲۴ و ۶۳۵/۲۶ متر عارضه است. براساس معیار طول، کامل‌بودن عوارض مجموعه داده داوطلبانه نسبت به عوارض مجموعه داده رسمی به میزان ۱ محاسبه می‌شود. با توجه به این‌که تعداد عوارض در مجموعه داده رسمی ۸ و در مجموعه داده رسمی ۵ عارضه موجود است، میزان کامل‌بودن توسط معیار تعداد عوارض برابر با ۰/۶۳۵ محاسبه می‌شود. حال در صورت ایجاد ساختار گراف در عوارض داوطلبانه تعداد عوارض، از ۵ به تعداد ۸ افزایش می‌یابد. در این حالت کامل‌بودن برابر با مقدار ۱ است.

نتایج بررسی کامل‌بودن مجموعه داده داوطلبانه در کل ناحیه مطالعاتی نیز توسط معیارهای تعداد و طول عوارض نشان می‌دهد، در صورتی که از تعداد عوارض به عنوان معیار بررسی کامل‌بودن بهره گرفته شود، اختلاف قابل توجهی با نتایج حاصل شده از معیار طول مشاهده می-گردد ولی در صورت ایجاد ساختار گراف در شبکه عوارض مجموعه داوطلبانه، نتایج ارزیابی کامل‌بودن معیار تعداد عوارض به مقادیر حاصل شده توسط معیار طول نزدیک‌تر می‌شود (افزایش میزان کامل‌بودن از مقدار ۰/۵۵۱ به مقدار ۱/۳۶۵ - جدول ۳).

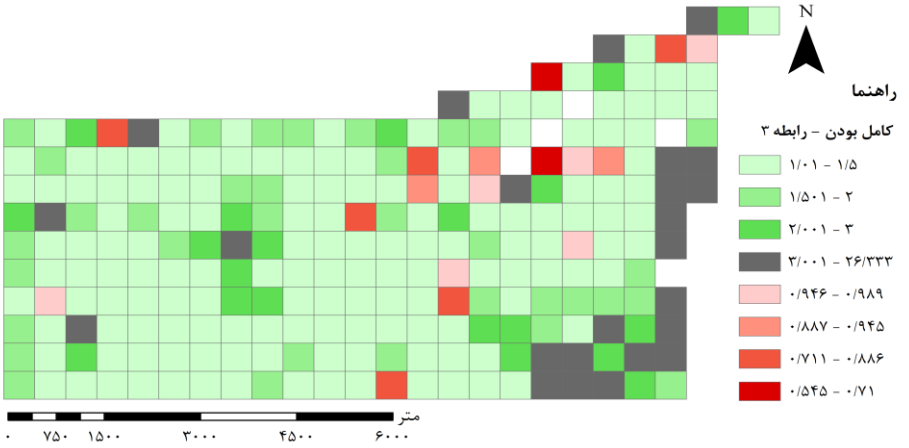


شکل ۱۴- ضعف معیار تعداد عوارض پیش از ایجاد ساختار گراف در شبکه عوارض در ارزیابی کامل‌بودن

در شکل ۱۵ و شکل ۱۶ نتایج استفاده از این پارامتر بدون ایجاد ساختار گراف در مجموعه داده و با استفاده از آن آورده شده است. در حالتی که عملیات ذکر شده بر روی شبکه داده داوطلبانه انجام نشود، مقدار محاسبه‌شده به‌عنوان معیار کامل‌بودن در کمترین و بیشترین حالت برابر با مقادیر ۰/۳۸ و ۱۳/۳۳ محاسبه می‌شود. این در حالی است که در صورت استفاده از این عملیات این مقادیر به ۰/۶۸ و ۲۵/۶۷ تغییر پیدا خواهد کرد. ازاین‌رو به دلیل اختلاف زیاد این مقادیر، توصیه مطالعه به ایجاد ساختار گراف در شبکه عوارض داوطلبانه پیش از به-کارگیری معیارهای وابسته به ساختار شبکه نظیر معیار تعداد عوارض در نتایج ارزیابی است.



