

УДК 519.711.3

UDC 519.711.3

**ТЕОРИЯ НЕЧЕТКИХ МНОЖЕСТВ И
КЛЕТОЧНЫХ АВТОМАТОВ КАК
ИНСТРУМЕНТАРИЙ ПРОГНОЗА И
АДЕКВАТНОГО ОТРАЖЕНИЯ
СТОХАСТИЧЕСКОЙ ПРИРОДЫ
ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ**

**THE THEORY OF INDISTINCT SETS AND
CELLULAR AUTOMATIC MACHINES AS
TOOLKIT OF FORECASTING AND
ADEQUATE REFLECTION OF THE
STOCHASTIC NATURE OF ECONOMIC
PROCESSES**

Попова Елена Витальевна
д.э.н., к.ф.-м.н., профессор

Popova Elena Vitalevna
Dr.Sci.Econ., Cand.Phys.-Math.Sci., professor

Позднышева Наталья Олеговна
ассистент кафедры экономической кибернетики

Pozdnisheva Natalia Olegovna
assistant of economic cybernetics sub department

Савинская Дина Николаевна
ассистент кафедры информационных систем
*Кубанский государственный аграрный
университет, Краснодар, Россия*

Savinskaya Dina Nikolaevna
assistant of information systems sub department
Kuban state agrarian university, Krasnodar, Russia

Кумратова Альфира Менлигуловна
к. э. н., доцент
*Карачаево-Черкесская государственная
технологическая академия, Черкесск, Россия*

Kymratova Alfira Menligulovna
Cand.Econ.Sci., associate professor
*Karachaevo-Circassian state technological academy,
Cherkessk, Russia*

Терехов Алексей Геннадьевич
ассистент кафедры информационных систем
*Кубанский государственный аграрный
университет, Краснодар, Россия*

Terekhov Aleksey Gennadievich
assistant of information systems sub department
Kuban state agrarian university, Krasnodar, Russia

В статье представлена прогнозная модель, базирующаяся на теории клеточных автоматов и математическом аппарате нечетких множеств. Продемонстрирована её работа на реальных данных временного ряда урожайностей сахарной свеклы по Мостовскому району Краснодарского края

In the article the forecasting model which is based on the theory of cellular automatic machines and mathematical apparatus of indistinct sets is presented. Its work on the real data of time number productivities of sugar beet in Mostovskoy area of Krasnodar territory is shown

Ключевые слова: КЛЕТОЧНЫЙ АВТОМАТ, ТЕОРИЯ НЕЧЕТКИХ МНОЖЕСТВ, ВРЕМЕННОЙ РЯД, ЛИНГВИСТИЧЕСКИЕ ПЕРЕМЕННЫЕ, ТЕРМ-МНОЖЕСТВО, ВАЛИДАЦИЯ И ВЕРИФИКАЦИЯ

Keywords: CELLULAR AUTOMATIC MACHINE, THEORY OF INDISTINCT SETS, TIME NUMBER, LINGUISTIC VARIABLES, TERM-SET, VALIDATION AND VERIFICATION

Практика показала, что в современных условиях, к примеру, для российской экономики с её нестабильностью и финансовыми кризисами, классическая экономическая теория и статистика, построенные на линейных равновесных моделях, оказались малопродуктивными или, более того, неадекватными.

Антиплановая идеология реформ 1991-1995 гг. совершенно игнорировала тот факт, что планирование – это, прежде всего, определение целей и возможных путей их достижения, а также неотъемлемая составляющая управленческого процесса в любой экономической системе.

Оказалась недееспособной в рыночных условиях хозяйствования и новая структура АПК. Только в 2000-ых годах у хозяйственников новой формации стало появляться понимание того, что аргументированное представление о будущем состоянии экономики в целом и агропромышленного комплекса в частности возможно только на основе применения различных современных методов прогнозирования.

Это и неудивительно, поскольку переход от социалистического планового хозяйства к свободной рыночной экономике является крупным, можно сказать «бифуркационным» поворотом, повлекшим за собой необозримую совокупность «нелинейностей». Именно понятие бифуркации является ключевым понятием нелинейной науки («nonlinear science», как ее называют в англоязычной литературе). По существу, это математический образ «перехода количественных изменений в качественные».

Экономическое прогнозирование имеет огромное значение для развития теории и практики управления экономикой. Необходимость предвидения вероятностного исхода событий в будущем никогда прежде не была столь актуальной, как сейчас. Это связано и с высокой причинно-следственной непредсказуемостью, и коренными социально-экономическими изменениями общества, и мировыми кризисными

явлениями. Сложность производственных систем, разнообразие и, во многих случаях, неопределённый характер поступающей информации придают творческий потенциал задаче управления, которая в свою очередь повышает, а зачастую даже ставит человеческий фактор в позицию решающего. Это непосредственно проявляется в возрастающей роли методов экспертных оценок в современном планировании и управлении производством.

Остановимся подробнее на проблеме эффективной подготовки, прогнозирования и формирования управленческих решений в сельском хозяйстве, где все эти действия осуществляются в условиях неопределённости и риска из-за сложности комплексного учёта трудно формализуемых факторов.

В настоящей работе предлагается концепция снижения агроэкономического риска за счет возможностей более точного прогноза урожайностей следующего года, а так же математическая модель и метод для прогнозирования ожидаемой в наступающем году урожайности сельскохозяйственной культуры, рассматриваемой в процессе решения задач землепользования для отдельного хозяйства, района, региона и т.д. Предлагаемая модель базируется на теории и инструментарии нечетких множеств и клеточных автоматов. Исходными данными для неё служат элементы временного ряда урожайностей. Результатом применения предлагаемого метода к указанному ряду является значение ожидаемой в наступающем году урожайности в виде нечеткого множества. К тому же данная работа нацелена не только на получение максимально более точного прогноза ожидаемой урожайности, но и на обеспечение возможно более адекватного отражения стохастической природы моделируемого процесса. Достижение этих целей становится исключительно актуальным в случае практического решения задач землепользования, относящихся к зоне рискованного земледелия. Причем, полученные более точные прогнозные

значения могут быть использованы в качестве входных данных для построения модели верхнего уровня. Так же они обеспечат реализацию главной задачи научно-информационного обеспечения, а именно эффективного функционирования сельскохозяйственных товаропроизводителей в условиях изменчивости внешней и внутренней среды, связанной с погодным разнообразием, динамикой рыночной конъюнктуры, возможным изменением экологической обстановки, инфляцией и другими внеплановыми факторами.

Неслучайным является и выбор инструментария для решения обозначенной задачи, ведь тема клеточных автоматов очень актуальна на данный момент, т.к. может привести к разгадкам многих вопросов, как в сфере хозяйствования, так и в окружающем мире в целом. В этом исследовании продемонстрирована работа предложенной прогнозной клеточно-автоматной модели на реальных данных, а именно, на данных временного ряда урожайностей сахарной свеклы по Мостовскому району Краснодарского края за период с 1950 по 2005 гг. причем, результат прогноза получается в терминах лингвистических переменных (Н – низкая, С – средняя, П – промежуточная и В – высокая урожайности).

