

ИСПОЛЬЗОВАНИЕ ГИС ДЛЯ СИСТЕМНОГО АНАЛИЗА ЭКОБИОМЕДИЦИНСКИХ ДАННЫХ В КОНТЕКСТЕ ОЦЕНИВАНИЯ РИСКОВ. Ч. 2. ПОСТАНОВКА ЗАДАЧИ И РЕЗУЛЬТАТЫ КОНКРЕТНОГО ПРИКЛАДНОГО ИССЛЕДОВАНИЯ*

Формулируются исходные экобиомедицинские гипотезы, характеризуется состав исходных данных и необходимых измерений. Дается информация об инструментально технологическом комплексе математического и программного обеспечения и его использовании. Результаты количественного анализа иллюстрируются в виде географических карт и таблиц.

Введение

Данная статья является непосредственным продолжением [1] и составляет вторую часть общего исследования. В первой части давалась вводная информация о предметной области и содержании задач оценивания и управления рисками. В этой части на более детальном и формализованном уровне приводится постановка конкретной задачи из изучаемой предметной области и экспериментальных количественных результатах.

1. Постановка задачи конкретного прикладного исследования

Исходные экобиомедицинские гипотезы, состав данных и измерений. Известно, что полиароматический углеводород *бензопирен* (газоподобное соединение, которое легко распространяется в воздухе) является продуктом неполного сгорания органического топлива и негативно действует на иммунную систему человека, нарушая общие функции иммунной системы и приводя к возрастанию частоты и продолжительности инфекционных заболеваний и неадекватной реакции организма на вакцины. Так, в исследовательской базе данных IRIS — интегрированной системе информации о рисках, поддерживаемой США и Канадой, — *бензопирен* занимает шестое место среди тысячи наиболее опасных со-

единений. В качестве иллюстрации проследим некоторую цепочку событий, в которой *бензопирен*:

- входит в состав выбросов из некоторого промышленного источника загрязнений — источника эмиссий;
- переносится ветром и достигает географической зоны, в которой находится (проживает) изучаемая когорта детей;
- вносится в организм путем вдыхания, о чем свидетельствуют измерения содержания другого вещества, а именно *гидроксипирена* в моче (биомаркер);
- предположительно приводит к определенной частоте заболеваний *бронхитом* (эффект), о чем свидетельствует статистика заболеваний у детей.

Подлежит проверке токсикологическая гипотеза о том, что существует положительная взаимосвязь между дозой и эффектом, хотя не известна и заранее не установлена количественная зависимость между ними. Еще одно предположение состоит в том, что имеется принципиальная возможность регулирования интенсивности источника эмиссий (управление). Таким образом:

- если существует принципиальная возможность снижения эмиссий;
- имеются достоверные свидетельства того, что загрязнитель проник в организм и "сработал" в определен-

* Работа частично поддерживалась US CRDF и Правительством Украины, грант UN — 436, 1997 — 2000 гг.
© А.И. Кукса, 2004

ном биологическом и медицинском смысле;

• удастся оценить и измерить уменьшение величины ожидаемого эффекта, —

тогда можно говорить о принципиальной возможности управления риском в данной практической ситуации.

С учетом изложенного (гипотез, условий и теоретических предпосылок) задача конкретного прикладного исследования, используемая в данной статье лишь для иллюстрации и в контексте определенного метода принятия решений, выглядит следующим образом:

1) в качестве источников промышленного загрязнения воздуха в г. Мариуполе, Украина, рассматриваются два крупных металлургических комбината, а также некоторые другие предприятия, по состоянию на 1996 год;

2) в роли загрязнителя (поллютанта — в данном случае) рассматривается *бензопирен*, интенсивность выбросов которого считается известной, как и географическое расположение заводских труб на карте местности;

3) в качестве группы риска рассматривается когорта 3-летних детей, проживающих в городском районе, непосредственно примыкающем к промышленной зоне;

4) распространение поллютанта в воздухе моделируется с учетом метеусловий, характерных для данной местности и имевших место, по данным Гидрометстата, в изучаемый период времени;

5) в роли биомаркера экспозиции рассматривается концентрация *гидроксипирена* в моче — соединения, которое является продуктом распада *бензопирена*, вдыхаемого вместе с воздухом;

6) в качестве медицинского эффекта рассматривается заболевание *бронхитом*;

7) в общем известна или может быть оценена стоимость комплекса производственно-технологических мер, направленных на снижение уровня эмиссий;

8) дополнительно полагается, что негативные последствия заболеваний в исследуемой группе риска могут быть количественно оценены.