شکل ۱۵- میزان کامل‌بودن داده‌های داوطلبانه با استفاده از معیار تعداد عوارض پیش از ایجاد ساختار گراف



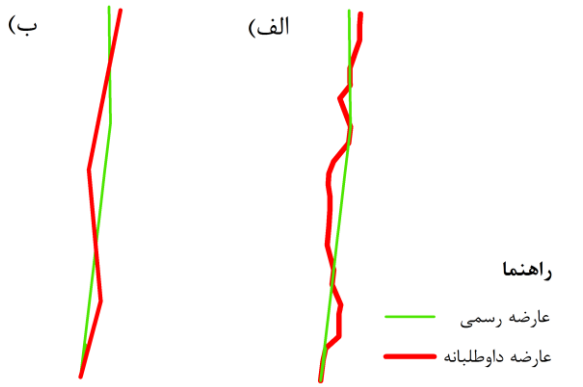
شکل ۱۶- میزان کامل‌بودن داده‌های داوطلبانه با استفاده از معیار تعداد عوارض پس از ایجاد ساختار گراف

جدول ۳- تعیین پارامتر کامل‌بودن در کل منطقه

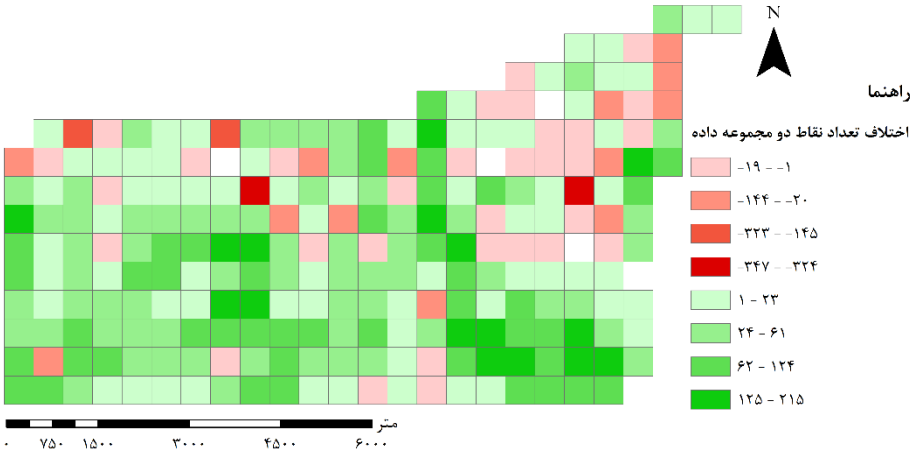
کامل‌بودن	مجموعه داده داوطلبانه	مجموعه داده رسمی	
۱/۴۷۲	۱۸۱۸	۱۲۳۵ کیلومتر	معیار طول
۰/۵۵۱	۱۶۱۲۱	۲۹۲۳۹	معیار تعداد – پیش از عملیات ایجاد ساختار گراف
۱/۳۶۵	۳۹۹۲۱	۲۹۲۳۹	معیار تعداد - پس از عملیات ایجاد ساختار گراف

در مجموعه داده‌های داوطلبانه به دلیل طیف متنوع روش‌های پیش‌بینی شده در ورود داده به پایگاه داده و سطح دانش و مهارت متفاوت افراد، این امکان وجود خواهد داشت، عوارض تقریبا مستقیم در دنیا واقعی به صورت عوارض دارای شکستگی زیاد در مجموعه داده مدل‌سازی شوند. شکل ۱۷ نمونه‌ای از عارضه در دنیای واقعی و متناظر آن در مجموعه داده داوطلبانه را نمایش می‌دهد که همراه با شکستگی زیاد مدل‌سازی شده است. بدیهی است در مورد چنین عوارضی، ارزیابی کامل‌بودن توسط تعداد نقاط تشکیل‌دهنده عارضه صحیح نخواهد بود.

اعمال فیلترهای کاهش پیچیدگی عوارض و ساده-سازی خطوط در این موارد می‌تواند مشکل به وجود آمده را برطرف نماید. حالت ب شکل فلان عارضه داوطلبانه پس از ساده‌سازی خطوط، توسط الگوریتم douglas peucker را نشان می‌دهد. همان‌گونه که نتایج جدول ۴ نشان می-دهد میزان کامل‌بودن عارضه داوطلبانه قبل و بعد از ساده-سازی خطوط از ۱/۹۳ به ۱ کاهش یافته است. توصیه مطالعه پس از بررسی عوارض مختلف این است پیش از بررسی کیفیت عوارض داوطلبانه نحوه مدل‌سازی این عوارض بررسی شود.