Для целей наглядной иллюстрации, валидации и верификации предлагаемой модели рассмотрим временной ряд

$$Y: y_i, i = \overline{1, n} \quad (1)$$

урожайностей сахарной свеклы Мостовского района Краснодарского края за период с 1950 по 2005 гг., которые перенумерованы индексом $i = 1, 2, \dots, n$, где $n = 2005 - 1950 + 1 = 56$; y_i – средняя урожайность (ц/га) сахарной свеклы в i -ом году. С целью визуализации этого ряда на рисунке 1 дано графическое представление этого ряда в виде гистограммы.

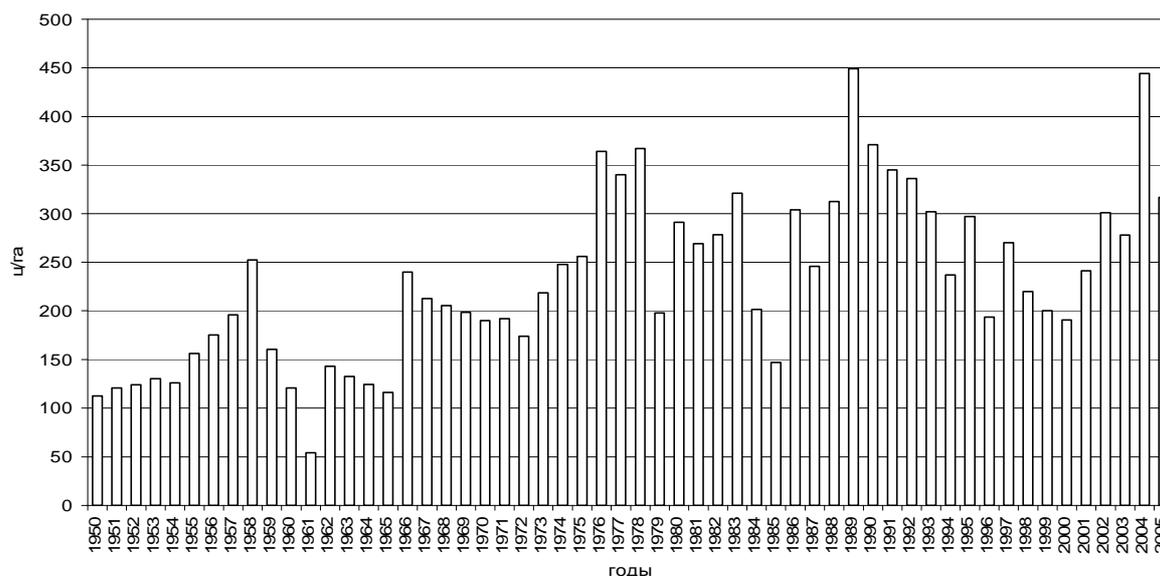


Рисунок 1. Гистограмма временного ряда (1) урожайности сахарной свеклы по Московскому району с 1950 по 2005 гг.

Для максимального учета долговременной памяти, присущей рассматриваемому временному ряду, предлагается использовать интервальные значения прогнозируемого показателя, для чего весь спектр наблюдаемых урожайностей разделен на четыре альтернативы: оптимистическую (высокий уровень), пессимистическую (низкий уровень), промежуточную и среднюю [3]. Если каждому числовому значению элементов данного временного ряда поставить в соответствие одну из этих альтернатив, то получим интервальный временной ряд или в другой терминологии, лингвистический временной ряд (ЛВР).

Преобразование временного ряда (1) в ЛВР – замена числовых элементов $y_i, i = \overline{1, n}$ лингвистическими переменными, называемыми терминами; совокупность этих термов принято называть терм-множеством [2], которое в настоящей главе обозначаем через $\bar{U} = \{u\}$. При этом принимаем, что множество \bar{U} состоит из трех элементов: $u = H$ – низкая урожайность, $u = C$ – средняя урожайность, $u = I$ – промежуточное значение урожайности, $u = B$ – высокая урожайность. Заменяя элементы y_i ряда (1) соответствующими термами из \bar{U} , получаем ЛВР

$$U: u_i, i = \overline{1, n}. \quad (2)$$

В работе [6] предлагается строить ЛВР вида (2), опираясь на такой показатель как скользящая средняя. Однако скользящие средние обладают одним принципиальным недостатком, что при их построении практически всегда остается открытым вопрос определения наилучшего порядка. Чаще всего на практике порядок средней определяется эвристически, т.е. интуитивно. В связи с этим в настоящей диссертационной работе предлагается алгоритм преобразования ряда (1) в ряд (2) на базе интервального подхода. Этот алгоритм состоит из трех этапов.

Первый этап начинается с визуализации гистограммы, представляющей ряд (1). На этой гистограмме выделяем жирными точками столбики, составляющие явно высокую урожайность, и столбики, представляющие низкую урожайность (см. рис.1). Далее, соединяя соседние жирные точки пунктирными отрезками, получаем, как показано на рисунке 2 верхнюю огибающую ломанную (ВОЛ) и нижнюю огибающую ломанную (НОЛ).

