ОРИГИНАЛЬНОЕ
ИССЛЕДОВАНИЕ

ВОПРОСЫ ЛЕСНОЙ НАУКИ, 2025, Т. 8. № 3. Статья № 175

DOI: 10.31509/2658-607x-202583-175
УДК 614.842; 630*96

РАЗРАБОТКА ПЛАГИНА OPEN SOURCE QGIS ДЛЯ ОЦЕНКИ КАЧЕСТВА ДАННЫХ ПО ДОРОЖНОЙ СЕТИ В РЕГИОНАЛЬНЫХ ЛЕСОТРАНСПОРТНЫХ ПРОЕКТАХ

© 2025

Е. С. Подольская[1*], И. М. Зиняев[2]

[1]Центр по проблемам экологии и продуктивности лесов им. А. С. Исаева Российской академии наук
Россия, 117997 Москва, ул. Профсоюзная, 84/32, стр. 14

[2]МИРЭА – Российский технологический университет
Россия, 119454 Москва, проспект Вернадского, 78

*E-mail: podols_kate@mail.ru

Поступила в редакцию: 02.06.2025
После рецензирования: 19.06.2025
Принята к печати: 15.07.2025

В статье представлена разработка плагина для Open Source QGIS для автоматизированной оценки качества данных по дорожной сети при решении логистических и транспортных задач лесного комплекса и инфраструктурных проектов региона России (Новосибирская область). Автоматизация оценки качества данных особенно актуальна в условиях использования открытых данных, разнообразия и увеличения количества источников. Были изучены национальные стандарты, отраслевые документы, кейсы российских компаний. Разработанный плагин «Compare_road» используется для анализа и верификации данных по дорогам в формате векторных линейных файлов. При анализе вычисляются следующие метрики: точность соответствия, полнота и актуальность данных. Модуль опубликован на GitHub под лицензией MIT. Целевой аудиторией плагина являются исследователи, студенты и другие пользователи. Для апробирования работы плагина проведён анализ данных по сети дорог Новосибирской области проектов Natural Earth, VMAP, Digital Chart of the World при использовании Open Street Map (OSM) в качестве эталона. Наилучшие показатели точности показал Digital Chart of the World – 270.7 м, лучший показатель полноты данных VMAP – 22.6%. Выполнен анализ качества данных по дорогам для лесничеств Новосибирской области: наилучшие показатели точности и полноты у лесничеств Татарское

(127.6 м, 35.9 %) и Маслянинское (154 м, 35.9 %), наихудшие – у Кыштовского (378.5 м, 0.7 %) и Сузунского (418 м, 1.5 %).

Ключевые слова: *ГИС, Open Source, QGIS, плагин, OSM, Natural Earth, дорожная сеть, оценка качества*

ВВЕДЕНИЕ

Качество пространственных данных – это степень соответствия совокупности собственных характеристик заявленным требованиям. В геоинформатике два понятия можно назвать основополагающими – «пространственные данные» и «пространственный объект» (Лурье, 2010).

Первому термину присущи два различных понятия. Первое из них включает в себя пространственные данные в широком понимании этого слова, то есть все данные, которые в своей основе содержат координатные данные, которые включают в себя, но не ограничиваются, описаниями объектов реальности, цифровые изображения и карты, каталоги координат опорной геодезической сети. Второе определение прямо связано с геоинформационными системами. Оно гласит, что это цифровые данные об объектах реальности. В своей основе они состоят из двух частей – координатных (пространственных) и атрибутивных (непространственных) данных.

Для оценки качества данных используются международные, гармонизированные и национальные стандарты: ISO (International Organization for Standardization), FGDC (Federal Geographic Data Committee), CEN (Comité Européen de Normalisation). В Российской Федерации для оценки качества пространственных данных принят стандарт ГОСТ Р 57773-2017 (https://files.stroyinf.ru/Data/655/65547.pdf). Он является модифицированным по отношению к международному стандарту ИСО 19157:2013 (https://docs.cntd.ru/document/1200157078) путём приведения его в соответствие международным и национальным стандартам. ГОСТ Р 57773-2017 приводит следующие критерии:

- полнота: определяется наличием и отсутствием объектов, их атрибутов и отношений;
- логическая согласованность: определяется степенью соответствия логическим правилам, представленным в структуре данных, атрибутах и отноше-

ниях между признаками (структура данных может быть концептуальной, логической или физической);

- позиционная точность: определяется точностью положения объектов внутри пространственной системы координат;
- временное качество: определяется качеством временных атрибутов и временных отношений объектов;
- тематическая точность: определяется точностью количественных атрибутов, корректностью неколичественных атрибутов и классификаций объектов и их соотношений.

У Росавтодора имеется отраслевой методический документ (ОДМ 218.9.008–2019), который носят рекомендательный характер. В данном документе указаны различные метрики для различных видов автомобильных дорог, включающие пункты, которые были указаны в ГОСТ. Также кроме этих метрик в документе есть схемы внесения данных на протяжении всего жизненного цикла дороги и технологии по сбору данных.

Качество итоговых данных прямо зависит от качества входных пространственных данных (Мартынова, 2023). Данный тезис не подлежит сомнению, поэтому необходимо задумываться о способах оценки качества пространственных данных. Работа Е. В. Мартыновой (2023) разделяет качество на две категории. Первая из них – внутреннее качество. Оно предполагает под собой данные, которые не допускают ошибки в целом или соответствуют определенным стандартам в используемой отрасли. Например, упомянутый выше отраслевой методический документ может являться стандартом для внутреннего качества. Оно должно достигать уровня подобия между «реальными» данными, на которых основывается пространственный объект, и итоговыми пространственными данными. Второй категорией является внешнее качество. Данный тип качества рассчитан на конечных пользователей и их потребности. Внешнее качество может иметь совсем другие критерии, которые будут варьироваться от одного пользователя к другому, а значит один и тот же набор данных может быть удовлетворителен для одних, но неприменим в работе других. Это означает, что даже набор пространственных данных с ошибками может быть использован в определенных работах при условии соответствия требованиям пользователя.

Работа Г. Г. Побединского и А. Н. Прусакова, (2019) обращает внимание на нормативно-правовой аспект, связанный с пространственными данными. На момент публикации статьи говорилось о необходимости совершенствования базы геодезии и картографии, включая такие шаги, как разработка и переработка нормативно-технических документов, объединение этих документов в своды правил и стандарты.

Согласно исследованию С. Г. Дышленко (2016), оценка качества данных является не только важным, но и очень трудоемким процессом. Общее время на проверку полученных пространственных данных может занимать до 30% от общего времени выполнения проекта. При обнаружении ошибок имеется два пути решения проблемы. Первый – создание нового материала с нуля. Такой вариант подойдет, когда готовый материал невозможно исправить или это потребует слишком больших человеческих затрат. Второй метод включает в себя редактирование уже созданного ошибочного объекта. Предлагаются два варианта векторизации объектов: ручной и автоматический. Ручная оцифровка является очень трудоёмким процессом, который занимает до 60% времени создания цифровых карт. Использование технологий автоматической дешифровки может ускорить процесс на 20-30%, но ценой ухудшения качества итогового продукта. Для оценки качества получаемых данных предлагаются два способа. Первый является бинарным, где единица обладает качественными характеристиками и полнотой, удовлетворяющей требованиям проекта, а ноль, наоборот, не обладает такими свойствами. Также рассматривается аддитивный показатель, который позволяет сравнивать два набора данных, находя из них более подходящий для работы. Также для пространственных данных предлагается дополнительный критерий – возможность внесения исправлений. Он описывает возможность редактирования конечного продукта без его кардинальной переработки. Данные с такими свойствами возможно использовать множество раз, редактируя отдельные части набора.

