

Академик НАН Украины Г. Г. Пивняк, Б. С. Бусыгин, С. Л. Никулин

Гис-технология интегрированного анализа разнородных и разноуровневых геоданных

The current state and prospects of using GIS in the scope of geoscience are considered. The structure of the specialized geoinformation system RAPID is described. The work technique in the system using heterogeneous and different level data is briefly characterized. Some examples of the solution of forecast tasks based on geophysical and high-accuracy space survey materials are given.

Стремительное развитие вычислительной техники, космических технологий и математических методов обработки пространственно-привязанных данных (геоданных) полностью изменили характер производственных и исследовательских процессов в недропользовании, геологии, экологии. В настоящее время они базируются на повсеместном использовании компьютерных технологий на базе геоинформационных систем (ГИС-технологий). Авторами этого сообщения рассмотрены современное состояние и перспективы использования ГИС в сфере наук о Земле, а также дано описание геоинформационной технологии интегрированного анализа разнородных и разноуровневых данных, основанной на инструментарии специализированной геоинформационной системы РАПИД (распознавание, автоматизированное прогнозирование, интерпретация данных). Технология позволяет с общих методических позиций эффективно решать широкий круг задач, связанных с обработкой и интерпретацией геоданных.

1. Состояние проблемы и тенденции развития. В настоящее время в практике применения геоинформационных систем в науках о Земле наблюдается несколько взаимодействующих тенденций. Во-первых, происходит быстрое и глубокое проникновение инструментария универсальных ГИС — крупных коммерческих программных продуктов, реализующих основные функции работы с пространственно-привязанными данными: сбор, упорядочение, хранение, манипулирование, первичную обработку, а также выполнение специфических операций ГИС-анализа. В число последних входит отбор пространственных объектов по заданным критериям, буферизация, оверлейный, геостатистический и сетевой анализ, картографическая алгебра и ряд других широко распространенных процедур. При безусловной полезности указанных операций, обеспечиваются лишь основные потребности анализа данных, относящихся к конкретным сферам наук о Земле. Подобная ситуация обусловила требования включения в состав ГИС отдельных инструментов для решения некоторых задач конкретных проблемных областей. Так, например, в ГИС ArcView включены специальные средства обработки гидрогеологических данных. И хотя открытая архитектура большинства крупных универсальных ГИС позволяет пользователям реализовывать некоторые необходимые профессиональные функции [1–4], это не устраняет основной проблемы универсальных ГИС, которые, как правило, представляют собой наборы разнородных инструментов. Обычно последовательность и способ применения такого инструментария не увязывается в достаточно конкретные технологические схемы и не оговариваются разработчиками.

В результате технологические схемы решения конкретных задач разрабатываются пользователями самостоятельно, на основе собственного опыта или с использованием литературных источников. Поскольку при решении различных прикладных задач технологические схемы существенно отличаются, то от специалистов предметных областей обычно требуется знание всего инструментария используемой ГИС. Это обстоятельство существенно ограничивает эффективность внедрения универсальных ГИС в практику решения прикладных задач.

Другая тенденция заключается в разработке специализированных ГИС, которые реализуют значительное число подходов, методов и моделей из конкретной предметной области [5–7] и обеспечивают работу в рамках одной или нескольких фиксированных технологических схем. При этом от пользователя не требуется универсальных знаний в области ГИС, а следовательно, применяются понятные ему терминология и интерфейс. Внедрение таких систем в научную и производственную практику происходит значительно эффективнее и требует меньших затрат на освоение.

Особенностями таких систем являются работа на одном гипсометрическом уровне, использование ограниченного числа шкал измерений, фиксированные и негибкие технологические схемы, специфический функциональный аппарат.

И, наконец, в качестве отдельной тенденции, связанной с повышением удельного веса аэрокосмических материалов при решении задач недропользования, можно отметить широкое распространение специализированных систем обработки аэрокосмоснимков. Яркими представителями таких систем являются Erdas Imagine (фирма Leica Geosystems), ErMapper (Earth Resource Mapping Pty), eCognition (Definiens Imaging) и др. Несмотря на несомненную полезность, они обладают всеми недостатками ГИС указанных выше типов: сложны в изучении, не позволяют выполнять специфические операции конкретных предметных областей и обеспечивать достаточными возможностями комплексного анализа данные разных типов.

