

На правах рукописи



Струков Денис Раймондович

**ИНФОРМАЦИОННАЯ СИСТЕМА ПРОСТРАНСТВЕННОГО
МОНИТОРИНГА СОСТОЯНИЯ ЗДОРОВЬЯ НАСЕЛЕНИЯ РЕГИОНА**

Специальность:

05.11.17 - Приборы, системы и изделия медицинского назначения

Автореферат диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург -2016

Работа выполнена в Федеральном государственном автономном образовательном учреждении высшего образования «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)» на кафедре Биотехнических систем.

Научный руководитель – доктор технических наук, профессор Горохов Владимир Леонидович

Официальные оппоненты:

1. Гельман Виктор Яковлевич, доктор технических наук, Федеральное государственное бюджетное образовательное учреждение высшего образования «Северо-Западный государственный медицинский университет им. И.И. Мечникова» Министерства здравоохранения Российской Федерации, профессор кафедры медицинской информатики и физики;
2. Шапиро Клара Ильинична, доктор медицинских наук, Государственное бюджетное образовательное учреждение высшего профессионального образования «Первый Санкт-Петербургский государственный медицинский университет имени академика И.П. Павлова» Министерства здравоохранения Российской Федерации, профессор кафедры общественного здоровья и здравоохранения

Ведущая организация - Федеральное государственное автономное образовательное учреждение высшего профессионального образования «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (г. Санкт-Петербург)

Защита состоится “26” октября 2016 г. в 16-00 ч. на заседании совета по защите докторских и кандидатских диссертаций Д212.238.10 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В. И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, 5.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского государственного электротехнического университета “ЛЭТИ” им. В. И. Ульянова (Ленина) и на сайте www.eltech.ru

Автореферат разослан в _____ июля 2016 г.

Ученый секретарь
совета по защите докторских
и кандидатских диссертаций
Д212.238.10



Садыкова Е. В.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. В распространенных сегодня системах экологического, санитарно-гигиенического, медико-экологического мониторинга существуют проблемы с совершенствованием методов сбора, обработки и анализа различных данных для устойчивого выявления статистически достоверных закономерностей распространения заболеваний населения в регионе.

Проблема заключается в том, что традиционные способы статистического пространственного прогноза основываются на нормальном гауссовском распределении, но этот подход может приводить к совершенно неверным результатам из-за появления данных подчиняющихся ненормальным распределениям. В данной работе разрабатываются алгоритмические и программные решения позволяющие выделять статистически надежные зоны пространственного распределения показателей здоровья населения (в т.ч. во времени), так же находить статистически устойчивые причинно-следственные связи между медико-демографическими показателями отдельных групп населения и факторами окружающей среды в условиях негауссовских процессов. Существенной особенностью развиваемых систем мониторинга является усиление устойчивости (робастности) средств выявления зон пространственной связи заболеваемости и загрязненности и средств визуализации результатов мониторинга для ЛПР в различных сферах городского управления: здравоохранении, санитарно-эпидемиологической службе, охраны окружающей среды (ОС) и других.

Целью диссертационной работы является повышение эффективности работы геоинформационных систем медицинского назначения за счет применения устойчивых алгоритмов выявления статистической зависимости между пространственно-временным распределением болезней и факторами. Исходя из поставленной цели необходимо решить следующие **задачи**:

- разработка кортежа устойчивых оценок вероятностей для системы показателей и алгоритмов статистически устойчивых оценок действия факторов окружающей среды региона на здоровье населения, проживающего на этой локации;
- разработка ранговых алгоритмов нормализации данных для повышения статистической надежности геостатистической интерполяции для получения пространственно-временного распределения болезней и влияния на них факторов окружающей среды.
- выявление статистически устойчивых зон загрязнения окружающей среды в пределах Санкт-Петербурга;
- получение экспериментальных результатов по выявлению географических локаций с наличием причинно-следственных связей между показателями здоровья населения и окружающей среды при помощи разработанных методов и алгоритмов;
- создание информационной системы пространственного мониторинга состояния здоровья населения города (ПМСЗН) с «дружественными интерфейсами» для специалистов, работающих с медицинской статистикой и экологической информацией и лиц, принимающих управленческие решения (ЛПР) в здравоохранении, санитарно-эпидемиологических службах.

Объектом исследования является информационная система пространственного мониторинга состояния здоровья населения региона.

Предметом исследования является информационное, методическое, программно-алгоритмическое обеспечение системы

Методы исследования.

Для решения поставленных задач использовались методы математического моделирования, геоинформационного анализа, математической статистики, геостатистики. Для анализа использовалась стандартная эпидемиологическая модель «Человек - окружающая городская среда», выбор показателей осуществлялся на основе баз данных медицинской статистики, социально-гигиенического и экологического мониторинга.

Новые научные результаты.

1. Разработана структура процесса мониторинга здоровья населения «Алгоритм ПАЗФ», учитывающая специфику медицинских данных диспансеризации детей, данных мониторинга атмосферного воздуха, а также - географическую привязку этих данных к территории города.

2. Разработаны алгоритмы оценки условных вероятностей для многомерных таблиц сопряженности, которые позволяют вычислять коэффициенты здоровья детей, коэффициенты отклонения и тяжести последствий, коэффициенты зависимости в различных городских ареалах проживания и ареалов загрязнения. Эти алгоритмы в составе модели ПАЗФ оценивают коэффициенты «силы эффекта», объединяющие характеристики здоровья, отклонения и тяжести и его зависимости от индекса загрязнения атмосферы по простейшим ареалам - административным районам Санкт-Петербурга.

3. Предложены алгоритмы пространственной вероятностной интерполяции (робастного кригинга и кокригинга) на основе рангового и квантильного нормирования данных, которые обеспечили моделирование для некоторых типов заболеваемости (МКБ-10, ПРР), и показателей распределения индекса загрязнения атмосферы (ИЗА) химическими веществами, а также – способствовали изучению взаимосвязей между этими двумя типами показателей.

3. Получены статистические устойчивые пространственные ГИС модели, привязанные к географической модели города:

- распределение устойчивых (робастных) зон загрязнения окружающей среды в пределах Санкт-Петербурга;
- распределения групп заболеваний у населения, зависимых от мутагенных факторов загрязнения окружающей среды, присутствующих в атмосфере мегаполиса;
- распределения групп заболеваний у населения, зависимых от канцерогенных веществ, присутствовавших в атмосферном воздухе мегаполиса,
- распределения групп заболеваний у населения, не зависимых от канцерогенных веществ, присутствовавших в атмосферном воздухе мегаполиса.

