

УДК 004:575:616.9

Мониторинг особо опасных инфекций (на примере проблемы чумы)

С.П. Соколова*, Е.А. Кузьмина*, В.З. Абдуллина**

*Санкт-Петербургский институт информатики и автоматизации, Российская академия наук, Санкт-Петербург, 199178, Россия

**Институт проблем информатики и управления МОН, Алматы, 480100, Республика Казахстан

Аннотация. Предложена интеллектуальная система мониторинга особо опасных инфекций (на примере проблемы чумы) с использованием интервального аналога подхода иммунокомпьютинга и интервальных первичных данных, характеризующих состояние чумной триады: носитель–переносчик–чумный микроб. Приведен пример прогнозирования эпизоотических процессов в ландшафтно-экологическом районе (равнина Акдала) Прибалхашского природного очага чумы.

Ключевые слова: чумная триада, эпизоотический процесс, подход иммунокомпьютинга, индекс риска чумы

ВВЕДЕНИЕ

Чума относится к особо опасным инфекциям. Природные очаги чумы существуют с древних времен и расположены в различных частях мира (Индия, Средняя Азия, Африка, Северная и Южная Америка). Пандемии чумы уносили огромное количество человеческих жизней. К примеру, вторая пандемия (под названием «Черная смерть») длилась с 14 до 17 века. Как отмечают историки, это было самое сильное биологическое бедствие в истории. За последние 150 лет вспышки чумы регулярно происходят в различных частях мира и приводят к значительным человеческим и экономическим потерям.

В Республике Казахстан природные очаги чумы покрывают площадь в 130 миллионов гектаров и за последние, более чем 50 лет они рассматриваются как наиболее активные очаги чумы в мире. Эти очаги характеризуются различными механизмами регулирования на популяционном и биоценоотическом уровнях. Чумный эпизоотический процесс является сложной многокомпонентной динамической системой. Поведение отдельных подсистем чумной эпизоотической триады (носитель–переносчик–чумный микроб); сложные взаимоотношения вышеупомянутых подсистем неоднократно исследовались микробиологами, биологами, эпидемиологами и т.д. [1-3]. Однако, несмотря на наличие большого экспериментального материала, характеризующего чумную триаду на популяционном, организменном и клеточном уровнях, до сих пор отсутствуют эффективные методики оценки состояния как отдельных подсистем, так и чумной триады в целом. Это справедливо для решения вопроса прогнозирования возникновения эпизоотических процессов в природных очагах чумы. Имеются работы [1–3 и др.], в которых представлены результаты изучения динамики взаимодействующих паразито-хозяйных отношений на популяционном уровне в чумной триаде.

В число функций системы мониторинга проблемы чумы входят следующие функции:

- проектирования реляционной базы данных с временной динамикой для природного очага чумы;
- организации и поддержки реляционной базы данных с временной динамикой по ретроспективным данным количественных и качественных параметров чумной триады (точечным и интервального типа);
- организация мониторинга эпизоотических процессов и возможной эпидемиологической ситуации на основе математических моделей и вычислительных процедур иммунокомпьютинга с адаптацией к вычислениям с интервальными данными.

В статье представлены результаты проектирования и поддержки реляционной базы данных с временной динамикой для природного очага чумы, а также результаты реализации вычислительных процедур иммунокомпьютинга на основе универсальной программной системы MATLAB и ее интервального расширения INTLAB. Применение математических моделей иммунокомпьютинга позволило решить задачи сжатия разнородных исходных первичных данных, классификации стадий эпизоотического процесса, кластеризации и оценивания состояния чумной триады.

Работа выполнялась в рамках двух проектов:

- проекта К-159-98 по линии Международного научно-технического центра на тему: «Мониторинг количественных и качественных параметров особо опасных инфекций в природных очагах чумы Республики Казахстан». – 1998–1999гг.;
- проекта INCO-COPERNICUS #ICA2-СТ-2000 на тему: «Чума в Центральной Азии – эпидемиологическое изучение на пространственно-временной динамике». – 2001–2005 гг.

Первичные данные по ландшафтно-экологическим районам (ЛЭР) — равнина Акдала за период с 1949 по 1998 гг. были предоставлены исследователями Казахского научного центра карантинных и зоонозных инфекций им. М. Айкимбаева (бывшего Казахского противочумного исследовательского института).

ОБЩАЯ ХАРАКТЕРИСТИКА ПРИРОДНОГО ОЧАГА ЧУМЫ

Как было отмечено выше, природные очаги чумы в Республике Казахстан занимают площадь около 130 миллионов га и являются крупнейшими в мире очагами чумы. На протяжении последних пятидесяти лет эти очаги являются и наиболее активными по чуме регионами мира. Природный очаг чумы является весьма сложной системой, в которой необходимо учитывать множество взаимозависимых и взаимодействующих факторов живой и неживой природы.

Эпизоотическая и эпидемическая активность природных очагов чумы трудно поддаются прогнозированию традиционно применяемыми методами и сравнительно мало изучены в системном плане. Это объясняется тем, что очаги имеют различные механизмы регуляции на популяционном, видовом и биоценотическом уровнях. Кроме того, недостаточно изучено поведение отдельных элементов чумной эпизоотической триады «возбудитель–переносчик–носитель» в многолетней динамике [1–3].

Ниже представлена совокупность значимых факторов воздействия и набора индикаторов процессов популяции в природном очаге чумы для возбудителя, переносчика и носителя [1–3].

Состояние возбудителя (чумного микроба) характеризуется следующими дифференциально-диагностическими признаками. К качественным признакам относятся: морфология колоний, чувствительность к бактериофагу, ферментация как глицерина, так и рамнозы, денитрификация/нитрификация, пестициногенность, чувствительность к пестициду, потребность в факторах роста. К количественным

признакам относятся: кальцийзависимость, содержание антигена Фракции 1 в реакции пассивной гемагглютинации, вирулентность для белых мышей и морских свинок.

Чаще всего состояние возбудителя может быть охарактеризовано его численностью, выраженной через такие косвенные показатели, как зараженность грызунов, блох или проб не только данного участка, но и соседних участков. Численность микроба косвенно может быть охарактеризована и долей зверьков, в крови у которых обнаружен чумной антиген.

Состояние переносчиков чумного микроба (блох) выражается их численностью, сезонной активностью нападения, половым и возрастным составом имагинальной фазы, сроками массового появления имаго на зверьках и во входах нор грызунов после зимовки, яйцекладки (начало, ее пик, прекращение), выноса имаго (начало, массовое количество, конец) и другими показателями. Кроме того, состояние переносчика характеризуется относительными показателями численности: индекс обилия, индекс встречаемости, индекс доминирования, индекс интенсивности.