2. Использование средств математического моделирования

Использованное математическое и программное обеспечение включает в себя целую цепочку как готовых средств, так и специально разработанных [2–4].

Industrial Source Complex Dispersion Models (ISC3) — это пакет прикладных программ для математического моделирования процессов распространения веществ в воздухе, обусловленных промышленными источниками. В результате оцениваются (измеряются) концентрации веществ в воздухе — в точках географической области, непосредственно примыкающей и находящейся в зоне воздействия источника загрязнения. В свою очередь точки измерения определяются географическими координатами "объектов воздействия" загрязнителя, в данном случае — местожительством детей. Другими словами, в терминах предметной области, устойчивые концентрации загрязнителя в воздухе в местах непосредственного проживания детей рассматриваются как уровни экспозиции.

Для моделирования с помощью пакета ISC3 и оценивания таким образом состояния окружающей среды необходимо подготовить достаточно обширный входной файл, включающий:

- опции компиляции файла;
- характеристики промышленных источников (географические координаты труб и их физические характеристики — высоту и внутренний диаметр, интенсивности выбросов);
- географические координаты рецепторов, установленные на карте на основе адресов проживания детей;
- метеорологические данные (так называемый STAR-файл — в терминологии пакета ISC3);
- опции вывода результатов.

STAR-файл — это стандартный файл, указывающий гелео- и метеословия для воздушных потоков на дан-

ной территории и в данное время года ("роза ветров"). Его атрибутами являются шесть категорий стабильности (зависящей, в основном, от солнечной активности), направление и скорость ветра. STAR-файл подготовлен на основе данных измерений метеорологических станций 1, 2, 25, 31 г. Мариуполя за 1996 год.

Математические модели, реализованные в пакете ISC3, представляют собой известные стохастические дифференциальные уравнения Эйлера и описывают процессы переноса частиц в воздухе. В результате моделирования, после компиляции входного файла, пакетом ISC3 в автоматическом режиме создается выходной файл с концентрациями загрязнителя в воздухе в точках — рецепторах, заданных во входном файле. Выходные данные пакета ISC3 можно получить в двух форматах: для явно указанного множества точек (points) или для прямоугольной сетки координат на местности (grid). Каждый из выходных форматов конвертируется посредством специально разработанных программ — конверторов в форматы данных DBASE (dbf), которые уже затем используются как входные данные в географической информационной системе ArcView.

Собственно, пакет ISC3, работающий в автоматическом режиме, несмотря на всю сложность моделируемых им физических процессов, реализует одну функцию, а именно: отобра-

жает заданные на входе интенсивности промышленных источников эмиссий в концентрации того же загрязнителя в интересующих нас точках исследуемой географической области. Поэтому, используя ISC3 как функциональный оператор, можно рассматривать уровни эмиссий как переменную — действие (в терминологии упомянутых выше процедур принятия решений) и, варьируя ею, получать для каждого уровня эмиссий концентрации загрязнителя на всей географической области. Таким образом, посредством вариаций уровнями эмиссий проведены вычислительные эксперименты и получены базы данных для концентрации бензопирена в воздухе в точках проживания детей на карте местности при потенциальном снижении промышленных выбросов на 25%, 50%, 75%, 90% от начального уровня (100%). Результаты представлены в табл. 1.

3. Использование географических информационных систем

Географическая информационная система ArcView — это настольная система для работы с картами, а также инструмент для быстрого выбора и отображения различных типов данных и визуализированного представления информации. ArcView имеет доступ к данным в своем собственном формате *shapefile*, в форматах других географических информационных систем, а также во многих других форматах дан-

Таблица 1. Зависимость концентрации бензопирена в воздухе от процента сокращения выбросов относительно начального уровня (100%) в местах проживания детей

Actions	100	0	0	0	0	5	2	3	11	12	7	3	0	0	1	1
	75	0	0	1	5	1	1	7	17	11	0	0	1	1	0	0
	50	0	0	6	2	1	1	22	11	0	1	1	0	0	0	0
	25	0	8	2	5	11	6	11	2	0	0	0	0	0	0	0
	10	9	17	14	3	0	1	1	0	0	0	0	0	0	0	0
	0	0.0009					0.0055					0.011				
	Benzo(a)pyrene conc. in air (ng/m ³)															

ных — графических и табличных.

Дополняющее систему *ArcView* мощное специализированное информационно-аналитическое программное средство *ArcView Spatial Analyst* позволяет из наборов данных, имеющих географическую привязку, формировать сложные запросы, создавать тематические карты, осуществлять многомерный анализ растровых данных или интегрированный, растрово-векторный анализ графических объектов.

