

О.А. Логинов, Л.В. Сарычева

Прогнозирование уровня грунтовых вод с применением клеточных автоматов

Предложена методика анализа инвестиционной привлекательности регионов на основе статистических показателей их эколого-социально-экономического мониторинга. Методика учитывает неравномерность распределения показателей и влияние соседства регионов. Приведены оценки инвестиционной привлекательности регионов Украины (2000–2003) и Европы (2004–2007).

The technique of the regions investment appeal analysis on the basis of regions ecological-social-economical monitoring statistics is suggested. The technique considers the non-uniformity of indicators distribution and the influence of the regions neighbourhood. The estimations of investment appeal of Ukraine regions (2000–2003) and Europe (2004–2007) are given.

Запропоновано методику аналізу інвестиційної привабливості регіонів на основі статистичних показників їх еколого-соціально-економічного моніторингу. Методика враховує нерівномірність розподілу показників і вплив сусідства регіонів. Наведено оцінки інвестиційної привабливості регіонів України (2000–2003) і Європи (2004–2007).

Введение. Постановка проблемы. Грунтовые воды – подземные воды первого от поверхности земли постоянного водоносного горизонта, не имеющего сплошной кровли водонепроницаемых пород, не обладают напором и подвержены сезонным колебаниям уровня и дебита [1].

Контроль за состоянием и изменением гидрогеологической обстановки осуществляется на основе стационарных наблюдений за режимом подземных вод. Особое внимание уделяется грунтовым водам, так как они более всего подвержены антропогенному воздействию. Поступающие от наблюдений данные обрабатываются с получением среднемесячных, среднегодовых и среднемноголетних значений глубины залегания уровня, сроков наступления и характеристик его экстремальных значений, сезонных и многолетних амплитуд колебаний и т.д.

Моделирование и прогнозирование уровня грунтовых вод (УГВ) представляет собой важную народнохозяйственную задачу. Для получения количественного прогноза используются различные методы, в том числе классические методы прогнозирования временных рядов, математическое моделирование динамики гидрогеологических процессов.

Большая часть подходов к прогнозированию УГВ базируется на предположении, что наблюдения, составляющие исходный временной ряд, точны, независимы и подчиняются нормальному закону.

В настоящее время при прогнозировании УГВ ни в одном из методов не принимается во внимание хаотическая природа процесса временных изменений УГВ. В то же время большинство естественных явлений следуют «смещенному случайному блужданию», т.е. тренду с шумом, что показал британский статистик-гидролог Херст, исследуя статистику объемов стоков рек [2].

Временные ряды (ВР) наблюдений УГВ обладают долговременной памятью – способностью к сохранению, накоплению и воспроизведению информации, характеризующейся практически неограниченным временем хранения вследствие систематического повторения семантически закодированной информации, приводящей к установлению ассоциативных связей между элементами, по которым отдельная информация может быть восстановлена. Это дает возможность применения клеточных автоматов к прогнозированию значений УГВ. В настоящее время подобных исследований, судя по публикациям, не проведено. В то же время появились работы, излагающие результаты успешного использования теории клеточных автоматов в экономике [2].

Клеточные автоматы (КА) – дискретные динамические системы, поведение которых полностью определяется в терминах локальных зависимостей. В значительной степени так же обстоит дело для большого класса непрерывных динамических систем, определенных уравнениями в частных производных. В этом смы-

сле клеточные автоматы в информатике являются аналогом физического понятия «поля» [3]. Целесообразность применения нечетких множеств [4] и клеточных автоматов объясняется погрешностями в исходных данных и сезонными колебаниями значений УГВ.

Цель работы – разработка методики и соответствующей информационной технологии прогнозирования УГВ с применением клеточных автоматов и нечетких множеств, проверка их адекватности на реальных мониторинговых данных Днепропетровской области.

Методика прогнозирования уровня грунтовых вод с использованием клеточных автоматов и нечетких множеств

Концептуальная схема разработанной методики прогнозирования УГВ и соответствующей информационной технологии приведена на рис. 1.

