



УДК 528.087.4

© 2001 г. Т.Г. Решетнева, канд. геол.-мин. наук
(Амурский государственный университет, Благовещенск)

ПРОГНОЗИРОВАНИЕ В ГИС НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

Вопросы научного обоснования освоения природной среды становятся все более актуальными. Одним из основных методов их решения является прогнозирование на основе применения ГИС-технологий и статистических методов исследований. В статье дается краткая характеристика состояния проблемы геоинформационного прогнозирования и рассматривается концепция комплексного решения указанной задачи, основанная на использовании моделей временных рядов.

Сложившаяся за многие годы хозяйственная специализация отдельных регионов России, их своеобразие, уникальность природы и ресурсов определили сегодня большой спектр сочетаний экологических условий. На повестку дня встали задачи исследования особенностей экологических условий территорий с целью разработки конкретных мероприятий по снижению экологического риска от ряда факторов, повышению экологической безопасности, разработки системы мер по оздоровлению обстановки, экологической дифференциации территорий и т.д.

К числу основных методов решения перечисленных задач относятся научное предсказание и собственно прогнозирование, основанные на всестороннем комплексном исследовании разнообразных данных.

Очевидно, что создание общей концепции оценки использования природных ресурсов, которая позволила бы выработать единую систему показателей для прогноза самых разнообразных природообразующих мероприятий, невозможно без проведения геоинформационных исследований.

В настоящее время в ГИС активно развивается подсистема поддержки принятия решений, в которой с помощью определенных методов на основе созданной базы данных (базы знаний) осуществляется доступ к информации, необходимой для прогнозирования, а также ее анализ и представление в заданном направлении. Это осуществляется путем построения технологии, в которую входит набор специальных процедур и алгоритмов сбора, анализа и моделирования, а также визуализации данных.

Предсказание – это суждение о будущем состоянии процесса или явления, основанное на субъективном «взвешивании» большого числа фак-

торов качественного и количественного характера. *Прогнозирование* – исследовательский процесс, в результате которого получают прогноз о состоянии объекта. Прогноз является вероятностным суждением о возможном состоянии объекта или об альтернативных путях его достижения.

Геоинформационное прогнозирование представляет собой набор методов получения и анализа информации об окружающей среде и разработки прогнозных оценок для поддержки принятия решений [3]. Целью геоинформационного прогнозирования является снижение уровня неопределенности при принятии решений. Оно решает две основные задачи: оценка тех или иных геоинформационных параметров для данного момента времени и получение прогнозных оценок на перспективу. Оценка параметров применяется для оперативного анализа существующей ситуации, прогноз-ные оценки используются при выборе и обосновании принимаемых решений или при изучении явлений и их будущих последствий.

Геоинформационное прогнозирование как технология интегрирует статистические методы прогнозирования, методы деловой графики, методы геоинформационного и цифрового моделирования. При геоинформационном прогнозировании выделяются три основных этапа обработки информации:

- сбор, группировка, обобщение и унификация первичных данных;
- анализ, моделирование вторичных (унифицированных) данных;
- получение прогнозных оценок и их верификация.

Решение задач прогнозирования реализуется путем целенаправленных геоинформационных исследований, что определяет важность выбора цели исследования. Ключевым аспектом определения целей исследования является выявления специфических типов информации, необходимой при решении проблем управления или исследовании природных явлений. Исходя из этого, цель сбора данных может быть:

- *разведочная*, направленная на сбор предварительной информации, предназначенной для определения проблем, проверки гипотез и выбора моделей;
- *описательная (дескриптивная)*, направленная на получение описательных характеристик тех или иных аспектов геоинформационной среды;
- *казуальная*, направленная на выявление и определение причинно-следственных связей и обоснование гипотез;
- *рекомендательная*, направленная на разработку и принятие решений для различных территориальных уровней.

Характер целей геоинформационного исследования предопределяет выбор конкретных типов исследования, носящих те же названия, а именно: разведочный, описательный, казуальный, рекомендательный. На практике при проведении конкретного геоинформационного исследования используются все типы исследований, причем в любой последовательности.

При решении задач геоинформационного прогнозирования вся используемая информация подразделяется на два класса: первичная и вторичная. *Первичные данные* получают в результате сбора информации об объектах местности различными технологиями: полевые методы, картометрические методы, дистанционные методы и т.д. На основе этих данных, после их унификации, создают наборы первичных цифровых моделей – местности, объектов, явлений.