4. Разработана система пространственного мониторинга состояния здоровья населения (ПМСЗН) Санкт-Петербурга на базе программного обеспечения ГИС и модернизированных методов анализа, визуализации и прогноза показателей состояния здоровья и окружающей среды в городском пространстве. Разработанная система обеспечивает оперативную передачу наглядных данных и количественно подтвержденных гипотез через механизмы веб-приложений для оперативного принятия решений в сфере организации здравоохранения и защиты окружающей среды. Данная система имеет подсистему интерактивного картографического веб-приложения, как современного ин-

формационно-телекоммуникационного средства отображения результатов медико-экологических геоданных для городских медицинских информационных служб.

Положения, выносимые на защиту:

Построение устойчивых к неоднородности данных (робастных) систем пространственного мониторинга состояния здоровья населения региона достигается тем что:

- разработанные непараметрические композиции оценок условных вероятностей обеспечивали при мониторинге медицинских фактов статистически устойчивое выявление пространственно временных зависимостей характеристик здоровья, отклонения, тяжести последствий и экологических факторов (ПАЗФ) от действия экологических факторов на население, проживающего на определенной территории города или региона;

- разработанные алгоритмы пространственной вероятностной интерполяции на основе рангового и квантильного нормирования обусловили возможность развития геостатистического метода пространственно-временного анализа ряда характеристик заболеваемости и показателей загрязнений окружающей среды для объективного выявления статистических зависимостей между этими характеристиками. Таким образом, удается обеспечить: 1) построение пространственных прогнозов (цифровые пространственные модели распределения); 2) экспериментально проверить точность пространственных моделей, 3) объективно выявить зоны устойчивого действия факторов, 4) объективно выявить пространственные зависимости между показателями отклика у населения (например определенными видами заболеваемости чувствительных групп) и химическими факторами в окружающей городской среде (в определенной ее компоненте). Разработанные методические подходы позволили в системе пространственного мониторинга выявлять и отображать определенные локации (ареалы), где есть исследуемые зависимости, причем, удалость это сделать с количественно определенной точностью пространственного прогноза.

- разработанная с использованием описанных новаций архитектура системы пространственного мониторинга (ПМСЗН) и разработанная методическая поддержка представления статистически устойчивых аналитических геоданных в виде веб-интерфейсов для ЛПР, работающих в сферах организации здравоохранения и управления охраной окружающей среды.

Внедрение результатов работы

1. Методы отображения, анализа и прогноза при помощи ГИС внедрены в деятельности СПбМИАЦ: отдела геоинформационных технологий, информационно-аналитического отдела при решении научно-исследовательских и задач управления здравоохранением города Санкт-Петербурга, а именно: в разработке Классификаторов «Учреждения здравоохранения г. Санкт-Петербурга, разработанного в СПбМИАЦ в 2007 году; в формировании Концепции информатизации Здравоохранения г. Санкт-Петербурга в 2005 – 2015 гг., разработанной в СПбМИАЦ для Комитета по здравоохранению (КЗ); подготовке ежегодного доклада для Председателя Комитета по здравоохранению Правительства г. Санкт-Петербурга, а также – подготовка аналитических материалов для КЗ; при реализации ГИС здравоохранения на базе программного обес-

печения ArcGIS (ESRI), при создании структуры баз данных и баз геоданных в СПб-МИАЦ

2. Методы и элементы «Системы пространственного мониторинга состояния здоровья населения» были использованы в деятельности отдела социально-гигиенического мониторинга окружающей среды ТУ Роспотребнадзора города Санкт-Петербурга при разработке баз данных (БД) санитарно-эпидемиологического и экологического назначения и методов пространственного анализа для задач социально-гигиенического мониторинга Санкт-Петербурга и защиты окружающей среды.

3. Результаты экспериментов, методы, алгоритмы, инструменты программного обеспечения, а также отдельные подходы диссертационной работы внедрены в учебном процессе СПб ГЭТУ («ЛЭТИ»), в СПб ГУПС, ВЭШ, СПб ГЭУ.

Публикации: по теме диссертации опубликовано 57 научных работ из них главы в 3х монографиях, 13 статей в журналах (из них 3 статьи, опубликованные в рецензируемом журнале, определенном ВАК Минобрнаука РФ), 40 работ в материалах Российских и международных конференций, а также автор Геоинформационной системы «Геоинтеллект» (Свидетельство о государственной регистрации ПО ЭВМ№ 2015614104 от 06.04.2015 г.)

Структура и объем диссертации:

Диссертация состоит из введения из 4 глав, заключения и списка использованной литературы, включающего 104 наименования. Основная часть диссертации изложена на 217 страницах машинописного текста. Содержит 88 рисунков, 23 таблицы и 3 Приложения.

СОДЕРЖАНИЕ ДИССЕРТАЦИИ

Во введении обоснована актуальность темы диссертационной работы, определена цель и сформулированы задачи исследований, приведены основные результаты и научные положения, кратко раскрыто содержание диссертационной работы по главам.

Глава 1 посвящена анализу методов и проблем использования данных медицинской статистики и экологического мониторинга городской среды при изучении распределения показателей населения и факторов загрязнения окружающей среды.

Существуют методики, позволяющие сравнивать различные города по медико-экологическим признакам, делать выводы о тенденциях влияния факторов на человека, опираясь на макро показатели. Такие выводы, по результатам исследования, носят общий характер, поскольку сами модели взаимодействия между объектами в системе «окружающая среда-человек-город» поливариантны - их сложно описать в терминах показателей и их взаимосвязей. Таким образом, встает задача разработки статистически устойчивых (робастных) аналитических моделей взаимодействий «окружающая среда-человек-город». Еще одна трудность сравнения и медико-экологических признаков связана с некачественным сбором данных по небольшим территориям, что потребовало решения задач по устойчивой (робастной) обработке неоднородных данных.

Важным аспектом является учет и регистрация показателей загрязнения ОС, здоровья населения в конкретных локациях, а также влияния факторов на здоровье населения в этих локациях (селитебных, деловых, промышленных зонах). В соответствии с медико-эпидемиологическими принципами при изучении влияния конкретных факторов на население необходимо описать модель «Человек – окружающая среда»: факторы

воздействия, показатели откликов воздействия, каналы воздействия. Так, например приоритетным загрязнителем атмосферы города Санкт-Петербурга является автотранспорт (78%), в выхлопах которого содержатся органические вещества (бензол, формальдегид), а также – тяжелые металлы (кадмий, свинец и пр.) [Доклад Комитета по охране окружающей среды, природопользованию и экологической безопасности «Охрана окружающей среды, природопользование и обеспечение экологической безопасности в 2005 году/ Под ред. Д. А. Голубева, Н. Д. Сорокина - СПб 1998, 2001, 2003]. Здесь встает актуальная задача выявления факторов техносферы, влияющих на здоровье человека и разработка набора соответствующих показателей и алгоритмов оценки действия факторов окружающей среды.

