نشریه علمی مهندسی فناوری اطلاعات مکانی

سال نهم، شماره نخست، بهار ۱۴۰۰

Vol.9, No.1, Spring 2021

۴۱ -۶۴

مقاله پژوهشی

DOR: 20.1001.1.20089635.1400.9.1.2.9

سنجش میزان کامل بودن داده‌های مکانی داوطلبانه با تمرکز بر داده‌های بلوک‌های ساختمانی (مطالعه موردی: کلانشهر تهران)

رقیه ادبی[۱]، رحیم علی عباسپور[۲]*، علیرضا چهرقان[۳]

۱- دانشجوی کارشناسی ارشد *GIS* – دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی – پردیس دانشکده‌های فنی – دانشگاه تهران

۲- دانشیار دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی – پردیس دانشکده‌های فنی - دانشگاه تهران

۳- استایار دانشکده مهندسی معدن – دانشگاه صنعتی سهند

تاریخ دریافت مقاله: ۱۳۹۹/۰۹/۱۲ تاریخ پذیرش مقاله: ۱۴۰۰/۰۳/۲۲

چکیده

در حال حاضر پروژه *OpenStreetMap (OSM)* یکی از بزرگترین مجموعه داده‌های مکانی داوطلبانه است که در بسیاری حوزه‌های کاربردی مختلف به عنوان مکمل یا جایگزین با داده‌های معتبر استفاده می‌شود. با این حال کیفیت این داده‌ها یکی از موانع استفاده گسترده از آنها بوده است. در این مقاله از میان عناصر مرتبط با کیفیت داده‌های مکانی داوطلبانه، سعی در بررسی کامل بودن داده‌های بلوک ساختمانی *OSM* کلانشهر تهران شده است. برای این منظور از دو رویکرد واحد مبنا و عارضه مبنا که غالبا در مطالعات مشابه بکار می‌روند استفاده شده است. نتایج حاصل از این مطالعه مشابه با سایر مطالعات نشان می‌دهد که برآورد کامل بودن داده‌های بلوک ساختمانی *OSM* بشدت بین روش‌ها متفاوت است. نتایج نشان‌دهنده سرعت بالای ارزیابی روش‌های واحد مبنا و دقت بالاتر رویکرد عارضه مبنا است. علاوه بر این روش‌های واحد مبنا نسبت به تفاوت مدلسازی عوارض بین دو مجموعه داده رسمی و داوطلبانه بسیار حساس هستند. درحالیکه روش‌های عارضه مبنا، از حساسیت کمتری برخوردار هستند و در صورتی که داده‌ها صحت مکانی مناسبی داشته باشند منجر به برآورد دقیق‌تری از کامل بودن خواهند شد. بنابراین توصیه این مقاله استفاده از روش‌های عارضه مبنا در مطالعات ارزیابی کیفیت است.در نهایت بر اساس رویکرد عارضه مبنا کامل بودن داده‌های بلوک ساختمانی *OSM* در کل منطقه مطالعاتی برابر با ۲٫۷ درصد محاسبه گردید که نرخ پایین کامل بودن را نشان می‌دهد. ارزیابی بیشتر نتایج نشان می‌دهد که بخش شمالی، مرکزی و شرق تهران نسبت به سایر بخش‌ها از کامل بودن بیشتری برخودار هستند. همچنین در اکثر مناطق تهران بیش از ٪۸۰ از داده‌های رسمی در مجموعه داده *OSM* وجود ندارند که نشان‌دهنده کامل‌تر بودن مجموعه داده رسمی نسبت به مجموعه داده *OSM* می‌باشد.

کلیدواژه‌ها : اطلاعات مکانی داوطلبانه، بلوک ساختمانی، کامل بودن، روش‌های واحد مبنا، روش‌های عارضه مبنا.

* نویسنده مکاتبه کننده: گروه سیستم‌های اطلاعات مکانی، دانشکده مهندسی نقشه‌برداری و اطلاعات مکانی، پردیس دانشکده‌های فنی، دانشگاه تهران، تهران، ایران.

تلفن: ۰۲۱۸۸۰۰۸۸۴۱

Email: abaspour@ut.ac.ir

۱- مقدمه

در سال‌های اخیر تهیه و به‌روزرسانی نقشه‌ها با استفاده از روش‌های متداول نقشه برداری کندتر از قبل صورت می‌پذیرد. یکی از مهمترین دلایل این امر بالا رفتن هزینه تهیه نقشه‌ها از طریق روش‌های مرسوم می‌باشد. در سال‌های اخیر با حضور فراگیر تکنولوژی وب ۲، دسترس بودن سیستم موقعیت‌یابی جهانی و رشد روزافزون نیاز مصرف‌کنندگان عادی به اطلاعات مکانی، فرصتی ایجاد شده است تا شهروندان صرف‌نظر از دانش و تخصص خود، حجم عظیمی از داده‌های مکانی را تولید و در پایگاه‌های داده مکانی به اشتراک بگذارند *[۶]*. این‌گونه اطلاعات مکانی که عموماً توسط مردم عادی تولید می‌شوند، برای اولین بار در سال ۲۰۰۷ توسط گودچایلد تحت عنوان اطلاعات مکانی داوطلبانه[۱](*VGI*) معرفی شد *[۷]*. از سال ۲۰۰۷ تاکنون پروژه‌های مختلفی در زمینه *VGI* به وجود آمده است که می‌توان از پروژه *OSM* به عنوان یکی از موفق‌ترین و بزرگترین پروژه‌ها در این زمینه نام برد. *OSM* یک نقشه بر خط، رایگان و قابل ویرایش توسط کابران با پوشش جهانی تحت وب برای عموم افراد فراهم می‌آورد *[۸]*. با توجه به مشارکت تعداد زیاد کاربران، این مجموعه داده‌ها به صورت پویا به روزرسانی می‌شوند. مزایای فراوان اطلاعات مکانی داوطلبانه همچون پوشش جهانی حتی برای شهرهای کمتر توسعه یافته، قابلیت به روزرسانی سریع و هزینه پایین دسترسی به اطلاعات همواره کاربران را به استفاده از اطلاعات مکانی داوطلبانه سوق داده است *[۷]*. به دلیل انعطاف‌پذیری و ساختار داده باز *OSM* امکان استفاده آن برای طیف گسترده‌ای از برنامه‌های کاربردی وجود خواهد. سیستم‌های ناوبری و مسیریابی *[۹]*، بهبود کارتوگرافی *[۱۰]*، توسعه خدمات مکان مبنا *[۱۱]*و توسعه مدل‌های شهری سه بعدی *[۱۲]* نمونه‌ای از طیف گسترده کاربردهای پیاده‌سازی شده بر داده‌های *OSM* هستند. بلوک‌های ساختمانی از جمله داده‌های *OSM* هستند که می‌توانند در بسیاری از حوزه‌های کاربردی از جمله مدیریت کاربری زمین، ایجاد مدل‌های سه‌بعدی ساختمان، مدلسازی سیستم‌های انرژی شهری و بسیاری از موارد دیگر مورد استفاده قرار گیرند.

با وجود مزایای فراوان محیط *OSM*، کیفیت داده‌های این محیط بزرگترین محدودیت آن است. یکی از دلایل اصلی این موضوع، عدم اطمینان در مورد نحوه جمع‌آوری این داده‌ها می‌باشد *[۱۳]*. علاوه بر این، فقدان استانداردهای یکپارچه، عدم وجود مکانیزم‌های کنترل کیفیت و استفاده از فن‌آوری‌ها و ابزارهای مختلف در محیط *OSM*، کیفیت ناهمگن در این محیط را افزایش می‌دهد *[۱۴]*. بنابراین مهم‌ترین چالش پیش روی این محیط‌ها تردید درمورد کیفیت داده‌های گردآوری شده می‌باشد. از این رو لازم است که کیفیت این اطلاعات بررسی و تضمین شود تا بتوان از آن‌ها در مقابل داده‌های رسمی در کاربردهای مختلف استفاده کرد. سازمان بین‌المللی استاندارد در قالب سری *ISO/TC211* دستورالعمل‌هایی را در زمینه کیفیت داده‌های مکانی تعریف کرده است. بر اساس استاندارد *ISO 19157* کیفیت داده‌های مکانی شامل المان‌های مختلفی می‌باشد که عبارت‌اند از: دقت موقعیتی، کامل-بودن، سازگاری منطقی، دقت معنایی و دقت زمانی *[۱۵]*. در این مطالعه کامل‌بودن که یکی از مهمترین عناصر کیفیت ذکر شده به شمار می‌رود، هدف بررسی این مطالعه در نظر گرفته شده است. کامل‌بودن یک مجموعه داده مکانی به ارتباط میان اشیا درون پایگاه داده با اشیا دنیا واقعی اشاره دارد. علاوه بر ماهیت هندسی عوارض، کامل بودن دامنه اطلاعات توصیفی عوارض و روابط آنها را نیز پوشش می‌دهد *[۲۹]*. این تحقیق با تمرکز بر کامل‌بودن اطلاعات هندسی عوارض ارائه می‌شود. تاکنون تحقیقات مختلفی در مورد روش‌های ارزیابی کامل بودن داده‌ها بر اساس مقایسه با

[۱] *Volunteered Geographic Information*

یک مجموعه داده رسمی انجام شده است [۱۵، ۱۷، ۱۹، ۲۰ و ۲۱]. بر اساس این مطالعات، دو رویکرد واحدمبنا۱ و عارضه‌مبنا۲ برای اندازه‌گیری دقت کامل بودن داده‌ها وجود دارد.

هدف اصلی این تحقیق ارزیابی کامل‌بودن داده‌های بلوک ساختمانی *OSM* شهر تهران با تمرکز بر تجمیع روش‌هایی است که در تحقیقات پیشین استفاده شده است تا با دیدی واقعی‌تر به محاسبه کامل بودن پرداخته شود. همچنین از آنجائیکه کامل بودن شهر تهران در تحقیقات پیشین مورد بررسی قرار نگرفته بود، در این مطالعه سعی شده است ضمن محاسبه کامل بودن، نحوه توزیع این پارامتر برای شهر تهران نشان داده شود. نتایج این مطالعه دید روشنی در خصوص کامل بودن داده‌های بلوک ساختمانی *OSM* مناطق مطالعه شده شهر تهران ارائه می‌دهد. بینش حاصل از این مطالعه می‌تواند در ایجاد آگاهی لازم برای در نظر گرفتن استفاده از این داده‌ها در مقابل داده‌های رسمی در حوزه‌های کاربردی مختلف کمک کند.

ادامه این تحقیق به این شرح است: در بخش دوم پیشینه تحقیق ارائه شده است. در بخش سوم رویکردهای محاسبه کامل‌بودن به تشریح بیان می‌شود. بخش چهارم شامل پیاده‌سازی روش‌های بیان شده در بخش ۳ بر روی منطقه مورد مطالعه می‌باشد و نتایج این پیاده سازی را نشان می‌دهد. در انتها نیز در بخش پنجم نتیجه گیری و پیشنهادها ارائه می‌گردد.

۲- پیشینه تحقیق

نخسین مطالعات مربوط به کامل‌بودن ساختمان‌ها در *VGI* توسط گوتز و زیپف در سال ۲۰۱۲ انجام شده است. آنها تحلیلی کمّی از ساختمان‌های *OSM* در کشور آلمان برای بررسی مناسب بودن داده‌های *OSM* جهت تولید مدل‌های ساختمانی سه‌بعدی ارائه دادند. آنها دریافتند که *OSM* فقط ۳۰ درصد از کل ساختمان‌ها را تحت پوشش خود قرار می‌دهد و فقط ۰/۵ درصد از این ساختمان‌ها اطلاعاتی در مورد ارتفاع یا تعداد طبقات ارائه می‌دهند [۱۶].