В работе Б. А. Дворкина (2014) выделены проблемы, которые влияют на пространственные данные с точки зрения регионального использования. К ним относят: разнородность данных, неактуальность картографических материалов, разрозненность пространственных данных.

Данные пункты сильно осложняют оценку качества данных и выбор эталонного набора, так как даже данные государственных ведомств могут включать данные недостатки. Они возникают из-за некоторой недостаточности технической составляющей, а также отсутствия общей системы, в которой изменения объекта передаются всем пользователям. В работе указан вывод, что на момент исследования была необходима дальнейшая работа по изучению данной темы для получения более конкретных результатов.

Оценка удовлетворительности качества может приниматься по различным методикам. ГОСТ Р 57773-2017 предлагает использование обобщенного показателя качества данных (Aggregated Data Quality Results). Данная методика придает каждому параметру логическое значение между нулем и единицей, соответственно подходящим и неподходящим. Также есть варианты использования взвешенной оценки пригодности или оценка данных на основе максимальных или минимальных данных. Для использования данных методик от пользователя ожидается достаточная компетентность, позволяющая ему оценить пригодность данных по полученным значениям.

Альтернативой данным из региональных государственных источников могут являться краудсорсинговые проекты. Работа китайских исследователей (Wang et al., 2013) рассматривает качество данных Open Street Map (OSM, www.openstreetmap.org). Исследователи пришли к выводу, что подобные решения уступают официальным источникам как в точности, так и в полноте данных. Несмотря на это, краудсорсинговые пространственные данные имеют больше актуальной информации в определенных типах данных. Использовать такие данные при проверке качества возможно, но в ограниченных объемах. Анализ и интерпретация результатов при использовании краудсорсинговых данных требует понимания контекста социальных процессов, приведших к созданию пространственных данных и их гетерогенности (Mocnik et al., 2018). В работе других китайских исследователей (Chen et al., 2023) изучена достоверность данных OSM, связанных с аэропортами мира. В результате изучения выяснено, что аэропорты, которые были категорированы как «большие», были оцифрованы достаточно хорошо по всем изучаемым пунктам. С уменьшением же размеров аэропортов уменьшалось и количество данных по ним. Также

была выявлено географическое влияние на качество и полноту данных. Другие работы, например, М. Моради с соавторами (Moradi et al., 2023), показывают заметное улучшение полноты и точности данных OSM с течением времени. Большее количество объектов OSM ухудшает общее качество данных, особенно в местах с меньшим населением, так как на одного волонтера приходится большее количество объектов. Данные OSM по зданиям используются в ряде исследований (микроклимат, морфология городской застройки), требующих базовой информации об объектах, сами данные являются фрагментированными (Biljecki et al., 2023). Платформа OSM пополняется волонтерами-участниками, для успешной работы проекта в долгосрочной перспективе предпочтительнее привлекать большее количество новых участников с небольшим количеством правок, нежели полагаться на узкий круг активных участников (Zhang et al., 2024). Другие работы (Sehra et al., 2014; Gokceoglu, 2020) в заключении указывают, что количество пространственных данных будет только расти, при этом качество вносимой информации, предоставленной участниками-любителями, должно проверяться в автоматическом режиме при помощи машинного обучения. В статье, изучавшей работу волонтеров на территорию города Сальвадор (Elias et al., 2021), при анализе правок за период с 2008 по 2020 годы было обнаружено то, что количество правок практически является линейной функцией. На количество правок также могут влиять такие дополнительные факторы, как добавление новых данных (Elias et al., 2021).

Модификация метода Бернерса-Ли для оценки открытых данных была использована в работе (Kin, Lazorenko, 2021). Данная модификация учитывает международный стандарт ИСО 19157. Данные подразделяются на 5 уровней. Минимальный уровень – первый – предполагает данные, которые можно географически идентифицировать и наличие свободной лицензии. Второй уровень требует структурированный вид данных с геометрией, а также координаты в системе координат, находящейся в реестре EPSG, возможность использование геосервисов. Третий уровень предполагает наличие непатентованного формата файла, а также геометрию в формате OGC. Четвертый и пятый уровни применяются при наличии более качественной топологии. Работа М. Вагнера и К. Хен-

зен (Wagner, Henzen, 2022) тоже использует данный метод, но уделяет больше времени вопросу зрелости данных, обращая внимание на то, что по данной тематике не имеется стандартов, в отличие от качества данных. Также в статье сказано о невозможности полной автоматизации оценки качества данных на момент написания статьи, но по мере развития технологий и правильной организации данных в датасеты возможно уменьшение количества ручного труда до минимума.

Актуальный российский производственный опыт оценки качества данных представлен разработками ряда компаний. Компания Лемана Тех в одной из статей в своем официальном блоге (Инструмент …, 2024) описывает процесс создания и внедрения платформы оценки качества данных для объединения работы нескольких систем. В результате создания архитектуры, способной намного быстрее проводить обработку данных, а также поддерживать работоспособность при сбоях, время, затрачиваемое на выполнение необходимых тестов, уменьшилось с недель до часов.

Данная статья не связана с проверкой векторных данных напрямую, но она дает анализ одного из множества успешных кейсов оптимизации оценки качества данных в работе реальных компаний. На том же сайте в блоге компании «СИБУР» (Как обеспечить …, 2024) размещен ряд статей о разработке корпоративного решения для оценки качества в процессе импортозамещения после ухода иностранных компаний. Дано описание как технических, так и бизнес-задач оценки качества данных. Указывается, что бизнес-задачи оценки качества данных ставятся в зависимости от оценки финансовых потерь компании. Также в компании существуют уровни проверки данных. Данные проверяются на уровне ввода, в процессе поступления и уже после загрузки на уровне физической базы данных. Важным развитием работы являются метрики качества. К уже имеющимся и названным ранее добавились консистентность, целостность, разумность. Эти метрики также увеличивают количество возможных проверок данных.

Пространственные данные используются для анализа и последующего принятия управленческих решений. Чтобы такие решения были эффективными, пространственные данные должны быть эталонными, то есть соответствовать определенным стандартам качества. Определение критериев эталонных данных зависит от

задачи конечного пользователя. Российские лесотранспортные проекты регионального уровня нуждаются в качественных данных по дорожной сети с топологически корректной геометрией и наличием набора атрибутов. Геометрию дорожной сети можно проверять при помощи инструментов автоматизированной проверки, которыми обладают современные геоинформационные системы (ГИС). Разрозненность данных источников разных масштабов усложняет процесс.

Целью статьи является создание автоматизированного программного решения для оценки качества пространственных данных по дорогам. Для достижения этой цели необходимы: изучение отраслевых стандартов, успешных кейсов автоматизации оценки качества данных компаниями и примеров аналогичных работ, разработка методики и практическая реализация в виде плагина для Open Source QGIS, а также апробирование работы плагина на примере датасетов для регионального примера – лесничеств Новосибирской области.

Из существующих в репозитории QGIS (https://plugins.qgis.org) решений были изучены плагины LineComparison (https://plugins.qgis.org/plugins/linecomparison/) и LineSimilarity (https://plugins. qgis.org/plugins/line_similarity/). Эти решения проверяют либо координатную точность, либо полноту. На начало июня 2025 года эти модули (оба в экспериментальных версиях) не были обновлены уже длительное время.

