

УДК 004.942

ЗАСТОСУВАННЯ ГЕОІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ESRI ДЛЯ МОДЕЛЮВАННЯ ЕПІДЕМІОЛОГІЧНИХ ПРОЦЕСІВ

Ковзар В. Б.¹, Філозоф Р. С.¹, Савченко М. О.²

¹ПрАТ «ЕКОММ Со», Київ, Україна

²Київський національний університет будівництва і архітектури, Київ, Україна

E-mail: info@ecomm.kiev.ua¹

В статті описаний досвід практичного застосування геоінформаційної технології Esri для моделювання епідеміологічних захворювань. Використання авторами вбудованого в ArcGIS інструменту експлораторної регресії дозволило визначити потенційні фактори, що впливають на розвиток захворюваності, та з'ясувати величину їх впливу.

Ключові слова: моделювання, статистичні дані, регресійний аналіз, епідеміологічні дослідження, ГІС.

ВСТУП

У світовій практиці при оцінці якості життя людей на перше місце висувається стан їхнього здоров'я, оскільки саме воно є основою повноцінного життя і діяльності, як кожної людини, так і суспільства в цілому.

З давніх часів людство зазнавало вплив багатьох інфекційних захворювань, а спалахи епідемій холери, чуми, тифу, грипу та ін. забирали сотні тисяч життів. В останні роки, вченими були виявлені нові види вірусів. В процесі глобальних змін клімату, глобалізації, освоєння територій проблема не зменшується, а навпаки загострюється. Збільшення міграції людей призводить до швидкого поширення інфекції не лише із країни в країну, але й між різними континентами. Надто сприятливі умови для розвитку епідемій виникають при неврожаях, стихійних лихах, війнах, коли відбуваються безперервні пересування людських мас (біженці, військові та ін.), оскільки проблематичним стає здійснення заходів оздоровчого та протиепідемічного характеру.

Тому, враховуючи ці фактори, перед сучасними епідеміологами гостро постала необхідність моделювання і прогнозування процесів поширення захворювань, щоб оперативно і адекватно реагувати на епідеміологічну ситуацію.

В останні десятиліття для контролю стану населення і прийняття найбільш ефективних заходів щодо підтримання та поліпшення здоров'я людей і запобігання загроз епідемій небезпечних хвороб, все більш широко застосовуються геоінформаційні технології. Базові терміни та методичні засади, на яких спиралося дослідження описані в роботах [1-5]. Метою даного дослідження є виявлення факторів, що впливають на захворюваність населення гострими респіраторними вірусними інфекціями та грипом, і застосування інструментів ArcGIS (Esri, США), зокрема експлораторної регресії, для визначення величини їх впливу. Об'єктом дослідження є територія України, в якості таксономічних одиниць виступають 24 адміністративні області та АРК Крим (без урахування міст обласного значення Київ

та Севастополь). Для досягнення поставленої мети необхідне вирішення наступних задач: оцінка та перетворення доступних статистичних даних з метою створення селективної статистичної вибірки даних, попередній аналіз факторів та величини їх впливу, пошук оптимальної комбінації факторів та моделей, що описують взаємозв'язок між незалежними змінними (факторами) та залежною змінною (захворюваністю населення).