Состояние носителей (грызунов) характеризуется совокупностью факторов воздействия, набором характеристик популяций на биоценотическом, популяционном и организменном уровнях. Биоценотический уровень зоологического мониторинга включает информацию по видовому составу носителей в очаге чумы, степени доминирования по численности и встречаемости видов на территории очага, соотношению грызунов, хищных, зайцеобразных, насекомоядных млекопитающих и птиц норвиков в очаге, а также по уровню межвидовых контактов в норах основного носителя.

Популяционный уровень зоологического мониторинга включает широкий круг вопросов по основному носителю, которым в очагах чумы песчаночьего типа является большая песчанка. Это популяционная структура ареала основного носителя в очаге, пространственная структура его поселений на территории очага, плотность семейных нор на 1 га, соотношение обитаемых и нежилых нор, численность грызунов на 1 нору и на 1 га, возрастная и половая структура популяции, продуктивность популяции, уровень смертности в популяции, оценка физиологического состояния популяции. Причем, динамика этих показателей должна отслеживаться по годам и сезонам (весна–осень).

Организменный уровень исследований предусматривает получение информации по определению возраста, пола и генеративного состояния носителей, оценке физиологического состояния особей разного возраста и пола, уровню индивидуальной инфекционной чувствительности в популяции основного носителя, подвижности и миграционной активности.

Показателями внешних факторов являются, в первую очередь, погодные условия. Существенными являются такие погодные характеристики, как температура воздуха и почвы по дням и месяцам, объем выпавших осадков по дням и декадам, число дней с осадками определенной градации, число дней с морозом и без оттепели, повторяемость и скорость ветров по румбам, повторяемость типов атмосферной циркуляции и др.

Одним из природных очагов чумы, расположенных на территории Казахстана, является Прибалхашский автономный очаг, включающий в свой состав 12 ландшафтно-экологических районов (ЛЭР) и являющийся очагом чумы песчаночьего типа [1, 2]. Каждый из ландшафтно-экологических районов имеет самостоятельную зону возникновения эпизоотий. В целом для этого очага эпизоотические циклы выражены достаточно четко. Конкретные же сроки возникновения и угасания эпизоотий в рамках одного цикла на разных участках очаговости могут отличаться. Начало и окончание эпизоотических процессов определяется изменением численности носителей, в первую очередь, больших песчанок и их блох. На протяжении всего эпизоотического цикла меняется вирулентность возбудителя.

РЕЛЯЦИОННАЯ МОДЕЛЬ ДАННЫХ С ВРЕМЕННОЙ ДИНАМИКОЙ

Одним из интенсивно разрабатываемых и исследуемых направлений в теории баз данных является создание математических основ баз данных, построенных на моделях представления данных, учитывающих фактор времени, так называемых темпоральных или временных моделях [4, 5]. Наличие временного фактора накладывает свой отпечаток как на проектирование, так и на функционирование информационной системы и базы данных. Для информационных систем с временной динамикой важно рассматривать предметную область реального мира с учетом временной динамики данных, корректно решать вопросы проектирования базы данных с временной динамикой и обрабатывать запросы к базе данных с параметрами времени.

Определение 1. Каждое отношение в реляционной модели данных с временной динамикой включает время как *обязательный* атрибут отношения.

Реляционная модель данных с временной динамикой формализована следующим образом [4, 5]:

$$M_{VD} = \langle G_s, G_c, O \rangle, \quad (1)$$

где G_s — правила построения структур данных, G_c — правила ограничений целостности, O — операции обработки данных.

Правила построения структур данных G_s включают два типа правил:

$$G_s = \langle G_s^A, G_s^R \rangle, \quad (2)$$

где G_s^A — правила формирования характеристик атрибутов данных, G_s^R — правила построения схем отношений данных.

Правила формирования характеристик атрибутов данных G_s^A формализуются следующим образом: каждый атрибут данных определяется своим именем и характеризуется описанием назначения, связанным с набором характеристик, значения которых определяют его временную динамику.

Правила ограничений целостности G_c включают следующие типы ограничений:

$$G_c = \langle G_c^A, G_c^R, G_c^F \rangle, \quad (3)$$

где G_c^A — ограничения целостности для атрибутов данных, G_c^R — структурные ограничения целостности для отношений данных, G_c^F — функциональные ограничения целостности для атрибутов данных.

Ограничения целостности G_c^A атрибутов данных включают тип значения и множество значений атрибута данных:

$$G_c^A = \langle TIP, D \rangle, \quad (4)$$

где TIP — множество используемых типов данных, а D — множество значений атрибута. Для каждого типа данных применяется свое множество операций.

Для представления динамики данных в РМДВД формализована временная логика в виде *ANU*-исчисления [4]. Аксиоматика этого исчисления основана на *AN*-исчислении фон Вригта и модернизированном метрическом *U*-исчислении Прайора.

Структура баз данных проектировалась на основе полного перечня всех данных, используемых в работе сотрудниками противочумной службы при оценке эпизоотического и эпидемиологического состояния отдельных ЛЭР и очага чумы в целом [4–6]. Технология создания базы данных включала следующие шаги:

- анализ требований пользователей, при котором были выявлены все актуальные и потенциальные потребности: качественные и количественные характеристики чумной эпизоотической триады «возбудитель–переносчик–носитель»; факторы, описывающие эпизоотический процесс и эпидемиологические характеристики природного очага чумы;

- построение концептуальной модели базы данных, эта модель является информационной моделью рассматриваемой предметной области (природный очаг чумы) и учитывает информационные потребности всех пользователей системы мониторинга;
- выбор инструментальных и технических средств системы мониторинга;
- построение логической модели базы данных, основанной на концептуальной модели базы данных, используя языковые средства выбранной СУБД;
- построение физической модели базы данных;
- проектирование пользовательских приложений и интерфейса для системы мониторинга;
- первоначальная загрузка баз данных;
- определение параметров временных рядов, способов их построения, источников данных и формы представления интегрированных данных;
- создание приложений, позволяющих сформировать временные ряды.

Анализ требований пользователей позволил определить состав данных, которые необходимо включить в базу: D_1 — данные по большой песчанке (основной носитель); D_2 — данные по журналу исследования полевого материала (носитель – переносчик); D_3 — данные по атласу (носитель – переносчик); D_4 — данные по чумному микробу (возбудитель); D_5 — метеорологические данные.

При проектировании форм был создан простой, удобный, легко осваиваемый интерфейс графического типа с меню, кнопками, окнами и подсказками, «дружественный» для пользователей системы мониторинга. Примером организации интерфейса могут служить: форма, позволяющая работать с базами данных и главным меню (рис. 1) и форма для ввода информации по журналу исследования полевого материала.