В силу указанных причин, задача создания специализированных ГИС, позволяющих эффективно интегрировать разнородные и разноуровневые геоданные, является исключительно актуальной.

2. Технология ГИС РАПИД.

2.1. Общие сведения. Специалистами кафедры геоинформационных систем Национального горного университета в течение ряда лет создавалась и совершенствовалась технология интегрированного анализа разнородных и разноуровневых данных, позволяющая на общих методических принципах решать широкий круг задач наук о Земле [8–10]. Технология реализована в рамках специализированной геоинформационной системы РАПИД. ГИС РАПИД — универсальный инструмент прогнозирования и принятия решений при поисках месторождений полезных ископаемых, при районировании территорий, мониторинге и прогнозе экологических ситуаций.

Технология базируется на методах теории распознавания образов, обработки изображений, геостатистики и пространственного анализа; реализует принцип многовариантного решения задач посредством имитационного моделирования и проведения вычислительных экспериментов [8]. Основное внимание уделяется установлению прямых связей между пространственными закономерностями расположения изучаемых объектов, явлений и структурой данных, описывающих их, — физических полей, геологических схем и карт, геохимических ореолов, аэрокосмических снимков, морфологическими и ландшафтными особенностями рельефа и др. (рис. 1).

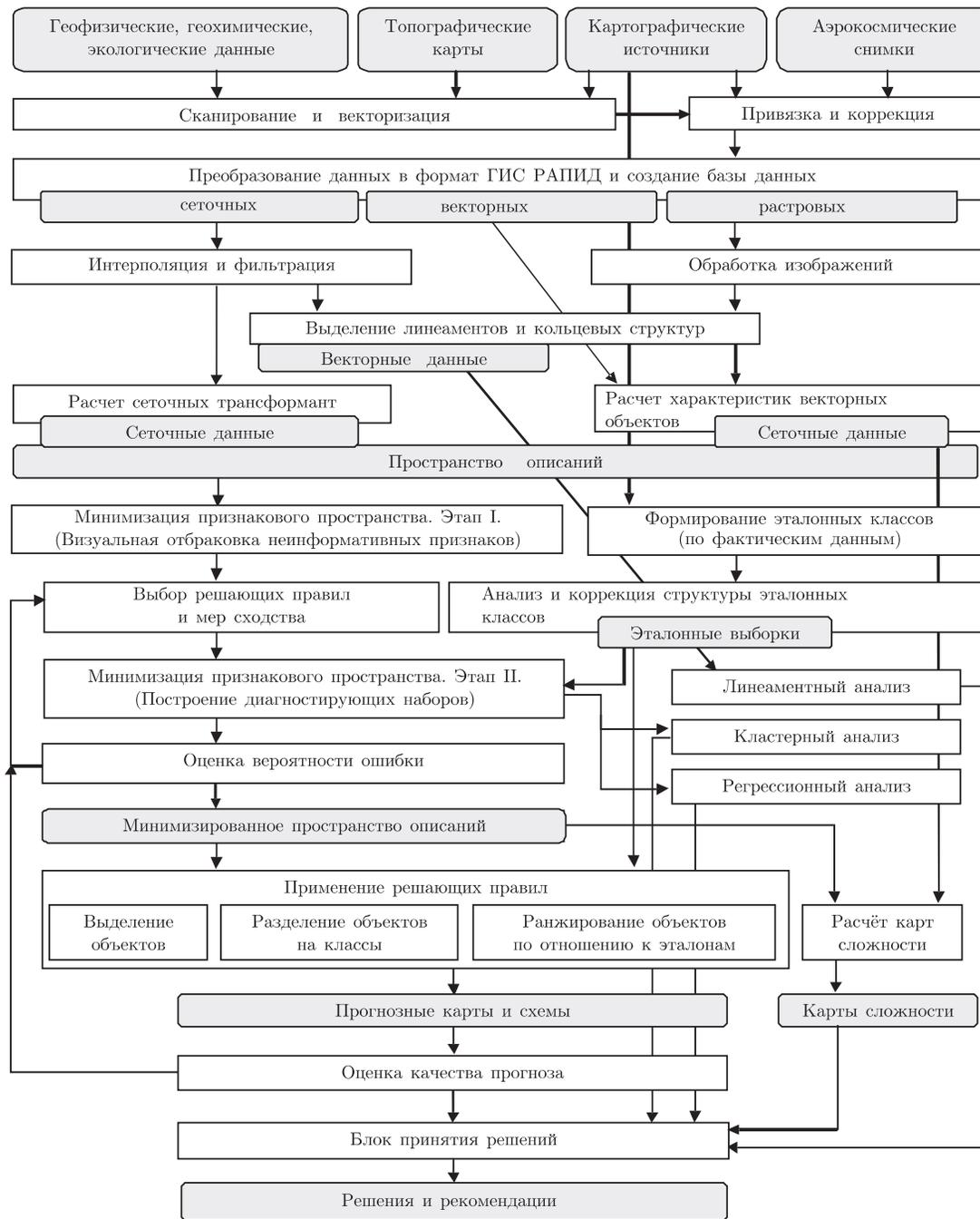


Рис. 1. Технологическая схема интегрированного анализа данных в ГИС РАПИД

РАПИД обеспечивает интеграцию растровых и векторных представлений данных, аналитические, информационно-справочные и измерительные функции; фильтрацию данных; формирование новых признаков и построение производных карт, исследование зависимостей и статистическую обработку; многомерное районирование, кластеризацию и классификацию; распознавание и картографирование ситуаций, многовариантную визуализацию данных, компоновку и печать выходных документов.

Система включает в себя ядро, обеспечивающее управление геоданными, а также совокупность прикладных модулей, решающих конкретные задачи обработки и анализа полученных результатов [11]. В качестве исходных используются данные, представленные в сеточной (физические поля и геохимические данные, приведенные к регулярной сети), векторной (картографические слои) и растровой (изображения, в т. ч. аэрокосмические) формах. Предварительно результаты, представленные на бумажных носителях, сканируются, геореференцируются и векторизуются в случае необходимости.