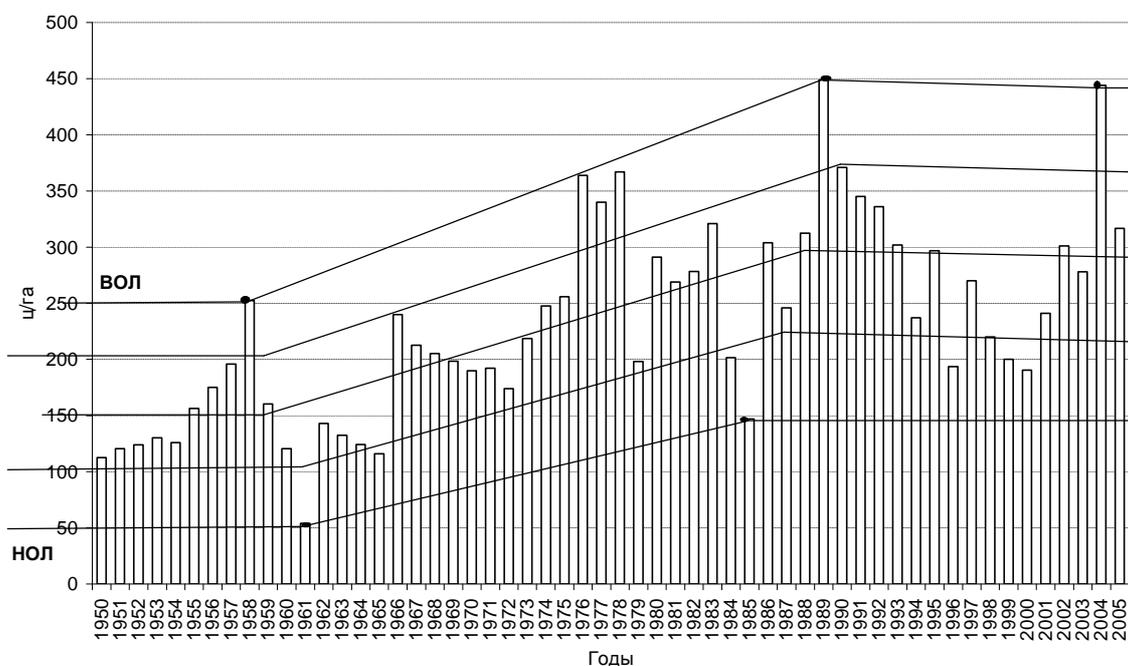


Рисунок 2. Гистограмма временного ряда (1) урожайности сахарной свеклы по Мостовскому району с 1950 по 2005 гг. после 1-го и 2-го этапов алгоритма

На втором этапе последовательно для каждого столбика гистограммы рассматриваем отрезок, соединяющий точку его пересечения с НОЛ точкой его пересечения с ВОЛ. Этот отрезок делим на четыре равновеликих интервала: нижний, средний, промежуточный и верхний. Отмечаем на каждом из таких отрезков концы среднего интервала, после чего каждую пару соседних верхних (нижних) концов средних интервалов соединяем пунктирным отрезком, и в результате получаем границы срединной области гистограммы (СОГ) и промежуточной области гистограммы (ПОГ).

На третьем этапе временной, ряд вида (1) преобразуем в ЛВР вида (2), осуществляя окрашивание каждого столбика гистограммы, как показано на рисунке 3. Рассматривая i -й столбик этой гистограммы, элемент y_i заменяем термом **Н**, если верхняя точка столбика находится ниже СОГ, иначе заменяем y_i термом **С** или **П**, если его верхняя точка принадлежит СОГ или ПОГ и, наконец, заменяем термом **В**, если верхняя точка этого столбика находится выше ПОГ. Работа третьего этапа, а вместе с ним и работа алгоритма заканчивается тогда, когда элемент y_i ряда (1) заменяется соответствующим термом. Тем самым ЛВР (2) считается построенным.

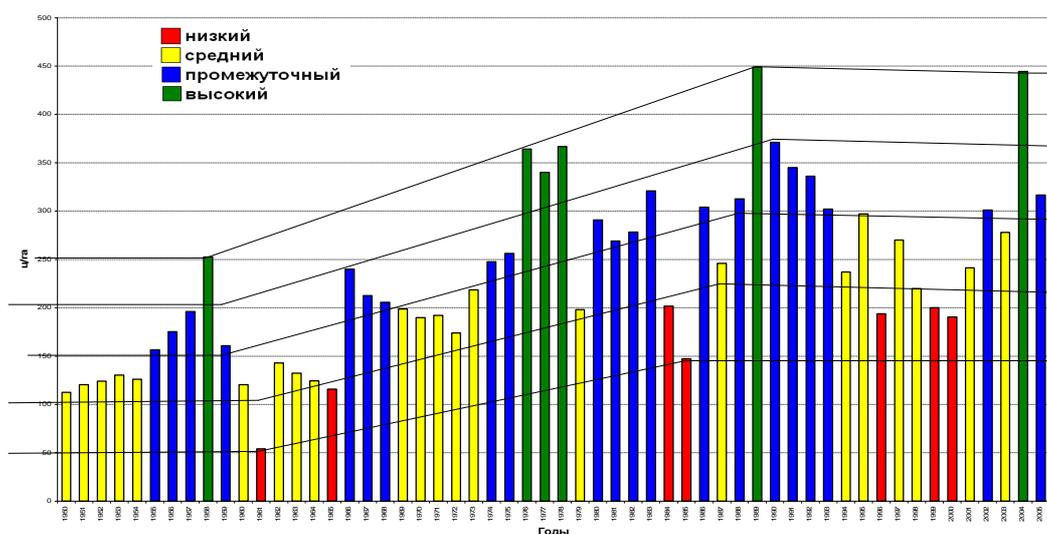


Рисунок 3. Гистограмма ЛВР (2) урожайности сахарной свеклы по Mostovskiy району с1950 по 2005 гг.

Примечание 1. Теоретически возможны случаи, когда вершина рассматриваемого столбика находится на верхней или на нижней границе СОГ. Тогда в первом случае элемент y_i заменяем термом **С**, во втором – **Н**.

Для временного ряда (1) в результате применения к нему алгоритма получен конкретный ЛВР, который представлен таблицей 1 и отражает урожайность сахарной свеклы по Мостовскому району.

Таблица 1 – Лингвистический временной ряд урожайности сахарной свеклы по Мостовскому району за период с 1950 по 2005 гг.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
u_i	С	С	С	С	С	П	П	П	В	П	С	Н	С	С
t_i	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963

i	15	16	17	18	19	20	21	22	23	24	25	26	27	28
u_i	С	Н	П	П	П	С	С	С	С	С	П	П	В	В
t_i	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977

i	29	30	31	32	33	34	35	36	37	38	39	40	41	42
u_i	В	С	П	П	П	П	Н	Н	П	С	П	В	П	П
t_i	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991

i	43	44	45	46	47	48	49	50	51	52	53	54	55	56
u_i	П	П	С	С	Н	С	С	Н	Н	С	П	С	В	П
t_i	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005

Как отмечается в [6,5], временные ряды вида (1) и ЛВР вида (2) обладают долговременной памятью [6]. Последнее означает, что такие ряды аккумулируют информацию о колебаниях погодных условий и их влияние на урожайность сельскохозяйственных культур. Иными словами, в этих рядах заключена информация об определенных закономерностях, которые в научной литературе принято относить к так называемой

долговременной памяти.

Её наличие у временного ряда (1) урожайности сахарной свеклы по Мостовскому району подтверждается результатами фрактального анализа [5] или, в более узком смысле, R/S -анализа [5], примененного к (1). Основная числовая характеристика этого результата заключается в том, что полученные значения показателя Херста H колеблются для ряда (1) в пределах от 0,7 до 0,9. Многолетний опыт, накопленный для рядов с таким значением H , свидетельствует о том, что в них имеют место долговременные корреляции между текущими событиями и будущими [5]. При этом особенно отметим, что такое поведение урожайности сахарной свеклы в зоне рискованного земледелия (в том числе и в Краснодарском крае) представляет собой типичное явление среди подавляющего большинства природных процессов и явлений [6].