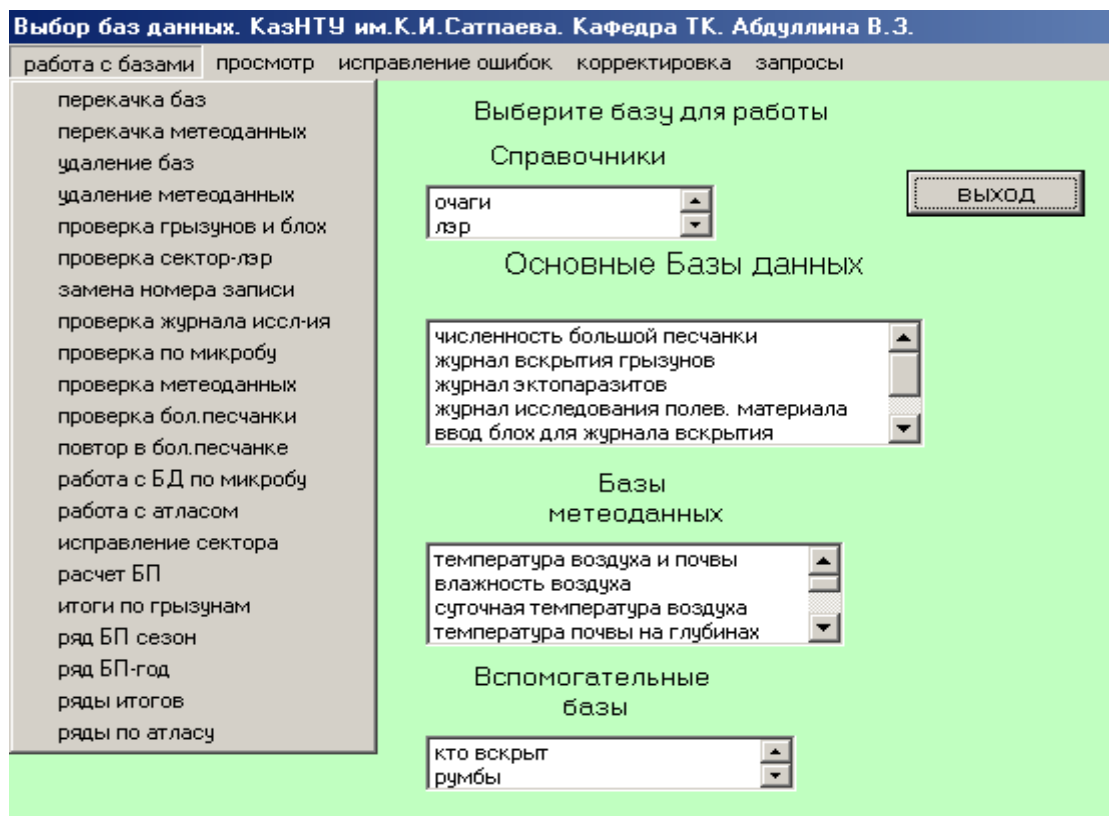


Рис. 1. Форма для работы с БД и меню.

Последняя форма является достаточно информативной, имеет довольно сложную структуру и изменяется в зависимости от того, какое действие выполняет пользователь в данный момент. Вид этой формы при вводе данных по носителям и переносчикам представлен на рис. 2.

Журнал исследования полевого материала

шифр: 10Г сектор: 3034306930 номер записи: 14548

ЛЭР: пески Бестас код ЛЭР: 30.3

год: 1994 день: 19 месяц: 10

Вид грызуна: краснохвостая песчанка код грызуна: 2

Ввод данных по блохам

вид блохи: X. gerbilli minax код блохи: 2

количество блох: 3

количество посевов: 1

результат бактер. исследования: -

Ввод данных по журналу вскрытия

номер грызуна: 1

вскрыто: взрослая самка

генеративное состояние: яловая

количество эмбрионов: 4 количество резорб. эмбрионов: 2

патологоанатомическое состояние: без видимых изменений

результат бактер. исследования: - результат биол. исследования: -

гельминты: РПГА: РНАГ:

По клещам: клещей: количество посевов: результат бактер. исследования:

Рис. 2. Форма для ввода данных по носителю и переносчику.

Просмотр журнала исследования полевого материала

Просмотр: год: 1998-1999

Журнал исследования (итоговые данные)

	nz	sectr	kodler	god	den	vc	vcitog	vcaitog	mcitog	nr
	12185	3034311744	30.2	1998	30	36	17	19	0	
	12186	3034311744	30.2	1998	7	59	31	28	0	
	12187	3034311744	30.2	1998	14	64	26	38	0	
	12188	3034311744	30.2	1998	22	58	28	28	0	
	12189	3034311744	30.2	1998	5	54	27	27	0	
	12190	3034311744	30.2	1998	12	52	21	31	0	
	12191	3034311822	30.2	1998	2	17	9	4	2	

данные о носителе

	nomer	nz	kto	gen	kem	krem	patlg	rezbakt	rezbiol	gelm
	272	12190	1	2			1	-	-	
	273	12190	1	2			1	-	-	
	274	12190	1	3	5		1	-	-	
	275	12190	1	1	8		1	-	-	

данные о блохах

	nz	ngr	kodbl	kbl	pcv	bkt
	12190	272	1	28	1	-

Рис. 3. Просмотр данных по журналу исследования полевого материала.

На основе информации, хранящейся в созданных базах данных, были сформированы временные ряды, позволяющие производить информационный анализ в нужном разрезе. Такие временные ряды отражают временную динамику хранимых данных и формируют итоговые характеристики по триаде носитель–переносчик–возбудитель совместно с метеорологическими данными, итоговые данные по большой песчанке по сезонам весна – осень, итоговые данные для секторов, в которых был обнаружен чумной микроб [4–6].

АППРОКСИМАЦИЯ РЕАЛЬНЫХ ПРОЦЕССОВ ГЕНЕТИЧЕСКИМИ АЛГОРИТМАМИ

Как известно, в настоящее время существуют две концепции математического представления математических моделей исследуемых систем. На основе первой концепции “пространство состояний” математическая модель исследуемой системы представляется с помощью дифференциальных или разностных (детерминированных или стохастических) уравнений. Второй — “вход–выход” описывает динамику систему с использованием либо детерминированных (стохастических) рядов Вольтерра или дискретным аналогом рядов Вольтерра, известным как полиномы Габора–Колмогорова (степенные ряды). Решив задачу реализации, можно на основе математической модели в виде характеристики “вход–выход” восстановить математическую модель в пространстве состояний.

Ниже представлены результаты решения задачи по аппроксимации реальных графиков функции с заданной точностью для получения полиномиальной модели в концепции “вход–выход”. Решение этой задачи основано на подходе генетического программирования [7] для моделирования динамики двух взаимодействующих популяций чумной триады: носителя (большой песчанки) и переносчика (блох).

