

## **Обзор современных инструментальных средств и методов для автоматизации процессов сбора и анализа геопространственных данных**

**А.И. Рогачев, А.И. Газов, А.А. Кузмицкий, Д.В. Федосенко, М.И. Труфанов**  
ФГБУН Центр информационных технологий в проектировании РАН  
РФ, г. Одинцово

### **Аннотация**

Постоянный рост количества аппаратов для аэрофотосъемки и, как следствие, увеличение объема геопространственных данных, содержащих высококачественные снимки, позволяет производить разведочный анализ данных (EDA), используя открытые источники и open-source платформы. Одной из ключевых проблем исследования природы тропических циклонов является отсутствие надёжных данных о локальных параметрах тропических циклонов, ибо современные приборы не выдерживают разрушительных воздействий тропического циклона, а дистанционные методы измерения параметров циклонов не позволяют получить необходимую информацию в полной мере.

**Ключевые слова:** Геоданные, анализ изображений, циклоны, сбор данных, API, моделирование

**Актуальность** Одним из эффективных методов разработки физической модели процессов развития мощных тропических циклонов является создание модельной экспериментальной установки, на которой можно проверить концепцию магнитогидродинамической природы изучаемых процессов, оценить роль атмосферного электричества в развитии тропических циклонов и выявить степень влияния отдельных факторов на общую картину протекания мощных тропических циклонов [3].

Модельный эксперимент возможно провести внутри цилиндрической камеры, боковые стенки которой изготовлены из диэлектрика (органического стекла), а торцевые фланцы из проводника, к которым будет подаваться разность электрических потенциалов. В установке возможно создать вертикальные электрические токи, протекающие по модельному рабочему телу (вода с примесью поваренной соли для повышения электрической проводимости воды), величина которых будет обеспечивать модельную величину индукции собственного азимутального магнитного поля в соответствии с величиной числа Гартмана.

Однако, подобный подход требует значительного количества временных ресурсов на проектирование и, непосредственно, изготовление установки.

Для сравнительного эксперимента необходимо предварительное моделирование поведения циклона. Одним из возможных способов проведение

моделирования может быть применение нейросетевой модели, например - генеративной сети Conditional Generative Adversarial Nets (CGAN). Идея такой модели сводится к обучению и использованию двух моделей – дискриминатора и генератора [2]. Задача генератора заключается в генерации изображений на основе входных параметров, таких как, например, температура и влажность воздуха, и скрытого состояния. Дискриминатор же должен отличать сгенерированные данные от настоящих, то есть тех, что используются при обучении. В случае использования Wasserstein GAN (WGAN) в качестве минимизируемого функционала выступает расстояние Вассерштейна [1], а генератор и дискриминатор являются сверточными нейронными сетями, для которых в качестве функции активации выступает Leaky ReLU на выходе всех слоев, кроме последнего. Для обучения подобной модели необходима как можно большая по объему обучающая выборка, что крайне важно для генеративных нейронных сетей. Сбор данных возможно реализовать с помощью открытых источников. Ниже рассмотрены некоторые из них.

**Результаты исследования.** Основные нововведения, влияющие на индустрию географических информационных систем, связаны с некоторыми прорывами в технологической и коммуникационной сферах. Так, например, влияние оказало внедрение лазерного сканирования, появление и развитие сервиса Google Earth, и, как наиболее актуальное в контексте обзора, развитие open-source решений. Повышению точности данных и росту числа собираемых параметров способствует распространение носимых устройств, информация с которых и собирается GIS-сервисами с согласия пользователей. Примерами коммерчески успешных проектов может быть, помимо упомянутого сервиса от Google, AWS, Global Mapper, Яндекс. Снижение стоимости производства и распространение дронов для сбора информации так же способствует росту объемов гео-данных, так как у геодезистов появилась возможность получать изображения с высоким разрешением с помощью БПЛА, однако наибольшая часть информации получается посредством дистанционного зондирования поверхности Земли (ДЗЗ) [4], которое проводится с помощью искусственных спутников, позволяющих получать снимки и измерения энергетических характеристик с наибольшей точностью [5]. Также для сбора информации могут быть использованы летательные аппараты, производящие сбор информации с более низкой высоты (до десятков километров), например - специально оснащённые самолеты с установленным на них оборудованием для проведения аэросъемки и датчики, позволяющие собирать данные точно, позволяя увеличить качество получаемой информации в тех случаях, когда использование спутников не предоставляет информации необходимой точности, либо же их применение вообще невозможно. В частности, таким образом возможно получить показания приборов непосредственно из эпицентра смерча.