2.2. Формирование пространства описаний. Одним из основных этапов работы технологии является создание пространства описаний — совокупности картографических слоев (признаков), описывающих территорию и отражающих те или иные аспекты изучаемого явления или процесса. Поскольку блок прогнозирования РАПИД работает с данными, заданными в точках регулярной сети, то возникает необходимость приведения исходных векторных данных к сеточному виду, при котором данные соотносятся с узлами сети, формирующими конечное множество объектов $X = \{X_1, X_2, \dots, X_n\}$. С этой целью для векторных материалов (геологических границ, разломов, водоразделов и т. п.) рассчитываются признаки-характеристики, отнесенные к узлам регулярной сети: расстояние от узла сети до ближайшего объекта определенного типа (например, источника загрязнения), количество взаимопересечений линейных объектов одного или нескольких типов (например, разломов и геологических границ), частота встречаемости объектов в скользящей окрестности и т. п. Обычно подобные характеристики отражают структурные взаимосвязи между пространственными объектами.

В свою очередь данные, изначально представленные в сеточном виде (геофизические поля, геохимические ореолы и т. п.), подвергаются различным трансформациям в рамках специальной подсистемы ГИС РАПИД. Как свидетельствует опыт авторов и других исследователей, использование трансформант в процессе прогнозирования и мониторинга дает лучшие результаты по сравнению с исходными признаками [8]. ГИС РАПИД позволяет выполнять более 200 трансформаций, в том числе дифференциальные, текстурные, гистограммные, морфологические, корреляционные, фрактальные и др.

Отдельной обработке при формировании пространства описаний подвергаются аэрокосмические материалы. Дело в том, что при решении многих задач недропользования непосредственное применение характеристик яркости аэрокосмических материалов не всегда дает удовлетворительные результаты. Значительно более информативными являются результаты анализа линейных и кольцевых элементов, предварительно выделяемых на изображениях (рис. 2, а, б). Эти элементы отражают структурно-тектоническую обстановку, в том числе и такие ее детали, которые не фиксируются классическими полевыми геологическими методами [12]. В ГИС РАПИД реализована оригинальная подсистема анализа аэрокосмических материалов, включающая средства автоматического обнаружения линейных и кольцевых элементов и расчета специальных характеристик сети линеаментов (рис. 2, в) в скользящей окрестности [10, 13].

Таким образом, каждой точке сети с координатами (m, n) ставится в соответствие вектор признаков $(m, n) \leftrightarrow X_{(m,n)}^{(P)}$; где P — размерность вектора $X^P = (X_1, X_2, \dots, X_P)$, которая зависит от количества имеющихся признаков.