Постановка задачи: для заданных векторов независимых переменных $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{id}) \in R^d$ и зависимых переменных $y_i \in R$ найти модели $p(\mathbf{x})$, оптимально близкие к неизвестной функции $f(\mathbf{x}) = y$.

Степенные полиномы являются универсальным средством для решения поставленной задачи, так как любое непрерывное отображение может быть приближено с достаточной точностью в среднеквадратическом смысле при использовании достаточно большого количества переменных. Были использованы полиномы Габора–Колмогорова, которые представляются в виде степенных рядов:

$$p(\mathbf{x}) = a_0 + \sum_{m=1}^M a_m \prod_{j=1}^d x_j^{r_{jm}}, \quad (5)$$

где a_m — коэффициентами при независимых переменных, $m \leq M$; x_j являются значениями независимых переменных входного вектора \mathbf{x} , $j \leq d$; и $r_{jm} = 0, 1, \dots$ являются степенями, с которыми x_j входит в m -ю переменную.

Предполагается, что r_{jm} является границей максимальной степени полинома s : $\sum_{j=1}^d r_{jm} \leq s$ для любого m .

Полиномы (5) представляют собой комбинации переменных, которые являются линейными по коэффициентам и нелинейными по переменным. Можно точно сказать, что векторное произведение и степени переменных формируются из стандартного базиса:

$$\phi = (1, x_1, x_2, \dots, x_d, x_1^2, x_1 x_2, x_1 x_3, \dots, x_1^3, x_1^2 x_2, x_1^2 x_3, \dots).$$

Используя этот базис ϕ степенной полином может быть представлен соотношением:

$$p(\mathbf{x}) = \sum_{m=1}^M a_m \phi_m(x), \quad (6)$$

где a_m являются коэффициентами, и ϕ_m являются базисными функциями из ϕ . Если рассматриваются достаточно малые подмножества ϕ то вместо функций ϕ_m можно

использовать базисные функции h_m следующего вида: $h_0(\mathbf{x})=1$, $h_1(\mathbf{x})=x_1$, $h_2(\mathbf{x})=x_2$, $h_3(\mathbf{x})=x_1x_2$, $h_4(\mathbf{x})=x_1^2$, $h_5(\mathbf{x})=x_2^2$, $h_6(\mathbf{x})=x_1^2x_2$, $h_7(\mathbf{x})=x_1x_2^2$, и $h_8(\mathbf{x})=x_2^3$.

Предположим, что имеется конечный набор N данных $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ независимых переменных векторов x_i и соответствующие выходные переменные y_i . Коэффициенты полинома a_m могут быть определены с помощью метода наименьших квадратов. Согласно этому методу наилучшими оценками a_m будут такие, которые обращают в минимум сумму квадратов невязок

$$\varepsilon(a_1, \dots, a_M) = \sum_{i=1}^N [y_i - p(\mathbf{x}_i)]^2.$$

Так как полином (6) является линейным по коэффициентам, то квадратичный критерий качества является выпуклым. Поэтому необходимое условие $\frac{\partial \varepsilon}{\partial a_m} = 0$, $m = 1, \dots, M$ минимума $\varepsilon(a_1, \dots, a_M)$ становится и достаточным. Выражение для определения неизвестных коэффициентов полинома представляется в виде:

$$\mathbf{a} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}, \tag{7}$$

где \mathbf{a} является $(m + 1) \times 1$ вектором, \mathbf{H} — $N \times (m + 1)$ матрица, строками которой являются векторы $\mathbf{h}(\mathbf{x}_i) = (h_0(\mathbf{x}_i), h_1(\mathbf{x}_i), \dots, h_m(\mathbf{x}_i))$, $m \leq M, i = 1, \dots, N$, \mathbf{y} является $N \times 1$ выходным вектором и $\mathbf{H}\mathbf{a}=\mathbf{p}$, $\mathbf{H}^T \mathbf{H}$ — ковариационная матрица. Системы генетического программирования рассматривают полиномы в виде структуры деревьев [7]. Каждое дерево содержит набор полиномов второй степени: $p(\mathbf{x}) = a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2$. Результирующий полином строится из дерева при помощи вертикальной техники. Коэффициенты составляющих полиномов в каждом функциональном узле вычисляются с помощью метода наименьших квадратов [7]. После этого полиномы используются для формирования выходных векторов, с помощью которых формируется матрица \mathbf{H} . Ниже в табл. 1 представлен набор составляющих квадратичных полиномов.

Таблица 1. Набор составляющих квадратичных полиномов.

$p_1(\mathbf{x}) = a_0 + a_1x_1 + a_2x_2$
$p_2(\mathbf{x}) = a_0 + a_1x_1 + a_2x_1^2x_2$
$p_3(\mathbf{x}) = a_0 + a_1x_1 + a_2x_1x_2^2$
$p_4(\mathbf{x}) = a_0 + a_1x_1 + a_2x_2^2$
$p_5(\mathbf{x}) = a_0 + a_1x_1 + a_2x_1x_2$
$p_6(\mathbf{x}) = a_0 + a_1x_1 + a_2x_2^3$

Предположим, что имеются временные ряды, характеризующие состояние взаимодействующих популяций носителя и переносчика, которые представлены в табл. 2. При этом использованы следующие обозначения: $chbr$ — численность носителя на гектар; $iobil$ — индекс обилия; $vv, vv1$ — приросты численности носителя от весны к весне текущего и предыдущего годов. В табл. 2 представлен результат построения полиномиальной модели для случая с 10 переменными. Для рассматриваемого случая получен следующий полином:

$$p(x) = 4.46x_1 - 610.208x_2 - 956.67x_3 - 993.686x_4 + 163.785x_5 - 0.002x_1^2 + \\ + 0.309x_1x_2 + 0.484x_1x_3 + 0.5x_1x_4 - 0.084x_1x_5.$$

Таблица 2. Результат построения полиномиальной модели.