ВИКЛАДЕННЯ ОСНОВНОГО МАТЕРІАЛУ

Технологічний прогрес привів людство до нового етапу в розвитку засобів і методів обробки даних. Потік нових ідей, які витікають із сфери комп'ютерних наук, утворених на межі статистики, теорії баз даних і штучного інтелекту, призводить до значного зростання об'єму інформації. З'являється необхідність аналітичної обробки надвеликих об'ємів даних, накопичуваних в інформаційних сховищах даних (Data Warehousing). Дана область отримала назву KDD (Knowledge Discovery In Databases – виявлення знань в базах даних). Сьогодні можна спостерігати зростання кількості програмних продуктів, в яких застосовані технології KDD, а також типів задач, де використання даних технологій дає вагомий економічний ефект. Елементи автоматичної обробки та аналізу даних стають невід'ємною частиною концепції електронних сховищ даних. Основним кроком KDD щодо опрацювання вмісту електронних сховищ даних є Data Mining (виявлення прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних). Не рідко, інтелектуальний аналіз даних і KDD вважають синонімами Data Mining, але фактично Data Mining є основним, але не єдиним елементом в множині KDD. Data Mining – це процес виявлення в сирих даних раніше невідомих, нетривіальних, практично корисних і доступних інтерпретацій знань, необхідних для прийняття рішень в різних сферах людської діяльності. Як правило, поділяється на задачі класифікації, моделювання та прогнозування.

Основу методів Data Mining складають різноманітні методи класифікації, моделювання і прогнозування, основані на використанні дерев рішень, нейронних мереж, генетичних алгоритмів, еволюційного програмування, асоціативної пам'яті, нечіткої логіки. До методів Data Mining нерідко відносять статистичні методи (дескриптивний аналіз, кореляційний та регресійний аналіз, факторний аналіз, дисперсійний аналіз, компонентний аналіз, дискримінантний аналіз, аналіз часових рядів). Одне з найважливіших призначень методів Data Mining полягає в наглядному представленні результатів обчислень, що дозволяє використовувати інструменти Data Mining особами, що не мають спеціальної математичної підготовки. В той же час, використання статистичних методів аналізу даних вимагає володіння знаннями теорії ймовірностей та математичної статистики.

В даному дослідженні було прийнято рішення використовувати один із методів Data Mining – регресійний аналіз. Регресійний аналіз – розділ математичної статистики, що об'єднує практичні методи дослідження регресійної залежності між величинами по статистичним даним. Його мета полягає у визначенні загального

вигляду рівняння регресії, побудові оцінок невідомих параметрів, що входять у рівняння регресії, і перевірці статистичних гіпотез про регресію. При вивченні залежності між двома змінними за результатами спостережень $(x_1, y_1), \dots, (x_n, y_n)$ у відповідності з теорією регресії передбачається, що одна з них Y має деякий розподіл ймовірностей при фіксованому значенні X іншої. На користь даного методу зіграв той факт, що починаючи з версії ArcGIS 10.1, в наборі Моделювання просторових взаємовідношень (Modeling Spatial Relationships) є набір інструментів, що дозволяє виконувати експлораторну/дослідницьку регресію (Exploratory regression) стосовно просторових об'єктів (тобто геокодованих або таких, що мають геопросторову прив'язку) та їх характеристик.

Регресійний аналіз може застосовуватись для розв'язання багатьох прикладних задач в різних сферах. Зокрема для медичної сфери – він може бути використаний для задач аналізу та прогнозування епідеміологічних захворюваностей, виявлення факторів, що впливають на її поширення і визначення величини їх впливу.

Існує три першопричини використання регресійного аналізу:

- змодельовати певне явище, щоб краще його зрозуміти, і можливо, використати це розуміння для впливу на політику прийняття рішення;
- змодельовати певне явище, щоб спрогнозувати його на іншій місцевості або в інший час (основною метою є побудова прогнозової моделі);
- для перевірки гіпотез (наприклад, чи є залежність між показником доходу людей і показником рівня захворюваності).

У контексті геоінформатики та на основі геоінформаційних систем (ГІС) регресійний аналіз дозволяє моделювати, досліджувати ситуацію, що склалася й аналізувати просторові взаємозв'язки, а також виявляти фактори, що впливають на ці просторові взаємини. Регресійний аналіз використовується також і для моделювання припущень про можливість здійснення будь-якої (просторової) події в майбутньому.