Созданное пространство описаний должно быть минимальным, но, в то же время, достаточным для решения поставленной задачи с необходимой точностью и надежностью. Поскольку в настоящее время не разработана формальная процедура, позволяющая заранее определить оптимальный набор признаков, то при решении конкретных задач вначале

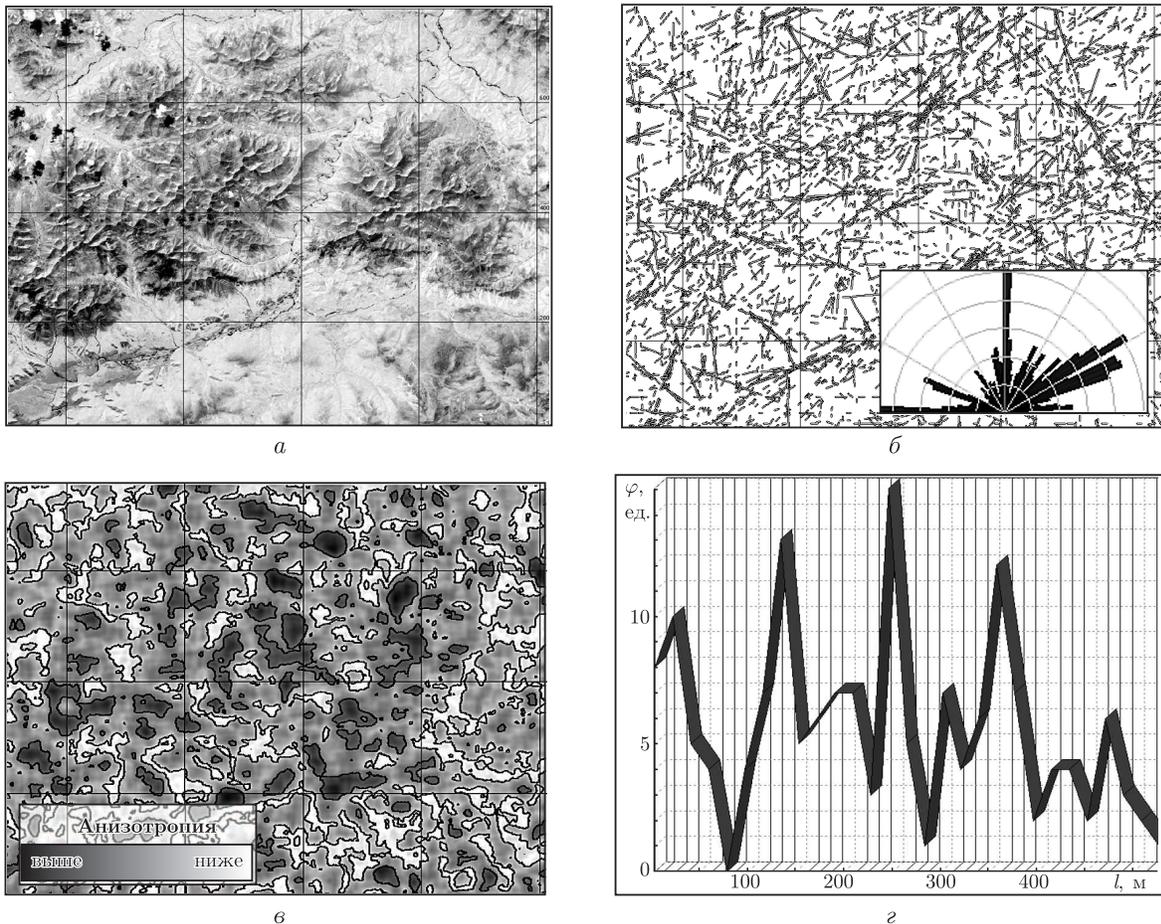


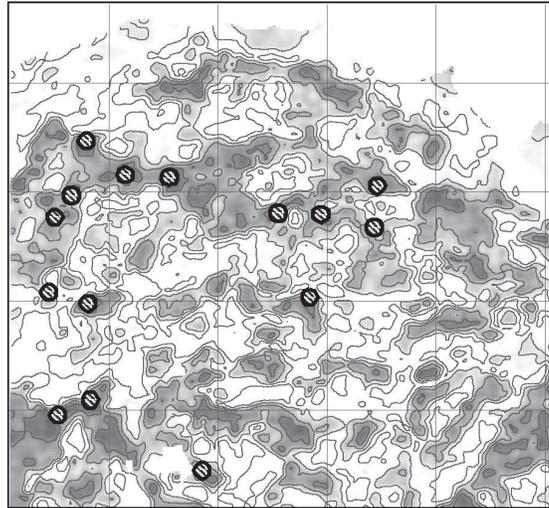
Рис. 2. Линеаментный анализ в ГИС РАПИД: *a* — исходный космический снимок (р-н Бурятии, сенсор Landsat-7, разрешение 30 м, синтезированный); *б* — построенная линеаментная сеть с розой-диаграммой направлений линеаментов; *в* — карта анизотропии линеаментной сети; *г* — гистограмма относительных расстояний между линейными элементами с азимутом (116 ± 5)°

создается исходная система признаков-описаний, а затем она минимизируется путем отбрасывания признаков, несущественных с точки зрения решения поставленной задачи [8]. Минимизация выполняется в два этапа. На первом, по результатам визуального анализа, отбрасываются заведомо неинформативные или дублирующие признаки.

Второй этап подразумевает сужающее отображение:

$$\{X^P\} \xrightarrow{F} \{X^r\}, \quad r < P, \quad F = F(\pi_1, \dots, \pi_K),$$

при котором достигается экстремум некоторого функционала качества $J_X(F)$. Здесь π_1, \dots, π_K — классы, в которые группируются точки сети, обладающие определенной общностью свойств. Минимизация функционала $J_X(F)$ предполагает использование эталонов — участков земной поверхности с известным характером проявления изучаемого процесса или явления (например, известные месторождения, эпицентры землетрясений и т. п.). Первоначально эталоны группируются в классы, согласно представлениям исследователя о близости их свойств. В дальнейшем необходимо оптимизировать состав и структуру эталонных классов, добиваясь их большей компактности в многомерном пространстве описаний. ГИС



Эпицентры землетрясений (1999–2006 гг.)

Рис. 3. Прогнозная карта риска возникновения землетрясений (результат ранжирования территории Северной Турции по мере сходства с эпицентрами землетрясений)