Год	chbp	Iobil	vv	vv1	Численность носителя (эксперимент)	Численность носителя (модель, 10 переменных)	Структура полинома	Коэффициенты
1975	1,67	2,15	0,23	0,43	4,49	32,7052	1 0 0 0 0	4.46095
1976	1,9	2,22	0,2	11,01	2,73	10,3816	0 1 0 0 0	-610.208
1977	2,1	1,08	10,81	10,52	2,16	-30,2353	0 0 1 0 0	-956.67
1978	12,91	2,05	-0,29	-3,54	28,24	55,0022	0 0 0 1 0	-993.686
1979	12,62	7,9	-3,25	-11,96	128,11	84,047	0 0 0 0 1	163.785
1980	9,37	11,1	-8,71	-9,33	92,3	87,8612	2 0 0 0 0	-0.00225
1981	0,66	12,67	-0,62	0	8,36	31,3687	1 1 0 0 0	0.309439
1982	0,04	3,53	0	0,66	0,16	10,4926	1 0 1 0 0	0.483686
1984	0,7	0,63	0	0	0,54	0,445526	1 0 0 1 0	0.500774
1985	0,28	0,44	-0,42	4,16	0,12	-16,9498	1 0 0 0 1	-0.0838382
1986	4,86	3	4,58	9,46	14,68	-1,89844		
1987	9,74	2,08	4,88	15,23	26,79	0,937886		
1988	20,09	5,06	10,35	1,16	82,18	118,264		
1989	10,9	18,11	-9,19	-13,85	177,5	147,922		
1990	6,24	17,69	-4,66	-7,59	100,43	117,506		
1991	3,31	7,14	-2,93	3,37	17,61	9,01304		
1992	9,61	5,22	6,3	9,95	45,68	47,5517		
1993	13,26	8,64	3,65	2,64	87,02	112,231		
1994	12,25	8,29	-1,01	-1,9	85,65	100,744		
1995	11,36	12,56	-0,89	1,12	139,85	122,874		
1996	13,37	13,43	2,01	-0,17	136,78	171,553		
1997	11,19	21,96	-2,18	7,35	178,4	186,698		
1998	20,72	15,32	9,53	0	353,08	314,346		

Графики реальной и модельной кривых представлены на рис 5. Увеличим количество переменных в полиноме до 20. Временные ряды, структура и параметры полиномиальной модели представлены в табл. 3.

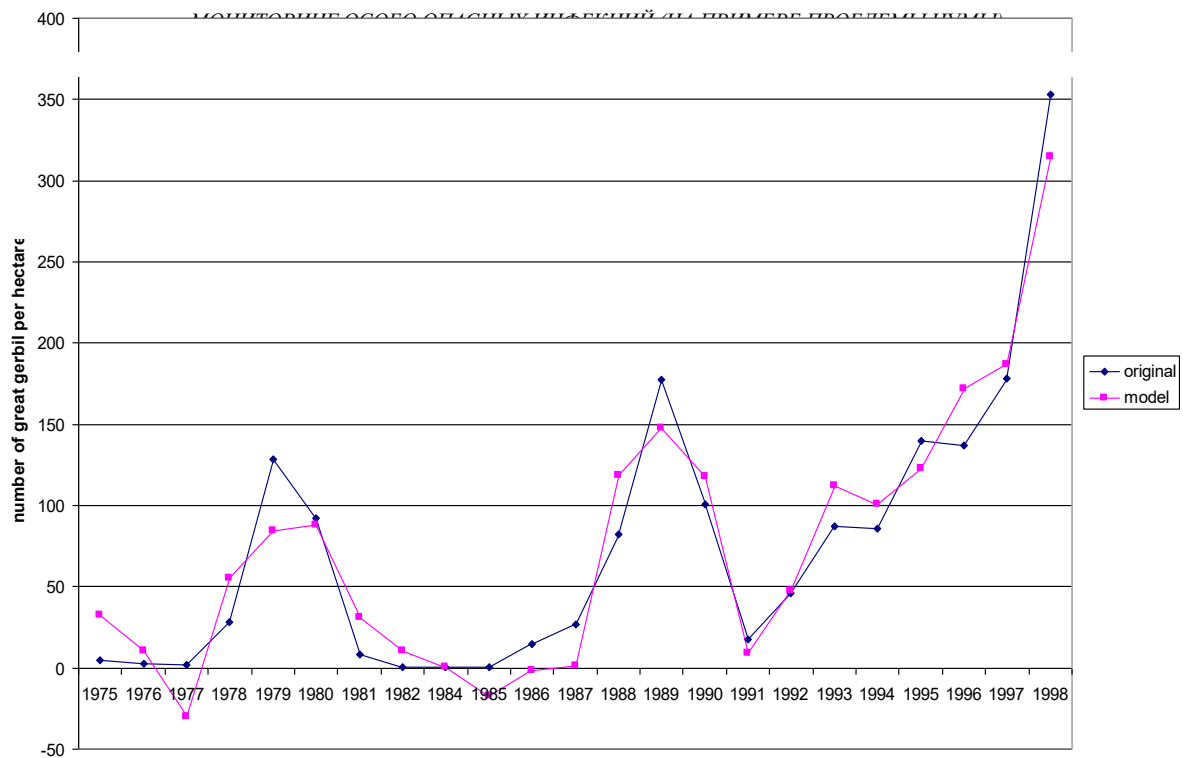


Рис. 5. Графики реальной и полиномиальной (с 10 переменными) кривых.

Таблица 3. Результат построения полиномиальной модели.

Год	chbp	Iobil	vv	vv1	Численность носителя (эксперимент)	Численность носителя (модель с 20 переменными)	Структура полинома	Коэффициенты
1975	1,67	2,15	0,23	0,43	4,49	6,53099	1 0 0 0 0	1.72029
1976	1,9	2,22	0,2	11,01	2,73	1,39399	0 1 0 0 0	-318.178
1977	2,1	1,08	10,81	10,52	2,16	3,94526	0 0 1 0 0	-341.747
1978	12,91	2,05	-0,29	-3,54	28,24	26,9237	0 0 0 1 0	469.111
1979	12,62	7,9	-3,25	-11,96	128,11	127,243	0 0 0 0 1	-664.691
1980	9,37	11,1	-8,71	-9,33	92,3	94,583	2 0 0 0 0	-0.0008685
1981	0,66	12,67	-0,62	0	8,36	8,46846	1 1 0 0 0	0.160206
1982	0,04	3,53	0	0,66	0,16	-0,659268	1 0 1 0 0	0.172455
1984	0,7	0,63	0	0	0,54	-4,26799	1 0 0 1 0	-0.238853
1985	0,28	0,44	-0,42	4,16	0,12	5,64518	1 0 0 0 1	0.335659
1986	4,86	3	4,58	9,46	14,68	6,14624	0 2 0 0 0	0.0779777
1987	9,74	2,08	4,88	15,23	26,79	29,9948	0 1 1 0 0	0.54044
1988	20,09	5,06	10,35	1,16	82,18	83,2839	0 1 0 1 0	-0.241786
1989	10,9	18,11	-9,19	-13,85	177,5	177,946	0 1 0 0 1	-0.0476934
1990	6,24	17,69	-4,66	-7,59	100,43	97,5393	0 0 2 0 0	0.0666727
1991	3,31	7,14	-2,93	3,37	17,61	24,1892	0 0 1 1 0	1.04355
1992	9,61	5,22	6,3	9,95	45,68	43,2588	0 0 1 0 1	-0.459014
1993	13,26	8,64	3,65	2,64	87,02	92,097	0 0 0 2 0	0.506574
1994	12,25	8,29	-1,01	-1,9	85,65	84,1213	0 0 0 1 1	-0.327171
1995	11,36	12,56	-0,89	1,12	139,85	116,01	0 0 0 0 2	0.193324
1996	13,37	13,43	2,01	-0,17	136,78	160,841		
1997	11,19	21,96	-2,18	7,35	178,4	179,782		
1998	20,72	15,32	9,53	0	353,08	347,845		