Інструменти з даного набору припускають два методи вирішення таких питань: метод найменших квадратів (МНК, англ. OLS – Ordinary Least Squares) і географічно зважена регресія (ГЗР, англ. GWR – Geographically Weighted Regression). МНК, найбільш поширений метод створення регресії, є відправною точкою для всіх методів просторового аналізу. Він дозволяє створити загальну модель події (ранні смерті/сильний дощ та ін), і створює одне загальне рівняння регресії для моделювання досліджуваної події. ГЗР – один з видів просторового аналізу, останнім часом все частіше використовується для вирішення аналітичних задач в географії та інших дисциплінах. Він створює локальну модель змінної або процесу, яку ви хочете передбачити, створюючи рівняння для кожного з параметрів. При коректному використанні ці методи надають надійний і потужний статистичний апарат для дослідження або оцінки лінійних взаємозв'язків.

Лінійний взаємозв'язок може бути прямим або зворотнім. Якщо при збільшенні температури повітря збільшується і кількість нещасних випадків, то це позитивне взаємовідношення і пряма кореляція. Інший спосіб описати цей же прямий взаємозв'язок – сказати, що кількість нещасних випадків зменшується при зниженні температури повітря.

Кореляційний аналіз та графіки взаємозв'язку явищ показують, наскільки сильно два явища залежать один від одного. Регресійний аналіз, в свою чергу, дозволяє отримати ще більше інформації про взаємозв'язок явищ. Цей аналіз дозволяє показати вплив, з яким одна або кілька змінних можуть потенційно викликати позитивну або негативну зміну іншої змінної.

Створення регресійної моделі – це послідовний процес, який включає пошук ефективних незалежних змінних для пояснення тих процесів, які ви намагаєтесь змодельовати або зрозуміти. Це управління інструментом регресії для визначення змінних, найбільш ефективних для пророкувань явищ, і, потім, видалення і додавання змінних з метою створення найкращої моделі для побудови припущення.

Для впровадження будь-якої ГІС необхідні дані, отримання яких часто є непростою задачею і вимагає значних зусиль. Оскільки джерелами даних для подібного моделювання в ідеальному випадку є значні масиви статистичних даних спостережень за станом певних об'єктів дослідження. При цьому, чим більша статистична вибірка береться до уваги і чим більше факторів впливу на залежні змінні буде виявлено, тим точнішою буде модель оцінки їх динаміки.

Переконались в цьому стала нагода з самого початку роботи. Державними статистичними органами для широкого загалу в мережі Інтернет публікуються статистичні дані за різними соціально-економічними критеріями. Але виявилось, що для даного дослідження таких даних недостатньо. Знаходження необхідних даних виявилось все ж можливим, але було ускладнене різноманітними процедурами звернень до офіційних інстанцій (санітарно-епідеміологічні служби, центру моніторингу захворюваностей та ін.) та супроводжувалась непродуктивною втратою часу.

В той же час, якщо звернути увагу на досвід інших розвинених держав, можна спостерігати, що дане питання може вирішуватись значно легше і швидше. Рішенням даної та багатьох інших проблем, було б створення в країні загальнодержавних геопорталів та сховищ даних, де користувач міг би знайти всі необхідні дані для аналізу. У якості прикладу можна зауважити ініціативний проект Африканського банку розвитку – 'Information Highway', яка суттєво розширить доступ до офіційної статистики 54-х країн і 16-ти міжнародних організацій Африки. Зміст проекту передбачає, що на основі даних, які акумулюються на аналітичних інтернет-порталах, всі зацікавлені особи і організації зможуть приймати ефективні й зважені політичні і економічні рішення (завершення проекту намічене на кінець липня 2013 року).

Традиційно дані для досліджень у медицині отримують з таких джерел:

1. Демографічна інформація. Є однією з ключових в медичних дослідженнях, оскільки містить інформацію про структуру та розподіл місць перебування людей, їх вікові характеристики людей і зайнятість.

2. Економічні та соціальні показники якості життя людей, розвитку соціальної інфраструктури.

3. Статистичні дані медичних закладів. Кількість спалахів і випадків захворюваностей за різними категоріями.