РАПИД реализует несколько подходов к решению этой задачи, определяющий качество результатов прогнозирования в целом.

Наличие эталонной выборки позволяет выполнить второй этап минимизации пространства описаний. Он основан на поиске диагностирующего набора признаков, наилучшим образом характеризующего территорию с точки зрения ее сходства с эталонной выборкой конкретного класса [8]. Диагностирующие наборы рассчитываются с использованием различных решающих правил и мер сходства; при этом оценивается вероятностная ошибка классификации (ВОК) набора. Оптимальным признается решающее правило, для которого минимизируется ВОК диагностирующего набора. Признаки, не вошедшие в этот набор, из дальнейшего рассмотрения исключаются.

2.3. Применение решающих правил. Найденное решающее правило и минимизированное пространство описаний позволяют перейти к процедурам распознавания образов. Последнее подразумевает отнесение исходных объектов (точек сети) $\{X_1, X_2, \dots, X_n\}$ к определенному классу π_k , $k = 1, 2, \dots, K$. Это означает, что нужно построить однозначное отображение множества X на множество классов:

$$\Pi = \{\pi_1, \pi_2, \dots, \pi_k\}: X \longrightarrow \Pi.$$

Основу технологии ГИС РАПИД составляют такие процедуры распознавания образов, как выделение искомым объектов, разделение объектов на классы и их ранжирование (рис. 3 и 4, а) по отношению к эталонам. Их результаты представляются в виде прогнозных карт и схем, отражающих перспективность территории с точки зрения наличия или возникновения изучаемых явлений и процессов. Блок оценки качества прогноза позволяет определить степень достоверности результатов, а также применимость использованных процедур в целом. Известно, что процедуры, предполагающие наличие эталонов (т. е. реализующие метод аналогий), не всегда эффективны. Это связано с уникальностью некоторых природных явлений (например, крупных рудных месторождений). В связи с этим технология ГИС РАПИД реализует подход к прогнозированию, основанный на построении карт сложности