Актуальной задачей является использование отклика на изменение окружающей среды и, как следствие, отклонением от нормального состояния каждого индивидуума человеческого общества, может являться заболеваемость. [Колосовская Е. Н., Мусийчук Ю. И., Красильников И. А., Мерабишвили В. М. и др.]. Для объективного решения этой задачи требуется предложить статистическое описание показателей, фиксируемых специальными медицинскими и социальными службами. К таким показателям относятся и временная утрата трудоспособности, смертность и многие другие. Однако для выявления вероятностных зависимостей между этими показателями и изменением окружающей среды требуется разработка специальных аналитических процедур и алгоритмов.

Кроме того актуальной задачей интеграции сбора и обработки геоданных по обширным территориям является построение цифровых пространственных моделей распределения показателей факторов и откликов.

Исследователи уже строили пространственные распределения (интерполяции) по результатам мониторинга в отдельных пунктах загрязнения атмосферного воздуха [Фридман К. Б., Мирский В. Е. и др.], однако эти распределения не учитывали статистические отклонения интерполяционных значений. Это часто могло приводить к ошибкам в оценках (прогнозах) медико-экологических рисков: они рассчитывались без предоставления распределения характеристик точности (средне-квадратичных ошибок) прогнозных значений интерполяции, что, в конечном счете, приводило к большим неопределенностям при формировании гипотез и решениям на основе них.

Отсюда актуальной задачей интеграции сбора и обработки геоданных по обширным территориям является построение цифровых пространственных моделей распределения показателей факторов и откликов, обеспечивающих статистическую устойчивость геостатистических инструментов пространственных картин отображений показателей здоровья населения. Требуется решение задач получения поверхности стандартных отклонений от прогнозных значений интерполяции, что позволит контролировать качество медико-экологических оценок и прогнозов.

Таким образом, в первой главе сформированы и обоснованы цели и задачи диссертационной работы по созданию системы пространственного мониторинга состояния здоровья населения.

Глава 2 посвящена разработке и модификации методов, алгоритмов и инструментов пространственного отображения и анализа медико-экологических данных для эффективной работы системы пространственного мониторинга.

1). Разрабатывается композиция алгоритмов вероятностных оценок пространственных устойчивых связей между факторами (определенными показателями загрязнения окружающей среды) и откликами (определенными показателями отклонений в состоянии здоровья населения) в определенных локациях. Прежде всего, для достоверного анализа связи между факторами (определенными показателями загрязнения окружающей среды) и откликами (определенными показателями отклонений в состоянии здоровья населения) предлагается многомерная композиция ряда вероятностных показателей $\{K_z, K_{от}, Power\}$ позволяющих количественно описать и выявлять важные медико-экологические причинно-следственные связи (алгоритм совместного анализа таблиц сопряженностей). В континуальном варианте речь идет о формировании эмпирических оценок непараметрических предметных (базовых) функционалов Ф.П. Тарасенко. Структура композиции разбивается на следующие позиции.

1-позиция. На первом этапе композиции на основе баз данных МИАЦ формируются выборки (многомерные таблицы сопряженности, таблицы частот) представляющие полные системы случайных событий $A_1 \dots A_K$ для характеристик здоровья населения (откликов) и условия выбора ареалов экологических факторов $\mathcal{E}_1 \dots \mathcal{E}_N$. Совокупности этих систем образуют множество событий, которые могут быть взаимно зависимыми или взаимно независимыми явлениями. Далее композиция (кортеж) состоит из следующих алгоритмов оценок частот.

2-позиция. Для выявления статистических связей предлагается процедуры вычисления условных и безусловных вероятностей появления этих событий, которые образуют композицию процедур статистических связей. Предлагаемые оценки частот является непараметрическими статистиками.

Конкретно для учета вышеприведенных факторов автором предложен и обоснован коэффициент «здоровья» $K_z = Z/N$, характеризующий долю здоровых Z на территории A_1, \dots, A_K , а так же коэффициент отклонения и тяжести $K_{от} = f(K_z)$ (характеризующий соотношение здоровых и больных на различных территориях A_1, \dots, A_K). Под «здоровыми» в результате анкетирования «Диспансеризации 2002» мы будем понимать тех детей, которые на момент обследования: не имеют диагноза, не имеют инвалидности, относятся к категории «I группа здоровья». Разработаны и реализованы программно процедуры их вычисления, а также выведен коэффициент «силы эффекта» $POWER$, обеспечивающий логическую увязку медико-экологических факторов и их объективацию. В соответствии с этим алгоритмом, для каждого из ареалов $\mathcal{E}_1 \dots \mathcal{E}_N$ (в нашем случае это административные районы Санкт-Петербурга) строятся таблицы сопряженностей (таблица 1):

Таблица 1. Характеристики выбранных ареалов.

Ареал	Численность (N)	Число больных (x)	Заболеваемость ($P(X)=x/N$)	Число инвалидов (y)	Коэффициент отклонения и тяжести ($K_{от}$)	Число здоровых (Z)	Коэффициент здоровья (K_z)
A1	409	43	0,1	4	0,11	66	0,16

A2	303	9	0,03	2	0,03	52	0,17
...
AK	Nk	Xk	P(X)k	Yk	(Kот)k	Зk	(Kз)k

3. Позиция. На основе этих N таблиц (по числу ареалов Э1,...,ЭN) и по каждой таблице оценивается N Функций зависимостей коэффициента отклонения и тяжести от коэффициента здоровья (Kот = f(KЗ)) для A1,...,AK.

4.Позиция. Оценивается частота: Power - коэффициент силы эффекта, характеризующий зависимость между отклонением, тяжестью и здоровьем популяции внутри ареалов и действия факторов Э1...ЭN, (Рис.2).

$$\text{Power} = (\text{отклонение} + \text{тяжесть}) / \text{здоровье}.$$

Предлагаемая оценка зависимости является оценкой частот событий и поэтому является непараметрической статистикой.

5. Позиция. Оценка зависимости вычисляется как тангенс угла наклона прямой Kот = f(KЗ). Чем больше фактор Э1, тем больше коэффициент силы эффекта.

$$\text{Power} = \text{tg}\alpha = \text{Kот} / \text{KЗ} = P(X \text{ не } Y + YX) / \sum N = (P(X) * (1 - P(Y)) + P(X) * P(Y/X)) / \sum N,$$

где вероятность $(P(X) * (1 - P(Y)) + P(X) * P(Y/X))$ вычисляется как оценка частот на основе сформированных выше таблиц сопряженности.