Исходный набор данных, импортируемый извне в систему *ArcView* и ее дополнение *ArcView Spatial Analyst*, включает в себя:

- базовую географическую карту местности достаточно большого масштаба;
- географические координаты на ней мест проживания изучаемой когорты детей;
- dbf — файлы концентрации загрязнителя в воздухе в местах проживания детей, для каждого потенциального действия по сокращению эмиссий (25%, 50%, 75%, 100%);
- dbf — файл концентрации *гидроксипирена* в моче;
- xls — файл, регистрирующий наличие или отсутствие заболевания *бронхитом*.

Задачи пространственного анализа (*spatial analysis*) экобиомедицинских данных с использованием *ArcView* и *ArcView Spatial Analyst* можно условно разделить на две группы:

- построение тематических карт местности, визуализирующих значения отдельного атрибута; для этого использовались соответствующие процедуры географической интерполяции или экстраполяции значений этого атрибута в соответствующих dbf — файлах;
- одновременная логическая обработка пар тем с последующей визуализацией результата и обнаружение статистических ассоциаций — взаимосвязей переменных; для этого использовались стандартные процедуры пакета *ArcView Spatial Analyst*, такие, как *Reclassify* и *Map Calculator*, для наход-

дения зон перекрытий соответствующих поверхностей.

Последовательность использования аналитических процедур применительно к указанным исходным наборам данных такова. Сначала осуществляется одномерный анализ, а именно для исходных объектов (концентраций *бензопирена* в воздухе, концентраций *гидроксипирена* в моче, наличия или отсутствия фактов заболевания *бронхитом*) строятся соответствующие им объекты-карты (рис. 1, 2, 3). Затем используются процедуры нахождения статистических ассоциаций-взаимосвязей — уже по отношению к парам объектов, а именно определяются ассоциации-взаимосвязи между концентрациями *бензопирена* в воздухе и *гидроксипирена* в моче (рис. 4, а также табл. 2, построенная на основе этой карты-рисунка 4), а также между концентрациями *гидроксипирена* в моче и наличием или отсутствием заболевания *бронхитом* (рис. 5, а также табл. 3, построенная на основе этой карты-рисунка 5).

Таким образом, конечный результат анализа многомерных экобиомедицинских данных (продукт) представлен в виде табл. 1, 2 и 3. В каждой ячейке указано количество случаев из общего числа рассмотренных, подтверждающих гипотезу взаимосвязи соответствующих пар переменных.

Заключение

В статье в общих чертах обоснована, описана и проиллюстрирована компьютерная ГИС-технология анализа многомерных экобиомедицинских данных. Конечный результат представлен в виде нескольких двумерных таблиц, отображающих связи пар переменных. По существу они являются таблицами сопряженности признаков в статистической выборке двумерных фактов. На этой основе могут строиться статистические гипотезы, оцениваться параметры распределений и приниматься статистические решения в отношении этих параметров ([5], [6]). Таким образом, мы имеем дело с так называемой KDDM-технологией (*Data Mining &*

Mariupol City Environmental Health GIS :
 Benzo(a)pyrene concentrations in air (microgram/m³)
 conditioned by emissions from three industrial sources in 1996