МАТЕРИАЛЫ И МЕТОДЫ

Разработанный плагин «Compare_road» используется для анализа и верификации дорожных слоёв региональных проектов лесотранспортного моделирования в формате векторных линейных файлов. При анализе необходимо вычисление следующих метрик: координатная точность, полнота данных, актуальность данных. Целевой аудиторией плагина являются исследователи, студенты и пользователи-любители. Входными данными являются два линейных векторных слоя, представляющие автодорожную сеть. Выходными данными являются вычисленные критерии качества, помещенные в таблицу для удобства восприятия пользователем. Был создан отдельный текстовый файл с ответами на частые вопросы. Список вопросов включает в себя вопросы о базовом функционале и причинах некорректных результатов работы плагина.

Для разработки плагина был использован модуль «Plugin Builder», шаблонная

структура которого упрощает начальный этап разработки. Для реализации плагина применялся язык программирования Python. Для написания программного кода использована среда разработки MS Visual Studio 2019. Для дальнейшей поддержки, а также получения обратной связи по багам используется GitHub. Дизайн плагина был создан в QtDesigner с дополнительными виджетами QGIS 3.28.1.

Методика расчета метрик оценки качества

Алгоритм оценки точности реализует методику, основанную на анализе пространственных отклонений между сравниваемыми осевыми линиями дорог. Алгоритм автоматически генерирует точки вдоль сравниваемой дорожной сети с заданным шагом, после чего для каждой точки вычисляется кратчайшее расстояние до линии дороги эталонного слоя. Особенностью реализации является использование медианы полученных расстояний, что обеспечивает устойчивость результатов к локальным выбросам и погрешностям данных, которые неизбежны при использовании открытых данных.

Функционал оценки полноты реализует методику, основанную на теории множеств и пространственном анализе. Система выполняет буферизацию обоих сравниваемых слоев с возможностью настройки радиуса буфера. Затем анализируется площадь пересечения полученных буферных зон, что позволяет количественно оценить степень покрытия эталонной сети сравниваемыми данными. Результат представляется в виде процентного отношения площади пересечения к площади эталонного буфера.

Функция оценки актуальности реализует подход, основанный на анализе метаданных исходных файлов. Плагин определяет актуальность данных путем сравнения текущей даты с датой последнего изменения файла слоя. Реализация предполагает возможность настройки отображения актуальности в разных единицах измерения времени: дни, недели, месяцы. Для более точного анализа актуальности имеется как проверка актуальности проверяемого слоя в сравнении с эталонным, так и проверка актуальности файла на текущую дату.

Техническая реализация плагина для Open Source QGIS

Базовая структура плагина создаётся плагином Plugin Builder (https://plugins.qgis.org/plugins/pluginbuilder/) автоматически в виде шаблона. Ключевыми составляющими базовой структуры являются следующие элементы:

каталог compare_roads: основной каталог плагина, содержащий все другие подкаталоги и файлы, название совпадает с названием плагина;

файл metadata.txt: файл метаданных, включающий в себя информацию о плагине, необходимую для его отображения в модулях QGIS; содержит название, информацию о версии, описание и другие метаданные плагина;

файл compare_road.py: основной файл плагина, включающий логику инициализации и работы плагина; в файле определяется класс плагина, который наследуется от QgsPlugin;

файл compare_road_dialog_base.ui: файл, включающий в себя графическую часть плагина.

Создание удобного пользовательского интерфейса является важной частью разработки плагина. Для создания интерфейса плагина QGIS используется фреймворк Qt, который является основой для графического интерфейса QGIS. Основным инструментом для проектирования пользовательского интерфейса является Qt Designer. Qt Designer генерирует файлы с расширением .ui. Такой подход позволяет разделить логику приложения и его визуальную часть, что упрощает разработку и поддержку кода.

Пользовательский интерфейс плагина был спроектирован с учетом вышеописанных требований. Основные элементы интерфейса включают выпадающие списки для выбора слоёв, табличный виджет для отображения результатов анализа, а также флаговые кнопки. Разработанный вариант интерфейса представлен на рисунке 1.

Для реализации вычислений основных метрик используется встроенный набор инструментов для анализа пространственных данных, предоставляемый QGIS. При разработке плагина для каждой отдельной операции была создана отдельная типовая функция, возвращающая выходной слой, который при необходимости

использовался в последующих этапах. Типовая функция автоматически определяет эталонный и сравниваемый слои, которые были выбраны пользователем, для этого в пользовательском интерфейсе были использованы mapLayerComboBox. Обращение к выбранным в них объектам позволяет сразу же получать слой, что облегчает процесс поиска. Если в процессе выполнения функции имеется настраиваемый параметр, то ему задается стандартное значение, которое перезаписывается при включении продвинутого режима на пользовательское через соответствующий элемент интерфейса.

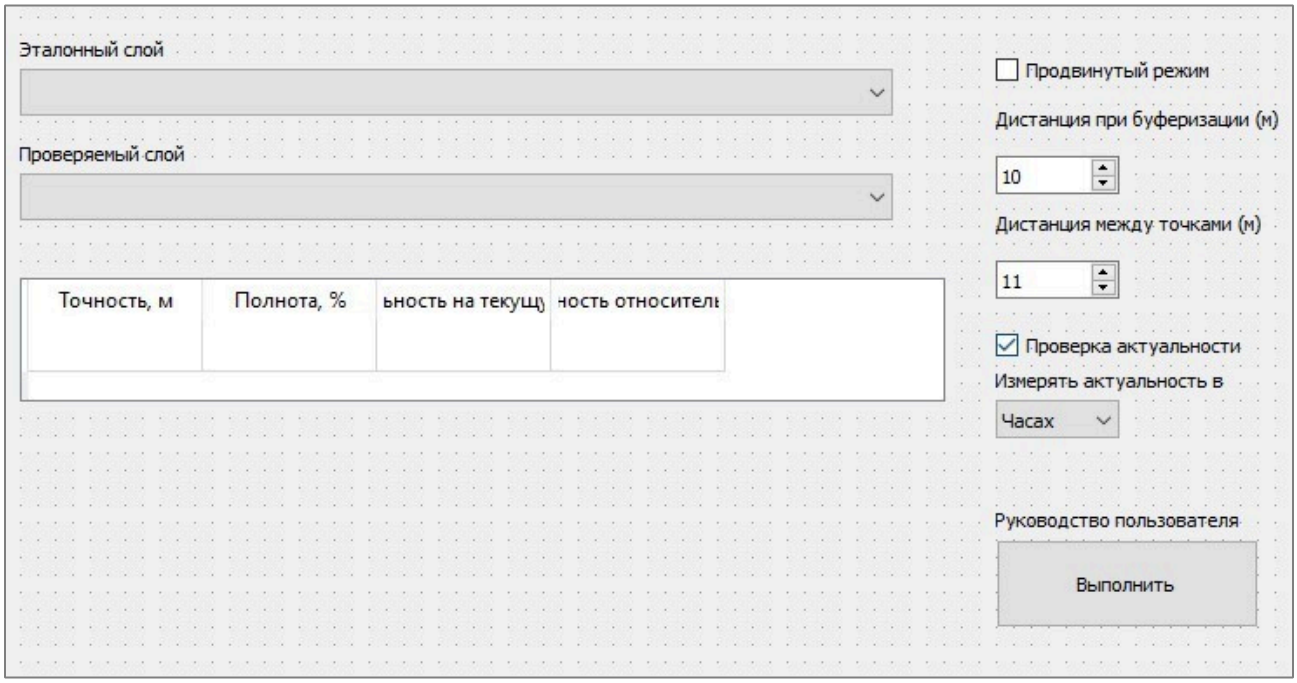

Рисунок 1. Разработанный вариант интерфейса плагина “Compare_road”

Далее формируется набор параметров, предназначенный для конкретной функции. При необходимости для каждого слоя создаётся свой отдельный набор. Например, при буферизации в этот набор входит указание входного слоя, расстояние буферизации с автоматической поправкой для сравниваемого слоя, параметры сглаживания в виде количества сегментов, флаг объединения результатов и стиль оформления краев буфера.