Вторичные данные – данные, полученные в результате обработки первичных данных или собранные ранее из внутренних или внешних источников для других целей. Это информация, хранимая в архивах, базах данных, других ГИС, а также распространяемая по сети Интернет. Кроме того, под вторичными данными понимают модели, построенные по результатам обобщения и преобразования первичных моделей. К таким моделям относятся динамические ряды, результаты классификации или кластерного анализа и т.д. Вторичные данные помогают выявить скрытые факторы, взаимосвязи, которые не были обнаружены в первичных данных.

После сбора данных необходимо их преобразовать к виду, отвечающему двум требованиям: удобному для обработки в геоинформационной системе; удобному для анализа при решении поставленной задачи прогнозирования. Выделяют следующие основные функции преобразования данных: обобщение, определение концепции, коммуникация (перевод результатов статистического анализа на понятный для пользователя язык), экстраполяция (определение степени соответствия полученных результатов всей совокупности) [4]. На основе этих данных создаются модели объектов прогнозирования, особенность которых заключается в том, что возникает необходимость исследования взаимодействия объекта прогнозирования с внешней средой.

Геоинформационное прогнозирование включает три типа задач, отличающих его от других видов прогнозирования:

- обработка статистических данных, основанная на широко применяемых методах математической статистики и теории вероятностей;
- задачи анализа изображений, для решения которых используются методы обработки изображений, распознавания образов, методы искусственного интеллекта и т.д. и требуется разработка и создание классификаторов, библиотек или баз данных;
- обработка последовательностей пространственных векторных или растровых объектов (например, фотоснимков или тематических карт, отображающих изменение объекта или явления с течением времени).

Геоинформационные данные носят временной характер. В том случае, когда прогноз необходим на определенный момент времени и связан с определением направления развития системы на перспективу, в качестве исходных данных целесообразно использовать временные ряды или временные модели.

При проведении периодически повторяемых наблюдений образуются наборы данных, характеризующиеся сбором информации об объектах земной поверхности через определенные периоды времени. В этом случае говорят о накоплении данных в виде временных рядов. Временные ряды представляют собой упорядоченные во времени наборы изменений каких-либо характеристик исследуемого объекта, процесса.

Временными рядами называют последовательности изменений

$$x(t_1), x(t_2), \dots, x(t_n),$$

упорядоченных в фиксированные моменты времени t_1, t_2, \dots, t_n [2].

В отличие от анализа случайных выборок анализ временных рядов основывается на предположении, что последовательные значения в наборе данных получены, как правило, через равные промежутки времени – периоды, которые могут быть различными: суточными, недельными, месячными, годовыми и т.д. По существу, временной ряд описывает непрерывное явление, протекающее во времени. Но информация об этом явлении поступает выборочно, вследствие чего реальный непрерывный процесс описывается в виде дискретных наборов данных.

Существуют три основные цели анализа временных рядов:

- построение формальной модели, позволяющей восстанавливать события в виде непрерывного процесса (переход от дискретной модели к непрерывной);
- установление механизма, определяющего динамику ряда (описание реального явления с помощью реально объясняющих и результативных переменных);
- прогнозирование явления или его анализ в ретроспективе (получение значений временного ряда по настоящим и прошлым значениям).

Эти цели требуют построения модели ряда и ее идентификации. Построение модели ряда означает, что ряд формально описан. При этом возможно формальное построение модели $\xi(t)$, основанное на подборе математических зависимостей между параметрами ряда, безотносительно к механизму их формирования

$$\xi(t) = \xi(x(t_1), x(t_2), \dots, x(t_n)).$$

Эта модель называется формальной, цель ее построения – связать наилучшим образом полученные дискретные данные.

Возможно неформальное построение модели, основанное на описании механизма z протекания реального явления с введением параметров, описывающих реальные характеристики a_i и b_j , влияющие на изменение процесса или явления.

$$F(a_i, t) = \xi(x(t_1), x(t_2), \dots, x(t_n), a_i, b_j), \quad i = 1, \dots, n; \quad j = 1, \dots, m.$$

Идентификация означает выбор в модели наблюдаемых b_j и вычисляемых a_i параметров, определение наблюдаемых параметров на основе экспериментальных данных. Как только модель ряда определена, с ее помощью можно интерпретировать рассматриваемые данные. Она позволяет,

не вникая в механизм формализации, экстраполировать в перспективу или в ретроспективу и тем самым проводить анализ будущих или прошедших параметров явления.