В [4] сформулировано предложение, представлять наличие в ЛВР долговременной памяти в терминах и понятиях клеточного автомата, в частности, линейного клеточного автомата. Теория клеточных автоматов утверждает, что «если клетки располагаются линейно вдоль прямой, и каждая клетка находится в определенном состоянии, то состояние соседей слева от рассматриваемой клетки влияют на состояние этой клетки на следующем временном шаге» [1]. В терминах клеточного автомата значение лингвистической переменной u_{i+k} в ЛВР (2) (см. таблицу 1) определяется -конфигурациями

$$u_{i+k-l}, u_{i+k-l+1}, \dots, u_{i+k}, l = \overline{1, k}, \quad (3)$$

т.е. конфигурациями длины l в отрезке этого ряда

$$u_{i+1}, u_{i+2}, \dots, u_{i+k}, l = \overline{1, n - k + 1}, \quad (4)$$

где через k обозначаем глубину памяти рассматриваемого ряда.

Из результатов проведенного R -анализа вытекает, что для урожайности сахарной свеклы по Мостовскому району значение k

ограничено сверху цифрой 5. Последнее означает, что для всякого $i = \overline{1, n - k + 1}$ значение лингвистической переменной u_{i+k} в (4) или в (2) определяется лишь такими l -конфигурациями вида (3), для которых $l \leq k = 5$. Алгоритм прогнозирования основывается на частотной статистике переходов в состояния H, C и B всех l -конфигураций, имеющих место в ЛВР (2).

Через $M(U)$ обозначим множество всех l -конфигураций $l \leq k, k = 5$, которые можно обнаружить в ЛВР (2); $M(U) = \bigcup_{l=1}^5 M_l$, где M_l – это подмножество всех l -конфигураций в ЛВР U при фиксированном l . Для рассматриваемых ВР Y (1) и ЛВР U (2) эти подмножества имеют следующий состав:

$$M_1 = \{H, C, П, B\},$$

$$M_2 = \{HH, HC, HP, CH, CC, CP, CB, PH, PC, PP, PB, BC, BP, BB\},$$

$$M_3 = \left\{ \begin{array}{l} HHC, HHP, HCC, HCP, HPC, HPP, CHH, CHC, CHP, CCH, \\ CCC, CCP, CPP, CPH, CPC, PCP, PCB, PPH, \\ PPS, PPP, PPB, PBP, PBB, BPP, BPP, BPC, BVB \end{array} \right\}.$$

$$M_5 = \{CCCCC, CCCCCP, PPPPP, PPPPB, PPPPH, PPPPS\}.$$

Примечание 2. Через N_l обозначим количество всех попарно различных l -конфигураций в ЛВР (2). Для принятого терм-множества $U = \{H, C, П, B\}$ теоретически возможное количество различных l -конфигураций, $l = 1, 2, \dots, k, k = 5$ составляет $\sum_{l=1}^k 4^l = 4 + 4^2 + 4^3 + 4^4 + 4^5 = 1364$, в то время как в реальном ЛВР (2), представленного в таблице 1, количество N_l всех таких попарно различных l -конфигураций, $l \leq 5$ составляет $N = \sum_{l=1}^5 N_l = 73$. Из них $N_1 = 4$, $N_2 = 14$, $N_3 = 29$, $N_4 = 20$, $N_5 = 6$. Тем самым фактически количество реальных l -конфигураций составляет менее 5,3% от количества теоретически возможных l -конфигураций.

Рассмотрим какую-либо фиксированную l -конфигурацию, которую обозначим в виде отрезка

$$u_1^0, u_2^0, \dots, u_j^0, \dots, u_l^0. \quad (5)$$

Если в ЛВР (2) выделен отрезок $u_{i+1}, u_{i+2}, \dots, u_{i+j}, \dots, u_{i+l}$, совпадающей с (5), т.е. $u_{i+j} = u_j^0, j = \overline{1, l}$, то по отношению к следующему элементу $u_{i+l+1} = u^0, u^0 \in U = \{H, C, П, В\}$, условимся говорить, что l -конфигурация (5) переходит в состояние u^0 , т.е. в лингвистическую переменную u_{i+l+1} , совпадающую с термом u^0 .

В используемом подходе базовым является следующее теоретическое предположение. Пусть последовательность (2) неограниченно растет, т.е. в ряду $u_i, i = \overline{1, n}$, значение параметра $n \rightarrow \infty$. Если в этой сколь угодно длинной последовательности некоторая конкретная фиксированная конфигурация (5) появляется и при этом всякий раз после нее следует переход в одно и то же состояние $u^0 \in \{H, C, П, В\}$, то говорим, что конфигурация (5) обладает памятью. Если имеют место перемежающиеся переходы в два фиксированных состояния, то говорим, что отрезок (5), т.е. l -конфигурация (5) обладает частичной памятью. Если же фиксированная конфигурация демонстрирует переходы в каждое из четырёх состояний $H, C, П, В$, то говорим, что память у данной конфигурации не обнаружена.

По составу представленной выше памяти клеточного автомата можно определённо сказать, что выявлены наличие и глубина памяти ЛВР (2). Длина отрезка лингвистического временного ряда, не превосходящая 5, определяет состояние прогнозируемого показателя на очередном временном шаге.

Анализ конкретного ЛВР, отражающего урожайность сахарной свеклы по Мостовскому району, позволяет сформулировать следующие утверждения:

1) для всякого отрезка длины 1 $(H, C, П, В)$ и всякого отрезка длины 2 $(HH, HC, HP, CH, CC, CP, CV, PH, PC, PP, PV, VC, VP, VV)$ в ряду $u_i, i = \overline{1, n}$, память отсутствует (только переход в одно состояние), т.к.

всякий раз находились случаи переходов из этих отрезков в 2 или 3 состояния из числа **Н, С, П и В** (первые «признаки» наличия памяти (частичной, т.е. переход в два состояния) обнаружались при $l = 2$: уже 64,28% 2-конфигураций, из числа встречающихся в ряду (2), демонстрируют частичную память);

2) для $l = 3$ конфигурации вида (5) демонстрируют переход только в одно состояние (память), т.е. с различной частотой переходы в какое-либо из четырех $u \in \{Н, С, П, В\}$, и 24,14% 3-конфигураций демонстрируют наличие частичной памяти;

3) для $l = 4$ 85% 4-конфигураций в ряду (2) демонстрирует наличие памяти и 15% демонстрирует наличие частичной памяти, другие случаи отсутствуют;

4) для $l = 5$ все 100% 5-конфигураций вида (5) демонстрируют наличие памяти.

Формирование памяти клеточного автомата завершается вычислением частотной статистики переходов -конфигураций (5) в определенное состояние $u^0 \in U = \{Н, С, П, В\}$. Эта статистика формируется следующим образом. Сначала, для каждой 1-конфигурации $u_1^0 \in U = \{Н, С, П, В\}$ подсчитываем количество ее переходов в каждое из четырех состояний **Н, С, П, В**.