В результаті на етапі збору інформації було сформовано початкову вибірку даних. Захворюваність населення, як предмет дослідження, та набір факторів, що ймовірно впливають на кількість випадків захворюваності. На основі наявних статистичних даних в якості проміжку часу було обрано інтервал в 5 років. В період з 1995 по 2010 рік вдалося отримати дані по таким факторам/показникам:

- кількість міського, сільського, дитячого та дорослого населення;
- кількість випадків захворюваностей ГРВІ та грипом;
- економічно активне населення віком 15-70 років;
- кількість незайнятих трудовою діяльністю громадян;
- середньорічна кількість поїздок в міському транспорті однієї особи;
- забезпеченість населення лікарями в розрахунку на одну особу;
- кількість викидів забруднюючих речовин в атмосферне повітря;
- рівень зареєстрованого безробіття (відношення кількості безробітних, зареєстрованих у державній службі зайнятості до середньорічної кількості населення працездатного віку, у відсотках);
- середньомісячний грошовий дохід населення;
- кількість міського населення;
- міграційний приріст;
- рівень смертності.

Перед використанням алгоритмів Data Mining необхідно провести підготовку набору даних для подальшого аналізу. Так як інтелектуальний аналіз даних може виявити лише присутні в даних закономірності, вихідні дані з однієї сторони повинні мати достатній обсяг, щоб ці закономірності в них були присутні, а з іншого – бути достатньо компактними, щоб аналіз займав прийнятний період часу. Найчастіше в якості джерел вихідних даних виступають сховища або вітрини даних. Підготовка необхідна для аналізу багатовимірних даних до кластеризації або інтелектуального аналізу даних.

Наступним етапом є очищення даних. Очищення видаляє шуми і вибірки з пропущеними даними. Ряд алгоритмів вміють обробляти пропущені дані, що мають прогностичну силу. Скажімо, при використанні методу асоціативних правил, обробляються не вектори ознак, а набори змінної розмірності. Вибір цільової функції буде залежати від того, що є метою аналізу; вибір «правильної» функції має основоположне значення для успішного інтелектуального аналізу даних. На даному етапі, було використано аналіз факторів за методом МНК для виявлення найкращої їх комбінації та відсіювання менш значущих факторів. Також було використано один із найефективніших методів аналізу інформації за допомогою ГІС, що дає змогу оцінити загальну картину і стан того, чи іншого явища – подання її у вигляді різноманітних карт, картодіаграм та графіків (рис.1).



Рис.1. Карта нев'язок МНК, що сприяє виявленню просторової автокореляції.

Esri ArcGIS ArcMap забезпечує доступ до великого числа схем класифікації та кольорів, які можна використовувати для виділення різних аспектів даних. Крім того, компанія Esri у 10-й версії ArcGIS запровадила підхід до обробки та візуалізації даних ГІС у часовому вимірі, що дозволяє застосовувати ці нововведення для моделювання просторово-розподілених динамічних систем.

В результаті інструменти регресійного аналізу виявили кореляційну залежність між первинними факторами та кількістю випадків захворюваності грипом і ГРВІ.

Статистично значимими факторами виявились:

- кількість міського населення,
- середньомісячний грошовий дохід населення,
- рівень зареєстрованого безробіття (відношення кількості безробітних, зареєстрованих у державній службі зайнятості, до середньорічної кількості населення працездатного віку, у відсотках),
- кількість викидів забруднюючих речовин в атмосферне повітря,
- забезпеченість населення лікарями в розрахунку на одну особу,
- середньорічна кількість поїздок в міському транспорті однієї особи.

Створення регресійної моделі являє собою ітераційний процес, спрямований на пошук ефективних незалежних змінних, щоб пояснити залежні змінні. Покрокове видалення та/або додавання змінних відбувається до тих пір, поки не буде знайдено найбільш прийнятну регресійну модель. Оскільки процес створення моделі часто має характер окремого дослідження, він має враховувати теоретичні аспекти, думку експертів у цій галузі і здоровий глузд (common sense).