شکل ۱۷- عارضه داوطلبانه با شکستگی زیاد در مقابل عارضه رسمی مستقیم



شکل ۱۸- اختلاف تعداد نقاط دو مجموعه داده

تاکنون میزان کامل بودن داده‌های داوطلبانه نسبت به داده‌های رسمی با استفاده از روش‌های مختلف مورد محاسبه قرار گرفته است. اکنون برای پاسخ به این سوال که کدام یک از دو مجموعه داده از لحاظ میزان کامل بودن از وضعیت بهتری برخوردار است از آزمون‌های آماری استفاده می‌گردد.

بررسی وجود تفاوت معنادار در تراکم شبکه عوارض داوطلبانه نسبت به عوارض مجموعه داده رسمی پس از انجام گام‌های زیر توسط آزمون آماری آزمایش شده است:

۱- از میان تعداد سلول‌های موجود در ناحیه مورد بررسی تعداد ۳۰ سلول از دو مجموعه داده به عنوان اعضای نمونه به صورت تصادفی انتخاب می‌شود (شکل ۱۹).

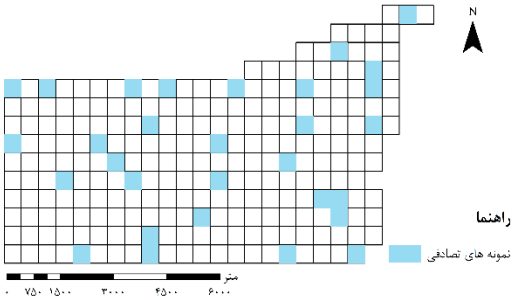
۲- فرض صفر این آزمون برتری تراکم شبکه عوارض مجموعه داده رسمی نسبت به شبکه عوارض مجموعه داده داوطلبانه درنظر گرفته شده است.

۳- این آزمون به صورت یک طرفه راست با سطح معناداری ۰/۰۵ و فرض مقابل، بیشتر بودن میزان چگالی عوارض داوطلبانه نسبت به چگالی عوارض مجموعه داده رسمی در نظر گرفته می‌شود.

۴- تعیین نتیجه آزمون آماری و تصمیم در مورد پذیرش و یا رد فرض صفر

جدول ۴- مقایسه نتایج حاصل از روش‌های مجموع طول و تعداد نقاط عوارض در تعیین میزان کامل‌بودن

رسمی	داوطلبانه	
۷۵۲/۴۶	۷۶۴/۴۰	معیار طول
۱۴	۲۷	معیار تعداد نقاط – بدون اعمال آنالیز ساده سازی خطوط
۴	۴	معیار تعداد نقاط – با اعمال آنالیز ساده سازی خطوط
۱/۰۱۶		کامل‌بودن معیار طول
۱/۹۳		کامل‌بودن معیار تعداد نقاط – بدون ساده سازی خطوط
۱		کامل‌بودن معیار تعداد نقاط – با استفاده از ساده سازی خطوط



شکل ۱۹- نمونه‌های تصادفی انتخاب شده از جامعه

شکل ۱۹، ۳۰ سلول تصادفی انتخاب شده به منظور انجام آزمایش آماری را نمایش می‌دهد. براساس p-value حاصل شده از آزمون برابر با ۰/۰۲۶ و کوچکتر بودن این مقدار از سطح معناداری آزمون، می‌توان نتیجه‌گیری کرد فرض صفر درنظرگرفته رد شده و وضعیت تراکم شبکه عوارض داوطلبانه در ناحیه بررسی شده نسبت به مجموعه رسمی مطلوب‌تر است.

مزیت دیگر داده‌های مکانی داوطلبانه نسبت به داده‌های رسمی وجود طیف گسترده‌ای از نوع کاربری‌های مختلف است. وجود این تفاوت در دو مجموعه داده به واسطه فلسفه مختلف ایجاد دو مجموعه داده است. به عنوان مثال پاسخگویی به مسیریابی کاربران دوچرخه سوار جز اولویت‌های مجموعه داده رسمی نیست. اهمیت موضوع اشاره شده به واسطه تاثیر آن در ارزیابی کامل‌بودن است. در صورتی که مجموعه داده داوطلبانه با عوارض دارای یک نوع کاربری وارد آنالیز شوند که این کاربری‌ها در استاندارد مجموعه داده رسمی وجود نداشته باشد، در این صورت برآورد انجام شده از کامل‌بودن صحیح نخواهد بود. توصیه مقاله در این موارد، بررسی پارامترهای کیفیت برای هر نوع عارضه به صورت مجزا است.