Функция использует временные слои

для хранения промежуточных результатов, что позволяет оптимизировать использование памяти. Возвращаются два временных слоя с результатами буферизации, которые не добавляются автоматически в проект, а передаются для дальнейшей работы алгоритма.

В интерфейс плагина были добавлены функции продвинутого режима и справки. При выборе продвинутого режима пользователю дается возможность изменять значения расстояния в функциях буферизации и распределения точек вдоль слоя осевых линий дорожной сети с помощью соответствующих элементов интерфейса, изображенных на рисунке 2.

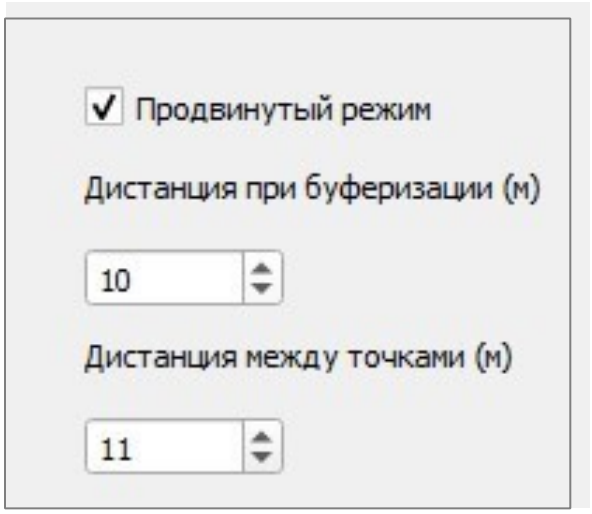

Рисунок 2. Настройки продвинутого режима работы плагина «Compare_road»

Стандартное значение дистанции между точками согласуется с методическим документом (Федеральное дорожное агентство, 2021). При большом количестве объектов можно увеличить дистанцию между точками для уменьшения времени выполнения алгоритма.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Проверка работоспособности плагина «Compare_road» на тестовом наборе данных

После разработки плагина был создан тестовый набор данных, имитирующий небольшой участок дорожной сети для проверки корректности работы разработанного программного решения. Дата изменения сравниваемого слоя была отредактирована вручную на более раннюю дату с помощью утилиты Total Commander. Также проверяемый слой намеренно был создан с несовершенствами, включающими в себя неполноту данных и координатную неточность. Дополнительно была проведена проверка при выборе двух одинаковых слоёв. Проверка показала корректную работу плагина, результат тестовой проверки показан на рисунках 3 и 4.

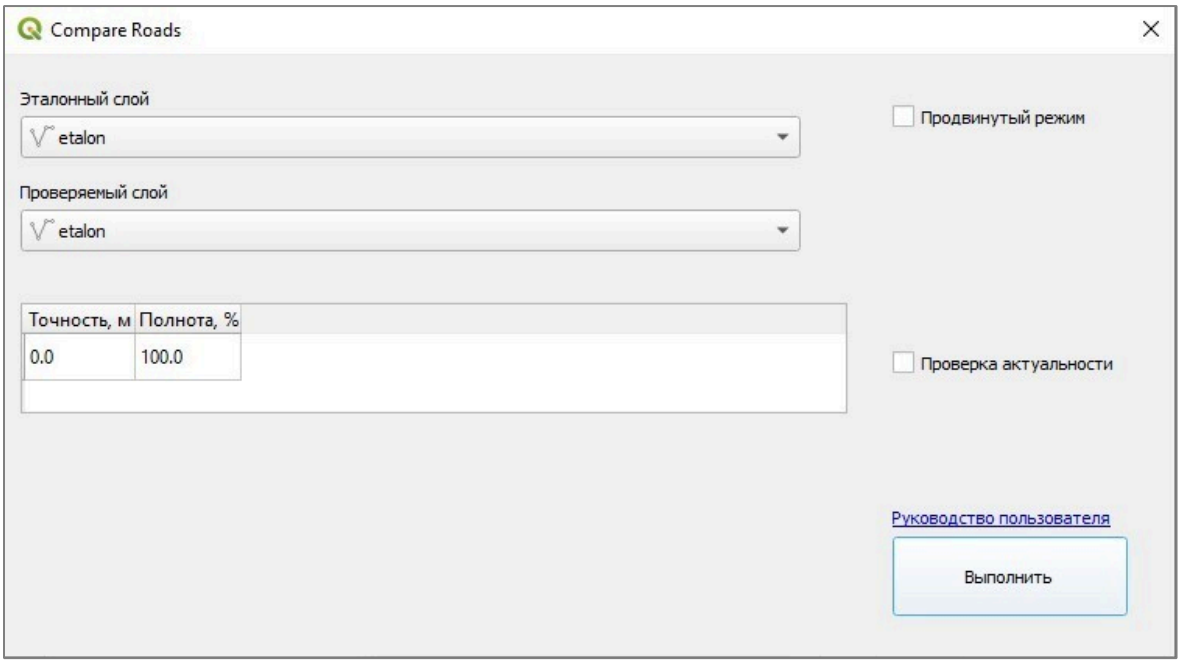

Рисунок 3. Анализ одинаковых слоёв при помощи плагина «Compare_road»

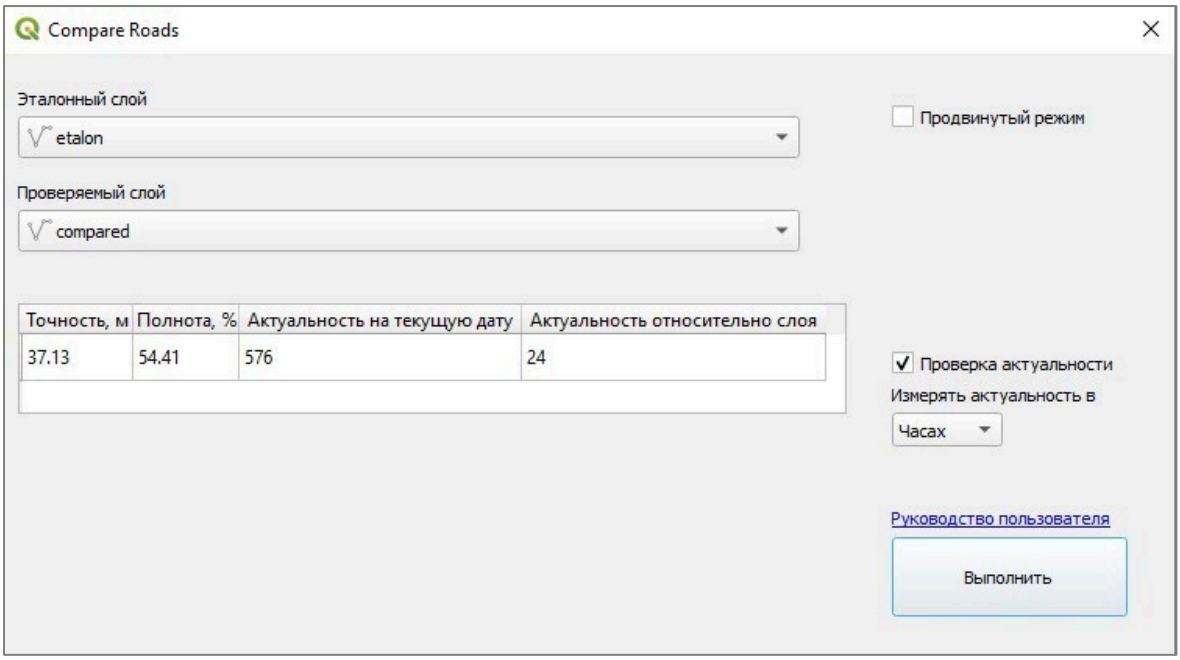

Рисунок 4. Анализ тестового набора данных при помощи плагина «Compare_road»

Использование плагина «Compare_road» для анализа открытых данных

После успешной проверки на тестовых данных был выполнен тест на проверку работоспособности плагина на открытых данных, для примера была выбрана территория Новосибирской области, данные представлены на рисунке 5.