Для рассматриваемого случая получен следующий полином:

$$p(x) = 1.72029x_1 - 318.178x_2 - 341.747x_3 + 469.111x_4 - 664.691x_5 - 0.00086825x_1^2 + \\ + 0.160206x_1x_2 + 0.172455x_1x_3 - 0.238853x_1x_4 + 0.335659x_1x_5 + 0.0779777x_2^2 + .. \\ + 0.54044x_2x_3 - 0.241786x_2x_4 - 0.0476934x_2x_5 + 0.0666727x_3^2 + 1.04355x_3x_4 - \\ - 0.459014x_3x_5 + 0.506574x_4^2 - 0.327171x_4x_5 + 0.19332 .$$

Графики реальной и полиномиальной (с 20 переменными) кривых представлены на рис 6.

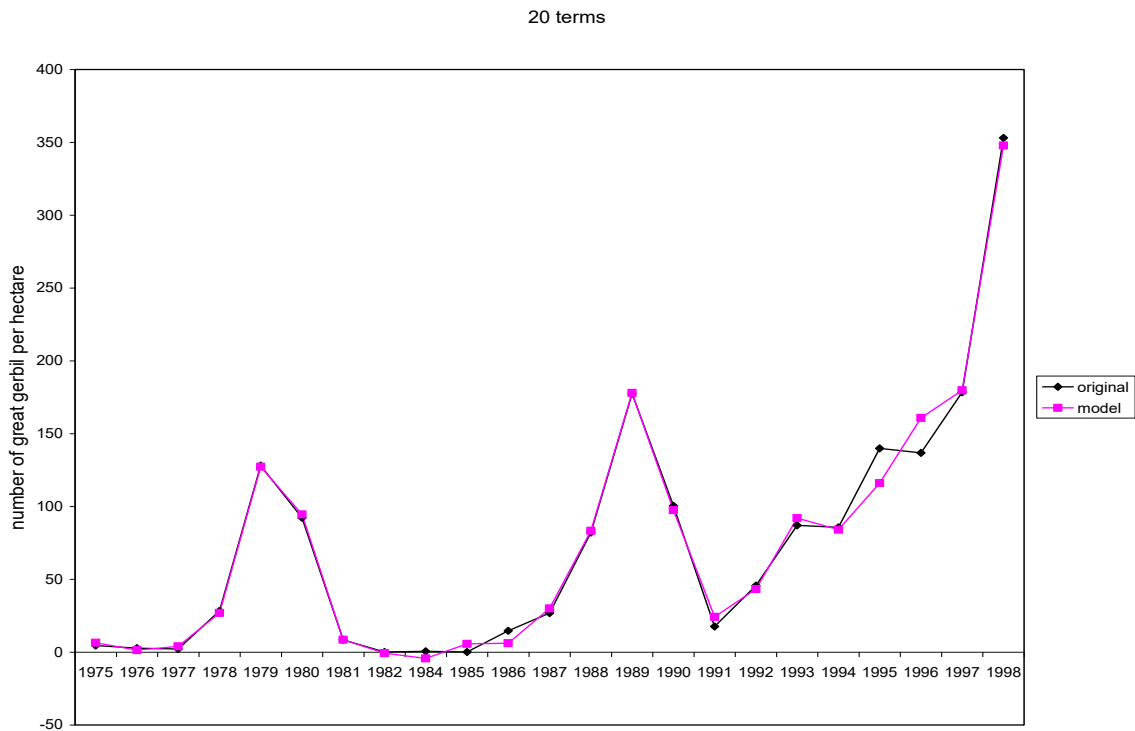


Рис. 6. Графики реальной и полиномиальной (с 20 переменными) кривых.

Таким образом, суммируя результаты представленных численных экспериментов, можно заключить, что использованный подход генетического программирования представляет собой эффективное средство для моделирования взаимодействующих популяций носителя и переносчика в исследуемой биологической системе. Чем больше переменных используется в полиномиальной модели, тем больше точность решения задачи аппроксимации.

РЕШЕНИЕ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ ЭПИЗООТИЧЕСКИХ ПРОЦЕССОВ НА ОСНОВЕ ПОДХОДА ИММУНОКОМПЬЮТИНГА

К нетрадиционным вычислительным процедурам на основе интеллектуальных информационных технологий относится подход иммунокомпьютинга [8–13]. Его строгий математический базис основан на биологическом прототипе иммунной сети, понятиях формального протеина и формальной иммунной сети [8–13].

Подход иммунокомпьютинга позволяет в высоко распределенной манере с большой эффективностью решать следующие задачи: распознавания образов, классификации и кластеризации, формирования различного типа индексов, прогнозирования стадий эпизоотического процесса и т.д. Архитектура системы мониторинга чумы на основе подхода иммунокомпьютинга представлена на рис. 7.



Рис. 7. Архитектура системы мониторинга чумы на основе подхода иммунокомпьютинга.

Следует отметить, что первичные данные о состоянии чумной триады и абиотических факторах исследуемого природного очага чумы могут представляться в виде точечных и интервальных значений индикаторов, в виде плоских и многомерных матриц (OLAP-кубы) и т.д.

База знаний содержит сведения, которые отражают закономерности, существующие в рассматриваемой предметной области, позволяют выводить новые знания и прогнозировать потенциально возможные состояния исследуемой области; сведения о структуре и содержании базы данных; сведения по языку общения; метазнания, определяющие способы представления и переработки знаний. В базу знаний помещаются как общедоступные данные, так и формализованные знания эксперта в данной предметной области, вычислительные алгоритмы реализации процедур иммунокомпьютинга и интерпретацию результатов вычислений.

Ниже представлен градиентный алгоритм для базовой процедуры иммунокомпьютинга — сингулярного разложения интервальной многомерной матрицы.