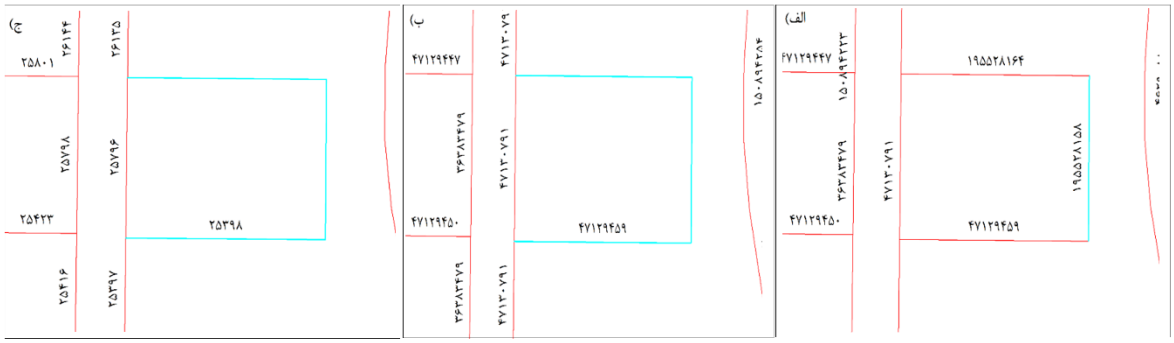
شکل ۲۰ نمونه‌ای از شبکه معابر دارای کاربری متفاوت نسبت به مجموعه داده رسمی را نمایش می‌دهد. مقادیر کامل‌بودن در اثر وجود و حذف این عوارض به ترتیب مقادیر ۱/۷۱ و ۱/۱۲ محاسبه شده است. همان‌طور که مشاهده می‌شود بهتر است در ارزیابی کامل‌بودن، پس از مقایسه مقایسه نوع عوارض موجود در دو مجموعه داده، عوارض دارای کاربری غیر مشترک حذف شده تا برآورد دقیق‌تر به عنوان کامل‌بودن مجموعه داده حاصل شود.

شکل ۲۰- نمونه‌ای از طیف متفاوت کاربری متنوع شبکه معابر داوطلبانه نسبت به مجموعه داده رسمی

۴-۲- ارزیابی روش عارضه‌مبنا

هنگام ایجاد ساختار گراف در داده‌های داوطلبانه این امکان وجود دارد که عارضه با شماره شناسایی مشخص به تعدادی عارضه برای رسیدن به ساختار گراف تقسیم گردد. به این صورت در ستون اطلاعات توصیفی عوارض شکسته شده، شماره شناسایی یکسانی ثبت خواهد شد. این موضوع به صورت ویژه به‌عنوان خطا پس از عملیات تناظریابی، در انجام عملیات پرس‌وجو از پایگاه داده جهت اندازه‌گیری پارامترهای کیفیت خود را نشان می‌دهد. برای جلوگیری از رخداد چنین مشکلی کافی است ستون جدیدی تحت عنوان شماره شناسایی جدید به اطلاعات توصیفی عوارض پس از عملیات تناظریابی اضافه گردد (حالت ج شکل ۲۱). همچنین در مورد عوارضی که به صورت مجزا از هم قرار گرفته‌اند و پس از فرآیند تبدیل به ساختار گراف به یک عارضه تبدیل می‌شوند نیز باید در ستون شماره شناسایی این عوارض تغییری حاصل شود (حالت الف شکل ۲۲). در الگوریتم پیاده شده پس از ساخت عارضه جدید از این عوارض، شماره شناسایی یکی از عوارض تشکیل دهنده آن به‌صورت تصادفی به‌کل عارضه نسبت داده شده است (حالت ب و ج شکل ۲۲).