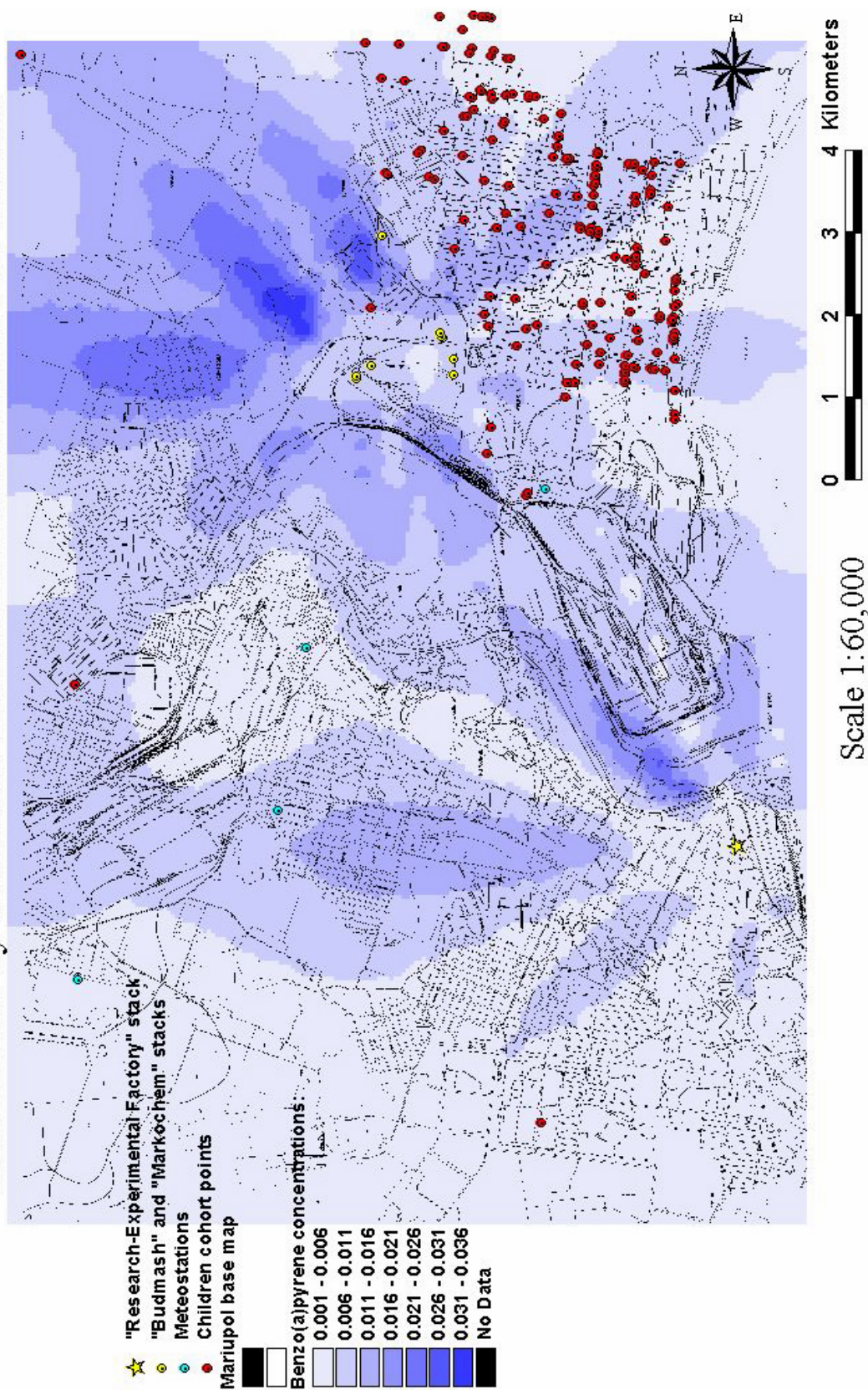


Рис. 1. Концентрации бензопирена в воздухе

Mariupol City Environmental Health GIS : 1-hydroxyperene concentrations in urine ($\mu\text{g/l}$ urine)