На основании переходов можно вычислить эмпирические значения частотей переходов 1-конфигураций в каждое из состояний **Н, С, П, В**:

$$\begin{aligned}
 w_1(H \rightarrow H) &= \frac{2}{7} & w_1(C \rightarrow H) &= \frac{4}{22} & w_1(\Pi \rightarrow H) &= \frac{1}{20} & w_1(B \rightarrow H) &= \frac{0}{6} \\
 w_1(H \rightarrow C) &= \frac{3}{7} & w_1(C \rightarrow C) &= \frac{12}{22} & w_1(\Pi \rightarrow C) &= \frac{5}{20} & w_1(B \rightarrow C) &= \frac{1}{6} \\
 w_1(H \rightarrow \Pi) &= \frac{2}{7} & w_1(C \rightarrow \Pi) &= \frac{5}{22} & w_1(\Pi \rightarrow \Pi) &= \frac{11}{20} & w_1(B \rightarrow \Pi) &= \frac{3}{6} \\
 w_1(H \rightarrow B) &= \frac{0}{7} & w_1(C \rightarrow B) &= \frac{1}{22} & w_1(\Pi \rightarrow B) &= \frac{3}{20} & w_1(B \rightarrow B) &= \frac{2}{6}
 \end{aligned} \quad (6)$$

Далее, для каждой 2-конфигурации $u_1^0 u_2^0 \in (\overline{U} \times \overline{U})$ подсчитываем количество переходов в каждое из трех состояний **Н, С, П, В**. Далее можно

вычислить эмпирические значения частот переходов из 2-конфигураций **НН,НС,НП** в состояния **Н,С,П** и **В**:

$$\begin{aligned}
 w_1(НН \rightarrow Н) &= \frac{0}{2} & w_1(СН \rightarrow Н) &= \frac{1}{4} & w_1(НП \rightarrow Н) &= \frac{1}{11} \\
 w_1(НН \rightarrow С) &= \frac{1}{2} & w_1(СН \rightarrow С) &= \frac{2}{4} & w_1(НП \rightarrow С) &= \frac{2}{11} \\
 w_1(НН \rightarrow П) &= \frac{1}{2} & w_1(СН \rightarrow П) &= \frac{1}{4} & w_1(НП \rightarrow П) &= \frac{6}{11} \quad (7) \\
 w_1(НН \rightarrow В) &= \frac{0}{2} & w_1(СН \rightarrow В) &= \frac{0}{4} & w_1(НП \rightarrow В) &= \frac{2}{11}
 \end{aligned}$$

Аналогичным образом вычисляются эмпирические значения частот переходов из 2-конфигураций **НС, НП, СС, СП, СВ, ПН, ПС, ПВ, ВС, ВП, ВВ** в **Н, С, П** и **В**.

Далее, для каждого значения $l \in \{3,4,5\}$ рассматриваем множество M_l всех l -конфигураций, встречающихся в ЛВР (2), мощность $|M_l| = N_l$. По аналогии с (6), (7) вычисляем эмпирические значения частот переходов из каждой конкретной l -конфигурации $u_1^0, u_2^0, \dots, u_l^0 \in M$ в состояние **Н, С, П** и **В**, $l = 3,4,5$:

$$\begin{aligned}
 &w_l(u_1^0 u_2^0 \dots u_l^0 \rightarrow Н), w_l(u_1^0 u_2^0 \dots u_l^0 \rightarrow С), \\
 &w_l(u_1^0 u_2^0 \dots u_l^0 \rightarrow П), w_l(u_1^0 u_2^0 \dots u_l^0 \rightarrow В). \quad (8)
 \end{aligned}$$

По результатам работы клеточного автомата была проведена верификация и валидация представленной прогнозной модели.

Сначала на примере исследуемого конкретного ЛВР (2), представляющего временной ряд урожайностей сахарной свеклы по Мостовскому району, приведем описание алгоритма прогнозирования, работающего на базе представленной выше прогнозной модели. Ставится задача прогнозирования для рассматриваемого ЛВР U неизвестного элемента u_{n+1} , на основании известных членов этого ряда $u_i, i = \overline{1, n}$, точнее, на основании вычисленных выше частот вида (6)-(8), для $l = 1,2, \dots, k$, где k – глубина памяти в ЛВР (2).

Прогноз u_{n+1} терма представляется в виде нечеткого термножества (НТМ) $U_{n+1} = \{(Н; \mu_H), (С; \mu_C), (П; \mu_P), (В; \mu_B)\}$, где

значение функции принадлежности μ удовлетворяет равенству $\mu_H + \mu_C + \mu_\Pi + \mu_B = 1$. Значения $\mu_H, \mu_C, \mu_\Pi, \mu_B$ вычисляются через значения частот вида (4)-(6), получаемых для различных l -конфигураций в следующем отрезке ЛВР

$$u_{n-l+1}, u_{n-k}, \dots, u_n \tag{9}$$

Сначала согласно (6) вычисляются частоты переходов l -конфигурации u_n в состояния H, C, Π, B : $w_1(u_n \rightarrow H), w_1(u_n \rightarrow C), w_1(u_n \rightarrow \Pi), w_1(u_n \rightarrow B)$. После чего, согласно (7), вычисляются эмпирические значения частот переходов из 2-конфигурации $u_{n-1}u_n$ в состояния H, C, Π и B : $w_2(u_{n-1}u_n \rightarrow H), w_2(u_{n-1}u_n \rightarrow C), w_2(u_{n-1}u_n \rightarrow \Pi), w_2(u_{n-1}u_n \rightarrow B)$. Далее согласно (8) вычисляем значение частот переходов из 3-конфигурации в $u_{n-2}u_{n-1}u_n$ в состояния H, C, Π, B . Если 3-конфигурация $u_{n-2}u_{n-1}u_n$ демонстрирует наличие памяти, например, $w_3(u_{n-2}u_{n-1}u_n \rightarrow C) = 1$, то переходим к вычислению искомым $\mu_H, \mu_C, \mu_\Pi, \mu_B$. Для этого сначала вычисляем ненормированные значения

$$\mu'_H = w_1(u_n \rightarrow H) + w_2(u_{n-1}u_n \rightarrow H) + 0,$$

$$\mu'_C = w_1(u_n \rightarrow C) + w_2(u_{n-1}u_n \rightarrow C) + 1,$$

$$\mu'_\Pi = w_1(u_n \rightarrow \Pi) + w_2(u_{n-1}u_n \rightarrow \Pi) + 1,$$

$$\mu'_B = w_1(u_n \rightarrow B) + w_2(u_{n-1}u_n \rightarrow B) + 0$$

и их сумму $\sigma_3 = \mu'_H + \mu'_C + \mu'_\Pi + \mu'_B$. После нормируем

$$\mu_H = \frac{\mu'_H}{\sigma_3}, \mu_C = \frac{\mu'_C}{\sigma_3}, \mu_\Pi = \frac{\mu'_\Pi}{\sigma_3}, \mu_B = \frac{\mu'_B}{\sigma_3}.$$