شکل ۲۱- شکست یک عارضه به چندین عارضه



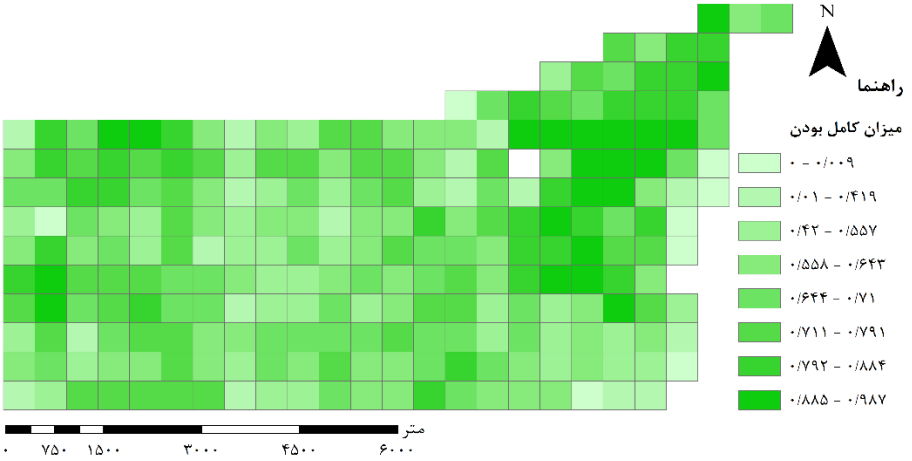
شکل ۲۲- ترکیب عوارض و تشکیل یک عارضه

پس از ایجاد ساختار گراف در داده‌های مکانی داوطلبانه عملیات تناظریابی بر روی دو مجموعه داده موجود صورت گرفت و از میان تعداد ۲۹۲۳۹ عارضه در مجموعه داده رسمی تعداد ۱۷۶۸۴ عارضه به‌عنوان عوارض دارای متناظر در مجموعه داده داوطلبانه انتخاب شد. ارزیابی کامل‌بودن عوارض داوطلبانه در دو سطح جزئیات کل منطقه و سلول‌های با ابعاد ۵۰۰ متر، نرخ پایین کامل‌بودن این عوارض را نشان داد (جدول ۵ و شکل ۲۳). طبیعی است استفاده از تنها یک معیار در تعیین شباهت هندسی عوارض در عملیات تناظریابی دو مجموعه داده منجر به شناسایی تعداد کمی از عوارض به عنوان عوارض متناظر گردد و در صورت بهره‌گیری از معیارهای دیگر تعیین شباهت هندسی عوارض نظیر طول، توجیه و مساحت، تعداد عوارض متناظر میان دو مجموعه داده و کامل‌بودن بهبود می‌یابد.

جدول ۵- ارزیابی معیار کامل‌بودن

کامل‌بودن	کل عوارض	دارای نظیر	
۶۰/۵	۲۹۲۳۹	۱۷۶۸۴	معیار تعداد
۶۶/۲	۱۲۳۵۷۰۵/۵۳	۸۱۸۳۳۲/۴۰	معیار طول (متر)

تناظریابی عوارض برداری علاوه بر تعیین اطلاعات نظیر دو مجموعه داده، در تعیین پارامترهای omission و commission داده‌های مکانی به عنوان دو بخش اصلی کامل‌بودن مفید خواهد بود. بر خلاف عوارض مجموعه داده رسمی بدون عضو متناظر در مجموعه داده داوطلبانه، که به عنوان نواقص مجموعه داده داوطلبانه شناسایی می‌شود استدلال در خصوص اطلاعات مازاد داده‌های مکانی داوطلبانه، پس از تناظریابی این مجموعه داده با داده‌های رسمی به سادگی کمیت omission نیست. تفاوت ذکر شده به این خاطر است که مجموعه داده داوطلبانه دارای طیف وسیعی از نوع کاربری‌ها است که استاندارد مجموعه داده رسمی ممکن است شامل آن‌ها نباشد. با این حال پس از تناظریابی مجموعه داده رسمی نسبت به داوطلبانه میزان omission برابر ۳۳/۸ و پس از تناظریابی مجموعه داده داوطلبانه و رسمی میزان commission برابر با ۴۷/۸ محاسبه شد. مشابه با آن‌چه در پاراگراف پیشین مطرح شد، انتظار می‌رود در صورت بهبود الگوریتم تناظریابی میزان این دو خطا کاهش یابد.



شکل ۲۳- بررسی میزان کامل‌بودن داده‌های داوطلبانه با استفاده از ترکیب دو رویکرد منطقه‌مبنا و عارضه‌مبنا

۴-۳- مقایسه روش‌های عارضه مبنا در مقابل منطقه‌مبنا

محاسبه پارامترهای omission و commission به کمک اطلاعات عوارض بدون عضو نظیر هر مجموعه داده درون مجموعه دیگر انجام پذیر است. از این رو در روش-های محاسباتی منطقه‌مبنا، به واسطه ترکیب عوارض حذف شده و اضافی مجموعه داده داوطلبانه، ارزیابی هر معیار به صورت مجزا امکان‌پذیر نخواهد بود.