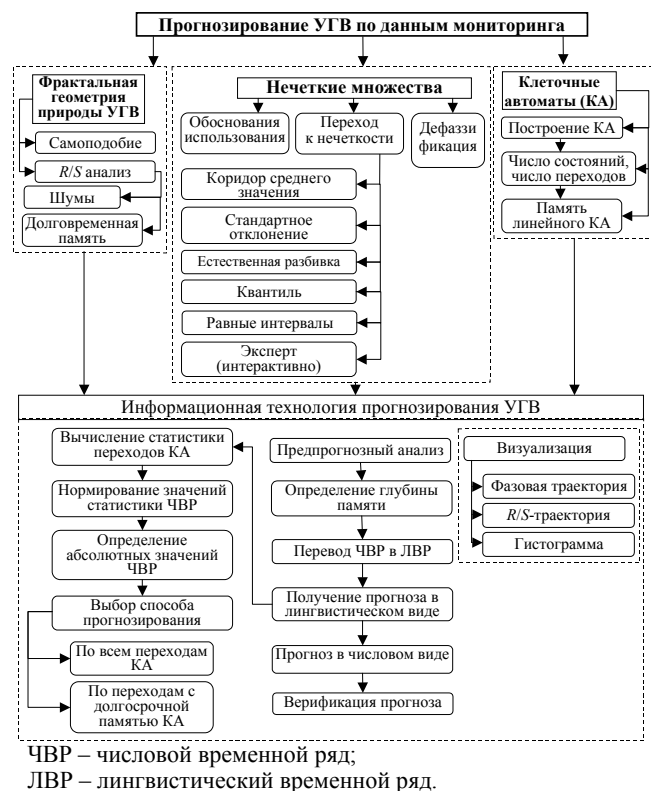


рис. 1. Концептуальная схема информационной технологии прогнозирования УГВ

Рассмотрим основные этапы получения прогнозного решения.

Этап 1. Предпрогнозный анализ наличия долговременной памяти ВР.

Инструментарий предпрогнозного анализа – фрактальный анализ ВР. Знание фрактальных характеристик ВР позволяет оценить перспективность надежного прогнозирования значений ВР с помощью прогнозной КА-модели.

Устойчивость тренда и уровень шума оцениваются тем, как изменяется с течением времени нормированный размах R/S (R – размах, S – стандартное отклонение) ВР или насколько показатель Херста превосходит значение 0,5 [2]. Если уровни ВР отражают «чисто случайный процесс» (являются независимыми случайными величинами), то значения показателя Херста $H \approx 0,5$. Если показатель Херста $H \geq 2/3$, то ВР обладает долговременной памятью.

Алгоритм R/S -анализа ВР следующий.

1. В исходном ВР $Z = \{z_1, z_2, \dots, z_n\}$ последовательно выделяем его начальные временные отрезки $Z_\tau = \{z_1, z_2, \dots, z_\tau\}$, $\tau = 3, 4, \dots, n$, и вычисляем текущие средние

$$\bar{Z}_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} z_i, \tau = 3, 4, \dots, n \quad (1)$$

2. Для каждого Z_τ , $\tau = 3, 4, \dots, n$, определяем накопленное отклонение для его отрезков длины t :

$$X_{\tau,t} = \sum_{i=1}^t (z_i - \bar{z}_\tau), t = 1, 2, \dots, \tau. \quad (2)$$

3. Вычисляем разность между максимальным и минимальным накопленными отклонениями – размах:

$$R = R(\tau) = \max_{1 \leq t \leq \tau} (X_{\tau,t}) - \min_{1 \leq t \leq \tau} (X_{\tau,t}). \quad (3)$$

4. Нормируем размах на стандартное отклонение $S = S(\tau)$ для отрезка ВР Z_τ , $3 \leq \tau < n$, получаем величину R/S .

5. Показатель Херста $H = H(\tau)$, характеризующий фрактальную размерность рассматриваемого ВР, получаем из соотношения

$$R/S = (a \cdot \tau)^H, \quad (4)$$

логарифмируя его и полагая $a = 1/2$ [2].

Определяем последовательность декартовых координат (x_τ, y_τ) точек H -траектории:

$$x_\tau = \tau, y_\tau = H(\tau) = \frac{\log(R(\tau)/S(\tau))}{\log(\tau/2)},$$

$$\tau = 3, 4, \dots, n. \quad (5)$$

R/S -траектория представляется в логарифмических координатах последовательностью точек $\{(\log(\tau/2), \log(R(\tau)/S(\tau)))\}$.