Если 3-конфигурация $u_{n-2}u_{n-1}u_n$ не демонстрирует наличие памяти то рассматриваем 4-конфигурацию $u_{n-3}u_{n-2}u_{n-1}u_n$, для которой вычисляем частоты ее переходов в состояния H, C, Π и B . Всякий раз к вычислению искомым $\mu_H, \mu_C, \mu_\Pi, \mu_B$ переходим тогда, когда встретится такая l -конфигурация, которая демонстрирует наличие памяти, например, получаем единичное значение частоты для термина B :

$w_1(u_{n-l+1}u_{n-l+2} \dots u_n \rightarrow B) = 1$. Тогда сначала вычисляем ненормированные значения

$$\mu'_H = w_1(u_n \rightarrow H) + w_2(u_{n-1}u_n \rightarrow H) + \dots + w_{l-1}(u_{n-l+2}u_{n-l+3} \dots u_n \rightarrow H) + 0,$$

$$\mu'_C = w_1(u_n \rightarrow C) + w_2(u_{n-1}u_n \rightarrow C) + \dots + w_{l-1}(u_{n-l+2}u_{n-l+3} \dots u_n \rightarrow C) + 0,$$

$$\mu'_\Pi = w_1(u_n \rightarrow \Pi) + w_2(u_{n-1}u_n \rightarrow \Pi) + \dots + w_{l-1}(u_{n-l+2}u_{n-l+3} \dots u_n \rightarrow \Pi) + 0,$$

$$\mu'_B = w_1(u_n \rightarrow B) + w_2(u_{n-1}u_n \rightarrow B) + \dots + w_{l-1}(u_{n-l+2}u_{n-l+3} \dots u_n \rightarrow B) + 1$$

и значения их суммы $\sigma_1 = \mu'_H + \mu'_C + \mu'_\Pi + \mu'_B$. После чего, вычисляем искомое значение функции принадлежности для НТМ U_{n+1} :

$$\mu_H = \frac{\mu'_H}{\sigma_1}, \mu_C = \frac{\mu'_C}{\sigma_1}, \mu_\Pi = \frac{\mu'_\Pi}{\sigma_1}, \mu_B = \frac{\mu'_B}{\sigma_1}.$$

Представленный таблицей 1 ЛВР урожайности сахарной свеклы заканчивается элементом $u_n = \Pi$, где $n = 56$ соответствует 2005 году. Осуществим прогноз этой урожайности на 2006-й год, т.е. построим для отсутствующего элемента u_{n+1} его нечеткое терм-множество $U_{n+1}^0 = \{(H; \mu_H^0), (C; \mu_C^0), (\Pi; \mu_\Pi^0), (B; \mu_B^0)\}$. Прогноз осуществляется на качественном уровне, т.е. определенно можно сказать, какая будет урожайность в следующем году: низкая, средняя, промежуточная или высокая.

Учитывая установленную глубину памяти $k = 5$, рассматриваем отрезок ЛВР

$$u_{n-4}u_{n-3}u_{n-2}u_{n-1}u_n = \text{СПСВП} \tag{10}$$

Для ряда (10) рассматриваем все его l -конфигурации, $l = \overline{1, k}, k = 5$: $\Pi, В\Pi, СВ\Pi, ПСВ\Pi, СПСВ\Pi$. Для $l = 1$ из перехода (6) получаем:

$$w_1(\Pi \rightarrow H) = \frac{1}{20} \quad w_1(\Pi \rightarrow C) = \frac{5}{20} \quad w_1(\Pi \rightarrow \Pi) = \frac{11}{20} \quad w_1(\Pi \rightarrow B) = \frac{3}{20}. \tag{11}$$

Для $l = 2$, по подобию соотношения (7) получаем значения частотей переходов из 2-конфигурации $В\Pi$ в термы H, C, Π, B :

$$w_2(В\Pi \rightarrow H) = \frac{0}{2} \quad w_2(В\Pi \rightarrow C) = \frac{1}{2} \quad w_2(В\Pi \rightarrow \Pi) = \frac{1}{1} \quad w_2(В\Pi \rightarrow B) = \frac{0}{2}. \tag{12}$$

Для $l = 3$, 3-конфигурация $СВ\Pi$, демонстрирует наличие памяти, в силу чего для ряда (2) процесс вычисления частотей можно прекратить,

но для достижения более высокой точности прогнозного значения можно продолжить частотный анализ до исчерпания памяти, т.е. до конфигурации длины 5.

На основании значений частот (11-12), вычисляем ненормированные значения функции принадлежности: $\mu'_H = \frac{1}{20} = 0,05$, $\mu'_C = \frac{5}{20} + \frac{1}{2} = 0,75$, $\mu'_П = \frac{11}{20} + \frac{1}{2} = 1,05$, $\mu'_B = \frac{3}{20} = 0,15$ и их сумму $\sigma = 0,05 + 0,75 + 1,05 + 0,15 = 2$. Далее осуществляя операцию нормирования получим искомое значение функции принадлежности:

$$\mu_H = \frac{\mu'_H}{\sigma} = \frac{0,05}{2} = 0,025, \mu_C = \frac{\mu'_C}{\sigma} = \frac{0,75}{2} = 0,375, \mu_{П} = \frac{\mu'_{П}}{\sigma} = \frac{1,05}{2} = 0,525, \mu_B = \frac{\mu'_B}{\sigma} = \frac{0,15}{2} = 0,075$$

Таким образом, прогноз урожайности сахарной свеклы по Мостовскому району на 2006 год представляется в виде НТМ $U^0_{n+1} = \{(H; 0,025), (C; 0,375), (П; 0,525), (B; 0,075)\}$. В лингвистических терминах этот прогноз можно сформулировать следующим образом: урожайность сахарной свеклы ожидается как промежуточное значение ($\mu_{П} = 0,525$) или, что менее вероятно низкой ($\mu_H = 0,025$). Промежуточные и окончательные результаты работы алгоритма прогнозирования представлены в таблице 2.