فن و همکاران، در سال ۲۰۱۴ کامل‌بودن داده‌های ساختمانی *OSM* آلمان را نسبت به مجموعه داده رسمی به عنوان اختلاف مساحت تحت پوشش ساختمان‌های دو مجموعه داده مورد ارزیابی قرار دادند. همچنین با شمارش تعداد ساختمان‌های *OSM* با اطلاعات توصیفی مانند نام، نوع، ارتفاع و غیره، کامل‌بودن اطلاعات توصیفی داده‌ها هم بررسی شد. نتایج حاصل از این روش نشان داد که داده‌های *OSM* از نظر مساحت تحت پوشش از کامل بودن بالایی برخوردار هستند. زیرا مساحت کل ساختمان‌ها در داده‌های *OSM* کمی بیشتر از داده‌های رسمی است. با این حال، هنوز فاقد اطلاعات توصیفی هستند [۱۷].

فرام و همکاران در سال ۲۰۱۵ کیفیت ساختمان‌های *OSM* در شهرهای شفیلد، لیدز و لندن (انگلستان) را با هدف بررسی کاربرد داده‌های *OSM* در برنامه‌های مدیریت ریسک از جمله برنامه‌های مواجهه با مخاطرات طبیعی ارزیابی کردند. این مطالعه با استفاده از یک رویکرد واحدمبنا از طریق مقایسه‌ای با داده‌های سازمان نقشه‌برداری انگلستان[۳] انجام شد و نشان داد که کامل‌بودن ساختمان‌های *OSM* در شهرهای انگلستان بسیار متغیر است [۲۰].

داکوستا در سال ۲۰۱۶ از دو شاخص ژاکارد[۴] و چکانوفسکی[۵] که میزان شباهت بین دو مجموعه را تعیین می‌کنند برای برآورد کامل‌بودن ساختمان‌های *OSM* استفاده کرده است. این مطالعه نشان داد که کامل‌بودن عوارض ساختمانی در مراکز شهر نسبتا زیاد(حدود ۹۷٪) است و مقدار آن (کمتر از ۶٪) با

[۱] *Unit-based*
[۲] *Object-based*
[۳] *Ordance Survey*
[۴] *Jaccard*
[۵] *Czekanowski*

دورشدن از مراکز شهری به سرعت کاهش می‌یابد [۲۱].

الحموی و همکاران در سال ۲۰۱۷ اولین ارزیابی مربوط به کیفیت داده‌های *OSM* جهت مدلسازی سیستم‌های انرژی شهری را انجام دادند. برای این منظور دو مجموعه داده کاربری زمین و ساختمان‌ها را مورد تجزیه و تحلیل قرار دادند. نتایج نشان داد اگرچه بعضی از اطلاعات همچون نوع فعالیت و کاربری ساختمان را نمی‌توان از مجموعه داده *OSM* استخراج کرد، این مجموعه داده منبع بزرگی از داده‌های باز را تشکیل می‌دهد که می‌تواند به یک مدل‌سازی پایدار کمک کند [۱۸].

تیان و همکاران در سال ۲۰۱۹ با استفاده از دو شاخص تعداد و تراکم ساختمان‌های *OSM*، کامل بودن، سیر تکامل و الگوی مکانی داده‌های ساختمان *OSM* در چین را طی سال‌های ۲۰۱۲ تا ۲۰۱۷ بررسی کردند. نتایج نشان داده است که داده‌های ساختمانی *OSM* تقریبا ۲۰ برابر افزایش یافته است با این حال داده‌های ساختمانی *OSM* در چین هنوز کاملاً کامل نیستند. توزیع تراکم داده‌های ساختمان *OSM* را برای شش کلانشهر در چین ، برای سال ۲۰۱۷ نشان می‌دهدکه بیشتر سلول‌های شبکه که تراکم ساختمان *OSM* آنها از ۲۵درصد بزرگتر است در مراکز خرید یا مراکز توریستی پراکنده شده‌اند [۲۹].

عشقی و آل شیخ در سال ۱۳۹۴ دقت معنایی عوارض خطی داده‌های داوطلبانه منطقه ۵ شهرداری شهر تهران را مورد ارزیابی قرار دادند. برای این منظور، برچسب‌های ارائه شده برای "نوع داده" در مجموعه داده رسمی و داوطلبانه بررسی شده و برچسب‌هایی که دارای بالاترین شباهت بودند به عنوان متناظر شناخته شده و از میان این شباهت‌ها دقت معنایی عوارض ارزیابی گردید. نتایج حاصل از این مطالعه نشان می‌دهد که از میان طیف گسترده‌ای از برچسب‌های تعریف شده از قبل هر کدام از پایگاه داده‌های رسمی و داوطلبانه، تنها تعدادی از برچسب‌ها در منطقه مطالعاتی متناظر یکدیگر هستند، برای مثال برچسب‌های تندرو و ترانک[۱] با ۶۰ عارضه متناظر(درصد تشابه ۸۵/۷٪) بعنوان برچسب‌های مشابه استخراج شده‌اند [۳].

چهرقان و علی‌عباسپور در سال ۱۳۹۶ داده‌های *VGI* مرتبط با شبکه راه‌های درون شهری منطقه شش تهران را از منظر هندسی با مجموعه داده رسمی مورد مقایسه قرار دادند. نتایج این تحقیق نشان داد که عوارض موجود در مجموعه داده *OSM* به طور میانگین ۷۸ درصد شباهت مکانی با مجموعه داده رسمی دارند. همچنین جهت ارزیابی سیر بهبود دقت هندسی مجموعه داده‌های *VGI*، شبکه راه‌های درون شهری در سال‌های ۲۰۱۴ تا ۲۰۱۷ مورد ارزیابی قرار گرفت. نتایج نشان داد که که ضمن افزایش ۲۷/۸درصدی مشارکت مردم در ترسیم عوارض، میانگین درصد شباهت مکانی نیز به مقدار ۵/۷۷ درصد بهبود یافته است. از این رو محققین در بسیاری از کاربردهای مرتبط با حوزه حمل و نقل می‌توانند از مجموعه دادهای *VGI* به عنوان جایگزینی برای مجموعه داده رسمی استفاده کنند [۵].

عسگری و همکاران در سال ۱۳۹۹ کیفیت اطلاعات مکانی داوطلبانه مربوط به تغییرات کاربری اراضی شهر قزوین را با استفاده از تصاویر ماهواره‌ای بعنوان مجموعه داده رسمی مورد ارزیابی قرار دادند. در این پژوهش یک روش واحد مبنا(براساس نسبت تعداد عوارض در دو مجموعه داده) و دو روش عارضه مبنا(روش مرکزی و همپوشانی) برای بررسی کامل بودن داده‌ها ارائه گردید. نتایج تحقیق حاکی از این است که میزان کامل‌بودن داده‌های داوطلبانه در سال‌های مختلف متفاوت می‌باشد و هر چقدر از زمان حال دورتر می‌شویم از دقت آن کاسته می‌شود به طوری که در سال ۱۳۹۷، ۷۲/۲ درصد از داده‌های داوطلبانه با

[۱] *trunk*

داده‌های مرجع کاملا همپوشانی داشت اما در سال ۱۳۸۶ این رقم ۲۵ درصد کاهش یافت [۴].

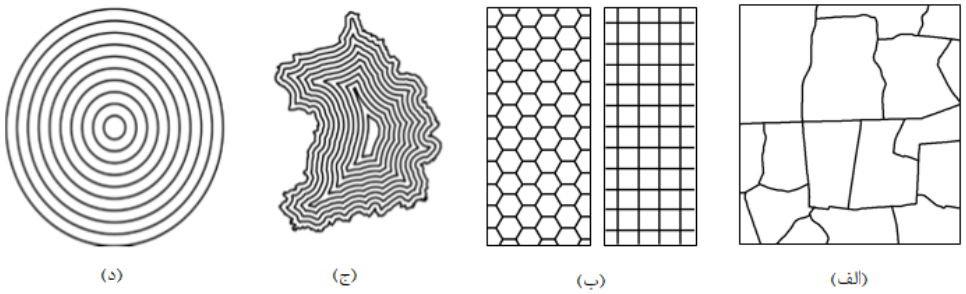
۳- رویکرد پیشنهادی

به طور کلی روش‌های مختلفی در ارتباط با کامل بودن اطلاعات مکانی داوطلبانه ارائه شده است که می‌توان در دو رویکرد کلی واحد مبنا و عارضه مبنا خلاصه کرد. در ادامه هر یک از این رویکردها به تفصیل تشریح شده‌اند.

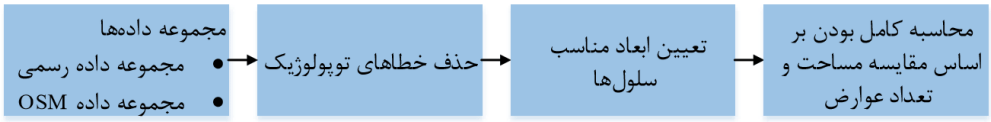
۳-۱- رویکرد واحد مبنا

برخلاف مجموعه داده‌های رسمی، در مجموعه داده داوطلبانه پارامترهای کیفیت به صورت ناهمگن در سراسر منطقه تغییر می‌کند. ازاین رو ارزیابی پارامتر کیفیت برای کل ناحیه مورد بررسی نمی‌تواند میزان قابل قبولی باشد. برای حل این مشکل رویکرد واحد مبنا معرفی شده است که پارامترهای کیفیت را در یک واحد مکانی مشخص با مساحت کوچکتر محاسبه می‌کند. شکل(۱) یک نمای کلی از واحدهای مکانی مختلف را نشان می‌دهد.

در بیشتر تحقیقات از الگوهای شبکه شش ضلعی منتظم و یا مرز مربوط به مناطق شهری استفاده شده است [۱۹، ۲۰، ۲۱و ۲۲]. از مزایای به کارگیری این دسته از روش‌ها می‌توان به محاسبات کمتر و سرعت پردازش بالاتر آنها اشاره نمود. شکل(۲) روند کلی رویکرد واحد مبنا را نشان می‌دهد.



شکل۱: (الف) مرزهای محلی و اداری (ب) شبکه منظم (مربعی_شش ضلعی) (ج) حریم مرزهای شهری (د) دوایر هم مرکز



شکل۲: روند کلی رویکرد واحد مبنا جهت ارزیابی کامل بودن

۳-۱-۱- حذف خطاهای توپولوژیک

یک مرحله حذف خطاهای توپولوژی بر روی هر دو مجموعه داده *OSM* و رسمی باید انجام گیرد. با توجه به تمرکز این تحقیق بر داده‌های چندضلعی این خطاها عبارت‌اند از: عوارض تکراری، همپوشانی عوارض و پلیگون‌های زائد. اینگونه خطاها باعث اختلال در نتایج ارزیابی می‌گردد؛ به همین دلیل نیاز است قبل از استفاده از داده‌ها، خطاها شناسایی و حذف شوند.

۳-۱-۲- تعیین ابعاد مناسب سلول‌ها

تعیین الگو و ابعاد مناسب سلول‌ها در عملیات ارزیابی واحدمبنا فرایندی کاملا تجربی است و رابطه ریاضی مشخصی برای آن وجود ندارد. در مقالات مختلف این مقدار در بازه عددی ۱۰۰ تا ۵۰۰ متر انتخاب شده است [۱۹و۲۰].

۳-۱-۳- محاسبه کامل بودن

در رویکرد واحد مبنا می‌توان از مقایسه مساحت و تعداد عوارض، کامل بودن عوارض چندضلعی را محاسبه کرد. در ادامه هر یک از معیارهای اندازه‌گیری کامل بودن به تفصیل تشریح شده‌اند.