6. Анализируем цвет шума, соответствующий показателю Херста H , используя известную его трактовку:

- $H \in (0; 0,2)$ – коричневый шум (полная неопределенность в отношении прогнозируемости ВР);
- $H \in [0,2; 0,4)$ – розовый шум (антиперсистентность [2] соответствующего отрезка ВР);
- $H \in [0,4; 0,6]$ – белый шум (наименьшая прогнозируемость);
- $H \in (0,6; 0,6(6))$ – серый шум (область нечеткого разграничения между черным и белым шумом);
- $H \in [0,6(6); 1]$ – черный шум (чем больше значение H , тем большая трендоустойчивость присуща соответствующему отрезку ВР, тем лучшая прогнозируемость).

Этап 2. Определение глубины памяти ВР.

Основанием для утверждения, что некоторый ВР обладает долговременной памятью, служит выполнение следующих условий [2]:

- H -траектория ВР оказывается в области черного шума через несколько своих начальных точек,
- для R/S -траектории ВР указанные точки вхождения в черный шум демонстрируют наличие тренда.

Глубину памяти определяет тот по порядку (в области черного шума) номер $q = \tau$, для которого выполняется следующие условия:

- H -траектория в q -й точке получает отрицательное приращение,
- R/S -траектория в этой точке демонстрирует так называемый «срыв с тренда» [2], т.е. резкое изменение тренда предшествующих точек траектории.

Этап 3. Преобразование числового ВР в лингвистический ВР.

Преобразование числового ВР в лингвистический временной ряд (ЛВР) с целью создания базиса памяти КА означает замену числовых

значений лингвистическими переменными – термами. Совокупность термов образует термножество U . Пусть множество U состоит из k элементов (например, $U = \{H, C, B\}$, H – низкий, C – средний, B – высокий УГВ для $k = 3$).

Алгоритмы преобразования числового ВР в ЛВР основаны на классификации элементов ряда Z на k классов (число классов равно числу элементов термножества). Заменяя элементы ряда Z соответствующими термами из U , получаем ЛВР. Предлагаются такие алгоритмы [5]:

1. Коридор среднего значения

- вычисляется \bar{Z} – среднее значение ВР;
- задаются граничные значения интервалов для терма C (средний);
- элементу ВР присваивается терм, в соответствии с тем, выше (терм B) или ниже (терм H) его значение, чем значение из интервала для терма C .

2. Стандартное отклонение

- вычисляются среднее \bar{Z} и стандартное отклонение s ВР;
- разбиение на интервалы начинается от \bar{Z} последовательным прибавлением и вычитанием m -й доли ($0 < m \leq 1$) стандартного отклонения s :

$$[\bar{Z} - m \cdot s; \bar{Z} + m \cdot s]; [\bar{Z} - 2m \cdot s; \bar{Z} - m \cdot s], [\bar{Z} + m \cdot s; \bar{Z} + 2m \cdot s]; \dots$$

- каждому элементу ВР, значение которого попало в интервал $[\bar{Z} - m \cdot s; \bar{Z} + m \cdot s]$, присваивается терм C (средний), в интервалы $[\bar{Z} - 2m \cdot s; \bar{Z} - m \cdot s]$, $[\bar{Z} + m \cdot s; \bar{Z} + 2m \cdot s)$ – соответствующие значения H, B .

3. Естественная разбивка (аналог кластеризации)

- ВР представляется в виде гистограммы;
- разбивка на k классов проводится в соответствии с резкими скачками в значениях частот;
- каждому элементу ВР присваивается терм, соответствующий номеру класса.

4. Квантиль

- для ВР подсчитываются относительные частоты v_i для гистограммы с числом интервалов, большим мощности терм-множества;

- находятся квантили Z_{p_i} , где $p_i = i/k$, $i = 1, 2, \dots, k-1$;

- промежуток $[Z_{\min}, Z_{\max}]$ (Z_{\min}, Z_{\max} – соответственно минимальное и максимальное значения ВР) разбивается точками Z_{p_i} , $i = 1, 2, \dots, k-1$;

- элементу ВР, значение которого попадает в i -й интервал, присваивается соответствующий i -й терм, $i = 1, 2, \dots, k$.