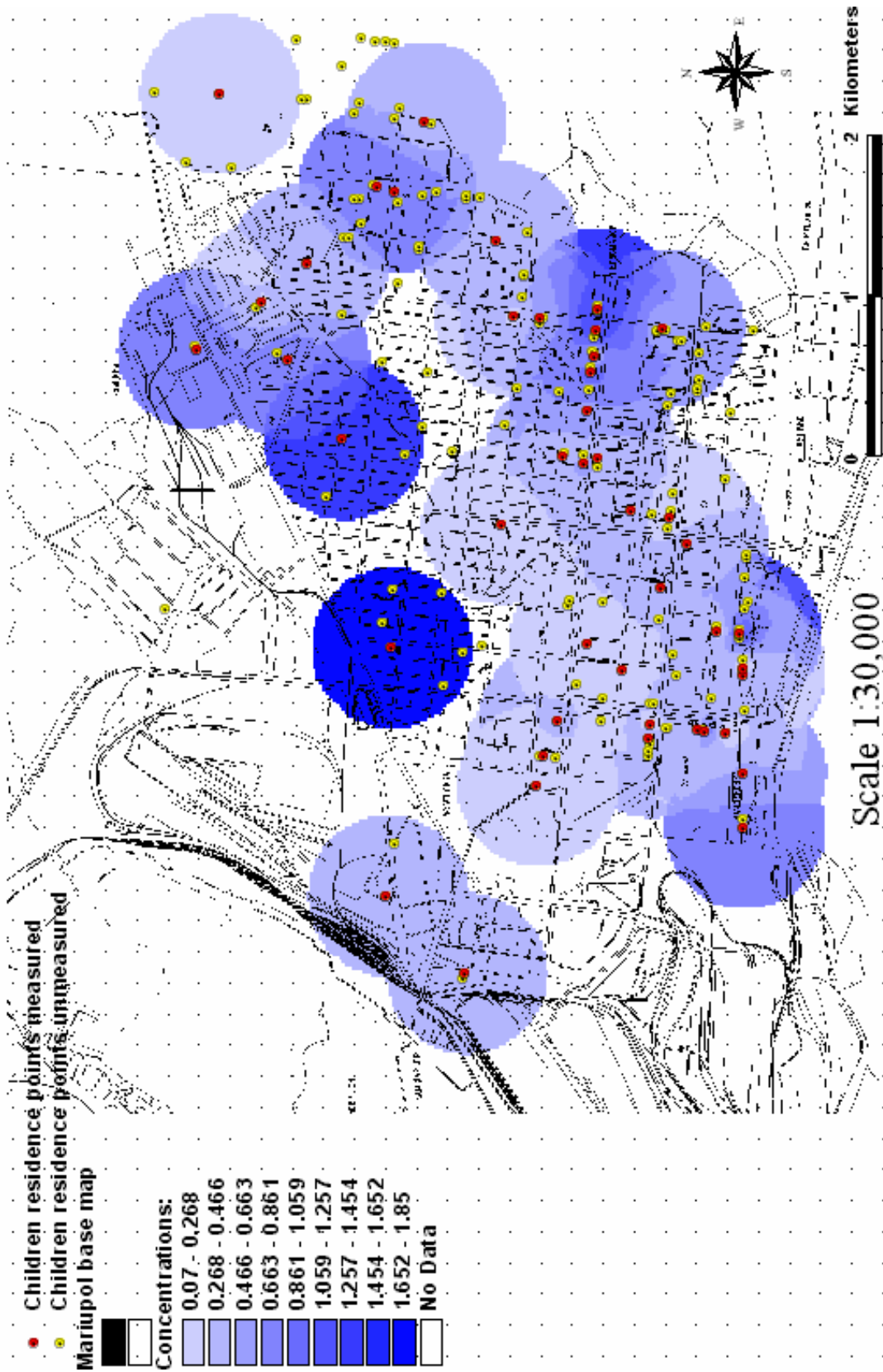


Рис. 2. Концентрации гидроксиперена в моче

Mariupol City Environmental Health GIS : Sickness rate of bronchitis among 3-year old children.