Анализ временных рядов предполагает, что данные содержат функциональную составляющую (обычно включающую несколько компонент) и случайный шум (ошибку), который затрудняет обнаружение регулярных компонент. Генезис наблюдений образует структура и классификация основных факторов, под воздействием которых формируются значения временного ряда. Выделяют четыре типа таких факторов [1].

Долговременные факторы, формирующие долговременную тенденцию в изменении анализируемого признака $x(t)$, которая обычно описывается с помощью той или иной неслучайной монотонной функции $f_{\text{ТР}}(t)$ – тренда.

Сезонные факторы, формирующие периодически повторяющиеся в определенное время года функциональные изменения анализируемого признака. Результат действия сезонных факторов описывают с помощью специальной неслучайной периодической (с периодами, кратными сезонам) функции $\varphi(t)$. Ее описание обусловлено содержательной сущностью задачи.

Конъюнктурные факторы, формирующие изменение анализируемого признака либо действием долговременных циклов, либо текущими конъюнктурными (разовыми) факторами. Результат действия циклических факторов обозначим с помощью неслучайной функции $Q(t)$.

Случайные (нерегулярные) факторы – факторы, не поддающиеся учету и регистрации. Их воздействие на формирование значений временного ряда обусловлено стохастической природой элементов $x(t)$. Результат воздействия случайных факторов обозначим с помощью функции случайных величин $\varepsilon(t)$. Случайные факторы могут быть разной природы. Они воздействуют на значения ряда как помехи, поэтому многие методы исследования временных рядов включают различные способы фильтрации, позволяющие выделить регулярную составляющую более отчетливо.

На практике не обязательно, чтобы в процессе формирования значений всякого временного ряда участвовали одновременно факторы всех четырех типов. Большинство функциональных составляющих временных рядов принадлежит к двум классам: они являются либо трендом, либо сезонной составляющей. *Тренд* представляет собой общую систематическую линейную или нелинейную компоненту, которая может изменяться во времени. *Сезонная составляющая* – периодически повторяющаяся компонента. Оба эти вида регулярных компонент часто присутствуют в ряде одновременно.

Исходный набор данных временного ряда всегда содержит погрешности. Поэтому одной из важных задач является устранение погрешностей для получения более точной информации об исследуемых объектах. В свя-

зи с этим решают задачу сглаживания и связанную с ней задачу построения ряда.

Сглаживание – процедура выделения случайной составляющей, т.е. способ построения нового ряда с измененными значениями наблюдений, в котором каждый член содержит какие-то поправки, уменьшающие эту случайную составляющую. Сглаживание всегда включает некоторый способ локального усреднения данных, при котором случайные компоненты взаимно погашают друг друга или фильтруются.

Различают два вида сглаживания [2]: алгоритмическое и аналитическое. *Алгоритмическое сглаживание* производится, если общий аналитический вид искомой функции $f_{\text{ТР}}(t) + \dots + \varphi(t)$, несущей информацию о временном процессе, не известен исследователю. Предполагается, что помеха в этом случае лежит в высокочастотной части спектра, а полезный сигнал – в низкочастотной. Не вникая в механизм образования ряда, осуществляют его сглаживание на основе фильтрации. В этом случае производится формальное построение модели $\zeta(t)$, основанное на подборе математических зависимостей между параметрами ряда, безотносительно к механизму их формирования.

Аналитическое сглаживание осуществляют, если известен механизм формирования членов ряда. Оно заключается в обработке, использующей аналитические описания, отражающие с меньшей или большей полнотой механизм явления. В этом случае построение ряда заключается в создании описания на основе дискретных данных, а именно: нахождение всех функций

$$F(t) = f_{\text{ТР}}(t) + \dots + \varphi(t) + Q(t) + \varepsilon(t).$$

Самым общим методом сглаживания является *скользящее среднее*, в котором каждый член нового ряда заменяется простым или взвешенным средним n соседних членов ряда, где n – ширина «окна». Вместо среднего можно использовать медиану значений, попавших в окно. Основное преимущество *медианного* сглаживания в сравнении со сглаживанием скользящим средним состоит в том, что результаты становятся более устойчивыми к выбросам внутри окна. Недостаток медианного сглаживания в том, что при отсутствии явных выбросов оно приводит к более «зубчатым» кривым, чем сглаживание скользящим средним, и не позволяет использовать веса.