МНК є відправною точкою для всіх видів просторового регресійного аналізу. Він забезпечує побудову глобальної моделі змінної або процесу і створює його рівняння регресії. Запускаючи комбінації по 6 факторів було виявлено, які з них є найбільш значущими.

Аналіз результатів МНК засвідчив, що розподіл величин не є нормальним, а також, що між обраними факторами присутнє явище мультиколінеарності (начебто незалежні змінні фактично між собою корелюють). Подібний розподіл негативно вплинув би на якість результату, тому дані необхідно було наблизити до нормального розподілу за допомогою логарифмічного трансформування (рис.2). В той же час, очевидно, що для кінцевої моделі необхідно обрати набір нескорельованих між собою факторів.

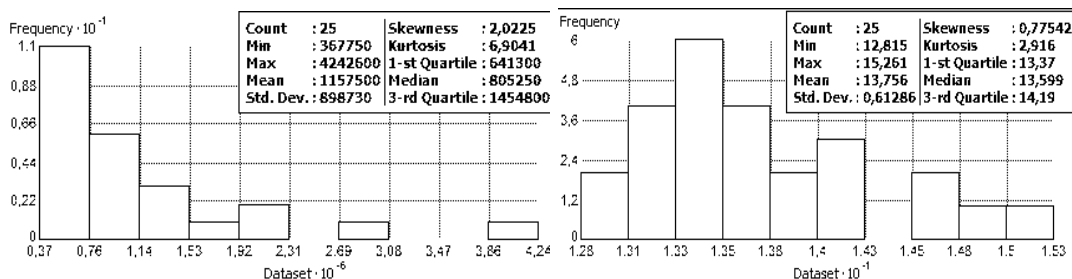


Рис.2. Аналіз розподілу даних міського населення за допомогою інструменту Гістограма до трансформації (зліва) та після трансформації (справа).

На наступному етапі, після дослідження того, які фактори мають бути трансформовані, та їх трансформування, було застосовано інструмент Дослідницької/Експлораторної регресії. Змінюючи вихідні параметри для проходження моделей, в результаті було визначено оптимальні вихідні параметри, при підстановці яких отримано 3 моделі. Із запропонованих обрано ту, яка мала найкращі показники. В результаті отримали модель, що з достовірністю 90% описує 50% випадків захворюваності на 1995 рік. Точність моделювання на пряму залежить від якості та достовірності вихідних даних. Отримана модель має бути перевірена на іншій статистичній вибірці, в даному випадку на даних по захворюваності за інші роки, щоб перевірити чи діють її залежності не лише для цього набору даних, а дійсно описують досліджуване явище. Якщо дані взаємозв'язки будуть підтверджені – з'явиться можливість прогнозу на майбутнє.

ВИСНОВКИ

Для отримання коректного результату моделювання необхідно виконати ряд умов, серед яких ключовими є створення селективної вибірки даних та підбір оптимальної моделі. Щодо першої умови, то, нажаль, з огляду на закритість або недоступність джерел інформації надзвичайно важко отримати всі необхідні дані та перевірити ймовірні фактори впливу на сумісність. Щодо другої умови – інструмент Дослідницької/Експлораторної регресії (Exploratory Regression) є цінним засобом аналізу даних, який дозволяє знайти правильну модель процесу. За умови вибору потенційно незалежних змінних дослідницької регресії, обґрунтованих теорією, думками експертів і здоровим глуздом, коректного трансформування даних цілком можливо виявити взаємозв'язки між змінними та підібрати правильну модель, що описує досліджуване явище, що є розподілене у просторі і часі.