Для построения гео-изображения стандартом является использование Shape-файлов, содержащих в себе векторное представление границ полигонов, соответствующих тем или иным географическим объектам.

Наиболее актуальным на сегодняшний день источником растровых данных из open-source проектов для работы с подобным форматом является OpenStreetMap (OSM) – картографический сервис, позволяющий каждому желающему добавлять недостающую информацию, или же редактировать существующую. Что более актуально – OSM имеет API, позволяющий автоматизировать процесс взаимодействия с данным сервисом посредством написания соответствующих скриптов на языке Python. Overpass API (или OSM3S) это доступный только для чтения API, который позволяет извлекать выборочные данные из базы OSM по пользовательскому запросу. Он выступает в качестве базы данных через Интернет: клиент посылает запрос к API и получает обратно набор данных, который соответствует запросу. Он действует в качестве базы данных Back-end для различных сервисов. Для обращения в Overpass API посредством Python следует использовать модуль overpy в качестве обертки.

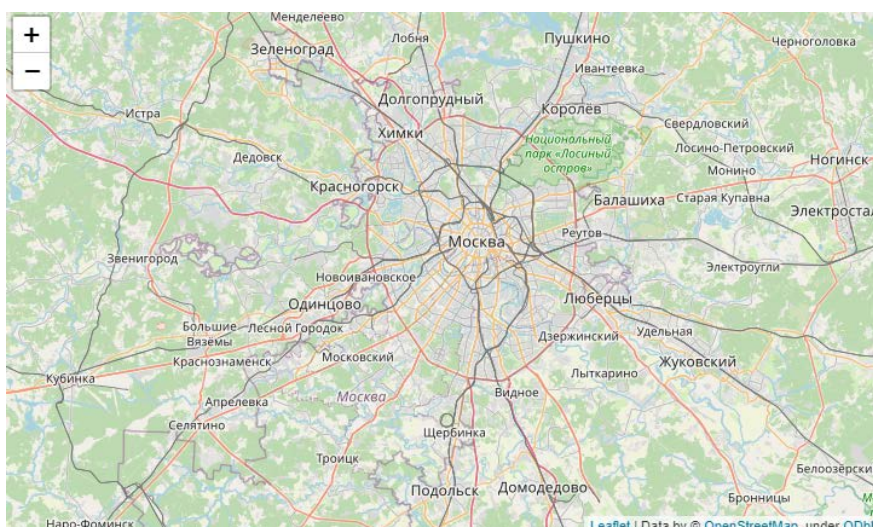
Для хранения и обработки различных данных в формате датафрейма возможно использовать georandas – библиотеку, работа с которой аналогична работе с пакетом pandas, что позволяет встраивать обработку геоданных в готовые пайплайны, сводя необходимые изменения к минимуму. Для визуализации объекта такого датафрейма существуют встроенные методы, использующие данные, хранящиеся в поле “geometry”. Пример визуализации, полученной с помощью API OSM и метода plot датафрейма georandas приведен на рисунке ниже.



**Рис.1.** Пример визуализированной карты Москвы с административным делением, полученной с помощью OSM и georandas

Для построения динамических визуализаций может быть использована библиотека Folium, экспортировать данные для которой возможно в формате GeoJSON. Соответствующие функции для экспорта присутствуют среди стандартных методов датафреймов georandas. При этом, в отличии от

стандартных визуализаций в `gpd`, Folium предоставляет интерактивную карту, которая может обновляться по мере обновления данных для построения.

