

ПРИМЕНЕНИЕ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА СТАТИСТИЧЕСКИХ ДАННЫХ ПРИ ПОСТРОЕНИИ КАРТОГРАФИЧЕСКИХ МОДЕЛЕЙ

А. С. Нечаусов, О. В. Чуприна

Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Харьков, Украина, a.nechausov@khai.edu, ksyusha.chuprina@gmail.com

Мониторинг опасных заболеваний является актуальной проблемой в области здравоохранения в любой стране. В эпоху глобализации здоровье человека зависит от ряда внешних факторов. На ряду с обострением экологической, энергетической, продовольственной, демографической и других жизненно важных проблем для всего человечества возникает задача контроля и урегулирования связанных с этим процессов. Для выполнения поставленных задач необходимы новые методы контроля распространения и развития болезней в глобальных масштабах на уровнях муниципального управления, методы изучения условий жизни человека, особенностей его географического расположения, методики и технологии, позволяющие отражать данные исследований в режиме реального времени благодаря существующим интерактивным web-ресурсам.

По данным ВОЗ за последние несколько лет наблюдается незначительное уменьшение показателей смертности от опасных болезней, что свидетельствует об эффективности созданных проектов и принимаемых решений. Однако, несмотря на то, что в последние годы отмечался заметный прогресс, во многих странах с наиболее высокими показателями смертности продолжают сохраняться существенные различия в состоянии здоровья как между странами, так и внутри стран. Тем не менее, современные тенденции продолжают усиливать потребность в обеспечении коллективного обмена информацией и распространения успешных подходов, требуемых для решения общих проблем в данном вопросе [3].

С целью получения полной базы данных о заболеваемости опасными болезнями каждая заинтересованная страна проводит сбор статистических данных, принимает на законодательном уровне программы для улучшения уровня жизни и здоровья населения. К сожалению, обычного сбора статистических данных для решения всего спектра задач, связанных с анализом текущей ситуации недостаточно. При этом отсутствует возможность в полном объеме оценить ситуацию и принять обоснованные решения по внедрению мероприятий для предотвращения заболевания или его локализации. С целью получения общей картины заболеваемости опасных заболеваний с учетом возможных факторов влияния необходимо использовать не только статистические данные, но и средства определения источников заболеваний и причины их возникновения, методы прогнозирования направления распространения. Применение комплексного анализа предоставит возможность выявления явных и неявных закономерностей, связанных как с географическим положением, так и с другим набором статистической информации, что позволит в дальнейшем строить подробные прогнозные оценки процесса.

Для решения задачи усовершенствования средств мониторинга случайных статистических процессов в сфере здравоохранения была предложена методика построения картографической модели кластерного анализа данных на основе статистических данных. В качестве тестового набора данных был взят отчет о заболеваемости туберкулезом в Харьковской области за 2013-2015 гг, на основании которого было проведено моделирование уровня заболеваемости с учетом распределения зарегистрированных случаев на местности для дальнейшего отслеживания взаимосвязей с другими пространственными факторами и прогнозирования возможных зон риска распространения заболевания.

Важным и существенным аспектом для проверки робастности методики является наличие достоверных данных за определенный период для отображения динамики изменения параметров в хронологическом разрезе. Однако данная методика может применяться не только в области здравоохранения, но и для геостатистического анализа случайных величин, таких как пожары, наводнения или землетрясения.

Современное программное обеспечение в области ГИС имеет достаточно широкий набор инструментов геопространственного анализа. Существующие методы имеют множество преимуществ для решения конкретных задач, но, если рассматривать их в контексте проблемы мониторинга с привязкой к местности и времени, их применение будет иметь низкую эффективность. Большинство систем работает с растровыми или векторными файлами, базами данных и слоями различных форматов, что в результате позволяет создать электронную карту, но не определить динамику изменения во времени. Для решения поставленной задачи предлагается использовать программное обеспечение, которое будет иметь возможность отображения динамики изменения процесса во времени, будет содержать встроенные методы визуализации статистических данных в виде тепловой карты и в качестве картографической подложки будет использовать интерактивную онлайн карту. Проанализировав требования к программным продуктам и сопоставив их со спецификацией различных офисных и онлайн решений, было принято решение, что для создания требуемой картографической модели целесообразно использовать географический трехмерный анализатор офисного пакета Microsoft Excel GeoFlow (Power Map). Инструментарий GeoFlow позволяет осуществлять выявление различного рода закономерностей распространения заболевания, наглядно представляя модель распространения в виде тепловой карты, что позволит своевременно проводить предупредительные мероприятия по факту распространения заболевания и выявлять их источники [2].