Таблица 2 – Промежуточные и окончательные результаты работы алгоритма прогнозирования

Прогнозируемый год	2006			
<i>l</i> -конфигурация	СПСВП			
Переходы <i>l</i> -конфигурации в состояния Н, С, П и В	Н	С	П	В
Ненормированные значения функции принадлежности $\mu'_H, \mu'_C, \mu'_{П}, \mu'_B$	1/20=0,05	5/20+1/2=0,75	11/20+1/2=1,05	3/20=0,15
Сумма ненормированных значений функций принадлежности	2			
Значение функции принадлежности $\mu_H, \mu_C, \mu_{П}, \mu_B$	0,025	0,375	0,525	0,075
Прогнозное НТМ $U = \{(H; \mu_H), (C; \mu_C), (П; \mu_{П}), (B; \mu_B)\}$	U={ (H;0,025), (C;0,375), (П;0,525), (B;0,075) }			

Применительно к понятию «модель», термин «верификация» означает проверку структуры и логики модели, а термин «валидация» означает проверку соответствия данных, полученных на основе модели, реальному процессу. Для реализации этих видов проверки построенной прогнозной модели последовательно рассматриваем лингвистические временные ряды

$$u_i, i = 1, 2, \dots, m, m = n - r, r = \overline{1, n - k}, \quad (13)$$

т.е. ряды (13) получают последовательно путем удаления из ЛВР (2) последних r его членов.

Для каждого фиксированного индекса m строим прогноз термина u_{m+1} , представляемого в виде НТМ $u_{m+1} = \{(H; \mu_H), (C; \mu_C), (П; \mu_P), (B; \mu_B)\}$.

Пусть, в полученном НТМ u_{m+1} , среди чисел $\mu_H, \mu_C, \mu_P, \mu_B$ максимальным является то число μ_Δ , $\Delta \in \{H, C, П, B\}$ у которого индекс Δ совпадает с термом u_{m+1} ряда (2). Тогда, говорим, что для рассматриваемого индекса m прогнозная нечеткая модель привела к непротиворечивому прогнозу. В противном случае, говорим о противоречивом прогнозе для термина u_{m+1} .

Для ЛВР (2), соответствующего ряду (1) урожайности сахарной свеклы по Мостовскому району, была проведена валидация прогнозной модели и был получен непротиворечивый прогноз для каждого $m = n - r, r = 1, 2, \dots, n - 5$. Иными словами, в процессе валидации прогнозной модели подтверждена адекватность предложенной прогнозной нечеткой модели реальным временным рядам урожайности сахарной свеклы по Мостовскому району.

Пусть получено лингвистическое прогнозное значение урожайности

$$U_{n+1}^0 = \{(H; 0,025), (C; 0,375), (П; 0,525), (B; 0,075)\}. \quad (14)$$

Приведем описание процесса преобразования лингвистического нечеткого множества (ЛНМ) в численное (классическое) НМ

$$Y_{n+1}^0 = \{(y_H^0; \mu_H), (y_C^0; \mu_C), (y_P^0; \mu_P), (y_B^0; \mu_B)\} \quad (15)$$

В качестве подходящих числовых значений элементов $y_u^0, u \in \{H, C, P, B\}$ выбираются в ВР Y (1) ближайšie к элементам y_u низкие, средние, промежуточные и высокие урожайности, которые затем усредняются:

$$y_H^0 = y_{50} + \sigma_{51} = \frac{1}{2}(200 + 190,5) = 195,25, y_C^0 = y_{54} = 278,$$

$$y_P^0 = y_{53} + \sigma_{56} = \frac{1}{2}(301,1 + 316,6) = 308,85, y_B^0 = \sigma_{55} = 444,3.$$

Отсюда, с учетом представленных в ЛНМ (14) значений функции принадлежности $\mu_H, \mu_C, \mu_P, \mu_B$ получаем искомый прогноз в виде НМ $Y_{n+1}^0 = \{(195,25; 0,025), (278,0; 0,375), (308,85; 0,525), (444,3; 0,075)\}$.

Применяя к НМ Y_{n+1}^0 операцию дефазификации [2], получаем прогнозное значение урожайности в обычном числовом виде, т.е. имеем:

$$Y_{n+1}^0 = \sum \mu_t y_t^0 = 195,25 \cdot 0,025 + 278,0 \cdot 0,375 + 308,85 \cdot 0,525 + 444,3 \cdot 0,075 = 304,6 \text{ ц/га},$$

где индексом $t = 1, 2, 3, 4$ перенумерованы соответственно термы H, C, P, B : $\mu_1 = \mu_H = 0,025, \mu_2 = \mu_C = 0,375, \mu_3 = \mu_P = 0,525, \mu_4 = \mu_B = 0,075$.

Согласно определению прогнозной модели на ее выходе можно получить ВР Y^0 прогнозных значений $y_i^0, i = L, L + 1, \dots, n$, занумерованных тем же индексом, которым были занумерованы значения урожайности в ВР (1). Тогда относительная погрешность прогнозирования для каждого наблюдения $i \in \{L, L + 1, \dots, n\}$ вычисляется по формуле $\varepsilon = \frac{|y_i - y_i^0|}{y_i}$. В качестве оценки точности прогнозирования принимаем

$$\text{среднее значение } \varepsilon = \frac{1}{n-L+1} \sum_{i=L}^n \varepsilon_i.$$

Примечание 3. На основании валидации результатов прогнозирования ВР урожайности сахарной свеклы по Мостовскому району получена оценка средней погрешности прогноза $\varepsilon \leq 12\%$.

Оценка погрешности результатов, полученных с помощью

предлагаемой прогнозной модели, обосновывается также по отношению такого результата валидации, как ВР лингвистических нечетких множеств $U: u_i, i = L, L + 1, \dots, n$. В этом случае погрешность $\bar{\varepsilon}_i$ лингвистического прогнозирования для каждого наблюдения i принимается равной нулю, если в ряду ЛНМ $U: u_i, i = L, L + 1, \dots, n$ для полученного ЛНМ $U^0 = \{(u_1^0; \mu_1), (u_2^0; \mu_2), (u_3^0; \mu_3), (u_4^0; \mu_4)\}$, где максимальное значение функции принадлежности $\mu = \max_{1 \leq t \leq 4} \mu_t$ достигается для такого индекса $t = t_0$, что в ЛВР (2) элемент u_i совпадает с термом $u_{t_0}^0$, т.е. $\bar{\varepsilon}_i = 0$, если выполняется равенство $u_i = u_{t_0}^0$, в противном случае значение $\bar{\varepsilon}_i = 1$. Погрешность лингвистического прогнозирования определяется как среднее значение $\bar{\varepsilon} = \frac{1}{n-L+1} \sum_{i=L}^n \bar{\varepsilon}_i$.

Примечание 4. На основании валидации результатов лингвистического прогнозирования ВР урожайности сахарной свеклы по Мостовскому району получена оценка средней погрешности прогноза $\bar{\varepsilon}_i = 0$, т.е. в процессе валидации прогнозная модель выдала точный прогноз в лингвистических термах для каждого года с 1950 по 2005 гг.

Примечание 5. Таким образом, нижний уровень реализации 2-уровневого подхода к моделированию представляет собой полную общепринятую последовательность устоявшегося стандарта набора этапов моделирования:

1⁰. Анализ объекта моделирования, включая структурирование, с целью формирования перечня параметров и показателей моделей.

2⁰. Выбор адекватного подхода и используемых математических методов моделирования.