5. Равные интервалы

- промежуток $[Z_{\min}, Z_{\max}]$ делится на k интервалов равной длины;

- элементу ВР присваивается терм, соответствующий номеру интервала, в который попадает его значение.

Этап 4. Частотный анализ памяти ЛВР. Формирование памяти КА.

Наличие долговременной памяти у ЛВР представляется в понятиях КА.

Обозначим: $M(U)$ – множество всех q -конфигураций, обнаруженных в ЛВР, Mq – подмножество всех q -конфигураций в ЛВР с фиксированным q . При переходе конфигурации в одно состояние несколько раз, считается, что эта конфигурация обладает памятью. Находим частоту переходов q -конфигураций.

Оперативная память КА строится в виде множества двудольных ориентированных графов. В каждом из них вершины правой доли представляют собой элементы терм-множества U и взаимно однозначно соответствуют вершинам левой доли, представляющим фиксированные q -конфигурации ($q = 1, 2, \dots, Q$, где Q – глубина памяти ЛВР).

Дугам графов приписываются веса – частоты переходов заданной конфигурации в соответствующие состояния из U .

Этап 5. Получение лингвистических прогнозных значений. Верификация модели.

Алгоритм прогнозирования основывается на подсчете частоты переходов в одно из состояний (например, H, C, B) всех q -конфигу-

раций. Прогноз представляется в виде нечеткого терм-множества [6]

$$U_{n+1}^0 = \{(u_1; \mu_1), (u_2; \mu_2), \dots, (u_k; \mu_k)\}, \quad (7)$$

где значения функции принадлежности удовлетворяют равенству:

$$\sum_{i=1}^k \mu_i = 1.$$

Вычисляются частоты переходов q -конфигураций. Когда встретится q -конфигурация, которая демонстрирует наличие памяти (единичное значение частоты для терма), тогда переходим к вычислению искомым μ_i , $i = 1, 2, \dots, k$. Сначала вычисляются ненормированные значения, их сумма, затем полученные значения частот нормируются, определяется функция принадлежности терм-множества (7). Прогнозирование возможно только при наличии долговременной памяти в клеточном автомате.

Верификация осуществляется путем проверки модели на том же временном ряду, но прогнозируя его последнее известное значение.

Этап 6. Получение числового прогноза и оценка его точности.

В качестве числовых прогнозных значений по полученному лингвистическому значению выбираем середину соответствующего интервала, по которому преобразовывали числовой временной ряд в ЛВР.

Уровень точности прогноза вычисляется следующим образом: начиная с длины q (например, $q = 3$) долговременной памяти последовательно прогнозируется $(q + 1)$ -й, $(q + 2)$ -й, и т.д. известный элемент ВР (рис. 2).

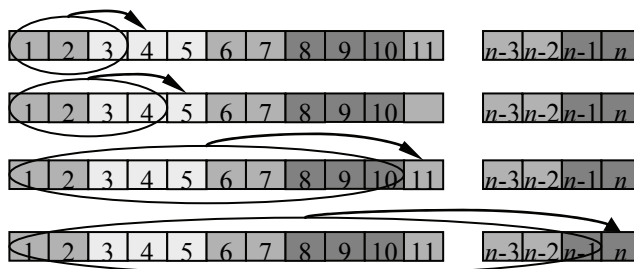


Рис. 2. Схема перебора к вычислению уровня точности прогноза

Находятся абсолютные ошибки, определяется их арифметическое среднее, которое дает точность прогноза.

Информационная технология прогнозирования уровня грунтовых вод

Предложенная методика положена в основу информационной технологии прогнозирования, для изложения которой используем данные по одной скважине за три года 1988–1990 гг. (строка «значение УГВ» в табл. 1).

Шаг 1. *Предпрогнозный анализ.* На основе ВР значений УГВ строятся *H*-траектория (рис. 3) и *R/S*-траектория (рис. 4), анализ которых говорит о наличии долговременной памяти ВР УГВ.

Шаг 2. Анализ *H*- и *R/S*-траектории дает глубину памяти ВР $q = 3$.

Шаг 3. Преобразование числового временного ряда в ЛВР с помощью алгоритма «коридор среднего значения» дает значения ЛВР, представленные в табл. 1 (строка «Лингвистическая оценка»).