برای روشن‌شدن موضوع، کامل‌بودن مجموعه داده داوطلبانه توسط دو رویکرد اشاره شده، ارزیابی و نتایج مورد تحلیل قرار گرفته است (شکل ۳). بررسی عوارض بدون عضو متناظر هر مجموعه داده نشان می‌دهد مجموعه داده داوطلبانه دارای ۱۲/۷ درصد اطلاعات اضافی (commission) و دارای ۱۲/۵ درصد دارای کمبود اطلاعات (omission) است (جدول ۶). از طرفی تفاضل اطلاعات مازاد و کمبود مجموعه داده داوطلبانه به میزان ۰/۲ درصد با نتایج حاصل از ارزیابی منطقه‌مبنا کامل‌بودن به میزان ۱/۰۰۲ سازگار است.

پیاده‌سازی و پردازش ساده‌تر در روش‌های منطقه‌مبنا فاکتور مهمی در چشم‌پوشی از محدودیت‌های رویکرد منطقه‌مبنا و بهره‌گیری مکرر آن در مطالعات ارزیابی کیفیت بوده است. البته سرعت پردازشی بالا در روش‌های منطقه‌مبنا در صورتی قابل دستیابی است که مجموعه داده داوطلبانه دارای خطاهای تاثیرگذار بر نتایج معیارهای کیفیت نباشد. ولی همان‌طور که بررسی شد به واسطه وجود خطاهای متفاوت در داده‌های داوطلبانه و نیاز به پیش‌پردازش‌هایی نظیر حذف عوارض دارای خطا توپولوژیک، کاهش پیچیدگی عوارض و ایجاد ساختار گراف در این داده‌ها، حجم محاسبات و زمان پردازش مورد تاثیر قرار می‌گیرد و ممکن است عملا تفاوتی در زمان اجرایی هر دو رویکرد مشاهده نگردد.

یکی از مزایا روش‌های عارضه‌مبنا، امکان تناظریابی و تعیین پارامترهای کیفیت داده، در مجموعه داده‌ها با استانداردهای متفاوت نام‌گذاری کاربری عوارض است. این موضوع همان‌طور که در روش‌های منطقه‌مبنا بررسی شد به شدت بر نتایج ارزیابی کیفیت تاثیرگذار است و لازم است با مقایسه دو مجموعه داده عوارض دارای کاربری غیر مشترک حذف گردد. البته حذف چنین عوارضی از مجموعه داده داوطلبانه به دلیل عدم شناخت کاربران از نوع کاربری آن‌ها در زمان ویرایش ممکن است منجر به حذف عوارض دارای کاربری یکسان شود.

برای تفهیم سرعت پردازش هر رویکرد، زمان پردازش در دو شیوه منطقه‌مبنا و عارضه‌مبنا محاسبه شده و نتایج آن در این قسمت ذکر شده است. زمان لازم برای اجرا الگوریتم تعیین میزان کامل‌بودن به صورت منطقه‌مبنا با استفاده از معیار طول به میزان ۶۵۱ ثانیه محاسبه شد. در صورت استفاده از معیار تعداد عوارض در روش‌های منطقه‌مبنا زمان سپری شده به میزان ۶۸۷ ثانیه محاسبه گردید. دلیل افزایش زمان محاسبات حالت دوم نسبت به حالت قبلی به واسطه استفاده از الگوریتم ایجاد ساختار گراف در داده‌های داوطلبانه است. در صورت بهره‌گیری از روش‌های عارضه‌مبنا در تعیین پارامتر کامل‌بودن میزان زمان سپری شده برای تعیین جواب به میزان ۳۵۷۹ ثانیه محاسبه شد. بخش زیادی از زمان سپری شده در این روش به واسطه استفاده از تناظریابی در این روش است.

جدول ۶- محاسبه دو خطا Omission و Commission در مجموعه داده شکل ۳

Omission	Commission	طول کلی عوارض - متر	طول عوارض بدون نظیر - متر	
-	۰/۱۲۷	۱۸۲۶/۳۵	۲۳۲/۸۴	مجموعه داده داوطلبانه
۰/۱۲۵	-	۱۸۲۱/۴۵	۲۲۷/۹۴	مجموعه داده رسمی

۵- نتیجه‌گیری و پیشنهادها

باتوجه به تنوع روش‌های ارزیابی کیفیت داده‌های مکانی و چالش‌های هر روش و نیاز هر تحلیل‌گر داده‌های مکانی به شناخت کافی در مورد هر روش، مطالعه کنونی با هدف ارزیابی روش‌های تعیین کامل‌بودن داده‌های مکانی نگارش شد. صرف نظر از نرخ پایین کامل‌بودن در روش-های عارضه‌مبنا به واسطه ضعف الگوریتم بهره گرفته شده، نتایج سایر معیارها سطح مطلوبی از کامل‌بودن داده‌های مکانی داوطلبانه در مناطق انتخابی را نمایش داد.