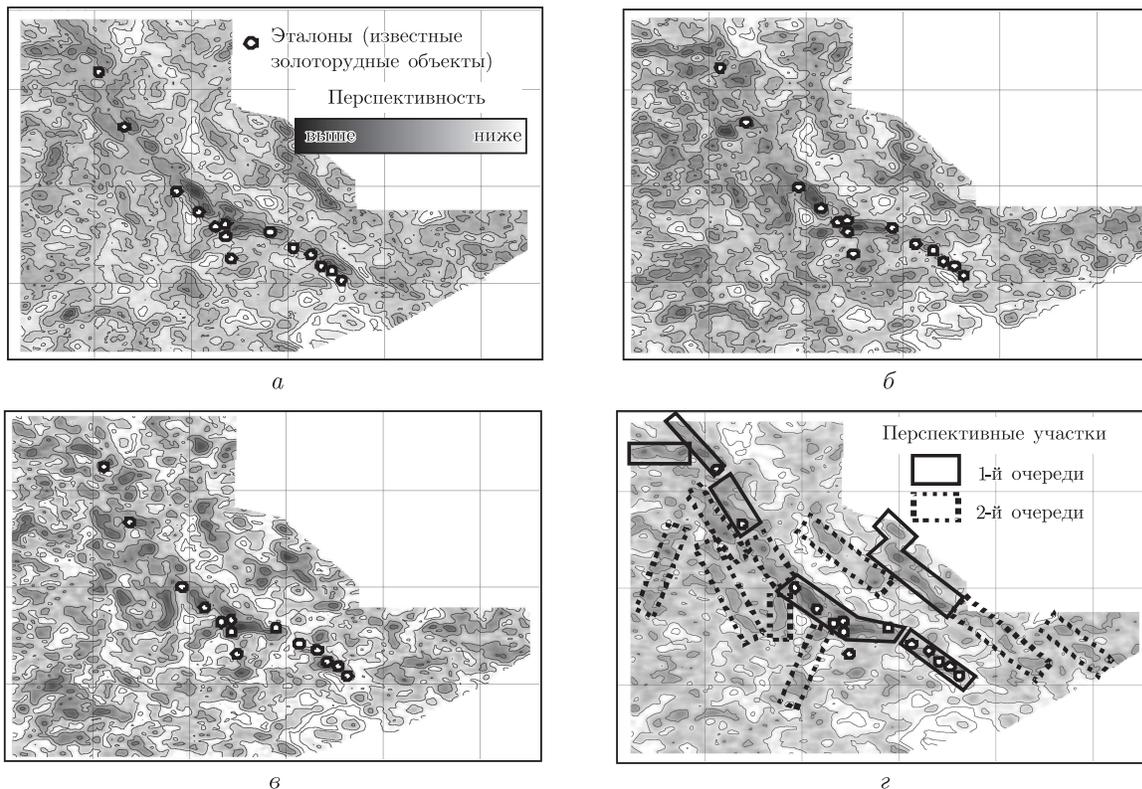


Рис. 4. Прогнозирование золоторудного оруденения (Сорокинская зеленокаменная структура, Украина) с использованием различных подходов: *а* — карта ранжирования территории по отношению к эталонам; *б* — карта показателя сложности геологического строения; *в* — карта показателя степени анизотропии линейной сети; *г* — комплексная прогнозная карта

геологического строения [13]. Подход базируется на предположении, что более сложное строение перспективно с точки зрения как наличия оруденения (см. рис. 4, *в*), так и возникновения опасных геологических явлений.

Кроме того, технология ГИС РАПИД дает возможность привлечения методов регрессионного, корреляционного, кластерного, а также линейного анализа. Так, последний позволяет решать целый комплекс задач, особенно в области поисков полезных ископаемых, а также предсказания чрезвычайных геологических ситуаций (оползней, карстов и т. п.). Например, известна приуроченность нефтегазовых месторождений к зонам повышенной анизотропии сети линейментов [14] (см. рис. 4, *в*), а золоторудных — к пересечениям линейментных зон субширотного и субмеридионального простирания [15]. Представляет большой интерес изучение закономерностей чередования линейных структур различных систем (см. рис. 2, *г*), а также случаи нарушения таких закономерностей.

Полученные вышеперечисленными способами прогнозные результаты подаются на вход блока принятия решений, позволяющего согласовать данную информацию с использованием различных подходов [13], создать результирующие карты и схемы (см. рис. 4, *г*), сформулировать рекомендации по проведению дальнейших исследований.