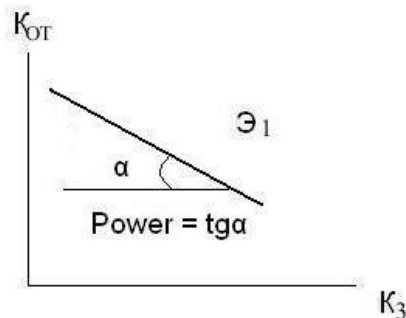


Рис 2.График зависимости Kот = f(KЗ)

Если зависимости Kот = f(KЗ) нет, то Power → 0, Это значит, что на той или иной территории Эi из Э1...ЭN экологический фактор (или их сумма) не вызывает явных откликов у популяции, проживающей на этой территории.

Разработанная композиция оценок представляет собой оригинальный алгоритм «увязки» оценок условных вероятностей (непараметрическая оценка функционала) на совокупности многомерных таблиц частот (таблиц сопряженностей). Как хорошо известно, метод частотных таблиц свободен от незнания функции распределения данных и поэтому предлагаемая композиция алгоритмов оценивания частот обладает принципиально устойчивостью (робастностью) к неоднородности данных (т.е. неизвестному и меняющемуся типу распределения данных), что и проявляется при ее практическом использовании и это отражено в последующем изложении.

2) Разрабатываются алгоритмы ранговой нормализации данных для пространственной вероятностной непараметрически устойчивой интерполяции (робастного кригинга и кокригинга) при нахождении статистически устойчивых локаций территорий, на которых характеристики загрязнения ОС постоянны по сезонам. Для пространственной

локализации статистических оценок медико-статистических характеристик обычно используются методы современной теории многомерного пространственного анализа (в геостатистике это кокригинг) основанные на нормальном распределении данных. Однако специфика медико-экологического мониторинга состоит (как уже упоминалось) в существенной статистической неоднородности данных. Для преодоления этой трудности в работе предлагается многомерные преобразования исходных данных в квантильные и ранговые последовательности. Ранг наблюдения x_i задается как его положение среди порядковых статистик: $R\{x_i\}=i$. Таким образом исходные данные заменяются ранговыми последовательностями (R_i) , где $i=1, \dots, n$. Здесь, реализуется непараметрический подход к выявлению и обнаружению локаций территорий, на которых характеристики загрязнения ОС статистически устойчиво коррелируют с медицинскими аномалиями. Он основан на ранговой модификации методов интерполяций и используется для формирования робастных цифровых пространственных моделей по набору значений ИЗА группы веществ, полученных в разные сезоны внутри одного календарного года. Данный подход приводит к следующим модификациям кригинга.

1. Модификация. При получении интерполяций, как результат, цифровых пространственных моделей значений «Индекса загрязнения атмосферы» рассчитывался по с учетом квантильных преобразований, по формуле $Z(s) = s/r^2$, где s – квантилизованное местоположение предсказываемой локации (как представление пары координат x, y). $Z(s)$ – предсказываемое значение, например ИЗА группы веществ, r – расстояние от истинного (замеренного) значения до искомой точки. В первый момент времени рассчитывается расстояние между искомой точкой и всеми окружающими ее точками. Далее рассчитывается медианное значение всех точек, окружающих искомую точку, с учетом расстояния до искомой точки. Переход к медианам и квантилям обеспечивает устойчивость оценок при неизвестном типе распределения данных. Этот метод в варианте вычисления средних довольно широко применялся для моделирования распространения загрязнений в компонентах окружающей среды одного какого-либо вещества или группы веществ и ранее, но не обнаружил робастных свойств при изменении типа распределения данных [Фридман К. Б., Мирский В. Е. и др.].
2. Модификация. Для случая ранжирования исходных наблюдений процедура универсального крикинга предлагается в виде : $V^*(R_i) = \sum_{i=1}^N w_i V(R_i)$, где w_i - веса приписываемые данным, R_i - преобразованные в ранги элементы исходной выборки.
3. Модификация. Помимо интерполяции и тренда, проводится анализ вариограмм с допущением, что чем ближе значения друг к другу (меньше лаг), тем более автокоррелированы между собой. После перехода от исходной выборки к рангам данных процедуры оценки вариограммы $\gamma(h)$ представляются в виде: $\gamma(h) = (1/2N(h)) \sum_{i=1}^N (V(R_i) - V(R_{i+h}))^2$, где N - число точек находящихся на расстоянии h . Подобные оценки обладают непараметрическими свойствами и следовательно являются и устойчивыми процедурами, что также подтверждается практическим использованием модифицируемых процедур. Таким образом использование подобных вариограмм, дает возможность пользователям подобрать к данным непараметрические

раметрически устойчивую (робастную) модель пространственной зависимости. Для расчета (прогноза) неизвестного значения переменной в заданном месте (интерполяции) процедура кригинга будет использовать подобранную модель вариограммы, конфигурацию пространственных данных и значения в точках измерений вокруг данного местоположения.

4. Модификация. Осуществляется свертка квантильных значений всех ячеек в растрах цифровой пространственной модели при помощи модуля Arc GIS Spatial Analyst. В результате – с учетом ранговых статистик получались устойчивые (робастные) зоны постоянного во времени нахождения повышенного ИЗА в тех или иных ареалах города. Метод удобен для нахождения робастных значений, здесь появилась возможность группировать факторы в зависимости от целей, например, в работе сгруппированы только те сезоны, в которых наличие фактора фильтрации зеленых насаждений отсутствует, чтобы вычислить робастные зоны с особой нагрузкой на окружающую городскую среду.
5. Модификация. Традиционный подход при ГИС анализе в медико-экологическом мониторинге не учитывал погрешность интерполяции (средне-квадратичные ошибки прогнозных значений), а соответственно, - погрешность суммарной результирующей цифровой модели, например загрязнения атмосферного воздуха. В данной работе осуществляется учет этих погрешностей и с использованием непараметрических оценок разброса.

Аналогичная модификация процедур кокригинга дала на практике устойчивое выявление зон заболеваемости. С учетом того, что кокригинг позволяет строить поверхности не только по одному набору данных, но и по нескольким наборам данных, модификация усиливает надежность и детальность результатов интерполяции и за счет использования дополнительной информации по взаимокоррелируемым данным.