За эталонные данные были приняты пространственные данные OSM. С помощью плагина QuickOSM (https://plugins.qgis.org/plugins/QuickOSM/) были получены данные по сети дорог, использовался модифицированный предустановленный набор ключей Дороги/Улицы. Ключи в OSM позволяют запрашивать только объекты, имеющие определенные теги. Предустановленный набор использует все значения «highway», поэтому он был отредактирован для отображения только дорог ранга областного значения и выше. Итоговый набор ключей для Новосибирской области изображен на рисунке 6.

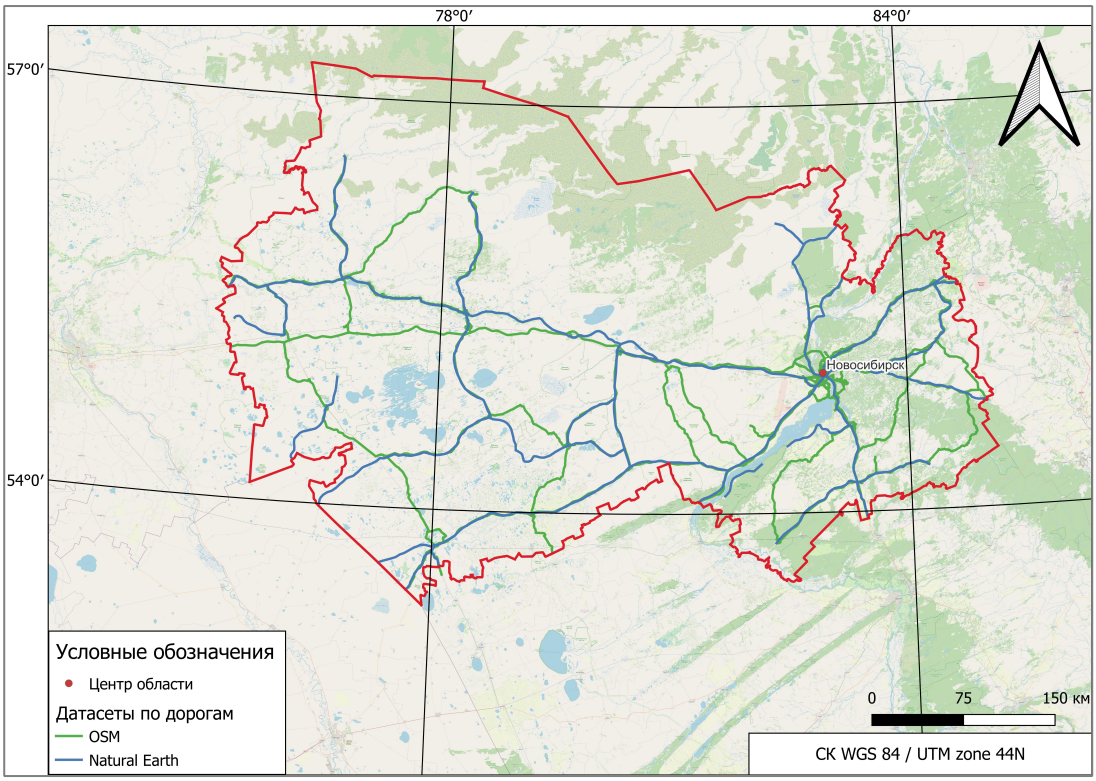

Рисунок 5. Отображение данных OSM и Natural Earth по дорожной сети для тестирования плагина «Compare_road»

		Ключ	Значение	Добавить	Удалить
1		highway	motorway		
2	Or	highway	motorway_link		
3	Or	highway	trunk		
4	Or	highway	trunk_link		
5	Or	highway	primary		
6	Or	highway	primary_link		
7	Or	highway	secondary		
8	Or	highway	secondary_link		

Рисунок 6. Набор ключей QuickOSM для дорог Новосибирской области

Подбор и анализ дополнительных пространственных данных

Для поиска открытых данных были использованы две различные поисковые системы: Яндекс и Google. Для обеих систем использовались операторы уточнения поиска для получения наиболее релевантных результатов. Поиск проводился

на русском и английском языках. Сайт Diva-GIS (https://diva-gis.org/data.html) обладает удобной фильтрацией по странам и наборам данных. Источником дорог в этом случае является цифровая карта мира (Digital Chart of the World) 1992 года. Следующим набором данных стал VMAP Level 1 (https://www.mapability.com/info/vmap1_intro.php). Также был найден один набор пространственных данных на сайте геологической службы США, который имеет название «Roads of the Former Soviet Union» (https://pubs.usgs.gov/of/2001/ofr-01-104/fsucoal/metadata/html/roads.htm). Он датируется 2001 годом и является совместной работой государственного геологического музея им. В. И. Вернадского и Геологической службы США.

Наиболее актуальным набором данных, находящихся в открытом доступе в предлагаемом исследовании, является датасет Natural Earth (https://www.naturalearthdata.com/), актуальная на июнь 2025 года версия которого датируется 2022 годом. Для анализа использовались следующие данные (https://www.naturalearthdata.com/downloads/10m-cultural-vectors/). Полученная точность анализа составила 270.7 м, а полнота 9.1%. Полученные результаты указывают на очень низкий уровень качества датасета.

Далее была произведена проверка ещё двух датасетов – VMAP и Digital Chart of the World. Результаты их анализа показали более приемлемые результаты. Ошибка точности у данных наборов данных составила 217 и 204 соответственно, оценка полноты – 22% и 18% соответственно.

После проверки для экстента всей области был проведен анализ данных по дорогам в каждом отдельном лесничестве региона, данные по которым взяты из Лесного плана Новосибирской области, сайт правительства Новосибирской области (Лесной план …, 2023). В качестве проверяемого слоя был выбран слой VMAP в силу наибольшей полноты. Большинство лесничеств показали результаты, схожие со средними значениями во всей области, но были и примеры отклонений от средних значений. Наилучшие результаты показало Татарское лесничество с показателями 127 м точности соответствия геометрии и 35.9% полноты, худшие –Кыштовское лесничество с 378 м точности и 0.7% полноты. Такой результат можно объяс-

нить выбором эталона, так как выбранные ключи OSM охватывают только дороги рангов областного значения и выше. Была предпринята попытка расширить количество дорог в эталоне, что привело к получению значений 252 метра точности и 6.2% полноты. Кирзинский заказник стал единственной территорией региона, где проверяемый слой не пересекался с изначальным эталоном. Увеличение количества рангов дорог для других лесничеств не имело значительного влияния на точность, но сильно уменьшало показатели полноты, поэтому для сохранения единообразия для всех лесничеств использовались ключи, указанные на рисунке 6.

Процесс оценки качества был проведен для дорожных данных на территорию Новосибирской области с датасетами Natural Earth, VMAP, Digital Chart of the World. Результаты оценки представлены в таблице 1.

Таблица 1. Результаты оценки источников данных по дорогам Новосибирской области

Название	Точность, м	Полнота, %
Natural Earth	270.7	9.1
VMAP	217.2	22.6
Digital Chart of the World	204.0	18.8

Проверка отдельных лесничеств показала результаты, схожие с результатами оценки в масштабах области, за исключением нескольких выбросов. Кыштовское и Татарское являются примерами лесничеств с выбросами. Болотнинское лесничество показало средние по региону значения, имея немного повышенное значение полноты. Результаты расчетов по этим лесничествам даны в таблице 2.