3⁰. Численная реализация выбранных математических методов на базе конкретных исходных статистических данных.

4⁰. Верификация, т.е. логический анализ модели и результатов методов моделирования.

5⁰. Валидация используемых методов, включая оценку погрешности.

б⁰. Доработка и представление окончательного варианта предлагаемого процесса моделирования на нижнем уровне.

Процесс обучения линейного клеточного автомата представляет собой достаточно творческий подход. Он заключается в виде подбора такой раскраски уровней ВР, предварительная валидация результатов которой даёт меньшую погрешность ошибки прогноза. Процесс предварительной валидации результатов выбранной раскраски клеточного автомата заключается в следующем:

1. фактически реализуется прогнозная модель на базе лингвистических переменных;
2. выявляются на уровне лингвистических переменных не угаданные значения.

Итак, если применять уже сформировавшийся и описанный в работе [3] подход к раскраске клеточного автомата мы имеем представленный на рисунке 4 вариант лингвистического ВР.

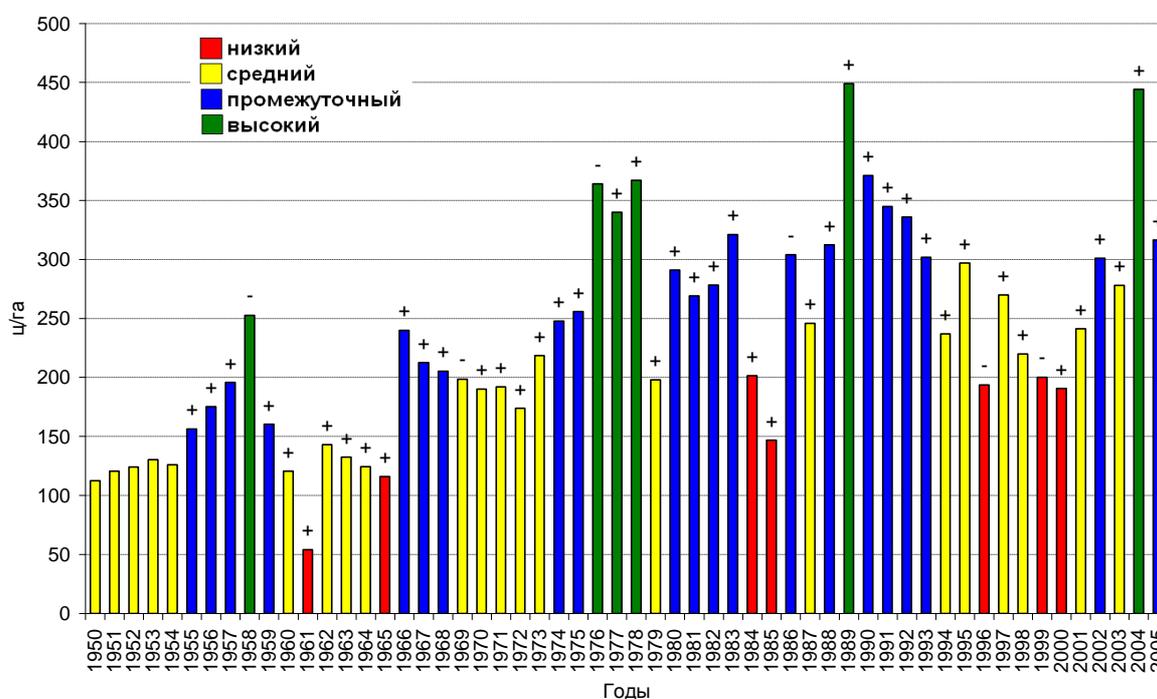


Рисунок 4 – Гистограмма ЛВР урожайности сахарной свеклы по Мостовскому району за период с 1950 по 2005 гг.

Результаты валидации: количество угаданных уровней – 45 шт., (на рисунке 4 обозначены знаком «+»), количество не угаданных уровней – 6 шт. (на рисунке 4 обозначены знаком «-»). Достижение цели на уровне лингвистических переменных дает возможность говорить о достаточности предложенного варианта обучения клеточного автомата.

Подводя итоги проделанной работы, нами исследованы 6 ВР урожайности сельскохозяйственных культур: озимая пшеница, подсолнечник, сахарная свекла и их приращения. Статистика ошибок прогноза приведена в таблице 3.

Таблица 3 – Статистика ошибок прогноза и приращения

ВР	Глубина памяти ряда l	Ошибка прогноза в %
Озимая пшеница	5	17
Подсолнечник	6	18
Сахарная свекла	5	12
Приращения ВР урожайности озимой пшеницы	5	16
Приращения ВР урожайности подсолнечника	6	14
Приращения ВР урожайности сахарной свеклы	5	24

Стоит отметить, процесс валидации ЛВР подтвердил адекватность предложенной прогнозной модели реальным временным рядам урожайности культур по Мостовскому району Краснодарского края.

В результате, используя подход, который базируется на применении клеточных автоматов и математического аппарата нечетких множеств, входные данные для построения моделей высшего уровня имеют достаточную степень достоверности. В свете коренного реформирования агропромышленного комплекса по-новому стоит вопрос о создании систем ведения сельского хозяйства в современных условиях. Ведь обеспечение эффективного управления возможно лишь при соблюдении соответствующего баланса функций, полномочий и ответственности, а так

же при возможности регулирования степени риска в процессе реализации. Определить последнее можно только в результате анализа и оценки вероятности ожидаемых экономических, погодных и иных условий хозяйствования.

Таким образом, практическое использование результатов, получаемых на выходе прогнозных моделей, и последующие построение и реализация адекватных моделей верхнего уровня позволит хозяйственнику принимать обоснованные управленческие решения и обеспечивать возможность маневрирования ресурсами и технологиями, стимулировать поиск гибких производственных решений.

Список использованных источников

1. Борисов А.Н., Алексеев А.В., Меркурьева Г.В. Обработка нечеткой информации в системах принятия решений. М.: Радио и связь, 1989. 304 с.
2. Жирабок А.Н. Нечеткие множества и их использование для принятия решений // Соровский образовательный журнал. 2001. Том 7. №2. С. 109-115.
3. Перепелица В.А., Касаева М.Д. Прогнозирование природного временного ряда на базе модели клеточного автомата // Современные аспекты экономики. 2002. № 9(22). С. 201-208.
4. Перепелица В.А., Тебуева Ф.Б., Темирова Л.Г., Касаева М.Д. Прогнозная модель урожайности на базе клеточных автоматов и нечетких множеств // Труды III международной конференции «Новые технологии в управлении, бизнесе и праве». НФ ИУБиП. Невинномысск, 2003. С. 163-167.
5. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. М.: Мир, 2000. 333 с.
6. Яновский Л.П. Принципы, методология и научное обоснование урожая по технологии «Зонт». Воронеж: ВГАУ, 2000. 379 с.