Таким образом, использование преобразования данных в ранговые и квантильные последовательности обеспечивает модификацию средств кригинга и кокригинга, которые создают не только саму (непараметрически устойчивую, свободную от распределения) интерполированную поверхность, но также поверхности ошибок и вероятности интерполированных значений, зависящих от значений факторов, что позволяет пользователю статистически достоверно и надежно (робастно) оценивать точность получаемых результатов интерполяции. Автором этот метод проверялся экспериментально для задачи нахождения конкретных локаций, где имеются взаимные корреляции между значениями группы веществ в атмосфере и определенными откликами в виде показателей определённых групп заболеваний у населения. Причем, эти выявленные зависимости были подтверждены результатами, полученными в лабораторных условиях и опубликованы в медицинской литературе.

3. На основе предложенных устойчивых процедур разрабатывается структура **процесса мониторинга здоровья населения «Алгоритм ПАЗФ»** для системы мониторинга здоровья населения, учитывающая взаимосвязи характеристик здоровья (не заболеваемости), отклонений от него и тяжести последствий этих отклонений у населения, проживающего на территории города и факторов среды в определенных локациях. Модель нацелена на исследование факторов, характеристик здоровья, отклонений от него и тяжести последствий этих отклонений у населения, проживающего на территории города.

Анализ экологических факторов нацелен на выявление причин реакций популяций, проживающих на той или иной территории города. Созданная автором модель процесса мониторинга представлена на рис 1. Она задает этапы робастного анализа данных на основе разработанных автором методик и рассмотренных выше алгоритмов вычисления коэффициентов отклонения и тяжести, коэффициентов здоровья и функции $f(K_3)$, зависимости $f(\text{ИЗА})$. В целом предложенная модель процесса мониторинга впервые позволяет осуществлять сравнение откликов населения крупного мегаполиса (на примере Санкт-Петербурга) с реакциями популяций других городов по специальным коэффициентам, учитывающим здоровье.

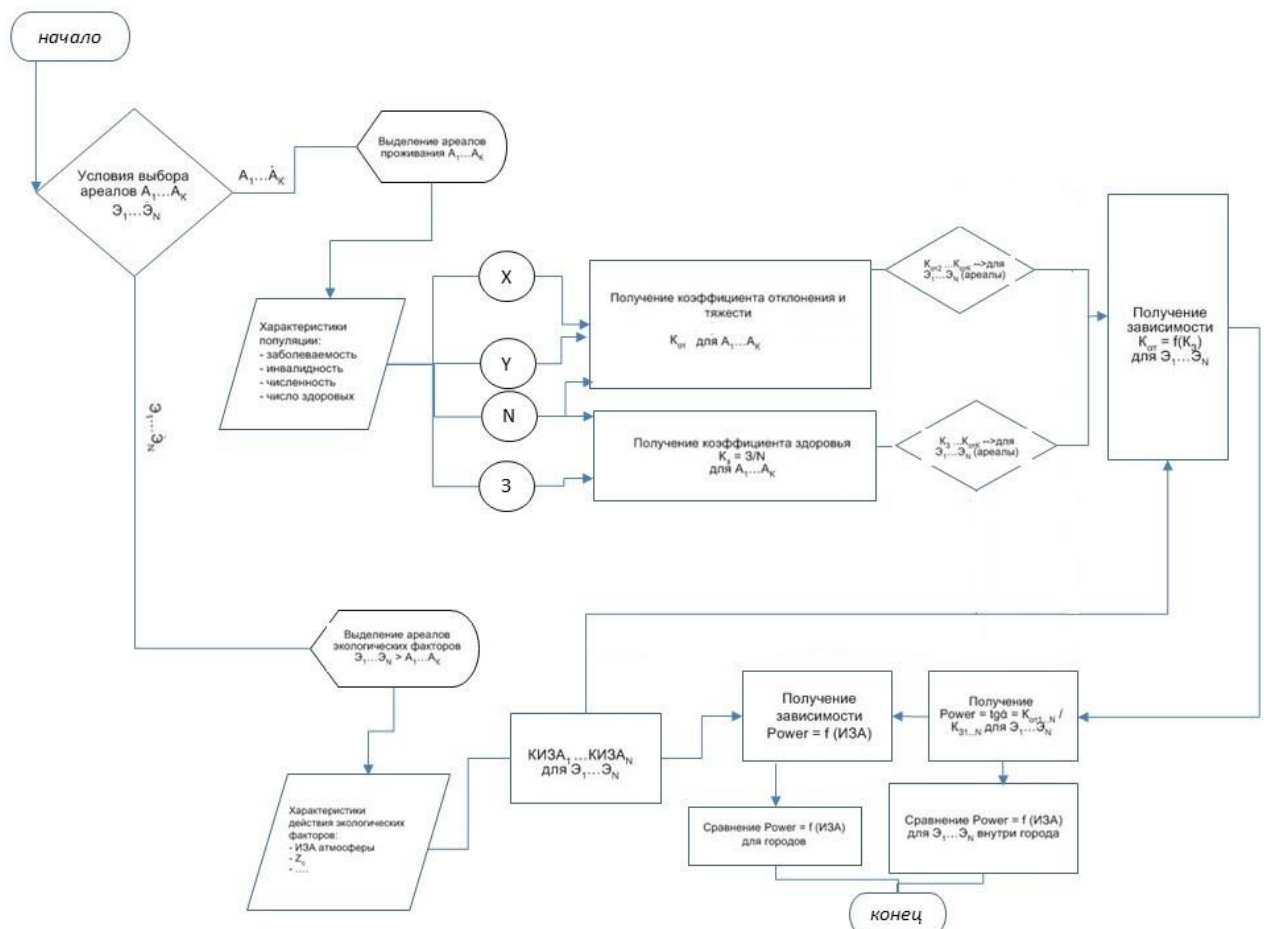


Рис. 1. Структура процесса мониторинга здоровья населения регионов «Алгоритм ПАЗФ»

Процесс состоит из следующих этапов.

1 этап. Проверка условий выбора ареалов.

2 этап. Выделение ареалов экологических факторов и ареалов проживания. Для этого в работе приведены и обоснованы условия и методика выбора ареалов A_1, \dots, A_K для характеристик здоровья населения (откликов) и условия выбора ареалов экологических факторов $\mathcal{E}_1, \dots, \mathcal{E}_N$. Для этого модернизированы: метод сетки, метод соседства с пунктами мониторинга и метод, учитывающий структуру города.

3 этап обеспечивает выбор характеристик популяций из баз медицинских данных и формирование полных систем событий и на их основе систем зависимых событий.

4. и 5 этапы реализуют вычисление коэффициентов отклонения и тяжести, коэффициентов здоровья и вычисление зависимости $f(Kз)$ с учетом экологических факторов КИЗА согласно выше разработанным автором процедурам. С точки зрения статистики здесь удается на строгой вероятностной основе вычислять меру статистической зависимости между многомерными факторами (медицинскими и экологическими) на многомерные таблицы частот. Это и позволяет статистически объективно вычислять $Pouer = f(ИЗА)$ и осуществлять их сравнение внутри города, для городов и регионов.