مقایسه مجموع طول عوارض در میان سایر روش‌های ارزیابی منطقه‌مبنا به سبب سادگی در پیاده‌سازی و عدم نیاز به ایجاد ساختار گراف در مجموعه داده، در دسته روش‌های ارزیابی مناسب طبقه‌بندی می‌شود. البته در برخی از موارد که منطقه مورد بررسی شامل تعداد کمی از عوارض با طول زیاد و تعداد زیادی از عوارض با طول کم است، نتایج ارزیابی براساس طول عوارض می‌تواند برآورد صحیحی از کامل‌بودن عوارض داوطلبانه نباشد. در این موارد بهتر است از معیار تعداد عوارض بهره گرفته شود.

در رابطه با روش سوم به‌کار برده شده روش‌های منطقه‌مبنا، ارزیابی کامل‌بودن با تبدیل عوارض دو مجموعه داده به نقاط سازنده آن، براساس دیدگاه مقاله بهتر است با توجه به طیف مختلف روش‌های جمع‌آوری داده‌های مکانی و امکان وجود عوارض دارای شکستگی زیاد در مجموعه داده داوطلبانه و نیاز به پردازش‌های اضافی در ساده‌سازی خطوط بهتر است از معیار معرفی شده استفاده نگردد.

در روش‌های تناظریابی به خصوص روش‌هایی که در آن از معیار هندسی نظیر ناحیه مشترک بین حریم عوارض استفاده می‌شود، حتماً لازم است این نکته را مدنظر قرار داد که بر خلاف روش‌های منطقه‌مبنا، وابستگی بسیار زیاد میان نتایج تناظریابی و میزان صحت مکانی عوارض ورودی برقرار است. به‌عبارت‌دیگر در صورت وجود جابجایی مکانی بین دو مجموعه داده (عدم تطابق سیستم مختصات دو مجموعه داده و عوامل دیگر) تضمینی برای به جواب رسیدن در این روش‌ها وجود نخواهد داشت. برخلاف روش‌های عارضه‌مبنا در صورتی که جابجایی دو مجموعه داده به صورت فاحش نباشد امکان استفاده از روش‌های منطقه‌مبنا منجر به خطا زیاد در نتایج این روش‌ها نخواهد شد. ولی در استفاده از روش-های منطقه‌مبنا توجه به حساسیت این روش‌ها با خطاهای موجود در فرآیند مدل‌سازی که نمونه‌ای از آن تحت عنوان اثر رفع خطاهای توپولوژیک بر نتایج مورد بررسی قرار گردید ضروری است.

روش‌های منطقه مبنا در تعیین میزان کامل‌بودن به خصوص معیار طول، به دلیل عدم نیاز به پیش‌پردازش‌ها نظیر رفع ایجاد ساختار گراف و تناظریابی، از سرعت محاسباتی به مراتب بالاتری نسبت به روش‌های عارضه-مبنا برخوردار هستند. بنابراین در مقاصدی که هدف تعیین این پارامتر به صورت سریع مورد نیاز است توصیه می‌گردد از این دسته از روش‌ها استفاده گردد. البته در به-کارگیری این روش ضروری است، داده‌های داوطلبانه عاری از خطاهای توپولوژیک باشند چراکه حذف این خطاها خود زمان‌بر خواهد بود.

در قسمت نهایی این مقاله پیشنهاد می‌گردد جهت تعیین میزان کامل‌بودن مناطق مورد بررسی از سایر معیارهای ارائه شده در مقاله [۳۱] جهت بهبود نتایج تناظریابی استفاده گردد. همچنین با توجه به اهمیت این معیار و تأثیر آن با پارامترهای دیگر پیشنهاد می‌گردد ارتباط این معیار بر سایر عناصر کیفیت نظیر صحت مکانی و صحت توصیفات مورد ارزیابی قرار گیرد.

مراجع

[1] Hecht, R. Kunze, C. and Hahmann, S. (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.

[2] Bruns, A. (2008). "The future is user-led: The path towards widespread produsage." Fibreculture journal. no. 11.