۳-۱-۳-۱- معیار مساحت

در بیشتر مقالات از شاخص نسبت مجموع مساحت عوارض در دو مجموعه داده برای محاسبه کامل‌بودن عوارض چندضلعی استفاده شده است [۱۵، ۱۹ و ۲۰]. در این روش کامل‌بودن داده‌های ساختمانی *OSM* نسبت به داده‌های رسمی پس از محاسبه مجموع مساحت ساختمان‌ها در هر ناحیه از طریق رابطه(۱) قابل محاسبه است.

$$C_{area} = \frac{\sum BuildingArea_{OSM}}{\sum BuildingArea_{\mathrm{Ref}}} \times 100 \qquad \text{رابطه(۱)}$$

در رابطه (۱)، $\sum BuildingArea_{OSM}$ مجموع مساحت ساختمان‌های مجموعه داده *OSM* و $\sum BuildingArea_{\mathrm{Ref}}$ مجموع مساحت ساختمان‌های مجموعه داده رسمی می‌باشد.

برخلاف کاربرد وسیع مقایسه مساحت دو مجموعه داده در ارزیابی کامل‌بودن، برخی خطاها در مدل‌سازی می‌تواند تاثیر نادرست بر ارزیابی نتایج داشته باشد. شکل(۳) نمونه‌ای از خطای تاثیرگذار بر معیار نسبت مساحت را نشان می‌دهد که ترسیم ساده ساختمان‌های پیچیده در *OSM* منجر به سطح بالاتری از کامل بودن می‌شود. این خطا در مدلسازی توسط سه مفهوم مثبت صحیح[۱] (مساحت ساختمان‌های مرجع که به درستی در *OSM* ترسیم شده‌اند)، منفی کاذب[۲] (مساحت ساختمان‌های مرجع که در *OSM* ترسیم نشده‌اند) و مثبت کاذب[۳] (مساحت ساختمان‌های *OSM* که در مجموعه داده‌های مرجع نقشه برداری نشده‌اند) بررسی می‌گردد [۱۹].

بر اساس سه مفهوم ذکرشده، می‌توان تخمین دقیق‌تری از پارامترهای کیفیت را بصورت رابطه‌های (۲) تا (۴) محاسبه کرد.

$$TP_{rate} = \frac{TP}{TP + FN} \times 100 \qquad \text{رابطه(۲)}$$

$$FP_{rate} = \frac{FP}{TP + FN} \times 100 \qquad \text{رابطه(۳)}$$

$$FN_{rate} = \frac{FN}{TP + FN} \times 100 \qquad \text{رابطه(۴)}$$

که در روابط(۲) تا (۴)، نرخ مثبت صحیح (TP_{rate})، نمایانگر داده‌های موجود در هر دو مجموعه داده *OSM* و مرجع است. نرخ منفی کاذب (FP_{rate})، داده‌های اضافی موجود در *OSM* را نشان می‌دهد. نرخ مثبت کاذب (FN_{rate})، نشانگر داده‌هایی است که در *OSM* موجود نیست.

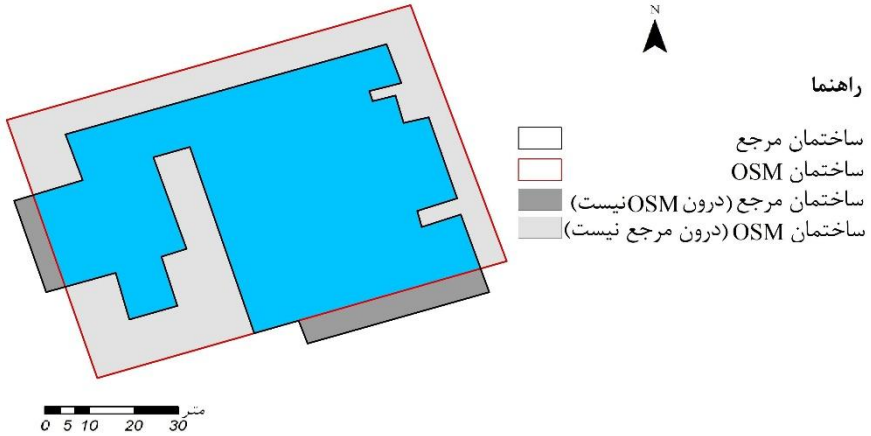
۳-۱-۳-۲- معیار تعداد عوارض

کامل بودن داده‌های چندضلعی را می‌توان از نسبت تعداد عوارض چندضلعی مجموعه داده داوطلبانه به رسمی بررسی کرد [۱۷]. در این روش پس از تقسیم ناحیه مورد بررسی به زیر نواحی کوچکتر، کامل‌بودن داده‌های ساختمانی *OSM* نسبت به داده‌های رسمی از طریق رابطه (۵) محاسبه می‌گردد.

$$C_{count} = \frac{\sum Building_{OSM}}{\sum Building_{\mathrm{Ref}}} \times 100 \qquad \text{رابطه(۵)}$$

در رابطه(۵) ، $\sum Building_{OSM}$ مجموع تعداد ساختمان‌های مجموعه داده *OSM* و $\sum Building_{\mathrm{Ref}}$ مجموع تعداد ساختمان‌های مجموعه داده رسمی می‌باشد.

[۱] *True Positive*
[۲] *False Negetive*
[۳] *False Positive*



شکل۳: نمونه‌ای از خطای تاثیرگذار بر معیار مساحت

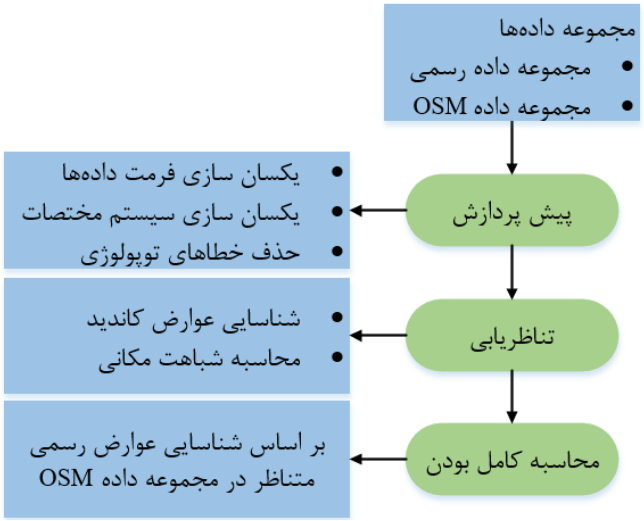
روش نسبت تعداد در مواردی می‌تواند کامل‌بودن را بیش از حد واقعی برآورد کند، به این دلیل که داده‌های رسمی معمولاً با درجه وضوح بالاتر اخذ می‌شوند. به طور خاص، ساختمان‌های منفرد یا مجتمع‌های ساختمانی که دارای بخش زیادی هستند، با جزئیات کافی در مجموعه داده رسمی ثبت می‌گردند که منجر به تعداد ساختمان‌های بیشتر می‌گردد. مثال شکل(۴) نشان می‌دهد که ساختمان‌های مجاورهم به عنوان ساختمان‌های جدا از هم در مجموعه داده‌های رسمی نشان داده شده‌اند (شکل (۴-ب)). با این حال، در *OSM*ساختمان‌های مجاور به عنوان یک ساختمان با طول طولانی ترسیم شده‌اند (شکل (۴-الف)). در مواردی از این دست، روش نسبت تعداد، کامل‌بودن را کمتر از حد واقعی برآورد می‌کند. بنابراین اگر قبل استفاده از این روش، ساختمان‌های مجاور هم در دو مجموعه داده ادغام شوند، تخمین دقیق‌تری بدست خواهد آمد. البته این رویکرد ممکن است مسئله دیگری ایجاد کند که ترسیم غیر دقیق یا نادرست داده‌های *OSM* ممکن است مانع شناسایی صحیح ساختمان‌های مجاور شود، در نتیجه سبب می‌گردد برآورد کامل بودن افزایش یابد.[۱۹].

ب الف

شکل۴: (الف) چند ضلعی‌های ساختمان *OSM* و (ب) چند ضلعی‌های ساختمان مرجع

۳-۲- رویکرد عارضه مبنا

مهم‌ترین گام در این روش، شناسایی عوارض مشابه در هر دو مجموعه داده مورد بررسی است. برای این منظور تناظریابی عوارض باید انجام شود. منظور از تناظریابی داده‌های مکانی فرآیند شناسایی عوارض مشابه بین مجموعه داده‌های مختلف می‌باشد که این عوارض مشابه نشان‌دهنده یک موجودیت یکسان در دنیای واقعی هستند. مسائل تناظریابی در مدل داده‌برداری بر اساس ماهیت داده تناظریابی شده، به سه دسته نقطه‌ای، خطی و چندضلعی تقسیم‌بندی می‌شوند [۲۳]. همچنین معیارهای اندازه‌گیری شباهت که جهت مقایسه میزان نزدیکی دو عارضه به یکدیگر در نظر گرفته می‌شوند، به سه دسته هندسی، معنایی و روابط مکانی تقسیم‌بندی می‌شوند [۲۴]. در این تحقیق جهت تناظریابی عوارض چندضلعی تنها از معیارهای هندسی استفاده می‌شود، زیرا در تمامی مجموعه داده‌ها امکان دسترسی به این معیارها وجود دارد و نیاز به اطلاعات اضافه به جز هندسه عوارض نیست. روند کلی رویکرد عارضه مبنا در شکل (۵) نمایان است که شامل سه مرحله اصلی می‌باشد.



شکل۵: روند کلی روش شی مبنا جهت ارزیابی کامل بودن

۳-۲-۱- پیش پردازش

طبق این ساختار ابتدا دو مجموعه داده رسمی و *OSM* به عنوان ورودی وارد فرآیند تناظریابی می‌شوند. سپس به دلیل تنوع منابع تولید داده نیاز به مرحله پیش پردازش می‌باشد. این مرحله شامل یکسان‌سازی قالب داده‌ها، تبدیل به سیستم مختصات یکسان و حذف خطاهای توپولوژیکی می‌باشد.

۳-۲-۲- تناظریابی

پس از مرحله پیش‌پردازش، درجه شباهت مکانی عوارض چندضلعی بر اساس معیارهای هندسی محاسبه می‌گردد، برای این منظور ابتدا نیاز به شناسایی عوارض کاندید مجموعه داده *OSM* برای هر عارضه از مجموعه داده رسمی می‌باشد تا شباهت مکانی تنها برای عوارض کاندید محاسبه شود و زمان محاسباتی اجرای الگوریتم کاهش یابد. در این پژوهش از کوچکترین مستطیل در برگیرنده عارضه(*MBR*[۱]) جهت محاسبه عوارض کاندید استفاده می‌شود. بدین ترتیب برای هر عارضه چندضلعی از مجموعه داده رسمی، کوچکترین مستطیل در برگیرنده در نظر گرفته می‌شود و عوارض چندضلعی از مجموعه داده *OSM* که حداقل یکی از

[۱] *Minimum Bunding Rectangle*

رئوس آنها درون محدوده این مستطیل قرار گیرد، به عنوان عوارض کاندید انتخاب شده و شباهت مکانی تنها برای آن عوارض با عارضه مذکور از مجموعه داده رسمی محاسبه می‌شود [۲۵].