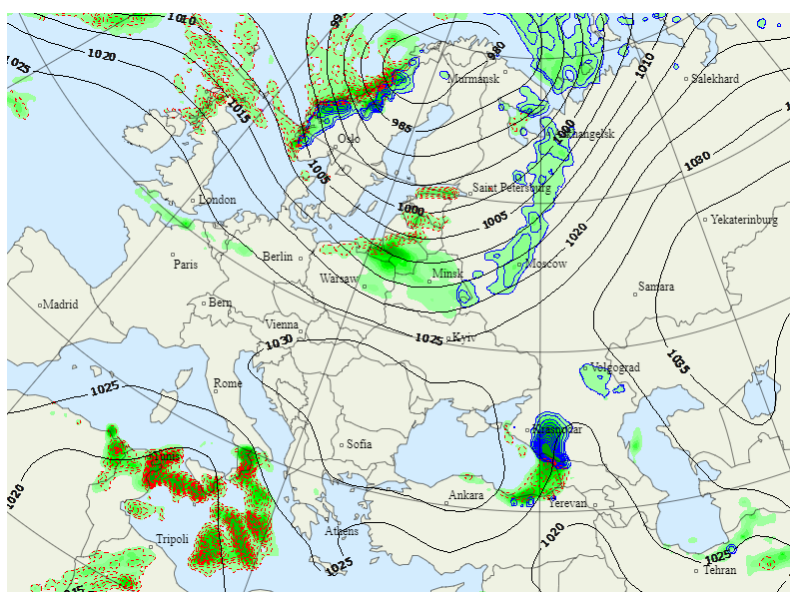


**Рис.2.** Пример визуализированной интерактивной карты Москвы, полученной с помощью OSM и Folium

Однако, помимо построения самих визуализаций, необходимы данные о различных значениях признаков, которые будут использованы в качестве условия при обучении CGAN, которые OSM не предоставляет. Решить данную задачу можно с помощью API от `gismeteo`. Среди данных, предоставляемых сервисом, можно получить информацию о таких признаках, как:

- Temperature - Температура
- Description - Словесное описание погоды
- Humidity – влажность
- Pressure – давление
- Cloudiness – облачность
- Storm – вероятность грозы, если запрашивается информация о будущей дате
- Precipitation – информация об осадках, содержащая их тип и количество
- Gm – метку описания геомагнитной обстановки
- Wind – направление и скорость ветра

На основе собранных данных возможно построить схематичную визуализацию с использованием значений параметров.



**Рис.3.** Пример возможной визуализации на основе данных, полученных через Gismeteo API

Используя собранные данные, а также фотоснимки различных областей, возможно обучить CGAN, в качестве целевой переменной выбрав, вообще говоря, любой из имеющихся параметров – это может быть как само изображение, так и значение, например, вероятности грозы. При этом обработкой изображения может заниматься отдельная сверточная нейронная сеть, выход которой будет связан с выходом полносвязной сети, обрабатывающей значения остальных признаков.

### Заключение

Рассмотрены наиболее популярные сервисы, предоставляющие возможности для сбора и обработки геопространственных данных. Благодаря наличию API процесс сбора может быть автоматизирован с помощью скриптов на языке Python, без необходимости создания дополнительных парсеров для различных сайтов. Хранение и визуализацию данных можно реализовать с использованием библиотек georandas и Folium. Полученные растровые данные могут быть использованы как в рамках EDA, так и для последующего использования в различных пайплайнах машинного обучения. Контроль за природными изменениями и моделирование с использованием данных, полученных с помощью ДЗЗ, способны минимизировать риски и оптимизировать процесс принятия решения о превентивных мерах в случае возникновения природных катастроф.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 19-07-00924 а*

## **Литература**

1. *Arjovsky, M., Chintala, S., and Bottou, L., “Wasserstein GAN”, arXiv e-prints. – 2017.*
2. *Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. Generative Adversarial Networks. – 2014.*
3. *Гридин В.Н., Смахтин А.П. Магнитогидродинамическая модель развития тропических циклонов и способов их подавления. Труды международной научно-технической конференции «Информационные технологии и математическое моделирование систем», М. – 2015. – С. 65-69.*
4. *Книжников Ю.Ф., Кравцова В.И., Тутубалина О.В. Аэрокосмические методы географических исследований, М. Издательский центр Академия. – 2011.*
5. *Некрасов Г. А., Поливода Д. Э., Прокофьева Е. Н. Анализ источников пространственных данных в развитии технологии недропользования // Горный информационно-аналитический бюллетень. – 2020. – № 5. – С. 164 – 176.*