Таким образом, подход к разработке устойчивых к отклонениям данных от нормального распределения алгоритмов выявления связей между медико-экологическими характеристиками и их пространственной локализации позволил реализовать процесс мониторинга здоровья населения «Алгоритм ПАЗФ» для системы мониторинга здоровья населения. Мониторинг учитывает взаимосвязи характеристик здоровья (не заболеваемости), отклонений от него и тяжести последствий этих отклонений у населения, проживающего на территории города и факторов среды в определенных локациях. Это позволило в содружестве с медиками создать системы пространственного мониторинга состояния здоровья населения города Санкт-Петербурга.

Глава 3 посвящена созданию системы пространственного мониторинга состояния здоровья населения города Санкт-Петербурга. На базе модели «Человек – окружающая среда - город» сформирована структурная схема пространственного мониторинга состояния здоровья населения: архитектура и концептуальная модель системы пространственного мониторинга состояния здоровья населения (ПМСЗН).

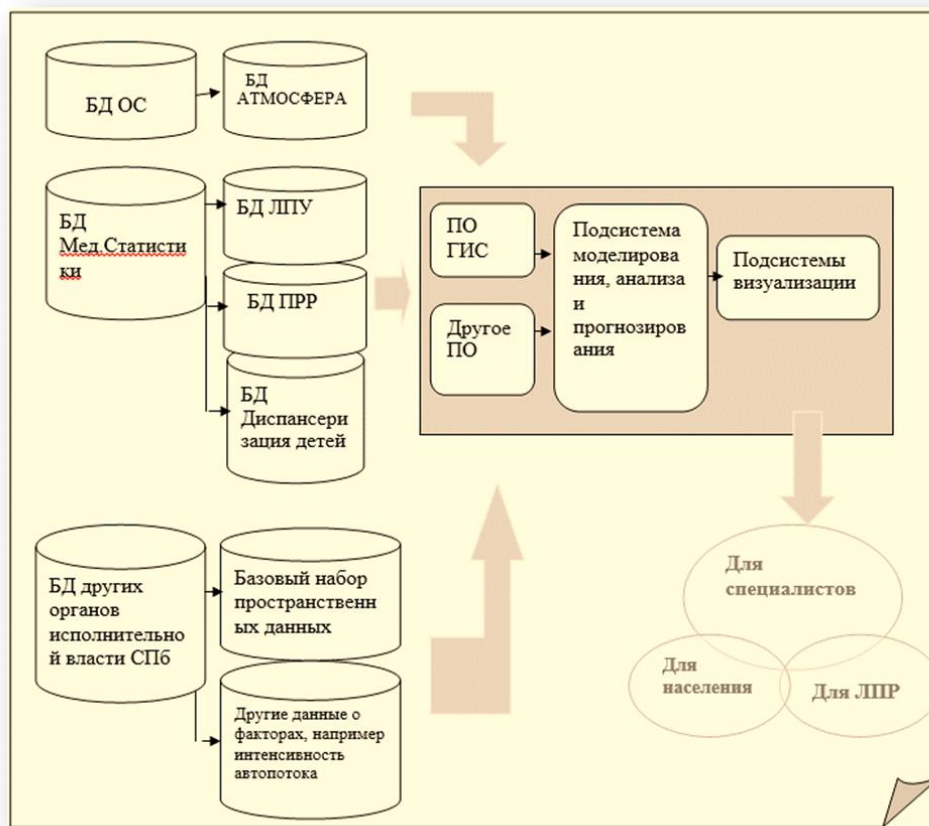


Рис. 3. Структурная схема ПМСЗН.

В главе раскрывается и обосновывается каждый блок структурной схемы, в т.ч. - информационное обеспечение ПМСЗН (базы данных медицинской статистики, регистры, проекты), программное обеспечение (ГИС, подсистемы принятия управленческих решений) и методическое обеспечение (классификация данных при визуализации геоданных, методы пространственного анализа в ГИС) и их взаимосвязи.

Помимо инструментариев и модулей, реализующих указанные в работе методы и алгоритмы обработки и анализа геоданных, в главе рассмотрены примеры технологической реализации ГИС здравоохранения, как части ПМСЗН, для ЛПР, для специалистов (врачей и медицинских статистиков) при помощи веб-ориентированных ГИС и сервисов (Web-based GIS- service или веб-приложений). В частности, для решения этих задач автором разработана ПО ЭВМ Геоинформационная система аналитика, зарегистрированная в Роспатенте соответствующим Свидетельством № 2015614104 от 06.04.2015 г. Приведен пример внедрения элемента ПМСЗН в органе управления здравоохранением Санкт-Петербурга.

Главе 4 посвящена экспериментальным результатам обработки данных здравоохранения с учетом загрязнения окружающей среды при помощи разработанных автором алгоритмов и методик пространственного анализа (ПМСЗН), модернизации методов геостатистического исследования. При помощи элементов системы пространственного мониторинга ПМСЗН получены распределения показателей здоровья населения и загрязнения атмосферы, устойчивые (робастные) оценки загрязнения в пределах административных делений Санкт-Петербурга, проведен анализ и выдвинуты гипотезы этиологий заболеваний в конкретных местах города, а также – получены первые результаты применения системы «алгоритм ПАЗФ» в трех районах Санкт-Петербурга.

1) Результаты применения системы ПМСЗН позволяют количественно оценить и сделать выводы о реакции детского населения на загрязнения по районам. Индикатором действия факторов в районах на выбранные отклики является коэффициент силы эффекта. Коэффициент силы эффекта (Power) равен тангенсу угла наклона прямой $Kot = f(K3)$. По графикам был посчитан коэффициент Power, данные сведены в таблицу сравнения «силы эффекта» воздействия факторов на популяцию между различными территориями. Для более подробного анализа реакций популяций в виде болезней органов дыхания (БОД) и врожденных деформаций и хромосомных нарушений у детей до 1 года жизни (ВАДХН), а также - тяжести последствий этих болезней, как откликов, на действие экологических факторов, удобно обратиться к коэффициенту силы эффекта (POWER). В работе подробно обсуждаются результаты по реакции детского населения, проживающих в районах Санкт-Петербурга.