Таким образом, на основании практического применения компьютерной технологии ГИС РАПИД в различных областях — геологии, экологии, предсказании чрезвычайных ситуаций, недропользовании — можно сделать следующие выводы:

технология позволяет решать большой круг задач различных предметных областей на общих методических принципах, а также обеспечивает возможность эффективного интегрированного анализа разнохарактерных и разноуровневых геоданных, полученных различными способами;

широкий спектр реализованных инструментов позволяет легко адаптировать технологическую схему в каждом конкретном случае, а также проводить междисциплинарные исследования;

ГИС РАПИД, ориентируясь на специалиста из области наук о Земле, использует понятную ему терминологию, и не требует универсальных знаний в области геоинформационных технологий, что обуславливает относительную легкость освоения и внедрения системы;

ГИС РАПИД не имеет аналогов в Украине и не уступает немногочисленным мировым аналогам. Это обеспечивает перспективу ее дальнейшего развития и совершенствования.

1. Шустин В. А., Фокина Л. А. Об использовании географических информационных систем для анализа и прогноза экологической ситуации // Вестн. ДВО РАН. – 2004. – № 1. – С. 45–52.
2. Thompson S., Hong Su, Bourdier J. GIS Applications in emergency response // Электр. ресурс. – Способ доступа: http://charlotte.utdallas.edu/mgis/ClassFiles/gisc6383/techassess_2004.
3. Іщуқ О. О., Середінін Є. С. Прогнозування та оцінка наслідків екстремальних поведених ситуацій засобами просторового аналізу ГІС // Вісн. геодезії та картографії. – 2000. – № 2. – С. 37–42.
4. Финкельштейн М. Я. ГИС-INTEGRO как инструмент геологических исследований // Геоинформатика. – 2002. – № 2. – С. 14–21.
5. Бодякин В. И. Разработка инструментального комплекса на базе нейросемантических структур для прогнозирования чрезвычайных ситуаций // Электр. ресурс. – Способ доступа: <http://www.ipu.ru/stran/bod/klimit/bezop.htm>.
6. Шахраманьян М., Ларионов В., Прошляков М. ГИС для прогнозирования чрезвычайных ситуаций // Компьютера. – 2001. – № 47. – С. 36–42.
7. Пархісенко Я. Прогнозування та оцінка наслідків надзвичайних паводкових ситуацій // Природ. камертон. Природа. Людина Суспільство. – 2002. – № 3. – С. 18–21. – [Вид. УІНСіР при Раді нац. безпеки і оборони України.].
8. Бусыгин Б. С., Мирошниченко Л. В. Распознавание образов при геолого-геофизическом прогнозировании. – Днепропетровск: Изд-во Днепр. ун-та, 1991. – 168 с.
9. Бусыгин Б. С., Никулин С. Л., Бойко В. А. Геоинформационная система РАПИД как средство мониторинга и прогнозирования чрезвычайных ситуаций // Сб. тр. 9 Междунар. конф. “Стихия-2006”. (Севастополь, сент., 2006). – Севастополь, 2006.
10. Бусыгин Б. С., Никулин С. Л., Бойко В. А. ГИС-технология поисков золота в Западном Узбекистане // Геоинформатика. – 2006. – № 1. – С. 44–49.
11. Бусыгин Б. С., Никулин С. Л., Залесский В. В. Компьютерная технология поисков золота на юго-западном склоне Украинского щита // Наука та інновації. – 2006. – № 2. – 10 с.
12. Кац Я. Г., Тевелев А. В., Полетаев А. И. Основы космической геологии. – Москва: Недра, 1988. – 430 с.
13. Бусыгин Б. С., Никулин С. Л., Залесский В. В. Прогнозирование золоторудной минерализации в районе Сорокинской зеленокаменной структуры // Наук. вісн. при Нац. гірн. ун-ті. – 2006. – № 3. – С. 68–72.
14. Кукушкин Д. А., Ян Г. Х. Некоторые вопросы методики анализа линеаментов (по данным дешифрирования космических снимков) // Исследование Земли из космоса. – 1983. – № 1. – С. 51–56.
15. Busygin B., Nikulin S. The integrated analysis of geological-geophysical and remote sensing data at the gold prospecting in Western Uzbekistan // 67-rd EAGE Conf.: Extended Abstracts. (Madrid, Spain, June 2005). – Madrid, 2005. – 1.

Національний горний університет,
Днепропетровск

Поступило в редакцию 25.01.2007