Шаг 4. Частотный анализ памяти лингвистического временного ряда:

$$M_1 = \{H, C, B\};$$

$$M_2 = \{HH, HC, CH, CC, CB, BC, BB\};$$

$$M_3 = \{HHH, HHC, HCC, CHH, CCH, CCC, CCB, CBC, CBV, VCC, BBB\};$$

$$M_4 = \{HHHH, HHHC, HCCC, CHHH, CCHH, CCCC, CCCC, CCBV, CBCC, CBVV, BVCC, BBBV\}.$$

Таблица 1. Исходные данные УГВ и их лингвистическая оценка

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
1988 год												
Значение УГВ (по модулю)	2,16	2,16	2,07	1,93	1,96	2,01	2,01	2,13	2,26	2,28	2,28	2,28
Лингвистическая оценка	C	C	H	H	H	H	H	C	C	C	C	C
1989 год												
Значение УГВ (по модулю)	1,97	1,84	1,8	1,83	1,92	2,09	2,3	2,53	2,65	2,64	2,63	2,63
Лингвистическая оценка	H	H	H	H	H	C	C	C	B	C	C	C
1990 год												
Значение УГВ (по модулю)	2,65	2,64	2,59	2,6	2,62	2,66	2,65	2,82	2,92	2,94	2,89	2,84
Лингвистическая оценка	B	C	C	C	C	B	B	B	B	B	B	B

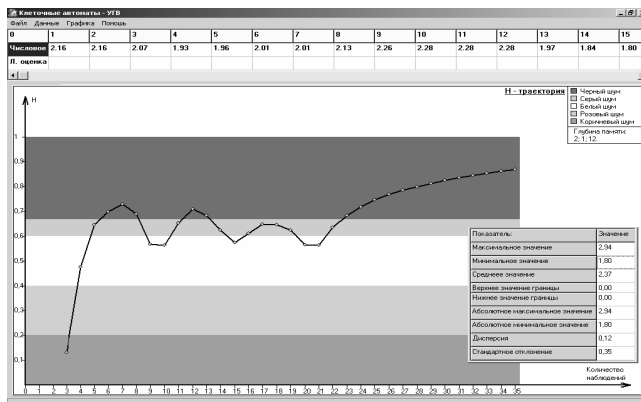


Рис. 3. *H*-траектория ВР наблюдений УГВ в 1988–1990 гг.

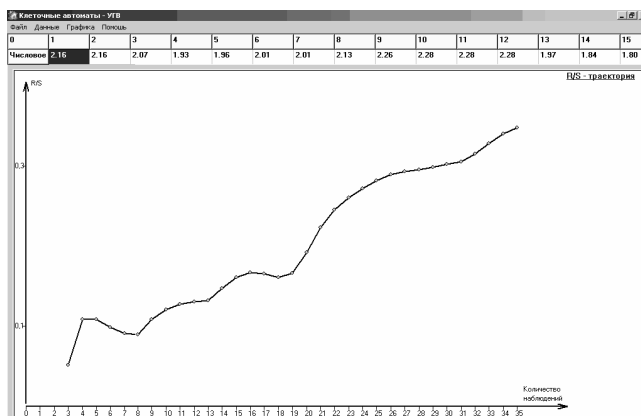


Рис. 4. *R/S*-траектория ВР наблюдений УГВ в 1988–1990 гг.

Для прогнозирования анализируем частоту переходов из q -конфигураций (q – число последовательных позиций в ЛВР) в соответствующие значения терм-множества (рис. 5). Например, наличие долговременной памяти в $q = 2$ конфигурации $W_2(BB \rightarrow B) = 1$.

Шаг 5. Получение лингвистических прогнозных значений.

Находим отношение числа переходов из данной конфигурации в одно из значений терм-множества M_1 к общему числу переходов:

$$W_1(B \rightarrow H) = \frac{0}{8}; W_1(B \rightarrow C) = \frac{2}{8};$$

$$W_1(B \rightarrow B) = \frac{6}{8};$$

$$W_2(BB \rightarrow H) = \frac{0}{5}; W_2(BB \rightarrow C) = \frac{0}{5};$$

$$W_2(BB \rightarrow B) = \frac{5}{5};$$

$$W_3(BBB \rightarrow H) = \frac{0}{4}; W_3(BBB \rightarrow C) = \frac{0}{4};$$

$$W_3(BBB \rightarrow B) = \frac{4}{4}.$$

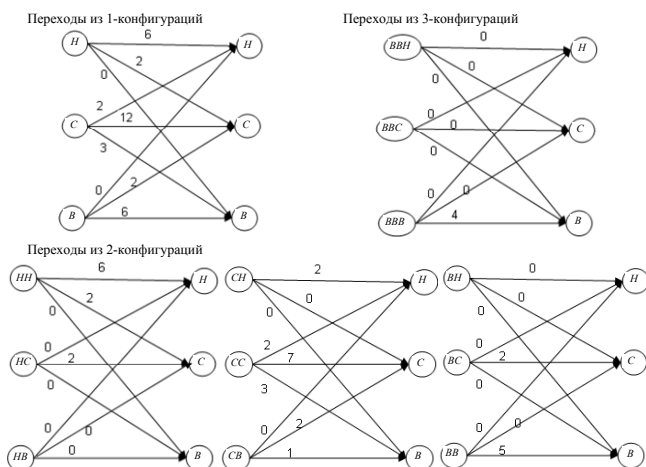


Рис. 5. Переходы из q -конфигураций ($q = 1, 2, 3$)

$$\mu'_H = 0 + 0 + 0 = 0; \mu'_C = \frac{2}{8} + 0 + 0 = 0,25;$$

$$\mu'_B = \frac{6}{8} + 1 + 1 = 2,75; \delta = \mu'_H + \mu'_C + \mu'_B = 3.$$

Нормируя μ'_H, μ'_C, μ'_B на δ , определяем:

$$\mu_H^0 = 0; \mu_C^0 = 0,08; \mu_B^0 = 0,92.$$

В результате получаем:

$$U_{n+1}^0 = \{(H; 0), (C; 0,08), (B; 0,92)\}.$$

УГВ ожидается высокий со степенью доверия $\mu_B^0 = 0,92$ или средний со степенью доверия $\mu_C^0 = 0,08$.

Верификация модели:

$$W_1(B \rightarrow H) = \frac{0}{7}; W_1(B \rightarrow C) = \frac{2}{7}; W_1(B \rightarrow B) = \frac{5}{7};$$

$$W_2(BB \rightarrow H) = \frac{0}{4}; W_2(BB \rightarrow C) = \frac{0}{4};$$

$$W_2(BB \rightarrow B) = \frac{4}{4}; W_3(BBB \rightarrow H) = \frac{0}{3};$$

$$W_3(BBB \rightarrow C) = \frac{0}{3}; W_3(BBB \rightarrow B) = \frac{3}{3}.$$

$$\mu'_H = 0 + 0 + 0 = 0; \mu'_C = \frac{2}{7} + 0 + 0 = 0,29;$$

$$\mu'_B = \frac{5}{7} + 1 + 1 = 2,71.$$

Нормируем: $\delta = \mu'_H + \mu'_C + \mu'_B = 3; \mu_H^0 = 0;$
 $\mu_C^0 = 0,1; \mu_B^0 = 0,9.$

В результате получаем:

$$U_{n+1}^0 = \{(H; 0), (C; 0,1), (B; 0,9)\}.$$

Уровень грунтовых вод в декабре 1990 г. ожидается высокий (со степенью доверия $\mu_B^0 = 0,9$), что совпадает с реальными данными.

Шаг 6. Получение числового прогноза.

Лингвистическим переменным ставим в соответствие числовые значения:

$$H \text{ (низкий)} = \frac{1,8 + 2,082}{2} = 1,941; \quad C \text{ (средний)} =$$

$$= 2,366; \quad B \text{ (высокий)} = \frac{2,65 + 2,94}{2} = 2,795.$$

Прогнозное значение (по модулю) в нечетком виде:

$$Z_{n+1}^0 = \{(1,941; 0), (2,366; 0,08), (2,795; 0,92)\},$$

и обычном виде:

$$Z_{n+1}^0 = 1,941 \cdot 0 + 2,366 \cdot 0,08 + 2,795 \cdot 0,92 = 2,76.$$

Таким образом, прогнозный УГВ на январь 1991 г. равен $-2,76$ м и совпадает с мониторинговым значением.