Таблица 2. Результаты анализа датасетов по отдельным лесничествам Новосибирской области

Название лесничества	Точность, м	Полнота, %
Кыштовское	378.5	0.7
Татарское	127.6	35.9
Болотнинское	219.8	26.8

Визуализация результатов анализа данных по дорогам лесничеств Новосибирской области представлена в виде картограмм на рис. 7-8. Наилучшие показатели – у лесничеств Татарское (127.6 м, 35.9 %) и Маслянинское (154 м, 35.9 %), наихудшие – у Кыштовского (378.5 м, 0.7 %) и Сузунского (418 м, 1.5 %). Территория Кирзинского заказника оказалась единственным местом, где данные эталона и проверяемого слоя не имели сходимости по соответствию геометрий дорог. Васюганский заповедник был единственной территорией, где дорожной сети в рассмотренных датасетах не было.

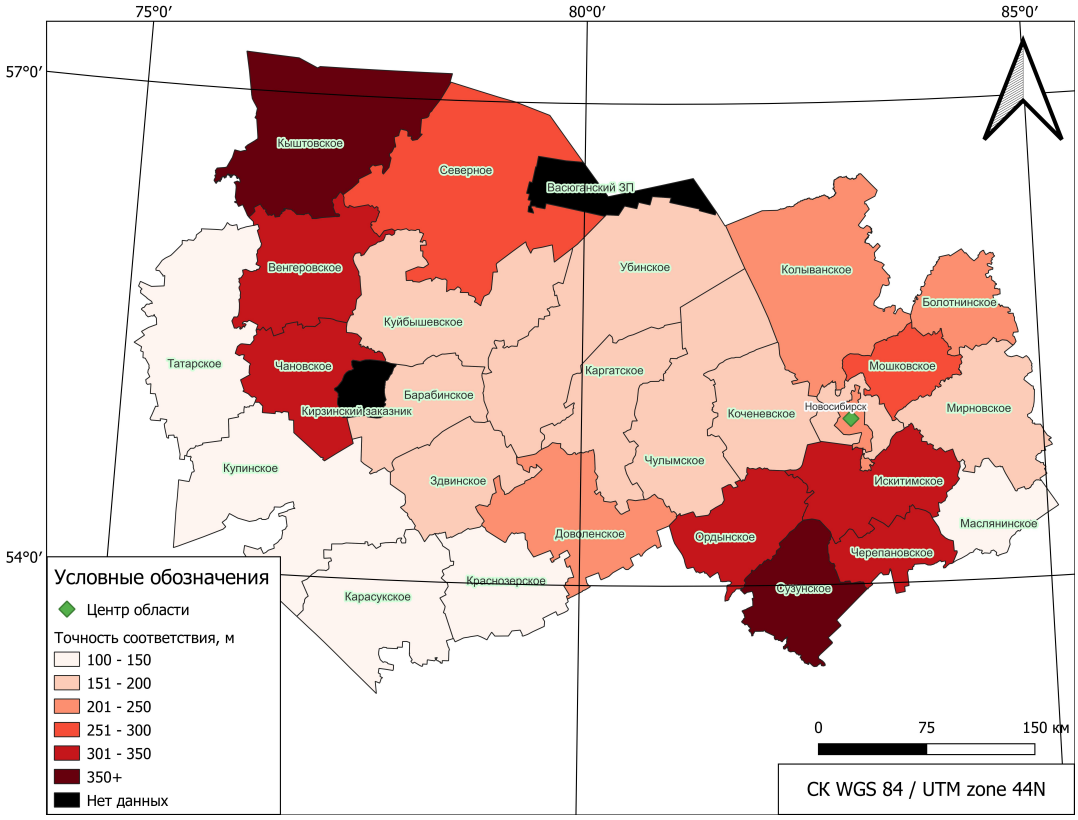

Рисунок 7. Картограмма оценки точности соответствия данных по дорогам лесничеств Новосибирской области

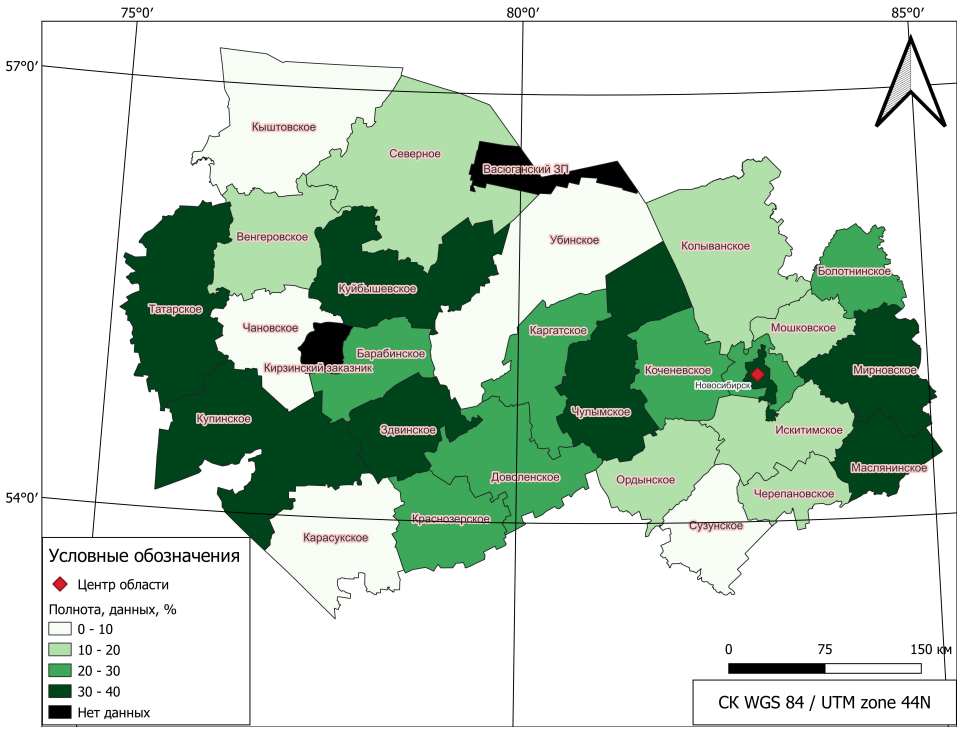

Рисунок 8. Картограмма оценки полноты данных по дорогам лесничеств Новосибирской области

Количество анализируемых метрик можно расширить, например, наличие атрибутивных данных может позволить более точно вычислять полноту данных, а также проверять неколичественные характеристики. Топологическую корректность по-прежнему можно отнести к автоматизированным процедурам контроля качества, решение которого требует дальнейших исследований. Необходимо отметить, что автоматизация контроля качества пространственных данных возможна, но имеет существенные ограничения, самым существенным из которых является отсутствие единообразия среди данных.

Публикация плагина «Compare_road»

Размещение разработанного плагина на хостинге GitHub предоставляет пользователям не только удобную платформу для загрузки, но и даёт возможность получения разработчиком обратной связи по багам, не найденным в процессе разработки и отладки. Также откры-

тый код позволяет конечным пользователям модифицировать его при необходимости.

Репозиторий расположен по ссылке https://github.com/IgorZinyaev/compare_road (2025). Конечный пользователь может установить плагин в настольном QGIS, загрузив репозиторий в .zip формате и установив через соответствующую функцию QGIS. Для плагина была выбрана одна из самых популярных для Open Source лицензия Massachusetts Institute of Technology License (или MIT). Данная лицензия позволяет свободно использовать разработанный плагин, копировать его, модифицировать код. Единственное условие использования – копия лицензии должна быть включены во все копии или значимые части разработки.