2) Результаты применения разработанных робастных процедур для выявления устойчивых загрязнений на территориях Санкт-Петербурга

В диссертационной работе при помощи инструментов пространственного анализа, реализованных система пространственного мониторинга, была разработана методика нахождения «устойчивых» (робастных) зон распределений тех или иных факторов в конкретных экологических исследованиях. Изучая совокупность полученных цифр-пространственных моделей, можно сделать следующие практические выводы: 1. Очевидно, что существует некоторая зависимость между характером пространственного распределения облаков загрязненности с сезонной розой ветров. Формальдегид являет-

ся более легким веществом по сравнению с бензолом и фенолом и поэтому сильнее испытывает влияние сезонного изменения направления ветров. 2. Для всех рассматриваемых в эксперименте распределений поллютантов выявились 2 области, где загрязненность максимальна и постоянна – это район Измайловского и Вознесенского проспектов и район Невского проспекта. Это согласуется с данными отчетов ГИБДД о местах сосредоточения наибольшего количества автомобильных «пробок» в Санкт-Петербурге.

3) При помощи модернизированного автором геостатистического метода, с применением модуля **Arc GIS Geostatistical Analyst** автором получены пространственные модели распределения факторов в атмосфере и откликов (патологий) отдельно, а также - распределения зависимостей между факторами и откликами, а именно: «Цифровые модели распределения мутагенных веществ в атмосфере (тяжелых металлов)», «Цифровые модели распределения врожденных деформаций и хромосомных нарушений у детей до 1 года жизни», «Цифровые модели зависимости распределения ВДАХН от мутагенных веществ в атмосфере», «Цифровые модели поверхности распределения отдельных отклонений, возникающих у женщин в перинатальный период», «Цифровые модели зависимости поверхности распределения отдельных отклонений, возникающих у женщин в перинатальный период развития плода и мутагенных факторов в атмосфере», «Пространственный анализ распределения канцерогенных факторов в атмосфере», «Пространственный анализ заболеваемости лейкозами», «Пространственный анализ зависимости распределения заболеваемости лейкозами от канцерогенных факторов в атмосфере». В работе приведены результаты моделирования и выводы как с т.з. распространения той или иной групп веществ в атмосфере города, заболеваемости, а также подробно изучается влияние рассматриваемых факторов на отклики в виде заболеваний чувствительных групп населения. Благодаря геоинформационным технологиям и анализу причинно-следственных связей были получены определенные локации и изучены исходные причины тех или иных заболеваний. Например - Рис 4.

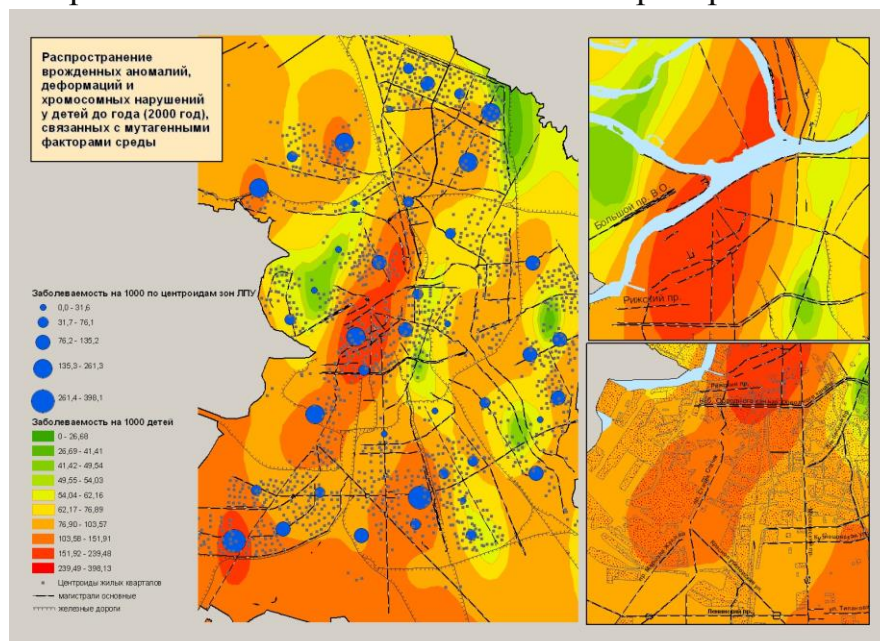


Рис.4. Единая географически привязанная цифровая модель поверхности распределения врожденных аномалий, деформаций и хромосомных нарушений у детей до года, зависимость от наличия мутагенов в атмосфере

Таким образом, ПМСЗН позволяет организовать систематизированный сбор, анализ и пространственный прогноз данных, необходимых для управления состоянием городской среды и здоровьем населения города, может быть интегрирована с системами экологического и санитарно-гигиенического мониторинга и решать задачи:

- прогнозирования состояния здоровья населения на перспективу;
- обоснования выбора ведущих (определяющих) факторов, влияющих на состояние здоровья населения;
- выявления очагов распространения эпидемий и прогнозировать их распространения, как на уровне страны, так и на уровне конкретного субъекта или крупного города;
- нахождения причинно-следственных связей между распространением или постоянным наличием того или иного показателя в отклонении населения и факторами, которые присутствуют на отдельных территориях;
- создания единой платформы для обмена данными между источниками информации (НИИ, МИАЦ, ЛПУ и пр.) в закрытом режиме для улучшения результатов выполнения НИР.

Помимо инструментариев и модулей, реализующих указанные в работе методы и алгоритмы обработки и анализа геоданных, в главе рассмотрены примеры технологической реализации ГИС здравоохранения, как части ПМСЗН, для ЛПР, для специалистов (врачей и медицинских статистиков) при помощи веб-ориентированных ГИС и сервисов (Web-based GIS-service или веб-приложений).

ОБЩИЕ ВЫВОДЫ:

1. Разработанные автором аналитические подсистемы и подсистемы отображения результатов исследования в ПМСЗН дают возможность специалистам оперативно получать справочную информацию по географической карте субъекта, строить гипотезы причинно-следственных связей, корректировать систему сбора данных при осуществлении мониторинга, интерпретировать результаты, а ЛПР – оперативно принимать решения и контролировать ключевые показатели здоровья человека в пространстве города, что значительно повышает эффективность любой системы мониторинга.

2. Автором модернизировано и разработано новое методическое обеспечение для анализа показателей здоровья населения в системе ПМСЗН:

2.1. Структура системы ПАЗФ демонстрирует объективность и полноту сбора и обработки данных диспансеризации детей, проводимых городским здравоохранением. Наряду с перманентно-фиксируемыми данными экологического мониторинга, детальные данные медицинских регистров и диспансеризации помогают с достаточной географической точностью принимать решения о неблагоприятных территориях города, о влияниях химических факторов окружающей среды на население, проживающего в конкретных местах города. Делать выводы не только по показателям отклонения от здоровья (заболеваемости), но и по показателям тяжести отклонения, коэффициенту здоровья, а также по показателям силы эффекта воздействия в определенном ареале.