За результатами проведеного дослідження можна зробити наступні висновки:

- пошук даних та створення нормальної статистичної вибірки є, мабуть найважливішим і в той же час найскладнішим етапом;
- за вибіркою, що вдалось скласти виявлені залежності різної сили впливу;
- більша частина незалежних змінних в тій чи іншій мірі корелює із залежною, в основному простежуються кореляції лінійного характеру;
- за результатами роботи було знайдено моделі, що описують досліджуване явище.

Визначення точних значень величин впливу незалежних змінних на залежну змінну, щодо знаходження вірної регресійної моделі і можливості складання епідеміологічного передбачення на майбутнє, на даний час залишається відкритим питанням, оскільки потребує більше часу і, можливо, додаткових даних для розширення статистичної вибірки. Крім того, в подальшому можливе застосування більш складних методів нелінійної регресії для більш точного визначення залежностей між факторами впливу та вихідною змінною.

Список літератури

1. Гохман В.В. ГИС в здравоохранении и медицине / Гохман В.В. [Электронный ресурс] – http://dataplus.ru/news/arcreview/detail.php?ID= 8329&SECTION_ID=265.
2. В'юн В.І., Кузьменко Г.С., Міхненко Ю.А., Архітектурні засади систем моніторингу та прогнозування // Математичні машини і системи, 2011, № 3, с.40-46.
3. Пшеничний О.Ю. Аналіз сучасних програмних засобів моделювання поширення вірусних захворювань / О.Ю. Пшеничний, І.М. Чорней, Н.Б. Шаховська, В.В. Литвин // Інформаційні системи та мережі : [збірник наукових праць] / – Львів.: Видавництво Національного університету "Львівська політехніка", 2010. – 396 с.: іл. – (Вісник / Національний університет "Львівська політехніка"; № 673). – С. 154-162.
4. Data Mining – интеллектуальный анализ данных / Информационные технологии. [Электронный ресурс] – <http://www.inftech.webservis.ru/it/database/datamining/ar2.html>.
5. Data mining - [Электронный ресурс] – http://ru.wikipedia.org/wiki/Data_mining.
6. Data Mining: What is Data Mining? [Электронный ресурс] – <http://www.anderson.ucla.edu/faculty/jason.frans/teacher/technologies/palace/datamining.htm>.
7. Big data. [Электронный ресурс] – http://en.wikipedia.org/wiki/Big_data.
8. Справка ArcGIS 10.1. Картографирование данных. [Электронный ресурс] – <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p00000002000000>.
9. Світличний О.О., Плотницький С.В. Основи геоінформатики. Візуалізація інформації в ГІС. Методи і технології візуалізації інформації в ГІС. [Электронный ресурс] – http://geoknigi.com/book_view.php?id=611.
10. Справка ArcGIS 10.1. Основы регрессионного анализа. [Электронный ресурс] – <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p0000000230000000/>.
11. Справка ArcGIS 10.1. Интерпретация результатов инструмента Исследовательская регрессия (Exploratory Regression). [Электронный ресурс] – <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p0000000520000000/>.

Ковгар В.Б. Применение геоинформационной технологии ESRI для моделирования эпидемиологических процессов / В.Б. Ковгар, Р.С. Филозоф, М.А. Савченко // Ученые записки Таврического национального университета имени В.И. Вернадского. Серия: География. – 2013. – Т. 26 (65). – № 1– С. 86-95.

В статье описано применение геоинформационной технологии Esri для моделирования эпидемиологических заболеваний. Предложено использование регрессионного анализа для определения факторов, потенциально влияющих на развитие заболеваемости, и определение степени их влияния.

Ключевые слова: моделирование, статистические данные, эпидемиологические исследования, регрессионный анализ, ГИС.