После публикации плагина на Github предполагается его дальнейшее развитие и поддержка. Во время фазы активного использования и обратной связи планируется добавление плагина в репозиторий QGIS для увеличения количества потенциальных пользователей, а также получение сообщений об ошибках. Общая продолжительность жизненного цикла плагина будет составлять 1 год с момента публикации, так как поддержка версий семейства QGIS 3 будет приостановлена 26 мая 2026 года, а долгосрочный релиз QGIS 4 запланирован на февраль 2026 года. При наличии откликов пользователей возможна переработка плагина для нового семейства версий QGIS 4 с учетом планируемого перехода на Qt6.

ЗАКЛЮЧЕНИЕ

Разработанный плагин контроля качества данных по дорогам «Compare_road» представляет собой Open Source-инструмент для решения задач регионального лесного хозяйства, логистики перевозок лесопромышленных предприятий, также может быть использован в инфраструктурных проектах. Плагин предназначен для использования на разных этапах лесотранспортных проектов, оценивая качество входных пространственных данных, например, при получении нескольких файлов дорожной сети из разных источников. Полученные результаты позволяют принимать обоснованное решение о возможности использования файлов дорог в дальнейшей работе. Плагин может использоваться как в региональных проектах, так и на локальных участках отдельных регионов.

БЛАГОДАРНОСТИ

Статья является частью дипломного проектирования в бакалавриате по направлению 09.03.02 «Информационные системы и технологии» кафедры ГИС, Института информатики и радиоэлектроники, РТУ МИРЭА, выполняемого в 2025 году.

ФИНАНСИРОВАНИЕ

Работа выполнена в рамках государственного задания ЦЭПЛ РАН по теме «Биоразнообразие и экосистемные функции лесов» (Регистрационный номер НИОКТР 124013000750-1).

СПИСОК ЛИТЕРАТУРЫ

ГОСТ Р 57773-2017. URL: https://files.stroyinf.ru/Data/655/65547.pdf (дата обращения 01.06.2025)

Дворкин Б. А. Инфраструктура пространственных данных: региональный аспект // Геоматика. 2014. № 1. С. 17–21.

Дышленко С. Г. Анализ и разработка характеристик качества геоданных // Перспективы науки и образования. 2016. № 2(20). С. 23–27.

Инструмент обеспечения качества данных: от теории к практике // Хабр URL: https://habr.com/ru/companies/lemana_tech/articles/857044/ (дата обращения 01.06.2025)

ИСО 19157:2013. URL: (https://docs.cntd.ru/document/1200157078) (дата обращения 01.06.2025)

Как обеспечить Data Quality терабайтов данных и зачем: опыт СИБУРа // Хабр URL: https://habr.com/ru/companies/sibur_official/articles/860372/ (дата обращения 01.06.2025)

Лесной план Новосибирской области // Министерство природных ресурсов и экологии Новосибирской области URL: https://mpr.nso.ru/page/653 (дата обращения 01.06.2025)

Лурье И. К. Геоинформационное картографирование. Методы геоинформатики и цифровой обработки космических снимков. Москва: КДУ, 2010. 424 с.

Мартынова Е. В. Методика оценки качества пространственных данных // Экономика и управление: проблемы, решения. 2023. № 7. Т. 2. С. 111–118.

ОДМ 218.9.008–2019. Геоинформационные системы автомобильных дорог: порядок сбора, хранения и обновле-

ния данных: отраслевой дорожный методический документ: издание официальное: издан Распоряжением Федерального дорожного агентства от 29.07.2019 № 1983-р: введен впервые / разработан ООО «ИндорСофт». Москва: «Информавтодор», 2021. 72 с.

Побединский Г. Г., Прусаков А. Н. О критериях качества государственных геопространственных данных Российской Федерации // Россия: тенденции и перспективы развития. Москва, 20–21 декабря 2018 года. Москва: Институт научной информации по общественным наукам РАН, 2019. № 14-1. С. 190–197.

Репозиторий. URL: https://github.com/IgorZinyaev/compare_road (дата обращения 01.06.2025)

Biljecki F., Chow Y. S., Lee K. Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes // Building and Environment. June 2023. Vol. 237. P. 1–19.

Chen Y., Wei Z., Zhou Q. Assessing completeness of global airport data in OSM // Advances in Cartography and GIScience of the ICA. August 2023. Vol. 4. P. 1–8.

Diva-GIS. URL: https://diva-gis.org/data.html (accessed on 01.06.2025)

Elias E., Amorim F., Silva L., Schmidt M., Camboim S., Fernandes V. A proposal for a QGIS plugin for spatio-temporal analysis of OSM data quality: the case study for the city of Salvador, Brazil. // State of The Map 2021. 9–11 July 2021. P. 27–30.

Gokceoglu C. Quality analysis of Big Geodata via Machine learning // International Conference on Data Science, Machine Learning and Statistics 2019. Van, Turkey. 26–29 June 2019. 30 slides.

Kin D., Lazorenko N. The method for the quality evaluation of open geospatial data for creation and updating of datasets for National Spatial Data Infrastructure in Ukraine // Polish Cartographical Review. September 2021. Vol. 53. P. 13–20.

LineComparison. URL: https://plugins.qgis.org/ plugins/linecomparison/ (accessed on 01.06.2025)

LineSimilarity. URL: https://plugins.qgis.org/plugins/line_similarity/ (accessed on 01.06.2025)

Mocnik F.-B., Mobasheri A., Zipf A. Open Source data mining infrastructure for exploring and analyzing OpenStreetMap // Open

Geospatial Data, Software and Standards. May 2018. Vol. 3. P. 1–15.

Moradi M., Roche S., Mostafavi M. Evaluating OSM Building Footprint Data Quality in Quebec Province, Canada from 2018 to 2023: A Comparative Study // Geomatics. December 2023. Vol. 3. P. 541–562.

Natural Earth. URL: https://www.naturalearthdata.com/ (accessed on 01.06.2025)

Open Street Map. URL: www.openstreetmap.org (дата обращения 01.06.2025)

Plugin Builder. URL: https://plugins.qgis.org/plugins/pluginbuilder/ (accessed on 01.06.2025)

QGIS. URL: https://plugins.qgis.org (accessed on 01.06.2025)

QuickOSM. URL: https://plugins.qgis.org/plugins/QuickOSM/ (accessed on 01.06.2025)

Roads of the Former Soviet Union. URL: clck.ru/3PKxcb (accessed on 01.06.2025)

Sehra S., Singh J., Rai H. A Systematic Study of OpenStreetMap Data Quality Assessment // 2014 Eleventh International Conference on Information Technology: New Generations. Las-Vegas, 7–9 April 2014. P. 377–380.

VMAP Level 1. URL: https://www.mapability.com/info/vmap1_intro.php (accessed on 01.06.2025)

Wagner M., Henzen C. Quality Assurance for Spatial Research Data // ISPRS International Journal of Geo-Information. June 2022. No. 11(6). P. 334.

Wang M., Li Q., Hu Q., Zhou, M. Quality analysis of Open Street map data // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. May 2013. Vol. 40. P. 155–158.

Zhang X., Anc J., Zhouc Y., Yang M, and Zhao X. How sustainable is OpenStreetMap? Tracking individual trajectories of editing behavior // International journal of digital Earth. February 2024. Vol. 17. No. 1. P. 1–21.

REFERENCES

Biljecki F., Chow Y. S., Lee K., Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes. *Building and Environment*, June 2023, Vol. 237, pp. 1–19.

Chen Y., Wei Z., Zhou Q., Assessing completeness of global airport data in OSM, *Advances in Cartography and GIScience of the ICA*, August 2023, Vol. 4, pp. 1–8.