پس از شناسایی عوارض کاندید، شباهت مکانی میان هر عارضه کاندید با عارضه مجموعه داده رسمی بصورت ترکیب وزنداری از معیارهای هندسی محاسبه می‌شود. همانطور که از رابطه(۶) ملاحظه می‌شود، شباهت نهایی دو عارضه چندضلعی مرجع و OSM برابر با ترکیب وزنداری از معیارهای شباهت مکانی σ_i است که در این تحقیق سه معیار مساحت همپوشانی، فاصله و جهت در نظر گرفته شده‌اند ($n=3$). همچنین وزن هر معیار (w_i) به کمک نظر کارشناسی انجام می‌پذیرد. پس از محاسبه شباهت مکانی برای تمام عوارض کاندید، عارضه‌ای به عنوان عارضه متناظر نهایی شناخته می‌شود که مقدار درجه شباهت مکانی آن بیشتر از حد آستانه تعریف شده باشد. درصورتیکه این مقدار از حد آستانه موردنظر کمتر باشد، عارضه بدون عضو متناظر شناخته می‌شود [۱].

رابطه(۶)
$$S(osm, ref) = \frac{\sum_{i=1}^{n} w_i \sigma_i (osm, ref)}{\sum_{i=1}^{n} w_i}$$

در انتها نیز روند ذکرشده برای تمامی چندضلعی‌های مجموعه داده رسمی اجرا می‌گردد تا خروجی که شامل عوارض متناظر از مجموعه داده OSM برای هر عارضه از مجموعه داده رسمی می‌باشد، حاصل شود. در ادامه هر یک از معیارهای اندازه‌گیری شباهت به تفصیل تشریح شده‌اند.

۳-۲-۲-۱- مساحت همپوشانی عوارض

مساحت همپوشانی دو چندضلعی یکی از معیارهایی است که در اکثر مطالعات تناظریابی در مرحله شناسایی شباهت دو عارضه یا در مرحله انتخاب عوارض کاندید استفاده شده است [۲۶و۲۲]. در این مقاله درجه شباهت بر اساس مساحت همپوشانی طبق رابطه(۷) تعریف شده است.

رابطه(۷)
$$\sigma_{Area} = \frac{Area_{overlap}}{min\left(Area_{osm}, Area_{ref}\right)}$$

که $Area_{overlap}$ مساحت ناحیه همپوشانی دو چندضلعی مرجع و OSM، $min(.,.)$ حداقل مقدار بین مساحت دو چندضلعی مرجع و OSM است.

۳-۲-۲-۲- معیار فاصله

فاصله اقلیدسی [۱و ۲۳] و فاصله هاسدروف[۲۷] از جمله پرکاربردترین معیارهای مورد استفاده در مسائل تناظریابی عوارض چندضلعی هستند. در این مطالعه از فاصله اقلیدسی بین دو عارضه چندضلعی (فاصله بین مراکز ثقل دو عارضه) استفاده شده است. در نتیجه با وجود مراکز ثقل دو عارضه چندضلعی OSM و رسمی که با مختصات $P_{osm} = (x_{osm}, y_{osm})$ و $P_{ref} = (x_{ref}, y_{ref})$ نشان داده می‌شوند، درجه شباهت این دو چند ضلعی بر اساس فاصله اقلیدسی طبق رابطه (۸) محاسبه می‌شود.

رابطه(۸)
$$\sigma_{Dis}\left(P_{osm}, P_{ref}\right) = 1 - \frac{\sqrt{\left(x_{osm} - x_{ref}\right)^2 + \left(y_{osm} - y_{ref}\right)^2}}{U}$$

در رابطه(۸) U حداکثر فاصله ممکن بین دو عارضه چندضلعی می‌باشد که بعنوان فاکتور نرمال‌سازی رابطه استفاده می‌شود.

۳-۲-۲-۳- معیار جهت

جهت یک چند ضلعی جزء معیارهایی است که در رابطه با ویژگی‌های هندسی یک عارضه تعریف می‌شود. با توجه به شکل(۶) جهت یک عارضه چندضلعی برابر با زاویه‌ای است که قطر *MBR* عارضه با محور افقی می‌سازد [۱و ۲۸]. از آنجایکه این زاویه در محدوده $(0, \frac{\pi}{2})$ قرار دارد، درجه شباهت دو عارضه بر اساس رابطه(۹) تعریف می‌شود.

رابطه(۹)
$$\sigma_{Dir}\left(osm, ref\right) = 1 - \frac{\left|\theta_{osm} - \theta_{ref}\right|}{\frac{\pi}{2}}$$

در رابطه(۹)، θ_{osm} و θ_{ref} به ترتیب برابر با جهت چندضلعی رسمی و *OSM* و اختلاف جهت دو عارضه می‌باشد.

۳-۲-۳- کامل بودن

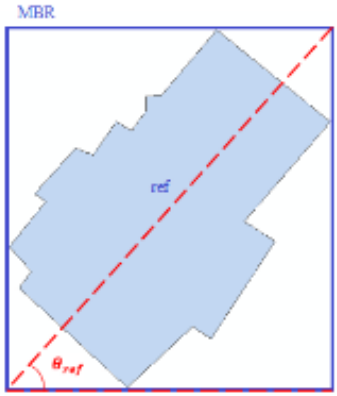
پس از انجام فرآیند تناظریابی، عوارض در دو مجموعه داده به دو دسته دارای متناظـر و بدون متناظـر تقسیم می‌شوند. از این داده‌ها برای تعیین پارامترهای کیفیت داده‌های مکانی استفاده می‌گردد. تعداد عوارض دارای متناظر در هر مجموعه داده به تعداد کل عوارض آن مجموعه داده، میزان حضور داده‌های یک مجموعه داده در مجموعه 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 مقادیر همان میزان کامل بودن داده‌های داوطلبانه نسبت به داده‌های رسمی می باشد.

براساس استاندارد *ISO19157* مولفه کامل بودن را می‌توان به دو مولفه خطای وجود[۱] و عدم وجود[۲] تقسیم کرد. خطای وجود به عوارضی اشاره دارد که در مجموعه داده مرجع وجود داشته ولی در مجموعه داده داوطلبانه وجود ندارد. خطای عدم وجود نماینده عوارضی از مجموعه داده داوطلبانه است که در مجموعه داده مرجع وجود ندارد (شکل (۷)). در جدول (۱) نحوه تعیین کامل بودن و دو خطای وجود و عدم وجود ذکر شده است.

۴- پیاده‌سازی

به منظور محاسبه کامل‌بودن مجموعه داده‌های داوطلبانه از مجموعه داده‌های رسمی مناطق ۱ تا ۲۰ تهران به سال ۱۳۹۶ در مقیاس ۱:۲۰۰۰ استفاده شده است. بدین منظور داده‌های داوطلبانه در تاریخ ۲۰۲۰/۰۹/۱۰ از وب سایت *Planet* (*https://planet.openstreetmap.org*) دانلود گردید. سپس با استفاده از نرم افزار مبدل[۳] داده‌های منطقه مورد مطالعه جدا گردید و سپس با استفاده از نرم افزار آرک جی آی اس[۴] داده‌های بلوک ساختمانی از آن استخراج گردید. شکل‌های (۸-الف) و (۸-ب) به ترتیب بلوک‌های ساختمانی مجموعه داده رسمی و *OSM* را در منطقه مورد مطالعه و جدول (۲) برخی از مشخصات کمّی این دو مجموعه داده نظیر وسعت منطقه، تعداد و مساحت عوارض موجود در ناحیه را نشان می‌دهد.

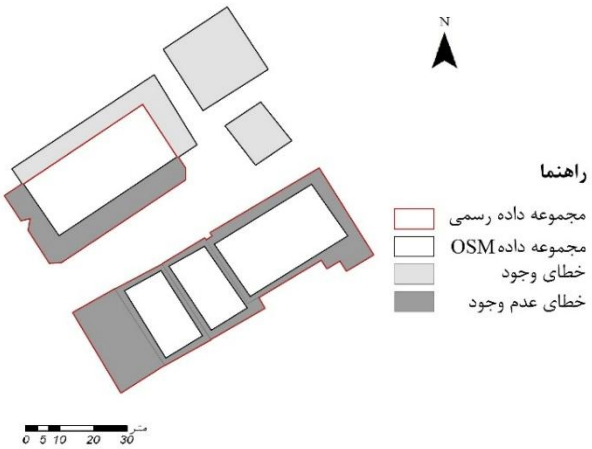
[۱] *Omission*
[۲] *Commission*
[۳] *OSMConvert*
[۴] *ArcGIS*



شکل۶: نمایش جهت چندضلعی مرجع

جدول۱: نحوه محاسبه کامل بودن،خطای وجود و عدم وجود در رویکرد عارضه مبنا

کامل بودن	تناظریابی داده‌های رسمی با داده‌های داوطلبانه استخراج عوارض رسمی دارای متناظر در مجموعه داده داوطلبانه محاسبه تعداد این عوارض به تعداد کل عوارض مجموعه داده رسمی
خطای وجود	تناظریابی داده‌های رسمی با داده‌های داوطلبانه استخراج عوارض رسمی بدون متناظر در مجموعه داده داوطلبانه محاسبه تعداد این عوارض به تعداد کل عوارض مجموعه داده رسمی
خطای عدم وجود	تناظریابی داده‌های داوطلبانه با داده‌های رسمی استخراج عوارض داوطلبانه بدون متناظر در مجموعه داده داوطلبانه محاسبه تعداد این عوارض به تعداد کل عوارض مجموعه داده داوطلبانه



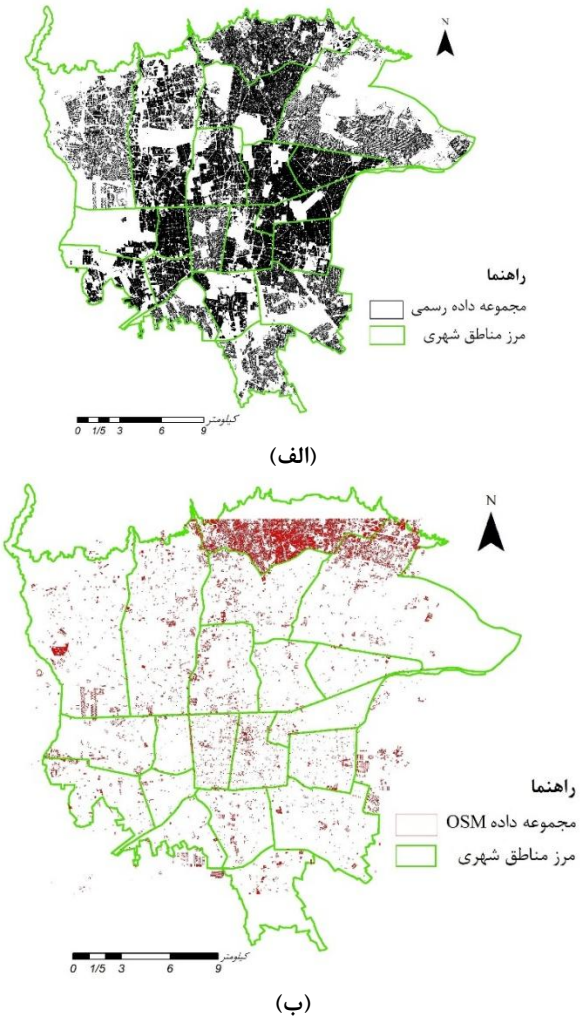
شکل۷: خطای وجود و عدم وجود

جدول۲: مشخصات کمی مجموعه داده رسمی و *OSM*

پارامترها	رسمی	داوطلبانه
مساحت ناحیه (کیلومتر مربع)	۴۹۷٫۳	۴۹۷٫۳
تعداد عوارض	۵۱۱۸۱۶	۳۰۱۹۶
مساحت عوارض (کیلومتر مربع)	۱۸۹٫۷	۱۷٫۸۵