Результаты

Разработанная информационная технология прогнозирования УГВ программно реализована. Проведено экспериментальное сравнение прогнозов, полученных разными методами, с наблюдаемыми мониторинговыми данными.

Исходные данные для экспериментов – ряды среднемесячных наблюдений УГВ за период 2002–2004 гг. по скважинам Днепропетровской области. Алгоритм перехода от ВР к ЛВР – «стандартное отклонение». Прогноз составлен по ряду из 36 значений (месяцев) на январь 2005 г. Результаты сравнения прогнозов методами «КА» и «Гусеница» представлены в табл. 2 (для ВР скважин № 10, 18, 19 наличие долговременной памяти не обнаружено).

Заключение. Разработана и программно реализована новая информационная технология прогнозирования УГВ. Экспериментальные результаты ее применения для большинства скважин наблюдений УГВ, расположенных в

Таблица 2. Результаты сравнения прогнозов методами «КА» и «Гусеница»

№ скв.	КА				ГУСЕНИЦА			Реальное значение, м
	Прогноз	Точность прогноза, м	Абсолютная ошибка, м	Относительная ошибка, %	Прогноз	Абсолютная ошибка, м	Относительная ошибка, %	
1	2,76	0,10	0,15	5,75	2,89	0,28	10,73	2,61
2	4,12	0,08	0,11	2,74	4,25	0,24	5,99	4,01
3	4,87	0,06	0,03	0,62	4,97	0,13	2,69	4,84
4	4,63	0,23	0,55	10,62	4,03	1,15	22,20	5,18
5	9,89	0,00	0,01	0,10	9,88	0,00	0,00	9,88
6	1,31	0,16	0,04	2,96	1,47	0,12	8,89	1,35
7	2,74	0,07	0,13	4,53	2,83	0,04	1,39	2,87
8	6,98	0,04	0,10	1,45	6,50	0,38	5,52	6,88
9	2,63	0,04	0,03	1,13	2,55	0,11	4,14	2,66
10	1,61	0,21	0,24	17,52	1,55	0,18	13,14	1,37
11	2,06	0,10	0,24	13,19	1,82	0,00	0,00	1,82
12	2,85	0,08	0,04	1,42	2,92	0,11	3,91	2,81
13	3,88	0,05	0,18	4,86	3,84	0,14	3,78	3,70
14	6,03	0,03	0,16	2,73	5,95	0,08	1,36	5,87
15	3,97	0,05	0,11	2,70	3,91	0,17	4,17	4,08
16	4,93	0,02	0,28	5,37	4,83	0,38	7,29	5,21
17	5,06	0,03	0,24	4,98	4,91	0,09	1,87	4,82
18	3,71	0,10	0,51	15,94	3,89	0,69	21,56	3,20
19	1,30	0,22	0,23	21,50	1,09	0,02	1,87	1,07
20	2,12	0,03	0,03	1,40	2,16	0,01	0,47	2,15
Среднее:	0,09	0,17	6,07		0,22	6,05		

Днепропетровской области, позволяют сделать прогноз УГВ с удовлетворительной точностью, во многих случаях более достоверный,

чем полученный другими методами прогнозирования.

1. Рубан С.А., Шинкаревский М.А. Гідрогеологічна оцінка і прогнози режиму підземних вод України. – Дніпропетровськ: Укрпівденгеологія. – 2005. – 372 с.

2. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. – М.: Мир, 2000. – 333 с.

3. Нейман Дж. Теория самовоспроизводящихся автоматов. – М.: Мир, 1971. – 378 с.

4. Борисов А.Н., Алексеев А.В., Меркурьева Г.В. Обработка нечеткой информации в системах принятия решений. – М.: Радио и связь, 1989. – 304 с.

5. Сарычева Л.В. Компьютерный эколого-социально-экономический мониторинг регионов. Математическое обеспечение: Монография. – Днепропетровск: НГУ, 2003. – 222 с.

6. Алтунин А.Е., Семухин М.В. Модели и алгоритмы принятия решений в нечетких условиях. – Тюмень: Изд-во ТюмГУ, 2000. – 352 с.

© Л.В. Сарычева, О.А. Логинов, 2009

Внимание !

Оформление подписки для желающих опубликовать статьи в нашем журнале обязательно.

В розничную продажу журнал не поступает.

Подписной индекс 71008