Возможности Power Map:

- отображение статистических данных на карте: есть возможность визуализации в трехмерном виде на Картах Bing для более миллиона строк данных из таблицы или модели данных Excel;

- анализ информации при просмотре данных в географическом пространстве и отслеживание их изменения в хронологическом разрезе;

- обмен интересными сведениями: создание снимков экрана и анимированных управляемых видеотуров, которыми можно делиться с другими пользователями посредством интернета;

- экспорт туров в видео;

- Power Map позволяет пользователям самостоятельно настраивать несколько слоев отображения данных (до миллиона строк в таблице) на трехмерной карте;

- управление слоями карты, настройка вариантов визуализации слоев в виде столбцов данных, круговых диаграмм, тепловых карт, а также визуализация изменений данных с течением времени [3].

Таким образом, используя трехмерные онлайн карты Bing совместно с средствами редактирования таблиц метаданных, можно различными способами представлять географические данные: в виде гистограмм, точечных диаграмм, диаграмм регионов или тепловых карт.

Для решения поставленной задачи динамического мониторинга явлений, использовались статистические данные о заболеваемости туберкулезом в Харьковской области за несколько лет, онлайн карты Bing и метод отображения в виде тепловой карты. В качестве актуальных источников выборки данных были использованы Государственная служба статистики Украины и Центр общественного здоровья, созданные Министерством здравоохранения Украины. На сайтах этих организаций присутствует информация о документах, являющихся действующими в данный момент времени, статистическая информация по уровню случаев на единицу времени по различным заболеваниям и много других категорий, связанных с охраной здоровья в государстве. В результате была определена выборка данных, на основании которых была сформирована БД районов Харькова с показателями заболеваемости, для которой проводится геостатистический анализ в хронологическом разрезе.

После анализа БД, по названиям улиц была произведена привязка с дальнейшим нанесением объектов на онлайн карты Bing (рис. 1)

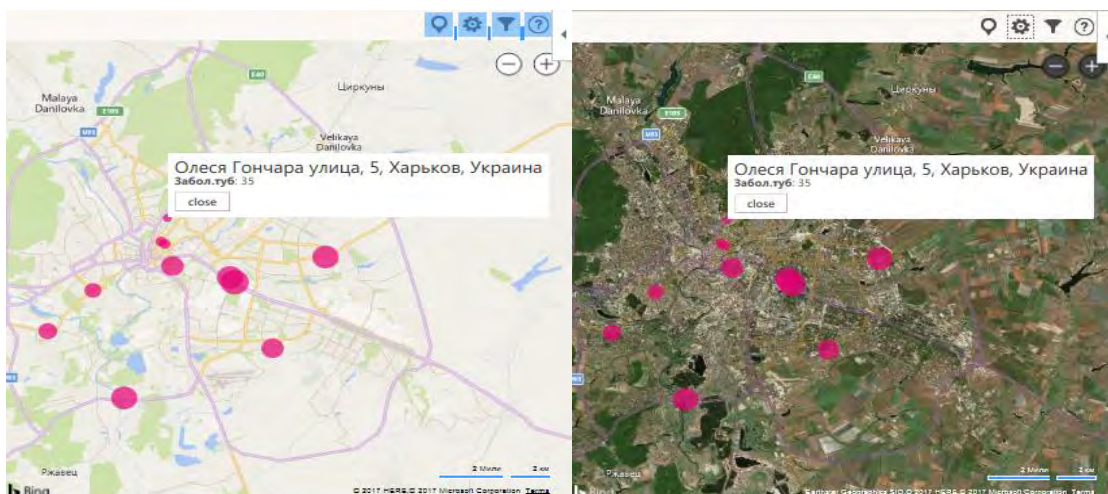


Рисунок 1 – Результаты привязки БД случаев заболеваемости туберкулезом к картографическому сервису Bing

Далее с помощью дополнительной встроенной надстройки в Excel Power Map была создана интерактивная трехмерная карта, на которой статистические данные представляются в виде тепловых точек, которые изменяют свой размер и цвет в соответствии с изменениями показателей заболеваемости на протяжении нескольких лет. Динамика представляется в виде анимации и может быть записана в форме видеофайла формата *.mp4.

Результат создания динамической картографической модели на основе статистических данных о заболеваемости туберкулезом, который был практически реализован в PowerMap, представлен на рис.2.