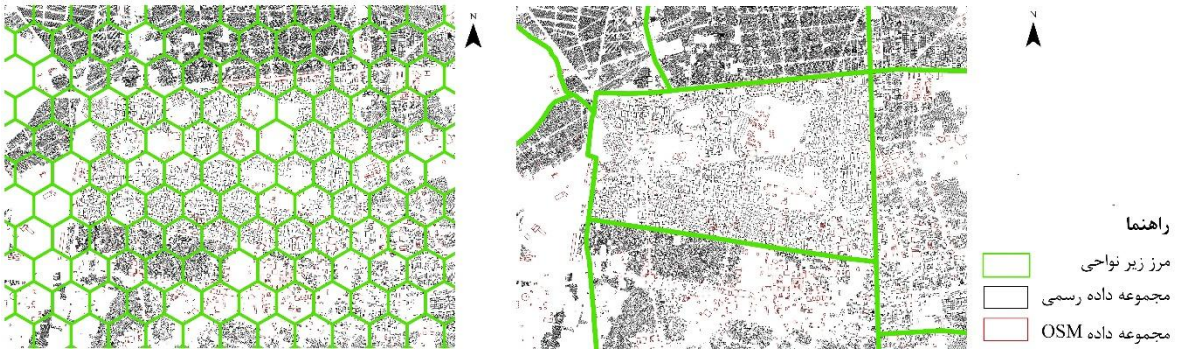
۴-۱- کامل بودن بر اساس رویکرد واحد مبنا

اولین گام در روش‌های واحد مبنا، تقسیم ناحیه به یکسری زیرنواحی می‌باشد. همانطورکه بیان شد، الگوهای متنوعی برای تقسیم ناحیه مورد بررسی به زیربخش‌های کوچکتر وجود دارد. نمایشی از نحوه تقسیم منطقه مورد مطالعه به زیر مناطق کوچکتر با استفاده از دو الگوی شبکه شش ضلعی منتظم و مرز مناطق شهری در شکل (۹) آورده شده است.



شکل۸: بلوک‌های ساختمانی در مجموعه داده‌های *OSM* (الف) و رسمی (ب)

با توجه به اینکه یک عارضه ممکن است در دو یا چند ناحیه قرار داشته باشد، برای محاسبه کامل بودن عوارض به روش نسبت تعداد، تعداد عوارض در هر ناحیه براساس مرکز ثقل عارضه بدست می‌آید(بر این اساس که یک عارضه در صورتی درون یک ناحیه قرار دارد که مرکز ثقل عارضه درون ناحیه قرار داشته باشد). از این رو، هر عارضه فقط یک بار شمارش می‌شود. همچنین هنگام محاسبه کامل بودن به روش نسبت مساحت، براینکه مساحت عارضه بیش از یکبار محاسبه نشوند عوارضی که در دو یا چند ناحیه قرار دارند، ابتدا در مرزها جدا می‌شوند. سپس مساحت عوارض درون ناحیه محاسبه می‌گردد.



شکل۹: تقسیم منطقه موردمطالعه به بخش‌های کوچکتر به منظور بررسی پارامترهای کیفیت براساس مرزهای شهری (الف) و شبکه شش ضلعی منتظم (ب)

۴-۱-۱- کامل بودن از طریق معیار مساحت

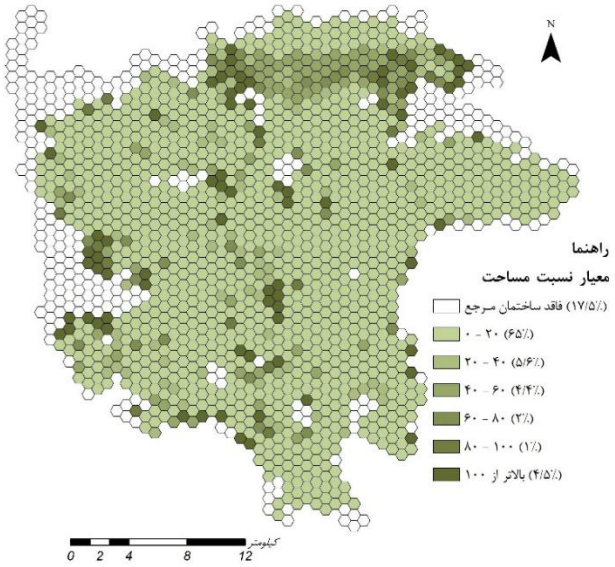
در این روش کامل‌بودن عوارض بر اساس نسبت مساحت عوارض داطلبانه به مساحت عوارض رسمی در هر ناحیه بدست می‌آید(رابطه (۱)). شکل (۱۰-الف) نتایج حاصل از محاسبه کامل‌بودن عوارض بر اساس مناطق شهری را نشان می‌دهد. همانطورکه در این شکل مشاهده می‌شود، منطقه ۱ و ۹ تهران دارای بیشترین مقدار کامل بودن و منطقه ۸، ۱۰ و ۱۳ تهران دارای کمترین مقدار کامل بودن (کمتر از ۳ درصد) هستند. جهت ارزیابی‌های دقیق‌تر، نمایش مقادیر کامل بودن در شبکه شش ضلعی منتظم در شکل (۱۰-ب) ارائه شده است. نتایج نشان می‌دهد بخش شمالی، مرکز و شرق تهران از کامل بودن بیشتری برخوردار هستند. ارزیابی نتایج با جزئیات دقیق‌تر نشان می‌دهد ۱۷/۵ درصد از شش ضلعی‌ها حاوی هیچ ساختمانی در مجموعه داده مرجع نیستند. اکثر شش ضلعی‌ها (حدود ۶۵ درصد) دارای کامل بودن صفر تا ۲۰ درصد هستند.

همچنین حدود ۴/۵ درصد از شش‌ضلعی‌ها دارای کامل‌بودن بیش از ۱۰۰ درصد هستند که نشان می‌دهد مجموعه داده *OSM* در این نواحی کامل‌تر از مجموعه داده مرجع می‌باشد. مقادیر بالاتر از ۱۰۰ درصد حداقل با سه دلیل قابل توضیح است:

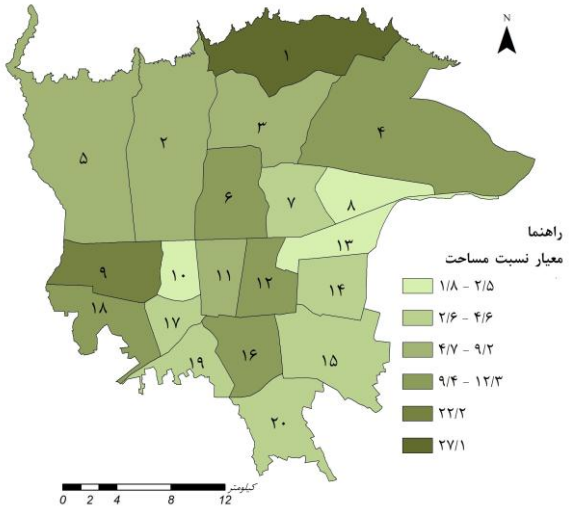
۱- به‌روز بودن داده‌های *OSM* نسبت به داده‌های رسمی. برخی از ساختمان‌های جدید ممکن است ساخته شده باشند که در مجموعه داده‌های رسمی ترسیم نشده‌اند.

۲- اگرچه مجموعه داده رسمی به عنوان واقعیت زمین در نظر گرفته می‌شود، ممکن است حاوی خطا باشد. درحقیقت بیانگر این موضوع است که احتمال دارد برخی از ساختمان‌ها درمجموعه داده رسمی ترسیم نشده باشند.

۳- در *OSM* برخی از عناصر که بلوک ساختمانی نیستند به اشتباه به عنوان بلوک ساختمانی ترسیم شوند.



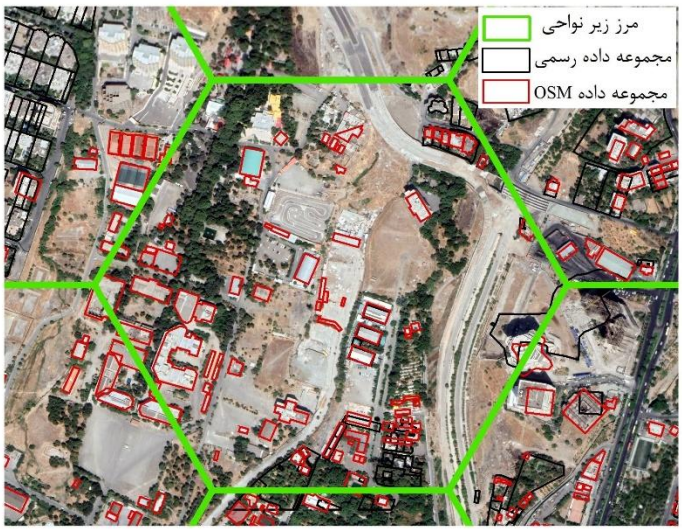
(الف)



(ب)

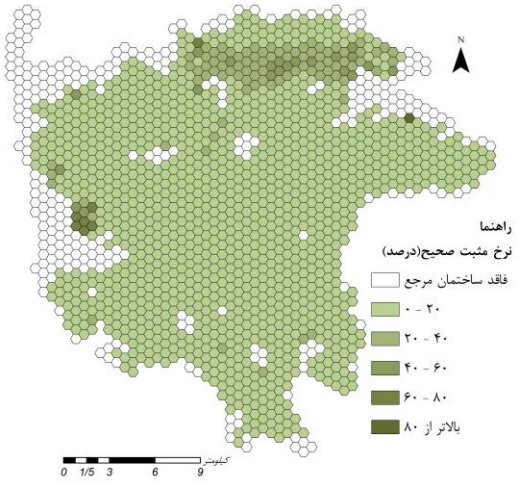
شکل۱۰: درصد کامل بودن بلوک‌های ساختمانی داده *OSM* توسط معیار مساحت در مناطق شهری (الف)، شبکه شش ضلعی (ب)

در این مطالعه کامل‌بودن بیش از ۱۰۰ بدلیل فاصله زمانی بین سال تهیه مجموعه داده رسمی و داوطلبانه می‌باشد. برای مثال شکل (۱۱) یک بررسی از مناطق با درجه کامل‌بودن بیش از ۱۰۰درصد را بر اساس تصویر گوگل ارث (زوم ۲۰ و قدرت تفکیک مکانی ۱متر) نشان می‌دهد. در این شکل داده‌های *OSM* حاوی اطلاعات به‌روزی می‌باشد که در مجموعه داده رسمی وجود ندارند.