Diva-GIS, URL: https://diva-gis.org/data.html (2025, 1 June)

Dvorkin B. A., Infrastruktura prostranstvennyh dannyh: regional'nyj aspekt (Spatial data infrastructure: regional perspective), *Geomatika*, 2014, No 1, pp. 17–21.

Dyshlenko S. G., Analiz i razrabotka harakteristik kachestva geodannyh (Analysis and development of geodata quality characteristics), *Perspektivy nauki i obrazovanija*, 2016, No. 2(20), pp. 23–27.

Elias E., Amorim F., Silva L., Schmidt M., Camboim S., Fernandes V., A proposal for a QGIS plugin for spatio-temporal analysis of OSM data quality: the case study for the city of Salvador, Brazil. *State of The Map 2021*, 9–11 July 2021, pp. 27–30.

Gokceoglu C. Quality analysis of Big Geodata via Machine learning. *International Conference on Data Science, Machine Learning and Statistics 2019*, Van, Turkey, 26–29 June 2019. 30 slides.

GOST P 57773-2017, URL: https://files.stroyinf.ru/Data/655/65547.pdf (2025, 1 June)

https://habr.com/ru/companies/lemana_tech/articles/857044/ (2025, 1 June)

https://habr.com/ru/companies/sibur_official/articles/860372/ (2025, 1 June)

https://mpr.nso.ru/page/653 (2025, 1 June)

ISO 19157:2013, URL: (https://docs.cntd.ru/document/1200157078 (2025, 1 June)

Kin D., Lazorenko N., The method for the quality evaluation of open geospatial data for creation and updating of datasets for National Spatial Data Infrastructure in Ukraine, *Polish Cartographical Review*, September 2021, Vol. 53, pp. 13–20.

LineComparison, URL: https://plugins.qgis.org/plugins/linecomparison/ (2025, 1 June)

LineSimilarity, URL: https://plugins.qgis.org/plugins/line_similarity/ (2025, 1 June)

Lur'e I. K., *Geoinformacionnoe kartografirovanie. Metody geoinformatiki i cifrovoj obrabotki kosmicheskih snimkov (Geoinformation mapping. Methods of geoinformatics and digital processing of space images),* Moscow: KDU, 2010, 424 p.

Martynova E. V., Metodika ocenki kachestva prostranstvennyh dannyh (Methodology for assessing the quality of spatial data), *Jekonomika i upravlenie: problemy, reshenija,* 2023, No. 7, Vol. 2, pp. 111–118.

Mocnik F.-B., Mobasheri A., Zipf A., Open Source data mining infrastructure for

exploring and analyzing OpenStreetMap. *Open Geospatial Data, Software and Standards*, May 2018, Vol. 3, pp. 1-15.

Moradi M., Roche S., Mostafavi M.A, Evaluating OSM Building Footprint Data Quality in Quebec Province, Canada from 2018 to 2023: A Comparative Study, *Geomatics*, December 2023, Vol. 3, pp. 541–562.

Natural Earth, URL: https://www.naturalearthdata.com/ (2025, 1 June)

ODM 218.9.008–2019, Geoinformacionnye sistemy avtomobil'nyh dorog: poryadok sbora, hraneniya i obnovleniya dannyh: otraslevoj dorozhnyj metodicheskij document (Geographic information systems for highways: procedure for collecting, storing and updating data: industry road methodological document), *Federal'noe dorozhnoe agentstvo*, July 29. No 1983-r, OOO «IndorSoft», Moscow: «Informavtodor», 2021, 72 p.

Open Street Map, URL: www.openstreetmap.org (2025, 1 June)

Plugin Builder, URL: https://plugins.qgis.org/plugins/pluginbuilder/ (2025, 1 June)

Pobedinskij G. G., Prusakov A. N., O kriterijah kachestva gosudarstvennyh geoprostranstvennyh dannyh Rossijskoj Federacii (On the quality criteria for state geospatial data of the Russian Federation) *Rossija: tendencii i perspektivy razvitija* (Russia: Trends and Development Prospects), Moscow, 20–21 December 2018, Moscow: Institut nauchnoj informacii po obshhestvennym naukam RAN, 2019, No. 14-1, pp. 190–197.

QGIS, URL: https://plugins.qgis.org (2025, 1 June)

QuickOSM, URL: https://plugins.qgis.org/plugins/QuickOSM/ (2025, 1 June)

Repozitorij, URL: https://github.com/Igor-Zinyaev/compare_road (2025, 1 June)

Roads of the Former Soviet Union, URL: clck.ru/3PKxcb (2025, 1 June)

Sehra S., Singh J., Rai H., A Systematic Study of OpenStreetMap Data Quality Assessment. *2014 Eleventh International Conference on Information Technology: New Generations*, Las-Vegas, NV, 7–9 April 2014, pp. 377–380.

VMAP Level 1, URL: https://www.mapability.com/info/vmap1_intro.php (2025, 1 June)

Wagner M., Henzen C., Quality Assurance for Spatial Research Data, *ISPRS International Journal of Geo-Information*, June 2022, No. 11(6), pp. 334.

Wang M., Li Q., Hu Q., Zhou M., Quality analysis of Open Street Map data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* May 2013, Vol. 40, pp. 155–158.

Zhang X., Anc J., Zhouc Y., Yang M, and Zhao X., How sustainable is OpenStreetMap? Tracking individual trajectories of editing behavior. *International journal of digital Earth*, February 2024, Vol. 17, No. 1, pp. 1–21.

DEVELOPMENT OF AN OPEN SOURCE QGIS PLUG-IN FOR ASSESSING THE QUALITY OF ROAD NETWORK DATA IN REGIONAL FORESTRY TRANSPORTATION PROJECTS

E. S. Podolskaia[1*], I. M. Zinyaev[2]

[1] Isaev Centre for Forest Ecology and Productivity of the RAS
Profsoyuznaya st. 84/32 bldg. 14, Moscow 117997, Russia

[2]MIREA – Russian Technological University
Vernadsky avenue, 78, Moscow 119454, Russia

*E-mail: podols_kate@mail.ru

Received: 02.06.2025
Revised: 19.06.2025
Accepted: 15.07.2025

Paper presents the development of a plugin for Open Source QGIS for automated assessment of data quality of road network when solving the logistics and transport problems of the forestry complex and infrastructure projects in a Russian region (Novosibirsk region). Automation of data quality assessment is especially relevant in the context of Open Data use, diversity and an increase in the number of sources. National standards, industry documents, cases of Russian companies were studied. The developed plug-in "Compare_road" is used to analyze and verify data on roads in the format of vector linear files. During the analysis, the following metrics are calculated: accuracy of compliance, completeness and data relevance. Plug-in is published on GitHub under the MIT license. The target audience of the plug-in is researchers, students and amateur-users. To test the work of the plug-in, an analysis of data on the road network of the Novosibirsk region of the Natural Earth, VMAP, Digital Chart of the World projects was carried out using OSM as a reference. The best accuracy indicators were shown by the Digital Chart of the World - 270.7 m, the best VMAP data completeness indicator - 22.6%. The analysis of roads data quality for the forestries of Novosibirsk region was carried out. Best quality resullts have Tatarskoye (127.6 m, 35.9 %) and Maslyaninskoye (154 m, 35.9 %), worst - Kushtovskoye (378.5 m, 0.7 %) and Suzunskoye (418 m, 1.5 %).

Keywords: *GIS, Open Source, QGIS, plug-in, OSM, Natural Earth, road network, quality assessment*

Рецензент: к. с.-х. н., и. о. директора ФБУ ВНИИЛМ Сидоренков В. М.