1. Градиентный алгоритм сингулярного разложения многомерной интервальной матрицы

Пусть анализируемые первичные данные оперативно представляются в виде интервальной многомерной матрицы (OLAP-куба) $\mathbf{A} \in IR^{n \times m \times k}$, $IR^{n \times m \times k}$ — интервальное пространство соответствующей размерности. Как известно, точечную многомерную матрицу $A = \{a_{ijl}\} \in R^{n \times m \times k}$, $A \in \mathbf{A} \in IR^{n \times m \times k}$ можно представить через компоненты сингулярного разложения в виде:

$$A = s_1 Z_1^B X_1 Y_1^T + s_2 Z_2^B X_2 Y_2^T + \dots + s_p Z_p^B X_p Y_p^T, \quad (8)$$

где

$$s_i = X_i^T A Y_i Z_i^B, \dots Y_i^T Y_i = 1, \dots X_i^T X_i = 1, \dots Z_i^{TB} Z_i^B = 1, i = 1, 2, \dots, p, \quad (9)$$

p — ранг матрицы, $s_1 \geq s_2 \geq s_3 \geq \dots \geq s_p \geq 0$ — ранжированные сингулярные числа матрицы A ; Y_i, X_i, Z_i — соответственно нормированные левый, правый и глубинный

сингулярные векторы; верхние индексы T и B , соответственно, означают операции транспонирования и транспонирования в глубину. Сингулярное число $s_i(\mathbf{A})$ интервальной многомерной матрицы представляется интервалом, ограничивающим множество сингулярных чисел любой матрицы $A \in \mathbf{A} \in \mathbb{R}^{n \times m \times k}$:

$$s_i(\mathbf{A}) = \text{hull} \{s_i(A) \mid (\forall A \in \mathbf{A})\}. \quad (10)$$

Под функцией $\text{hull}(M)$ понимается наиболее плотный интервал, ограничивающий множество M , т. е. $\text{hull}(M) = [\inf\{M\}, \sup\{M\}]$.

Задача. Найти минимум и максимум матричной целевой функции — сингулярного числа $s(\mathbf{A})$ в области определения элементов интервальной матрицы $A \in \mathbf{A} \in \mathbb{R}^{n \times m \times k}$.

Предложена процедура минимизации и максимизации i -го сингулярного числа поэлементно в интервале $\mathbf{A} \in \mathbb{R}^{n \times m \times k}$ в виде следующей итеративной схемы:

$$\begin{aligned} \hat{A}_k &= A_k + d \cdot Z_k^B(A_k) \cdot X_k(A_k) \cdot Y_k^T(A_k), \\ A_{k+1}(i, j) &= \begin{cases} A(i, j), \hat{A}_k(i, j) < \underline{A}(i, j) \\ \bar{A}(i, j), \hat{A}_k(i, j) > \bar{A}(i, j) \\ \hat{A}_k(i, j), \hat{A}_k(i, j) \in \mathbf{A}(i, j) \end{cases}, \\ &A_0 \in \mathbf{A}, \end{aligned} \quad (11)$$

где $X_k(A_k), Y_k^T(A_k), Z_k^B(A_k)$ — i -й левый, правый и глубинный сингулярные векторы матрицы A_k , $d \in [-1, 0) \cup (0, 1]$ — коэффициент, характеризующий направление (минимизация, максимизация) и длину шага оптимизации, A_0 — начальная матрица итерации.

С помощью итерации (11) можно определить нижнюю и верхнюю границы интервальных сингулярных чисел $\sigma_i(\mathbf{A})$ интервальной сингулярной матрицы \mathbf{A} с заданной точностью.

Для формализации критерия останова очередной процедуры (11) использованы следующие переменные:

$$M = \frac{1}{l} \sum_{j=k-l+1}^k \sigma_i(A_j), \quad J = \frac{1}{2} \sum_{j=k-l+1}^k \left| \text{sign}(M - \sigma_i(A_j)) - \text{sign}(M - \sigma_i(A_{j-1})) \right|, \quad (12)$$

где A_j — j -ая матрица процедуры (11), k — номер текущего цикла итерации, M — среднее арифметическое i сингулярных чисел, полученных на l последних циклах.

Условие останова итерации (11) имеет вид:

$$(J > l - 2) \text{ or } (J = 0) \text{ or } (\sigma_i(A_k)). \quad (13)$$

Пусть $A_0 \in \{\inf(\mathbf{A}), \text{mid}(\mathbf{A}), \sup(\mathbf{A})\}$, величина шага оптимизации равна $d = -1$ при вычислении нижней границы, $d = 1$ — верхней границы.

Предложенный градиентный алгоритм содержит следующие шаги:

Шаг 1. Провести итерацию (11), останов осуществляется при выполнении неравенства (13).

⋮

Шаг k . Провести итерацию (11), но в качестве начальной матрицы принять матрицу, на которой остановилась итерация (11) на предыдущем шаге. При этом коэффициент d уменьшить в два раза.

2. Формирование индекса риска чумы

Предположим, для исследуемого ландшафтно-экологического района (ЛЭР — равнина Акдала) возможны следующие ситуации: стандартная ситуация (или нормальное поведение) и нестандартные ситуации, характеризующиеся ходом эпизоотического процесса триады (носитель–переносчик–чумный микроб).

Под индексом риска чумы понимается обобщенный показатель нестандартной ситуации на исследуемом объекте.

Как показали исследования [11–13], применение математических моделей иммунокомпьютинга при формировании индекса риска чумы является эффективным методом сокращения огромного числа разнородных, динамических непрерывных и дискретных переменных в легко понятный индекс. В результате обеспечивается сжатие исходной информации, т.е. интегрирование сложных разнородных многомерных величин в одно число, посредством которого можно произвести комплексную оценку состояния сложной динамической системы и выявить нестандартные ситуации для исследуемых природных очагов.

Процедура формирования индекса риска чумы, вычисляемого на основе подхода иммунокомпьютинга, позволяет существенно сжимать набор разнородных индикаторов в одно число. Это число особенно наглядно сигнализирует об опасных ситуациях и тенденциях.

Эпизоотический процесс характеризуется, как правило, высокой восприимчивостью популяции носителей к чумному микробу, высокой плотностью и большой численностью носителей и переносчиков. Накопление чумных микробов повышает их вирулентность. Их увеличение вызывает специфическую реакцию в популяции носителей, что приводит к уменьшению плотности, как популяции носителя, так и популяции переносчика, и уменьшению вирулентности чумного микроба.

Таким образом, индикаторы процессов в чумной триаде являются весьма многочисленными и разнородными. Поэтому прямое использование их значений, получаемых в результате мониторинга, не позволяет формировать оценку текущего состояния чумной триады и прогнозирования риска эпизоотии.

На основе математического аппарата иммунокомпьютинга предложен качественно новый подход к прогнозированию эпизоотий в природных очагах чумы как к задаче распознавания образов. Этот подход уже продемонстрировал свою эффективность и робастность, а также перспективность использования в интеллектуальных автоматизированных системах мониторинга чумы.