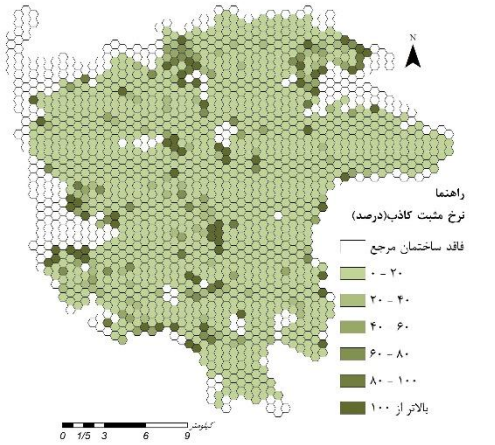


شکل۱۱: نمایش عوارضی از داده‌های مرجع و *OSM* در گوگل ارث

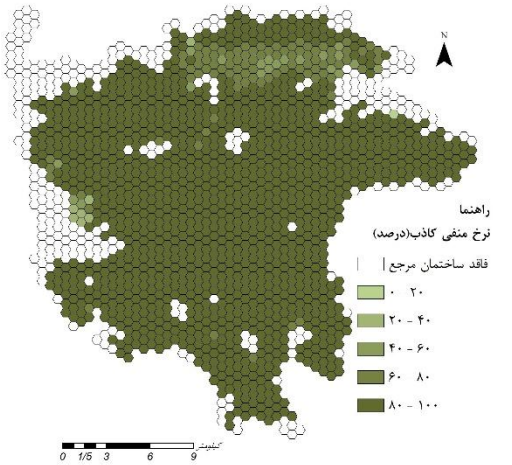
همانطورکه بیان شد، برخلاف کاربرد وسیع مقایسه مساحت دو مجموعه داده در ارزیابی کامل‌بودن، برخی خطاها در مدل‌سازی می‌تواند تاثیر نادرست بر نتایج ارزیابی داشته باشد. در نتیجه با استفاده از سه مفهوم نرخ مثبت صحیح، نرخ مثبت کاذب و منفی کاذب می‌توان تخمین دقیق‌تری از کامل بودن را برآورد کرد. نتایج حاصل از این سه مفهوم در شکل ۱۲ نشان داده شده است. نرخ مثبت صحیح (شکل ۱۲-الف) نمایانگر داده‌های موجود در هر دو مجموعه داده *OSM* و مرجع است. همانطورکه در این شکل مشاهده می‌شود، نرخ مثبت صحیح در بخش شمالی و شرق تهران بیشتر است، با این حال مقدار کمتری را نسبت به کامل بودن در شکل(۱۰_ب) نشان می‌دهد. این واقعیت بیان می‌کند که تمام ساختمان‌های *OSM* در مقایسه با مجموعه داده‌های مرجع به درستی ترسیم نشدند یا اینکه داده‌های *OSM* به روزتر از مجموعه داده مرجع هستند. با وجود این، در بعضی موارد که داده‌های *OSM* دارای صحت مناسبی نباشند، نرخ مثبت صحیح تخمین واقع بینانه‌تری از کامل بودن را نسبت به مقدار برآوردشده با روش نسبت مساحت ارائه می‌دهد. نرخ مثبت کاذب (شکل ۱۲-ب) نمایانگر داده‌های اضافی مجموعه داده داوطلبانه است. همانطور که مشاهده می‌شود بخش شمالی، شرق و مرکز تهران از نرخ مثبت کاذب بیشتری برخوردار هستند. اگر داده‌های *OSM* دارای صحت مناسبی باشند نرخ مثبت کاذب نشان-دهنده به روز بودن داده‌های *OSM* می‌باشد. شکل (۱۲-ج) درصد داده‌هایی که در مجموعه داده مرجع هستند ولی در *OSM* نیستند را نشان می‌دهد. همانطورکه از شکل مشاهده می‌شود اکثر مناطق دارای نرخ منفی کاذب بیش از ۸۰ درصد هستند.



(الف)



(ب)



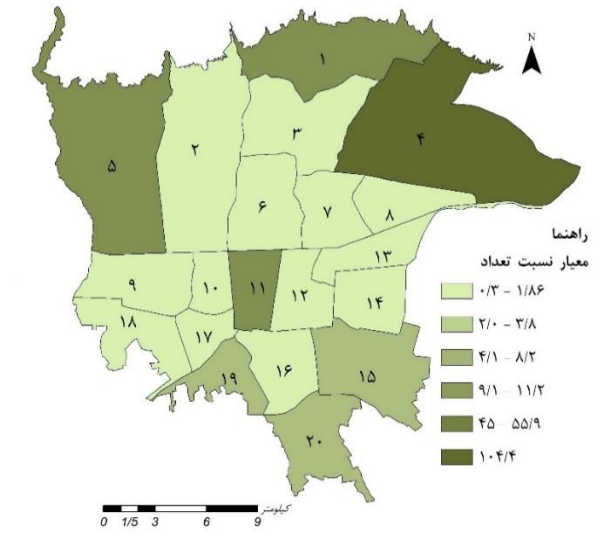
(ج)

شکل۱۲: نرخ مثبت صحیح (الف)،نرخ مثبت کاذب (ب)،نرخ منفی کاذب (ج)

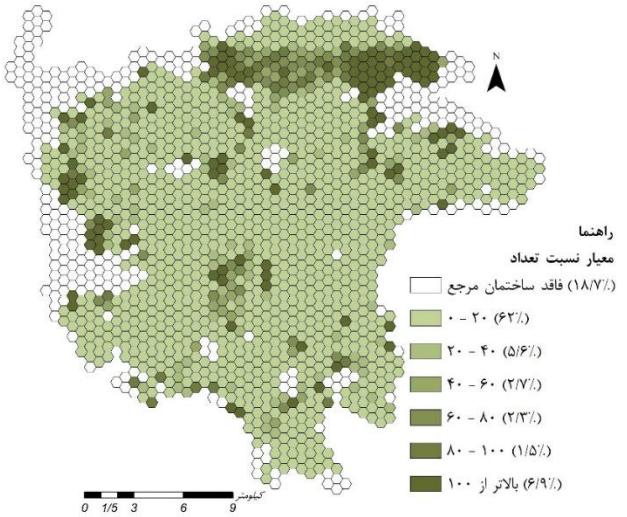
۴-۱-۲-کامل بودن با استفاده از معیار تعداد عوارض

در این روش کامل‌بودن عوارض بر اساس نسبت تعداد عوارض داوطلبانه به تعداد عوارض رسمی در هر ناحیه بدست می‌آید(رابطه (۲)).

شکل (۱۳) نتایج حاصل از محاسبه کامل بودن عوارض را توسط معیار نسبت تعداد عوارض (رابطه (۵)) در مناطق شهری و شبکه شش‌ضلعی منتظم نشان می‌دهد.



(الف)



(ب)

شکل۱۳: درصد کامل بودن مجموعه داده‌های *OSM* توسط معیار تعداد عوارض در مناطق شهری (الف) و شبکه شش ضلعی (ب)

ارزیابی‌های انجام شده بر اساس معیار تعداد عوارض هم‌راستا با نتایج حاصل شده از معیار مساحت (شکل ۱۰-ب) می‌باشد. بخش شمالی، مرکز و شرق تهران ازکامل‌بودن بیشتری برخودار هستند، همچنین اکثر شش ضلعی‌ها (حدود ۶۲ درصد) دارای کامل بودن ۰ تا ۲۰ درصد هستند. ۶/۹ درصد از شش ضلعی‌ها دارای کامل بودن بیش از ۱۰۰٪ می‌باشند. ارزیابی‌های انجام شده در مقیاس مناطق شهری نشان می‌دهد منطقه ۱، ۴، ۵ و ۱۱ تهران از کامل بودن بیشتری نسبت به سایر مناطق شهری برخوردار هستند.

با بررسی روش‌های واحدمبنا بر اساس هر دو معیار مساحت و تعداد عوارض کامل‌بودن در بخش شمالی و مرکز شرق تهران بیشتر و در سایر مناطق بصورت پراکنده می‌باشد. علاوه بر این در بعضی از مناطق تفاوت زیادی در کامل بودن داده‌های *OSM* براساس این دو معیار مشاهده می‌شود که همانطورکه قبلا گفته شد به دلیل تفاوت در مدل‌سازی مجموعه داده‌های *OSM* و مرجع می‌باشد. همچنین نتایج بررسی کامل بودن مجموعه داده داوطلبانه در کل ناحیه مطالعاتی توسط معیارهای مساحت و تعداد عوارض نیز نشان می‌دهد در صورتی‌که از مساحت عوارض به عنوان معیار بررسی کامل بودن بهره گرفته شود، اختلاف ۳/۵درصد با نتایج حاصل شده از معیار تعداد در جدول(۳) مشاهده می‌گردد.

جدول۳: تعیین پارامتر کامل بودن در کل منطقه

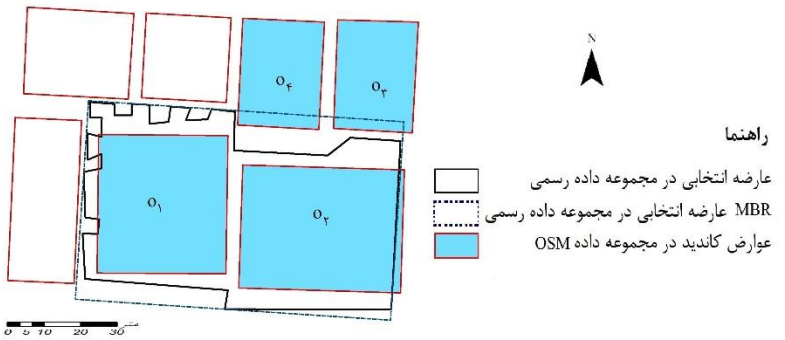
معیار کامل بودن	مجموعه داده رسمی	مجموعه داده *OSM*	کامل بودن (درصد)
معیار مساحت	۱۸۹/۷ کیلو متر مربع	۱۷/۸۵ کیلو متر مربع	۹/۴
معیار تعداد	۵۱۱۸۱۶	۳۰۱۹۶	۵/۹

۴-۲- کامل بودن بر اساس رویکرد عارضه مبنا

مهم‌ترین گام در استفاده از این روش‌ها شناسایی عوارض متناظر در دو مجموعه داده‌های مورد بررسی است. بنابراین عملیات تناظریابی بر روی دو مجموعه داده موجود بر اساس سه معیار هندسی مساحت همپوشانی، فاصله و جهت انجام می‌گیرد. برای تناظریابی، وزن نهایی هر یک از معیارهای شباهت و حدآستانه شباهت با توجه به مطالعه*[۱]* تعیین گردید. جدول (۴) وزن نهایی هر یک از ۳ معیار شباهت را نشان می‌دهد، همچنین مقدار حد آستانه شباهت برابر ۰/۷ در نظر گرفته شد. در ادامه برای هریک از عوارض در مجموعه داده مرجع کاندیدهای ممکن در مجموعه داده *OSM* شناسایی می‌گردد. سپس مقدار درجه شباهت مکانی هر جفت عارضه محاسبه می‌شود. برای مثال شکل(۱۴)، عارضه‌ای در مجموعه داده مرجع به همراه کاندیداهای ممکن در مجموعه داده *OSM* را نشان می‌دهد. همانطورکه در شکل مشاهده می‌شود، با ترسیم کوچکترین مستطیل دربرگیرنده برای عارضه انتخابی در مجموعه داده رسمی، ۴ عارضه در مجموعه داده *OSM* درون این محدوده قرار می‌گیرند و به عنوان عوارض کاندید انتخاب می‌گردند. سپس شباهت مکانی بر اساس سه معیار معرفی شده مساحت همپوشانی، فاصله و جهت برای جفت عوارض محاسبه می‌شود.