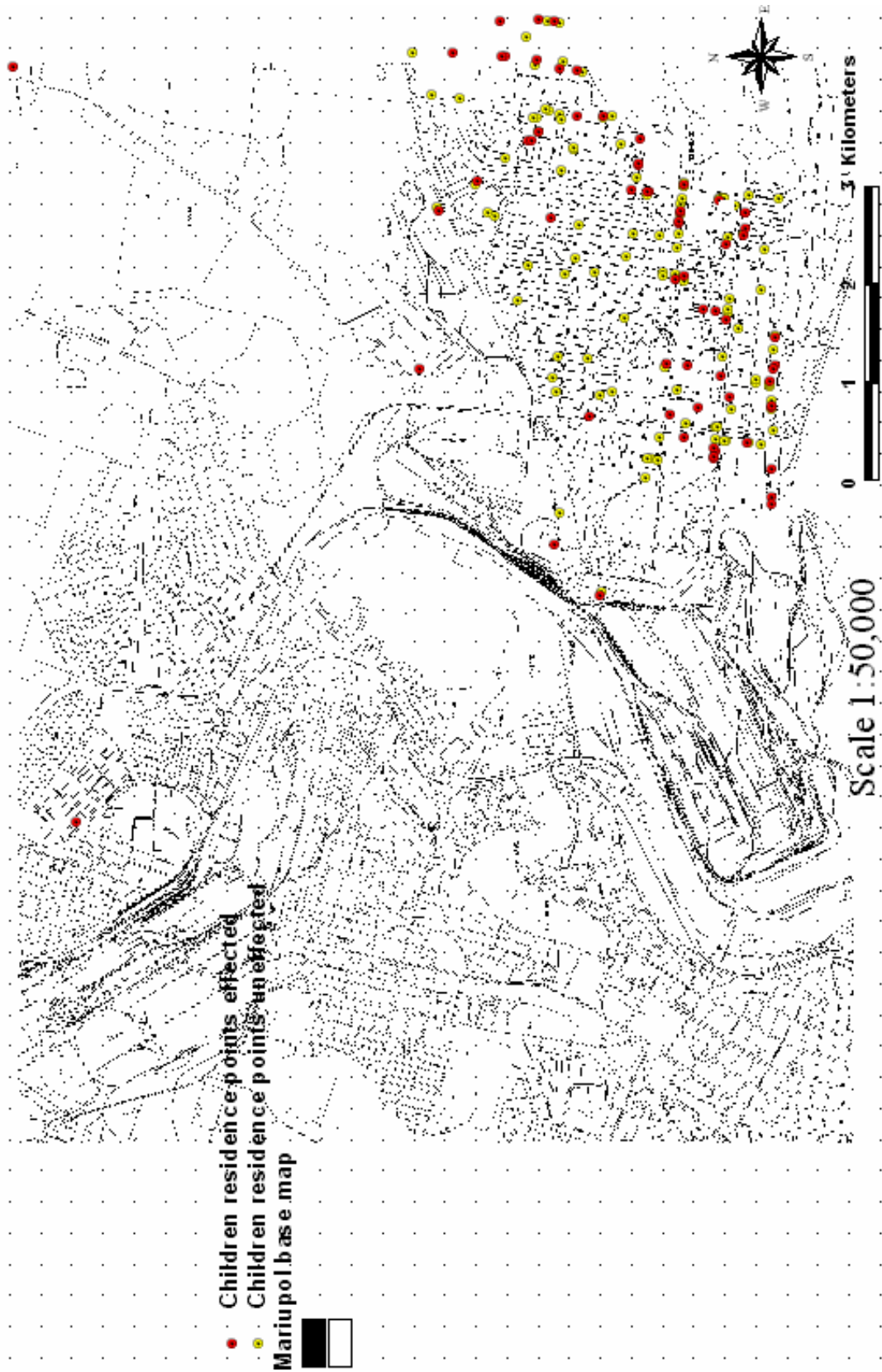


Рис. 3. Заболеваемость бронхитом

Mariupol City Environmental Health GIS:

Statistical associations between 1-hydroxypyrene concentrations in urine (mg/l) and benzo(a)pyrene concentrations in air (mg/m³)