[3] Goodchild, M. F. (2007). "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0." International Journal of Spatial Data Infrastructures Research. vol. 2, no. 2, pp. 27-29.

[4] Goodchild, M. F. and Li, L. (2012). "Assuring the quality of volunteered geographic information." Elsevier. vol. 1, pp. 110-120.

[5] Haklay, M. Basiouka, S. Antoniou, V. and Ather, A. (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.

[6] Sehra, S. S. Singh, J. and Rai, H. S. (2017). "Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox." Future Internet. vol. 9, no. 2, p. 15

[7] Veregin, H. (1999). "Data quality parameters." Geographical information systems. vol. 1, no. 5, pp. 177-189.

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

[9] Kresse, W. and Fadaie, K. (2004). "ISO standards for geographic information." Springer Science & Business Media.

[10] Zhang, H. and Malczewski, J. (2017). "Quality evaluation of volunteered geographic information: The case of OpenStreetMap." IGI Global. pp. 19-46.

[11] Zielstra, D. and Zipf, A. (2010). "A comparative study of proprietary geodata and volunteered geographic information for Germany." 13th AGILE international conference on geographic information science.

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

[13] Neis, P. Zielstra, D. and Zipf, A. (2011). "The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011." Future Internet. vol. 4, no. 1, pp. 1-21.

[14] Ludwig, I. Voss, A. and Krause-Traudes, M. (2011). "A Comparison of the Street Networks of Navteq and OSM in Germany." Springer. pp. 65-84.

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

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

[17] Camboim, S. Bravo, J. and Sluter, C. (2015). "An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil." ISPRS International Journal of Geo-Information. vol. 4, no. 3, pp. 1366-1388.

[18] Hochmair, H. H. Zielstra, D. and Neis, P. (2015). "Assessing the Completeness of Bicycle Trail and Lane Features in Open Street M ap for the U nited S tates." Transactions in GIS. vol. 19, no. 1, pp. 63-81.

[19] Tenney, M. (2014). "Quality evaluations on Canadian OpenStreetMap data." Spatial knowledge and information.

[20] Meier, J. C. H. (2015). "An Analysis of Quality for Volunteered Geographic Information." Master of Science, Faculty of Arts Geography and Environmental Studies, Wilfrid Laurier University.

[21] Zhang, H. and Malczewski, J. (2018). "Accuracy evaluation of the Canadian OpenStreetMap road networks." International Journal of Geospatial and Environmental Research. vol. 5, no. 2, pp.145-162

[22] Chehreghan, A. and Ali Abbaspour, R. (2018). "An evaluation of data completeness of VGI through geometric similarity assessment." International Journal of Image and Data Fusion. vol. 9, no. 4, pp. 319-337.

[23] Zhou, Q. and Lin, H. (2019). "Investigating the completeness and omission roads of OpenStreetMap data in Hubei, China by comparing with Street Map and Street View." arXiv preprint arXiv:1909.04323.

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

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

[26] Zhou, Q. (2018). "Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation." International Journal of Geographical Information Science. vol. 32, no. 2, pp. 257-281.

[27] Nasiri, A. Ali Abbaspour, R. Chehreghan, A. and Jokar Arsanjani, J. (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, p. 253.

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

[29] Welch, B. L. (1947). "The generalization ofstudent's' problem when several different population variances are involved." Biometrika. vol. 34, no. 1/2, pp. 28-35.

[30] Zhang, M. (2009). "Methods and implementations of road-network matching," doctor engineer, Institute of Photogrammetry and Cartography, Technische Universität München.

[31] Chehreghan, A. R. and Ali Abbaspour, R. (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.

[32] Safra, E. Kanza, Y. Sagiv, Y. and Doytsher, Y. (2013). "Ad hoc matching of vectorial road networks." International Journal of Geographical Information Science. vol. 27, no. 1, pp. 114-153.

[33] Chehreghan, A. and Ali Abbaspour, R. (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.

[34] Fu, Z. and Wu, J. (2008). "Entity matching in vector spatial data." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. 37, no. B4, pp. 1467-72.

[35] Brassel, K. Bucher, F. Stephan, E.-M. and Vckovski, A. (1995). "Completeness." Elements of spatial data quality Elsevier. pp. 81-108.

[36] Kunze, C. Hecht, R. and Hahmann, S. (2013). "Zur Vollständigkeit des Gebäudedatenbestandes von OpenStreetMap ." Kartographische Nachrichten. vol. 63, no. 2/3, pp. 73-81.