MODELLING OF EPIDEMIOLOGICAL PROCESSES USING GIS- TECHNOLOGY ESRI

Kovgar V. B.¹, Filozof R. S.¹, Savchenko M. O.²

¹*ПрАТ «ECOMM Co», Kiev, Ukraine*

²*Kiev National University of construction and architecture, Kiev, Ukraine*

E-mail: info@ecomm.kiev.ua¹, savchenko.mariya.a@gmail.com²

The article describes GIS-application for modelling of epidemiological processes based on Esri-technology. The purpose of the research was to identify factors that affect Ukrainian people acute respiratory viral infections and flu morbidity by ArcGIS (Esri, USA) tools, including Exploratory Regression, and to determine the magnitude of their impact. The object of research is located on territory of Ukraine. As taxonomic units there are 24 administrative regions and the Autonomous Republic of Crimea (excluding Kyiv City and Sevastopol). Creation of a regression model is an iterative process, aimed at finding effective independent variables, to explain the target dependency. The process goes on until an appropriate high quality model be found. Ordinary Least Squares (OLS) is the starting point for spatial regression analysis. It provides a global model of the phenomenon/process and creates a single regression equation to represent that phenomenon/process. Analysis of OLS showed that the distribution of values is not normal and between selected factors present phenomenon of multicollinearity (independent variables are correlated each other). Such distribution would have negatively affected the quality of the results, and the data had to be lead to the normal distribution by using logarithmic transformation. To achieve good results for the final model had to be selected a set of factors without a correlation each other. After researching what factors need to be transformed, and their transformation tool Exploratory Regression tool was applied. By varying the initial parameters for passing models, optimal output parameters were found and 3 models were received. Of the received models was selected one with the best options. The model should be tested in other statistical sample, on the statistic data for other years to see if its dependencies are correct not just for prime data set, and really describe the phenomenon. For getting correct results of modelling, a number of conditions must be performed, among them are the creation of selective data retrieval and the selection of the optimal model. According to the first condition, unfortunately, due to the closure or unavailability of information sources, to obtain all the necessary data and to check the likely impacts on compatibility is extremely difficult. As for the second condition – tool Exploratory Regression is a valuable tool data analysis that allows finding the correct model of the phenomenon/process. Determining accurate values of variables impact of independent variables on the dependent variable, finding the correct regression model and the possibility of drawing epidemiological predictions for the future, requires a lot of time and more quality statistic data. In the further possible to use more sophisticated methods like nonlinear regression to a more precise definition of dependence between factors of influence and output variables.

Keywords: modelling, statistics, epidemiological research, regression analysis, GIS.

References

1. Hochman V.V., GIS for public health and medicine Retrieved from http://dataplus.ru/news/arcreview/detail.php?ID=8329&SECTION_ID=265.
2. Vyun V.I., Kuz'menko G.E., Mihnenko Y.A., Architectural Principles of monitoring and forecasting. Mathematical Machines and Systems, No. 3, p.40-46 (2011).
3. Pshenchnuy O.U., Chorney I.N., Shakhovska N.B., V.V. Litvin, Analysis of modern software tools for modelling the spread of viral diseases .Information Systems and Networks. Lvivska politechnika. p.396
4. Data Mining . Information Technologies. Retrieved from <http://www.inftech.webservis.ru/it/database/datamining/ar2.html>.
5. Data_mining. Retrieved from http://ru.wikipedia.org/wiki/Data_mining.
6. Data Mining: What is Data Mining? Retrieved from – <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>.
7. Big data. Retrieved from http://en.wikipedia.org/wiki/Big_data.
8. ArcGIS Help 10.1. Map the data. Retrieved from <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p00000002000000>.
9. Svitlychnyi A.A., Plotnitskuy S.V., Basics of Geoinformatics. Information Visualization in GIS. Information visualization methods and technologies in GIS. Retrieved from http://geoknigi.com/book_view.php?id=611.
10. ArcGIS Help 10.1. Regression analysis basics. Retrieved from <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p000000023000000/>.
11. ArcGIS Help 10.1. Interpreting Exploratory Regression results (Exploratory Regression). Retrieved from <http://resources.arcgis.com/ru/help/main/10.1/index.html#na/005p000000052000000/>.

Поступила в редакцию 25.04.2013 г.