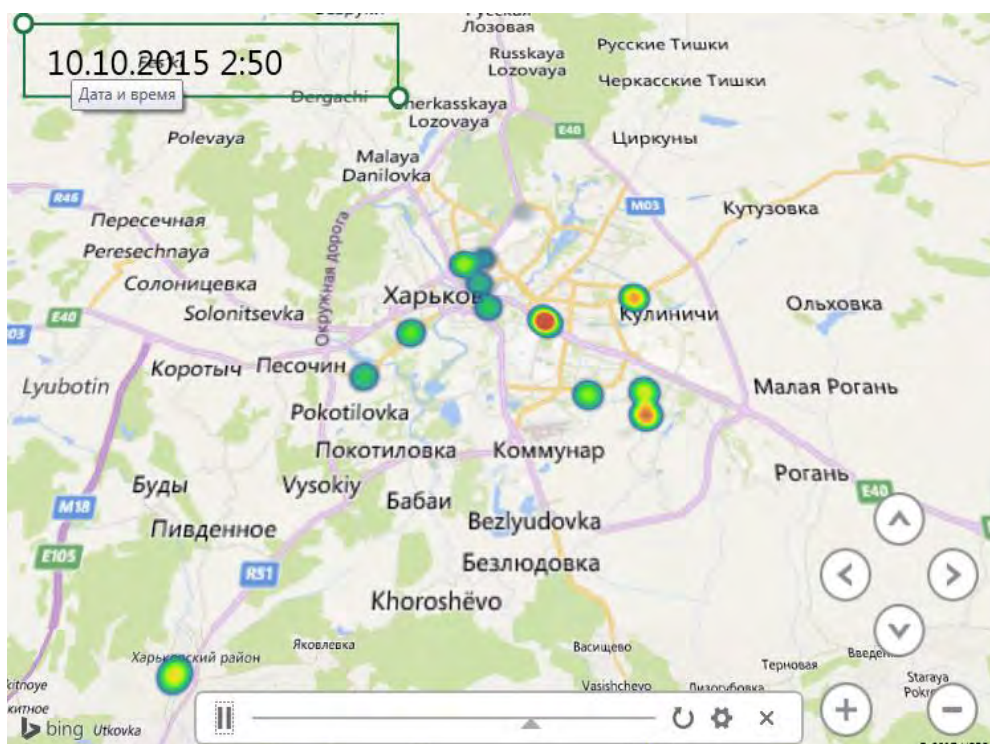


Рисунок 2 – Фрагмент видеофайла динамической картографической модели показателей заболеваемости туберкулезом в Харьковской области, созданный в PowerMap

На примере реализации модели динамического анализа распространения заболевания туберкулезом в Харьковской области была представлена методика, которая может иметь

применение в различных областях для определения закономерностей распространения различных случайных показателей, таких как пожары, рассеяние опасных газозвушных примесей в атмосфере [6], наводнения, выявления оползней или землетрясений. С учетом дополнительных факторов, определяющих эти случайные величины и на основании динамической карты, можно составлять прогнозные оценки и рассчитывать их вероятности. Выше перечисленные результаты можно использовать не только для мониторинга и анализа распространения показателей, но и для более наглядной визуализации результатов прогнозирования, выявления очагов и причин возникновения случайных величин, что позволит принимать обоснованные решения по устранению источников или последствий негативного влияния рассматриваемых величин и показателей.

Условия, необходимые для применения данной методики:

- актуальные достоверные источники данных;
- статистические данные за определенный период для создания динамической картографической модели;
- БД на основе статистических данных с учетом различных показателей;
- наличие соответствующего ПО Microsoft Excel с встроенными надстройками;
- наличие подключения к сети Интернет и картам Bing.

Целесообразность использования предложенной методики обоснована успешным практическим применением выбранных программных средств для решения задач динамического моделирования статистико-аналитических картографических данных на реальном примере. По результатам работы был предложен алгоритм методики построения картографической модели анализа статистических данных с дальнейшей перспективой использования инструментов для составления прогнозных оценок.

Алгоритм методики приведен на рис. 3.