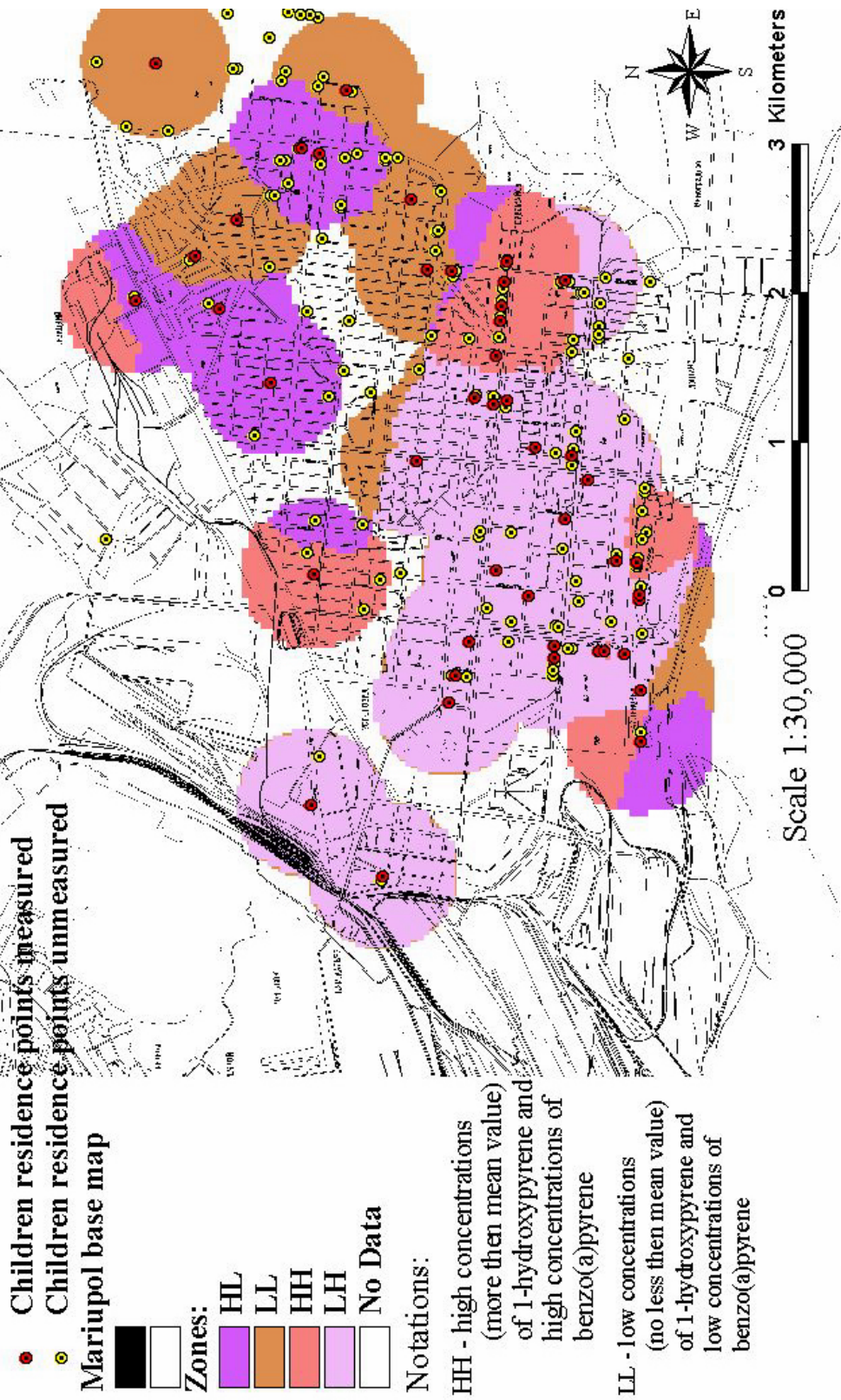


Рис. 4. Статистические ассоциации – взаимосвязи между концентрациями бензо(a)пирена в воздухе и гидроксипирена в моче

Mariupol City Environmental Health GIS :

Statistical associations between 1-hydroxypyrene concentrations in urine ($\mu\text{g/l}$) and sickness rate of bronchitis among 3-year old children