2.2 Применение ГИС и существующих детерминированных робастных процедур позволяют оперативно искать устойчивые во времени зоны загрязнения, в т. ч. внутри года (по сезонам), которые сами по себе могут быть источниками риска здоровью населения.

2.3. Модернизированные геостатистические методы и методы отображения, дают возможность, при их соответствующей адаптации под данные медицинской статистики (включая регистры) и загрязнения воздуха, проводить пространственный прогноз распределения медико-экологических показателей, а также - строить сложные зависимости между факторами воздействия на человека в городской среде и откликами, выраженными характеристиками заболеваемости. Некоторые цифровые пространственные модели распределения средне-квадратичных ошибок интерполяционных значений индекса загрязнения атмосферы по пунктам мониторинга воздуха дали конкретные количественные и качественные инструкции по оптимизации сети пунктов сбора данных атмосферного мониторинга для ЛПР в системе защиты окружающей среды.

3. В работе автором получен ряд значимых для дальнейшего изучения и развития выводов по результатам обработки данных с применением методов и алгоритмов в т. ч. разработанных автором:

3.1. Нахождение устойчивых загрязнений атмосферного воздуха в разные сезоны года, и нахождения неблагоприятных районов проживания в Санкт-Петербурге по модели «Алгоритм ПАЗФ» построенной по данным детской диспансеризации.

3.2. Полученные распределения тяжелых металлов и ошибок интерполяций, цифровые модели распределения чувствительных групп населения и их зависимости от наличия в том или ином ареале мутагенных веществ в воздухе.

3.3. Нахождение причинно-следственных связей между канцерогенными веществами и лейкозами у детей, с учетом латентных периодов, и абсолютная независимость действия химических факторов атмосферы на появления рака молочной железы у женщин .

4. На основе экспериментальных результатов, полученных при помощи разработанных методов и алгоритмов анализа, прогноза и отображения показателей, а также - реализованной системы пространственного мониторинга, целесообразно принятие решений по созданию/перемещению промышленных предприятий города, созданию дополнительных медицинских учреждений в зонах с повышенной заболеваемостью, созданию зелёных зон в проблемных районах города, обеспечение профилактических мер и предупреждений по дислокации беременных женщин и детей до года и др. Разработанные и модернизированные в работу методы, алгоритмы должны лечь в основу системы пространственного мониторинга состояния здоровья населения, увеличивая ее эффективность с т.з. информационного, методического и аналитического обеспечения.

5. Разработанная и внедренная в СПб МИАЦ при непосредственном участии автора ГИС здравоохранения Санкт-Петербурга, как элемент системы ПМСЗН, уже способна демонстрировать населению показатели состояния здоровья населения, информировать о быстро-меняющейся санитарно-эпидемиологической ситуации на карте города благодаря веб-приложениям, доступным через Интернет, что также улучшает эффективность системы пространственного мониторинга и особенно актуально при оперативных решениях в современном мире. Она может быть легко интегрирована в системы для оперативного представления информации (ЧС, ГИБДД и многие другие), поскольку главной интегрирующим звеном является картографическая основа по конкретному городу или региону.

Основные публикации по теме диссертации

В изданиях, рекомендованных ВАК России

1. **Струков Д. Р.** Итоги 1-ой всероссийской конференции «Геоинформационные системы в здравоохранении РФ: данные, аналитика, решения» / Струков Д.Р., Красильников И.А. // Врач и Информационные технологии. 2012. №2. С. 25-29.
2. **Струков Д. Р.** Геопортал Геоинформационная система здравоохранения Санкт-Петербурга в сети Интернет, как пример системы для управления территорией здравоохранения. /Струков Д.Р., Декстер А.П.// Врач и информационные технологии. 2012. №3. С. 58-63
3. **Струков Д. Р.** Геоинформационные системы и многомерные статистические методы пространственного анализа для исследования заболеваемости / Струков Д.Р., Горохов В. Л. // Информационно-управляющие системы. 2009. №3 (40). С. 56-62

Главы в монографиях:

4. **Струков Д. Р.** Геоинформационные системы и методы пространственного анализа для исследования заболеваемости населения злокачественными новообразованиями. / Струков Д.Р., Красильников И.А., Мерабишвили В.М.// Глава IV к сборнику «Злокачественные новообразования в Северо-Западном федеральном округе России» под ред. Проф. В. М. Мерабишвили, СПб МИАЦ, ПРР. 2005. С. 314-317.
5. **Струков Д. Р.** «Геоинформационные системы», глава №8 в книге Мерабишвили В. М., Щербук Ю. А. М. Современное развитие информационных систем онкологической службы. – СПб.: ООО «ИПК «КОСТА». 2009. С. 244.
6. **Strukov D. R.** APPROACHES TO ASSESSMENT OF THE URBAN ENVIRONMENT AND ITS IMPACT ON THE POPULATION HEALTH (A CASE STUDY OF ST PETERSBURG)/ D.R.Strukov, A.N. Grebenyuk, V.M. Mtrabishvili, Y.I. Musiichuk, G.A.Popov// Ecology and hydrometeorology of big cities and industrial zone/ (Russia-Mexico).- Monograph. Vol. 3.Possible ways for solving of environmental problems of big cities and industrial zones. Environmental management.– St.-Petersburg^ RSHU. 2010. P. 26-50. Русский вариант главы в книге Экология и гидрометеорология больших городов и промышленных зон (Россия-Мексика). Том III. Пути решения проблем окружающей среды в больших городах и промышленных зонах. Управление окружающей средой./ Струков Д.Р., Гебенюк А.Н., Мерабишвили В.М., Мусийчук Ю.И., Попов Г.А.// Подходы к оценке окружающей среды урбанизированных территорий и влияние ее на здоровье населения (на примере Санкт-Петербурга). 2010. С. 29-54.
7. **Свидетельство о государственной регистрации ПО ЭВМ:** Струков Д.Р., Лабодин А.В., Эпштейн М.М., Мельник М.А.; № 2015614104 от 06.04.2015 г. «Геоинформационная система «Геоинтеллект»

Другие статьи и материалы конференций:

Всего 50 публикаций в материалах, трудах конференций на данную тематику, а также в популярных журналах по теме диссертации, в т.ч. выпущен Сборник «Труды 1-ой и 2-ой Всероссийской конференции «ГИС в здравоохранении РФ», изданной при участии автора как результат конференций, организованных автором. (Под ред. И.А. Красильникова, д.м.н., - СПб Изд БЕРЕСТА -2013.