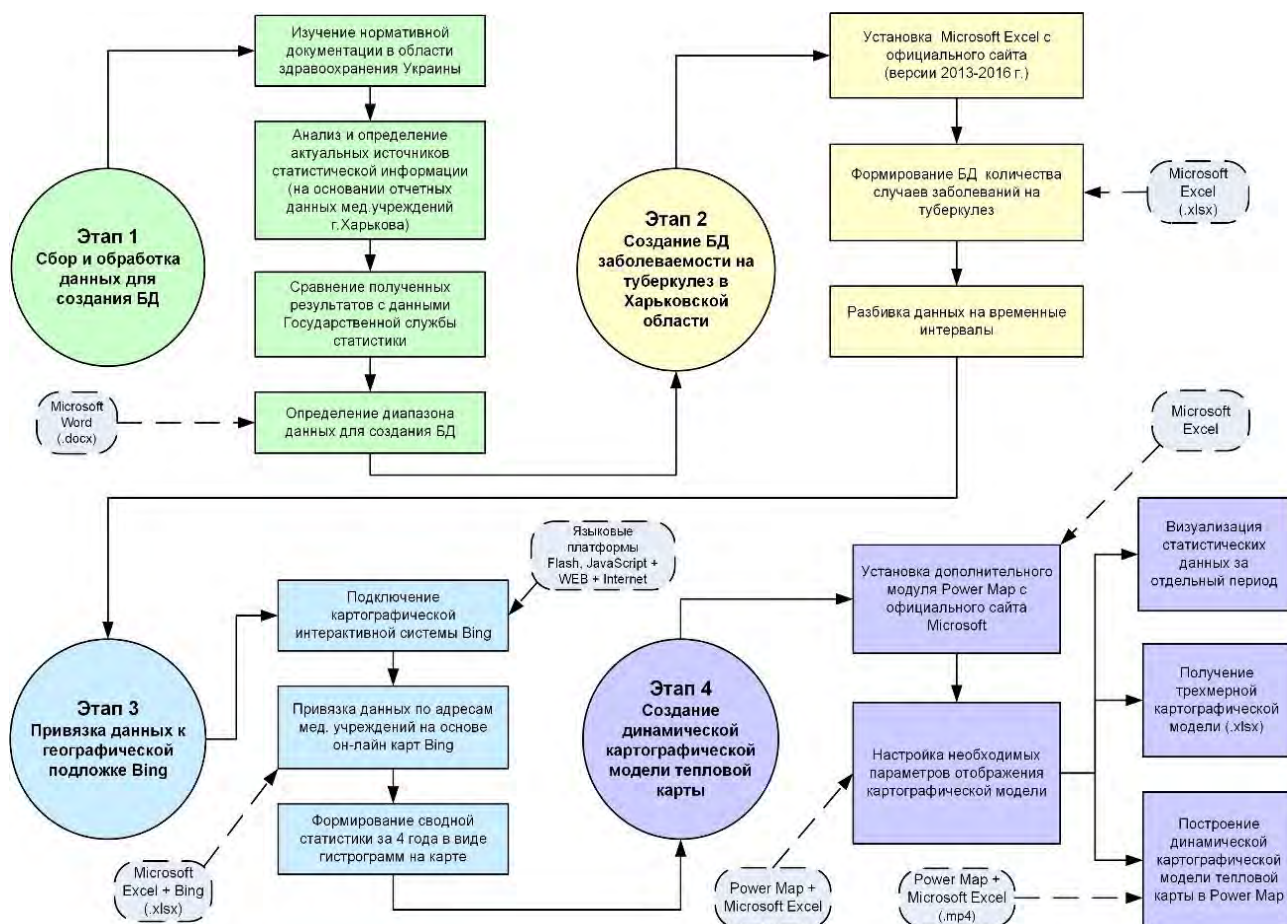


Рисунок 3 – Алгоритм методики построения картографической модели данных

ПЕРЕЧЕНЬ ИСТОЧНИКОВ

1. Официальный веб–портал Верховной Рады Украины [Электронный ресурс] / Законодательные акты. URL <http://zakon1.rada.gov.ua/laws/> (дата обращения 10.09.17).
2. Краак, М. Картография: визуализация геопространственных данных [Текст] / М. Краак, Ф. Ормелинг// Пер. с англ. под ред. В. С. Тикунова, Научный мир – Москва, 2005. – с. 325.
3. Всемирная организация здравоохранения [Электронный ресурс] / Мировая статистика здравоохранения 2013 – 2014. URL http://www.who.int/gho/publications/world_health_statistics/2013/ru/ (дата обращения 22.03.17)
4. Библиотека технической тематики. Техническая литература [Электронный ресурс] / Уокенбах, Д. Microsoft Excel 2013. Библия пользователя – 2015. URL [https://www.htbook.ru/samouchitel_excel_2013 – 2015](https://www.htbook.ru/samouchitel_excel_2013-2015) (дата обращения 13.04.17)
5. Леонтьев, В. П. Office 2016. Новейший самоучитель [Текст] / В.П. Леонтьев // MS Office. Офисные программы Майкрософт: Эксмо – Москва, 2015 – с. 368.
6. Nechausov, A. S. The information model of the system for local atmospheric air pollution monitoring [Text] / A. S. Nechausov // Системи обробки інформації : зб. наук. пр. Харк. ун-ту Повітр. Сил ім. Івана Кожедуба. – Харків, 2016. – Вип. 2 (139). – с. 190 – 195.