Когда ошибка измерения очень большая, лучше использовать метод наименьших квадратов, взвешенных относительно расстояния, или метод экспоненциального взвешенного сглаживания.

Все эти методы отфильтровывают шум и преобразуют измеренные исходные данные в данные, расположенные на относительно гладкой кривой.

Отрезок времени Δt от момента времени, для которого имеются последние статистические данные об изучаемом процессе, до момента, к ко-

торому относится прогноз, называется *периодом упреждения*, в зависимости от его длительности различают три вида прогноза [4]:

- краткосрочные – с периодом упреждения от нескольких дней до 3 лет;
- среднесрочные – от 3 до 5 лет;
- долгосрочные – от 5 лет и длительнее.

Каждый из методов прогнозирования обладает определенными достоинствами и недостатками. Их применение более эффективно в краткосрочном прогнозировании, поскольку они упрощают реальные процессы. Степень упрощения и период «краткосрочности» определяется изученностью процессов и явлений, объемом накопленных знаний и данных об исследуемых объектах. Это неизбежно приводит к необходимости создания для целей прогнозирования экспертных систем или накопления информации для их создания. Экспертные знания и оценки являются важными инструментами восприятия и анализа реальности, дополняя количественные подходы качественными оценками и моделями.

Получение прогнозных оценок должно быть подвергнуто верификации для их надежности. Цель верификации прогноза – оценка его функциональной полноты, точности и достоверности. Применяют различные виды верификации (прямая и косвенная, инверсная и консеквентная и др.).

После выполнения верификации прогноза необходимо оценить его качество. Для этой цели используются следующие показатели прогноза: полнота, точность, достоверность, ошибка, источник ошибки.

При выборе метода прогнозирования необходимо учитывать, для какого типа управления предполагается использовать прогноз. Это объясняется тем, что различные типы управления предъявляют разные требования к виду результатов и точности прогнозирования.

Результатом геоинформационного прогнозирования являются не только статистические данные (обычный прогноз), но и тематические карты, цифровые модели динамики явлений, трехмерные модели объектов, векторные изображения полей или динамики явлений и т.д., реализованные в ГИС.

В заключение следует отметить, что изложенный в статье материал дает лишь общее представление о методах геоинформационного прогнозирования, основанных на использовании моделей временных рядов, и практически не затрагивает многих важных методических вопросов. К их числу следует отнести приемы построения, обоснования и редактирования модели, обоснование граничных условий и количественных характеристик – параметров прогноза, технической и программной реализации, возможности интерпретации результатов. Все эти вопросы автор предполагает рассмотреть в последующих публикациях, посвященных методике геоинформационного прогнозирования.



ЛИТЕРАТУРА

1. Бугаевский Л.М., Цветков В.Я. Геоинформационные системы: Учебное пособие для вузов. М.:Златоуст, 2000.
2. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. Т. 2: Айвозян С.А. Основы эконометрики. М.: ЮНИТИ-ДАНА, 2001.
3. Савиных В.П., Цветков В.Я. Геоинформационный анализ данных дистанционного зондирования. М.: Картгеоцентр-Геодезиздат, 2001.
4. Цветков В.Я. Геоинформационные системы и технологии. М.: Финансы и статистика, 1998.

Статья представлена к публикации членом редколлегии А.Д. Плутенко.

УДК 528.087.4

© 2001 г. **В.Я. Цветков**, д-р техн. наук
(Московский государственный университет геодезии и картографии)

ГЕОИНФОРМАТИКА И СИНЕРГЕТИКА

Рассматриваются основы применения методов синергетики для анализа эффективности геоинформационных технологий и систем еще на уровне их проектирования и конструирования, что повысит надежность и качество геоинформационных исследований и геоинформационных решений.

В последние годы наблюдается рост интереса к междисциплинарному направлению, получившему название "синергетика" [11, 14].

Системы, составляющие предмет изучения синергетики, могут быть самыми различными и изучаться различными науками: социологией, лингвистикой, физикой, химией, биологией, математикой, нейрофизиологией, экономикой. Каждая из наук изучает системы своими методами и формулирует результаты на "своем" языке.

В отличие от специальных наук синергетика исследует общие закономерности развития систем разной природы. Абстрагируясь от специфики систем, синергетика описывает их эволюцию обобщенно, устанавливая связь явлений, изучаемых специфическими средствами различных наук.

Выявление общих свойств моделей разных систем на основе синергетического подхода позволяет использовать достижения одной области науки в других.