جدول۴: ورن نهایی هر یک از معیارها

W_{Area}	W_{Dis}	W_{Dir}
۰/۷۳	۰/۴۵	**۰/۵**



شکل۱۴: نمایش عوارض کاندید در مجموعه *OSM* برای عارضه منتخب در مجموعه داده رسمی

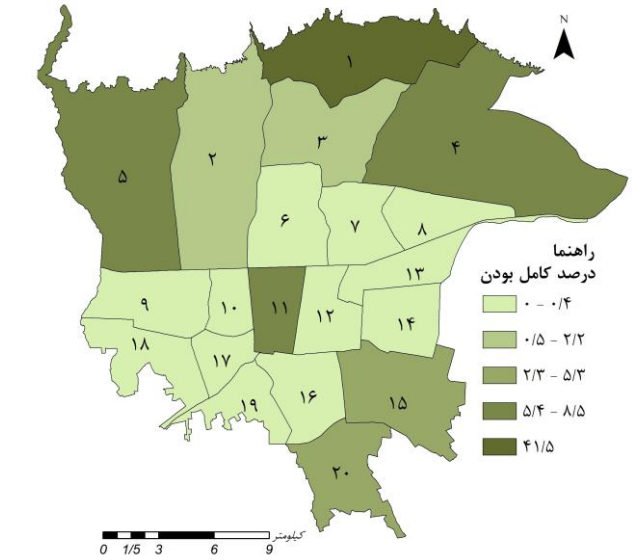
جدول (۵) نتایج محاسبه درجه شباهت مکانی به تفکیک هر معیار را برای هر کدام از این عوارض کاندید در مقایسه با عارضه در مجموعه داده رسمی نشان می‌دهد. بر این اساس عارضه‌های O_1 و O_2 میزان شباهت بیش از ۰/۷ دارند و از میان عوارض کاندید انتخاب می‌شوند.

در نهایت پس از محاسبه درجه شباهت برای تمام عوارض موجود در مجموعه داده رسمی، تعداد ۱۴۱۱۸ عارضه به‌عنوان عوارض متناظر در مجموعه داده داوطلبانه انتخاب شد. با توجه به تعداد عوارض رسمی متناظر در مجموعه داده داوطلبانه، کامل‌بودن برای کل منطقه برابر ۲/۷ درصد محاسبه شد که نرخ پایین کامل‌بودن این عوارض را نشان می‌دهد. تناظریابی عوارض برداری علاوه بر تعیین اطلاعات نظیر دو مجموعه داده، در تعیین پارامترهای خطای وجود و عدم‌وجود داده‌های مکانی به عنوان دو بخش اصلی کامل بودن مفید خواهد بود. از این رو پس از تناظریابی مجموعه داده رسمی نسبت به داوطلبانه، میزان خطای وجود برابر با ۹۷درصد و پس از تناظریابی مجموعه داده داوطلبانه نسبت به رسمی، میزان خطای عدم وجود برابر با ۵۶ درصد محاسبه گردید.

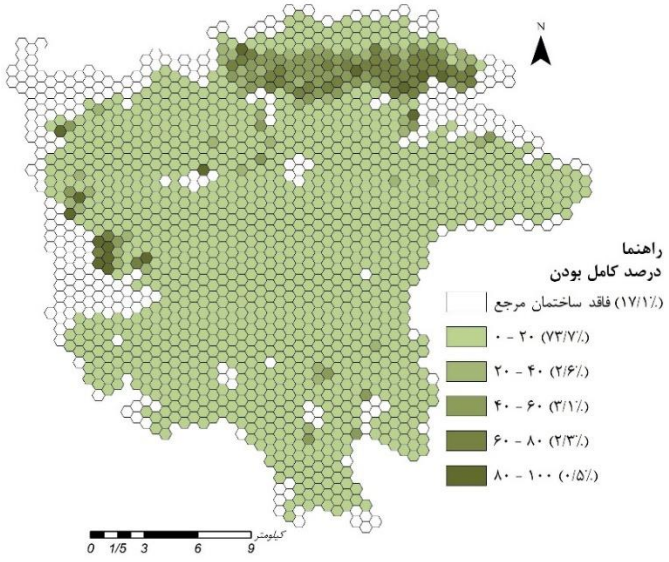
در شکل (۱۵) کامل‌بودن عوارض داوطلبانه در دو سطح جزئیات مناطق شهری و شبکه شش ضلعی منتظم نشان داده شده است. همانطورکه در شکل مشاهده می‌شود اکثر مناطق (۷۳/۷درصد) دارای کامل بودن ۰ تا ۲۰ درصد هستند. مناطق ۱، ۴، ۵ و ۱۱ از کامل بودن بیشتری نسبت به سایر مناطق شهری برخوردار هستند. نتایج حاصل از دو خطای وجود و عدم‌وجود در شکل (۱۶) نشان داده شده است. همانطور که از شکل مشاهده می‌شود، درصد پارامتر خطای عدم‌وجود در بخش شمالی و شرق تهران بیشتر از سایر بخش‌ها است. پارامتر خطای وجود در اکثر مناطق تهران بیش از ۸۰ درصد است که نشان‌دهنده کامل‌تر بودن مجموعه داده رسمی نسبت به مجموعه داده داوطلبانه می‌باشد.

جدول۵: مقادیر محاسبه شده هر یک از معیارها برای عوارض کاندید در مقایسه با عارضه انتخابی

	مقادیر معیارهای هندسی			
جفت عوارض	مساحت همپوشانی	فاصله	جهت	شباهت
R- O1	۰/۹۹۹۷	۰/۵۴۲۱	۰/۸۴۸۷	**۰/۸۳۰۹**
R- O2	۰/۹۵۷۶	۰/۴۹۸۹	۰/۹۴۹۲	**۰/۸۳۲۲**
R- O3	۰	۰/۳۵۷۴	۰/۷۷۸۴	**۰/۳۲۷۴**
R- O4	۰	۰/۳۶۴۶	۰/۷۸۹۹	**۰/۳۳۲۷**



(الف)



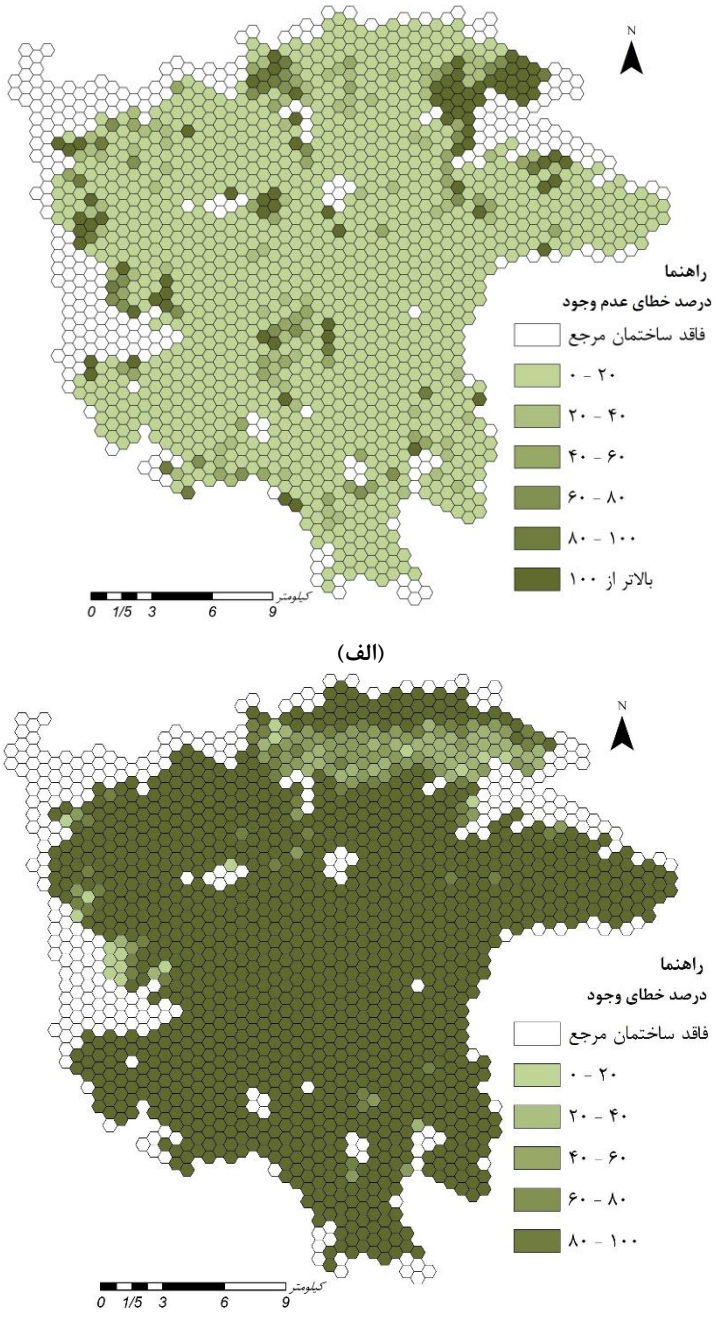
(ب)

شکل۱۵: درصد کامل بودن داده‌های داوطلبانه بر اساس ترکیب دو رویکرد عارضه مبنا و واحد مبنا در مناطق شهری (الف)، شبکه شش ضلعی(ب)

۴-۳- مقایسه رویکرد عارضه مبنا در برابر رویکرد واحد مبنا

با بررسی روش‌های واحدمبنا، در بعضی از مناطق تفاوت زیادی در کامل بودن داده‌های *OSM* بر اساس هر دو معیار مساحت و تعداد عوارض مشاهده گردید. همانطور که گفته شد بدلیل تفاوت مدلسازی عوارض در دو مجموعه داده رسمی و *OSM* می‌باشد. روش تعداد عوارض می‌تواند کامل‌بودن را کمتر از حد واقعی برآورد کند به این دلیل که داده‌های رسمی معمولاً با درجه وضوح بالاتر اخذ می‌شوند. به طور خاص، ساختمان‌های منفرد یا مجتمع‌های ساختمانی که دارای بخش زیادی هستند، با جزئیات کافی در

مجموعه داده رسمی ثبت می‌گردند که منجر به تعداد ساختمان‌های بیشتر می‌گردد. روش مساحت نیز می‌تواند کامل بودن را بیشتر از حد واقعی برآورد کند به این دلیل که در *OSM* عوارض از روی تصاویر هوایی ترسیم می‌شوند.



شکل۱۶: درصد دو پارامتر خطای وجود (الف) و عدم وجود (ب)

به همین خاطر مساحت بام ساختمان بجای کف ساختمان محاسبه می‌گردد، همچنین به دلیل وضوح پایین تصاویر، ممکن است ساختمان‌های پیچیده بصورت ساده ترسیم شوند. با این وجود پیاده‌سازی و پردازش ساده‌تر در رویکرد واحد مبنا فاکتور مهمی در چشم پوشی از محدودیت‌های این رویکرد و بهره‌گیری مکرر آن در مطالعات ارزیابی کیفیت بوده است. البته سرعت پردازشی بالا در رویکرد واحد مبنا در صورتی قابل دستیابی است که مجموعه داده داوطلبانه دارای خطاهای تاثیرگذار بر نتایج معیارهای کیفیت نباشد. برای تفهیم سرعت پردازش هر رویکرد، زمان پردازش هر رویکرد در نرم افزار متلب نسخه ۲۰۲۰ و سخت افزار با حافظه *4GB* و پردازنده مرکزی (*Intel Core i3-5005U 2GHz*) محاسبه شده و نتایج آن در جدول(۶) نشان داده شده است.

جدول ۶: زمان جهت محاسبه هریک از روش‌ها

رویکرد	نسبت تعداد	نسبت مساحت	عارضه مبنا
زمان اجرای برنامه(ثانیه)	۲۰۱	۳۲۳	**۴۴۲۵**

در روش‌های تناظریابی حتما لازم است این نکته را مدنظر قرار داد که، وابستگی بسیار زیاد میان نتایج تناظریابی و میزان صحت مکانی عوارض ورودی برقرار است. به عبارت دیگر در صورت وجود جابجایی مکانی بین دو مجموعه داده تضمینی برای به جواب درست رسیدن در این روش‌ها وجود نخواهد داشت. برخلاف رویکرد عارضه‌مبنا در صورتی که جابجایی دو مجموعه داده به صورت فاحش نباشد امکان استفاده از روش‌های واحد مبنا منجر به خطای زیاد در نتایج این روش‌ها نخواهد شد. رویکرد عارضه مبنا با امکان محاسبه دو خطای وجود و عدم‌وجود، درک دقیق‌تری از پارامتر کامل‌بودن را ارائه می‌دهد. در حالیکه در رویکرد واحد مبنا به واسطه ترکیب عوارض حذف شده و اضافی مجموعه داده داوطلبانه، ارزیابی پارامتر کامل‌بودن بصورت جداگانه با استفاده از معیار نسبت تعداد عوارض، امکان پذیر نخواهد بود.