Алгоритм формирования и расчета индекса риска чумы (ИРЧ) на основе подхода иммунокомпьютинга, состоит из следующих шагов [13]:

- Шаг 1.** По результатам мониторинга сформировать матрицу M размерности $m \times n$, где m — годы мониторинга, которые соответствуют строкам матрицы, а n — значения параметров мониторинга, которые соответствуют столбцам матрицы.
- Шаг 2.** Осуществить сингулярное разложение матрицы M , выбрать правый сингулярный вектор V_1 , соответствующий максимальному сингулярному числу матрицы M .
- Шаг 3.** Вычислить значения индекса риска чумы.

Продemonстрируем результат формирования индекса риска чумы на примере ЛЭР — равнина Акдала Прибалхашского природного очага чумы за 23 года.

Пример. Пусть состояние чумной триады в исследуемом ЛЭР характеризуется 45 временными рядами, каждая компонента которых может иметь, как точечные, так и интервальные значения индикаторов. Воспользуемся градиентным алгоритмом сингулярного разложения интервальной матрицы и вычислительным алгоритмом формирования индекса риска чумы.

Сформируем матрицу M размерности (11×45) , и рассчитаем значения индекса риска чумы по приведенному вычислительному алгоритму. Динамика изменения индекса риска чумы представлена на рис. 8.

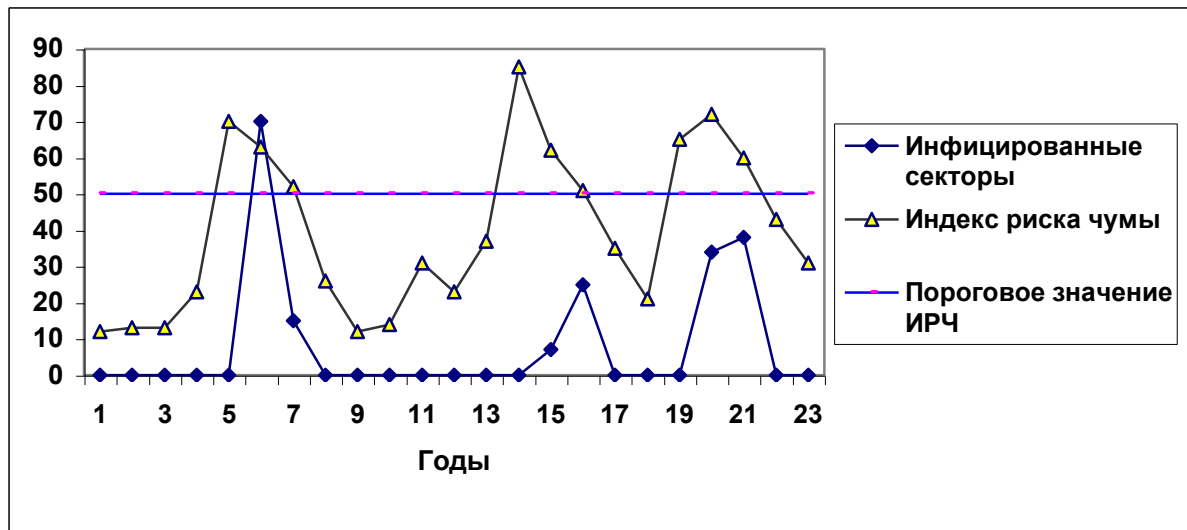


Рис. 8. Динамика изменения индекса риска чумы

На рис. 8. представлены два графика: график зараженности секторов (инфицированные секторы - кривая с ромбами) и график индекса риска чумы (кривая с треугольниками). Значения индекса риска чумы, лежащие выше порогового значения $I_h=50$, свидетельствуют о возможном возникновении эпизоотических процессов в исследуемом ландшафтно-экологическом районе.

Сравнение полученных результатов с реальными данными показывает, что индекс риска чумы, сформированный на основе математических моделей и вычислительных алгоритмов иммунокомпьютинга, с достаточной степенью точности позволяет отслеживать начало эпизоотических процессов в природном очаге чумы.

СПИСОК ЛИТЕРАТУРЫ

1. Айкимбаев А.М., Литвак Я.И., Шварц А.В. и др. Руководство по эпидемиологическому надзору в горных очагах чумы Тянь-Шаня и Алтая. Алма-Ата. 1991. 125 с.
2. Айкимбаев А.М., Мусагалиев Ж.С., Хамзин С.Х. и др. Эпидемиологический надзор за чумой в Урало-Эмбенском и Предустюртском автономных очагах. Алматы. 1994. 129 с.
3. Аубакиров С.А., Сержанов О.С., Фомушкин В.М. и др. Руководство по ландшафтно-эпизоотологическому районированию природных очагов чумы Средней Азии и Казахстана. Алма-Ата. 1991. 29 с.
4. Абдуллина В.З. Особенности проектирования и обработки реляционных баз данных в информационных системах с временной динамикой. Автореф. канд. дисс. Алматы: КазНТУ. 2003. 30 с.
5. Абдуллина В.З. Проектирование баз данных для противочумной службы Казахстана. Сборник материалов международной конференции *Менеджмент и новые технологии*. Алматы. 2001. с. 25–28.
6. Sokolova S.P., Abdullina V.Z., Abramov B.A. et al. *Artificial Immune System for the gerbil natural plagues focus.* /Edited by A.O. Tarakanov. Almaty. 2002. 180 p.
7. Sokolova S.P. and Ivlev R.S. *Mathematical Modeling and Investigation of Dynamic Properties of Biological Systems at Population Level.* Almaty. 2003. 123 p.

8. Tarakanov A.O., Skormin V.A., Sokolova S.P. *Immunocomputing: Principles and Applications*. N.Y. Springer. 2003. 193 p.
9. Tarakanov A. and Dasgupta D. A formal model of an artificial immune system. *BioSystems*. 2000. **55**(1–3). pp. 151–158.
10. Tarakanov A., Sokolova S., Abramov B.A. et al. Immunocomputing of the natural plague foci. In: *Proc. of the Genetic and Evolutionary Computation Conference (GECCO-2000). Workshop on Artificial Immune Systems*. Las Vegas. USA. 2000. pp. 38–41.
11. Sokolova L. A. Index design by immunocomputing. *Lecture Notes in Computer Science*. Springer: Berlin. 2003. **2787**. pp. 120–127.
12. Соколова Л.А. Индекс риска чумы на основе иммунокомпьютинга. *Труды СПИИРАН (под ред. Р.М. Юсупова)*. СПб.: СПИИРАН, 2003. Вып. 1. Т. **3**. с. 137–141.
13. Соколова Л.А. *Комплексная оценка многомерных систем методами иммунокомпьютинга*. Автореф. канд. дисс. Санкт-Петербург. 2005. 16 с.

Материал поступил в редакцию 03.04.2007, опубликован 25.04.2007