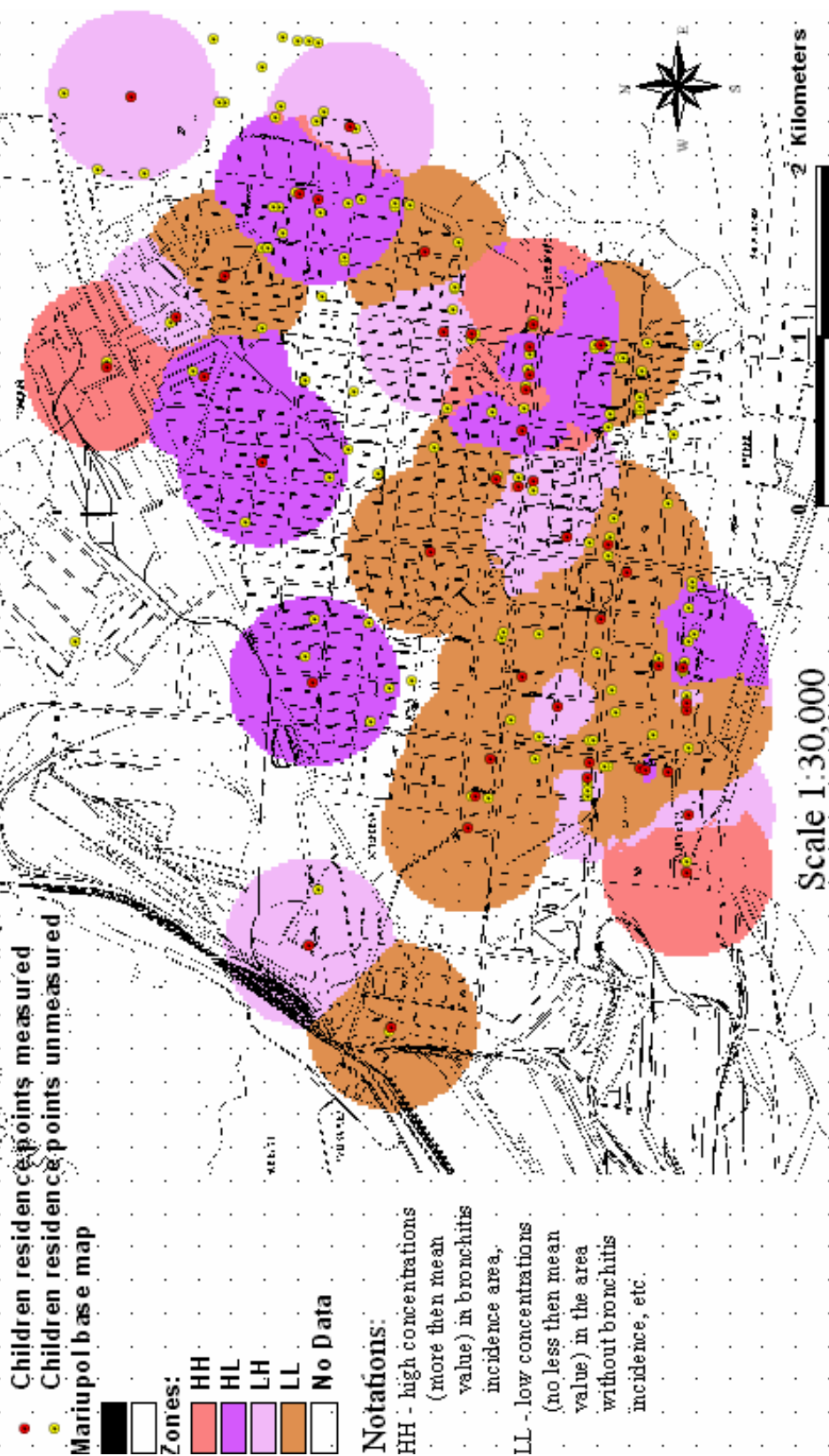


Рис. 4. Статистические ассоциации – взаимосвязи между концентрациями гидроксипиrena в моче и заболеваемостью бронхитом

Таблица 2. Зависимость концентрации *гидроксипирена* в моче от концентрации *бензопирена* в воздухе в местах проживания детей

($\mu\text{g/L}$ urine)

1-hydroxypyrene	2.0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	1.6	0	0	0	0	1	0	0	2	0	0	0	0	0	0
	1.2	0	0	0	0	3	0	0	3	2	0	0	0	0	0
	0.8	0	0	0	0	0	0	1	4	2	2	0	0	0	1
	0.4	0	0	0	0	1	2	2	1	8	5	3	0	0	1
		0	0.0009					0.0055					0.011		

Benzo(a)pyrene conc. in air (ng/m^3)

Knowledge Discovery) [7], применяемой в нашем случае к предметной области "окружающая среда – здоровье" (Environmental Health Domain).

Таблица 3. Зависимость количества случаев заболевания *бронхитом* в когорте 3-летних детей от концентрации *гидроксипирена* в моче

($\mu\text{g/L}$ urine)

1-hydroxypyrene	2.0	1	0
	1.6	2	1
	1.2	5	3
	0.8	6	4
	0.4	14	9
		0	1

bronchitis

Благодарности

Видение предметной области "окружающая среда – здоровье" и проблем, связанных с ее компьютерным моделированием, сложилось в результате общения со многими специалистами (которые, конечно, ни в коем случае не несут ответственности за возможные ошибки и огрубления). Прежде всего, это д-р D. Nryhorczuk

(Un-ty of Illinois at Chicago) и д-р мед. наук. Н.Г. Проданчук (Ин-т экогигиены и токсикологии, Киев), которые, собственно, изначально и инициировали работу по применению ГИС. Это также д-р Р. Scheff (Un-ty of Illinois at Chicago & US EPA, R.5, Chicago), который предоставил в распоряжение пакет ISC3 и консультировал в подготовке STAR-файла; д-р В.Р. Sonawane (US EPA, NCEA, Washington), приславший огромный пакет нормативно-методических материалов US EPA; наконец, проф. Ю.Н. Антипенко (Ин-т экогигиены и токсикологии, Киев) внес большой вклад в экобиомедицинский ликбез автора.

1. Кукса А.И. Использование ГИС для системного анализа экобиомедицинских данных в контексте оценивания рисков. Ч.1: Общая характеристика предметной области и возникающих проблем // Пробл. программирования. — 2003. — № 4. — С. 72–81.
2. *Industrial Source Complex (ICS3) dispersion models.* — Research Triangle Park, NC: US EPA, Office of Air Quality Planning and Standards Emissions, Monitoring, and Analysis Division. — 1995. — 300 p.
3. *ArcView GIS.* — Redlands, CA, USA.: Environmental Systems Research Institute, Inc. — 1996. — 350 p.
4. *ArcView Spatial Analyst.* — Redlands, CA, USA.: Environmental Systems Research Institute, Inc. — 1996. — 147 p.
5. Kuxa A. Robust decision making in the field of environmental health risk management (EHDM) // EURO XVII: 17th European Conference

on Operational Research. — Budapest: — July 16–19, 2000. — P. 108.

6. *Kuksa A.* Bayesian statistical decisions in environmental health risk management // Міжнар. конф. “Прогнозування та прийняття рішень в умовах невизначеності” (PDMU-2001). — Київ, 2001. — С. 47.
7. *Heckerman D.* Bayesian Networks for Data Mining // Data Mining and Knowledge Discovery. — 1997. — **1**, № 1. — P. 79–119.

Получено 04.07.03

Об авторе

Кукса Анатолий Иванович,
доктор физ.-мат. наук

Место работы автора

Институт программных систем НАН Украины,
просп. Академика Глушкова, 40,
Київ-187, 03680, Україна
Тел. (044) 266 3420
E-mail: kuksa@isofts.kiev.ua