۵- نتیجه‌گیری

پایگاه داده *OSM* هم از نظر تعداد کاربران و هم از نظر حجم داده‌های مکانی به سرعت در حال توسعه است. از این رو، به یک مجموعه داده جهانی پرکاربرد تبدیل شده است. پایگاه داده *OSM* دارای جنبه‌های مثبتی بوده که می‌توان به دسترس بودن اطلاعات، قابلیت به روزرسانی سریع و هزینه پایین دسترسی به اطلاعات اشاره کرد. *OSM* در حال حاضر یکی از بزرگترین مجموعه داده‌های مکانی داوطلبانه است که در بسیاری حوزه‌های کاربردی مختلف به عنوان مکمل یا جایگزین با داده‌های رسمی استفاده می‌شود. با این حال کیفیت این داده‌ها به عنوان یک چالش مهم می‌باشد. در این مقاله از میان عناصر مرتبط با کیفیت داده‌های مکانی داوطلبانه، سعی در بررسی کامل بودن داده‌های بلوک ساختمانی *OSM* کلانشهر تهران شده است. برای این منظور از دو رویکرد واحد مبنا و عارضه مبنا که غالبا در مطالعات مشابه بکار می‌روند استفاده شده است. نتایج حاصل از این مطالعه نشان داده است که مشابه با سایر مطالعات انجام شده، رویکرد عارضه مبنا منجر به برآورد دقیق‌تری از کامل‌بودن خواهد شد. در نهایت بر اساس رویکرد عارضه‌مبنا کامل‌بودن داده‌های بلوک ساختمانی *OSM* در کل منطقه مطالعاتی برابر با ۲/۷ درصد محاسبه گردید که نرخ پایین کامل بودن این داده‌ها را نشان می‌دهد. ارزیابی بیشتر نتایج نشان می‌دهد اکثر مناطق (۷۳/۷درصد) دارای کامل بودن ۰ تا ۲۰ درصد هستند و مناطق ۱، ۴، ۵ و ۱۱ تهران از کامل‌بودن بیشتری نسبت به سایر مناطق شهری برخوردار هستند. به عبارتی دیگر بخش شمالی،مرکز و شرق تهران نسبت

به سایر بخش‌ها از کامل بودن بیشتری برخودار هستند. همچنین در اکثر مناطق تهران بیش از ۸۰ درصد از داده‌های رسمی در مجموعه داده *OSM* وجود ندارند که نشان‌دهنده کامل‌تر بودن مجموعه داده رسمی نسبت به مجموعه داده *OSM* می‌باشد. بینش حاصل از این مطالعه می‌تواند در ایجاد آگاهی لازم برای در نظر گرفتن استفاده از این داده‌ها در مقابل داده‌های رسمی در حوزه‌های کاربردی مختلف کمک کند.

به منظور انجام تحقیقات آتی، می‌توان به ارزیابی روند کامل‌بودن داده‌ها و الگوی مشارکت کاربران از ابتدا تا کنون، الگوی مکانی و عوامل موثر بر توزیع مکانی داده‌های *OSM* اشاره پرداخت.

مراجع

[1] M. Chamani, R. Ali Abbaspour, and A. Chehreghan, "Matching of Polygon Objects Based on Geometric Measures in a Multi-scale Dataset," JGST, vol. 7, pp. 73-87, 2018 (Persian).

[2] M. Yazdannik, A. Chehreghan, and R. Ali Abbaspour, "An assessment of spatial similarity degree between polylines on multi-scale, multi-source maps," JGST, vol. 9, pp. 93-113, 2017 (Persian).

[3]M.Eshghi and A. Alesheikh, "An Assessment of the completeness of Volunteered Geographic Information: A Case Study of Tehran, Region 5,"pp.1-7, 2016 (Persian).

[4]O.Asgari, N Samani and, S Attarchi, "Assessing the quality of volunteered geographic information landuse change using satellite imagery," JGST, vol. 10, pp. 199-212, 2020 (Persian).

[5] A. Chehreghan, and R. Ali Abbaspour, "Assessment of Geometric Similarity of Volunteered Geographic Information in Intra-city Roads Network," JGST, vol. 10, pp. 357-370, 2017 (Persian).

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

[7] M. F. Goodchild, and L. Li, "Assuring the quality of volunteered geographic information," Spatial statistics, vol. 1, pp. 110-120, 2012.

[8] M.Jilani, M. Bertolotto, P. Corcoran, and A. Alghanim. Traditional vs. Machine-Learning Techniques for OSM Quality Assessment. In Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications,pp. 47-48, 2019.

[9] S. Schmitz, A. Zipf, and P. Neis, "New applications based on collaborative geodata—the case of routing", in Proceedings of XXVIII INCA international congress on collaborative mapping and space technology, 2008.

[10] M. A. Rylov and A. W. Reimer. "A comprehensive multi-criteria model for high cartographic quality point-feature label placement", Cartographica: The International Journal for Geographic Information and Geovisualization, vol. 49, pp. 52-68, 2014.

[11] A. Krek, M. Rumor, S. Zlatanova, and E. M. Fendel. "Interoperable Location Based Services for 3D cities on the Web using user generated content from OpenStreetMap", in Urban and Regional Data Management: CRC Press, pp. 87-96, 2009.

[12] M. Over, A.Schilling,S. Neubauer, and A. Zipf ."Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany", Computers, Environment and urban systems, vol. 34,pp. 496-507,2010.

[13] H. Chen, W. Zhang, C. Deng, N. Nie, and L. Yi, "Volunteered Geographic Information for Disaster Management with Application to Earthquake Disaster Databank & Sharing Platform," in IOP Conference Series: Earth and Environmental Science, p. 012015 , 2017.

[14] R. Feick and S. Roche, "Understanding the Value of VGI," in Crowdsourcing geographic knowledge, ed: Springer, pp. 29 -15, 2013.

[15] R. Hecht, C. Kunze, and S. Hahmann, "Measuring completeness of building footprints in OpenStreetMap over space and time," ISPRS International Journal of Geo-Information, vol. 2, pp. 1066-1091, 2013.

[16] M. Goetz and A. Zipf, "OpenStreetMap in 3D – Detailed Insights on the Current Situation in Germany,"In: Proceedings of the 15th AGILE International Conference on Geographic Information Science,2012.

[17] C. Fram, K. Chistopoulou, and C. Ellul, "Assessing the quality of OpenStreetMap building data and searching for a proxy variable to estimate OSM building data completeness," In: Proceedings of the 23rd GIS Research UK (GISRUK) conference, Leeds, UK, pp.17-15 ,2015.

[18] A.Alhamwi, W.Medjroubi, T.Vogt, and C. Agert. OpenStreetMap data in modelling the urban energy infrastructure: a first assessment and analysis. Energy Procedia,vol.142, pp.1968-1976, 2017.

[19] T. Törnros, H. Dorn, S. Hahmann, and A. Zipf, "Uncertainties of completeness measures in OpenStreetMap–A case study for buildings in a medium-sized German city, " ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, pp.2, 2015.

[20] M. A. Brovelli, M. Minghini, M. E. Molinari, and G. Zamboni," Positional accuracy assessment of the OpenStreetMap buildings layer through automatic homologous pairs detection:The method and a case study." 2016.

[21] H. Fan, A. Zipf, , Q. Fu, and P. Neis, "Quality assessment for building footprints data on OpenStreetMap, " International Journal of Geographical Information Science, vol. 28, , pp.700-719, 2014.

[22] Y. Xu, Z. Chen, Z. Xie, and L. Wu, "Quality assessment of building footprint data using a deep autoencoder network," International Journal of Geographical Information Science, vol. 31, pp. 1929-1951, 2017.

[23] Y. Wang, D. Chen, Z. Zhao, F. Ren, and Q. Du, "A Back- Propagation Neural Network- Based Approach for Multi- Represented Feature Matching in Update Propagation," Transactions in GIS, vol. 19, pp. 933-964, 2015.

[24] H. Du, "Matching disparate geospatial datasets and validating matches using spatial logic," PhD Thesis, University of Nottingham, 2015.

[25] W. Tang, Y. Hao, Y. Zhao, and N. Li , "Feature matching algorithm based on spatial similarity," In: Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Classification of Remote Sensing Images. International Society for Optics and Photonics, 2008.

[26] F. Zhonglianga and W. Jianhuaa, "Entity matching in vector spatial data," in Proceedings of the XXI ISPRS Congress, pp. 1467-147, 2008.

[24] Y. Huh, K. Yu, and J. Heo, "Detecting conjugate-point pairs for map alignment between two polygon datasets," Computers, Environment and Urban Systems, vol. 35, pp. 250-262, 2011.

[28] Y. Wang, Q. Du, F. Ren, and Z. Zhao, "A propagating update method of multi-represented vector map data based on spatial objective similarity and unified geographic entity code," in Cartography from Pole to Pole, ed: Springer, pp. 139-153, 2014.

[29] Veregin, H. (1999). Data quality parameters. Geographical information systems, 1, 177-189.

K. N. TOOSI UNIVERSITY OF TECHNOLOGY
FACULTY OF GEODESY AND GEOMATICS ENGINEERING

Journal of Geospatial Information Technology
Vol.9 No.1, Spring 2021

Research Paper

Assessment of the completeness of Volunteered Geographic Information focusing on building blocks data (Case Study: Tehran metropolis)

Roghayeh Adabi [1], Rahim Ali Abbaspour [2]*, Alireza Chehreghan[3]

1- MSc Student of GIS, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran, Iran.
2- Associate professor, School of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran, Iran.
3- Assistant professor, Mining Engineering Faculty, Sahand University of Technology, Tabriz, Iran.

Abstract

Open Street Map (OSM) is currently the largest collection of volunteered geographic data, widely used in many projects as an alternative to or integrated with authoritative data. However, the quality of these data has been one of the obstacles to the widely use of it. In this article, from among the elements related to the quality of volunteered geographic data, we have tried to examine the completeness of building block data in the OSM for Tehran metropolis. In order to do it, we have applied two object-based and unit-based approaches that are frequently applied in similar studies. The findings of this study demonstrate that the estimation of OSM building completeness strongly differs from these two approaches. The results demonstrate the high speed of evaluation of the unit-based approach and the higher accuracy of object-based method. Moreover, the results for the unit-based method indicate that a unit-based comparison of the total number or area of buildings is highly sensitive to disparities in modeling. While the object-based methods have lower sensitivity than disparities in modeling and if the data have an appropriate location accuracy, they will provide more accurate results for the completeness parameter. Therefore, the recommendation of this paper is to use an object-based method in quality assessment studies. In the end, based on the object-based approach, the parameter of completeness in the whole study area was calculated as 2.7%, which shows a low rate of completeness. Evaluating the results in closer detail, shows that the northern, central and eastern parts of Tehran are more complete than the other parts. Also, in most regions of Tehran, more than 80% of the official data is not in the OSM dataset, which indicates that the official dataset is more complete than the OSM one.

Key words: *Volunteered Geographic Information (VGI), building blocks, completeness, unit-based method, object-based method.*

Correspondence Address: *GIS Department, school of Surveying and Geospatial Information Engineering, College of Engineering, University of Tehran, Tehran.*
Tel : *+2188008841*
Email*: abaspour@ut.ac.ir*