УДК 519.711.2

Н.В. Рылова, И.Г. Оксанич

Кременчугский национальный университет имени Михаила Остроградского, Кременчуг

СИНТЕЗ ARIMA-МОДЕЛЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ЗНАЧЕНИЙ КОЭФФИЦИЕНТОВ ВЫХОДА КОНДИЦИОННЫХ ПОЛУПРОВОДНИКОВЫХ МАТЕРИАЛОВ

В статье предложена математическая модель для моделирования последовательностей коэффициентов выхода кондиционной продукции при производстве полупроводниковых материалов, проведена идентификация структуры модели, оценка параметров и проверена адекватность модели. А также разработан метод построения комбинированного прогноза, что даст возможность повысить точность прогнозирования по сравнению с применением частных моделей прогнозирования. Разработанная модель и метод используются для оперативного управления производством при прогнозировании значений коэффициентов выхода кондиционной продукции.

Ключевые слова: *ARIMA-модель, идентификация модели, временная последовательность, гипотеза, прогнозирование.*

Введение

Важным элементом оперативного управления производством является прогнозирование выхода кондиционной продукции. В связи с этим блок прогнозирования является неотъемлемой частью общей системы контроля и управления ходом производства. С точки зрения управления производственным подразделением главное задание прогнозирования заключается в предвидении изменений со временем основных технико-экономических характеристик рассматриваемого объекта, выявлении причин, обуславливающих характер этих изменений, а также в оценке возможных последствий предусмотренного поведения. На основании прогнозных значений могут быть сформированы дополнительные управляющие воздействия, предназначенные для ликвидации отклонений от намеченной планом траектории развития процесса.

Для прогнозирования производственного процесса необходимо сначала собрать ретроспективную информацию о фактических наблюдениях за процессом и влияющими на него факторами за определенный интервал времени в прошлом, а затем построить по этой информации математическую модель прогнозирования. Точность и достоверность получаемого прогноза зависит от объема выборки, точности и достоверности исходной информации, адекватности выбранной модели и глубины прогноза [1].

Технологический процесс производства полупроводниковых материалов отличается наличием большого числа исходных параметров, большим числом производимых технологических операций и значительной длительностью многих из них. Он имеет вероятностный характер, ему свойственно низкое отображение результатов, причиной которого является влияние многих неконтролируемых факторов [2], поэтому часто источником информации в таких условиях является выход кондиционной продукции.

Коэффициент выхода кондиционной продукции это отношение количества полученных кондиционных полуфабрикатов к общему количеству обработанных полуфабрикатов. Зная закон распределения или закономерность появления значения этого коэффициента для всех единиц оборудования производственного процесса, мы сможем генерировать его значения при проведении имитационного моделирования. Кроме того, наличие модели временной последовательности значений коэффициента дает возможность прогнозировать будущие значения этого показателя.

Поскольку при решении задачи оптимального распределения оборудования между различными видами полуфабрикатов (синтез локальных управлений) возникает необходимость в прогнозировании значений, то рассмотрим некоторые методы прогнозирования временных последовательностей значений коэффициентов выхода кондиционной продукции.

Целью данной работы является проведение экспериментальных исследований временных последовательностей (X1, X2, X3), образованных суточными наблюдениями на протяжении трех месяцев за значениями коэффициентов выхода кондиционной продукции для трех марок арсенида галлия, полученных при выращивании методом Чохральского, а также разработка математической модели для прогнозирования значений коэффициентов выхода кондиционной продукции при оперативном управлении производством.

Графики, показывающие динамику изменения значений коэффициентов выхода кондиционной продукции, представлены на рис. 1 – 3.

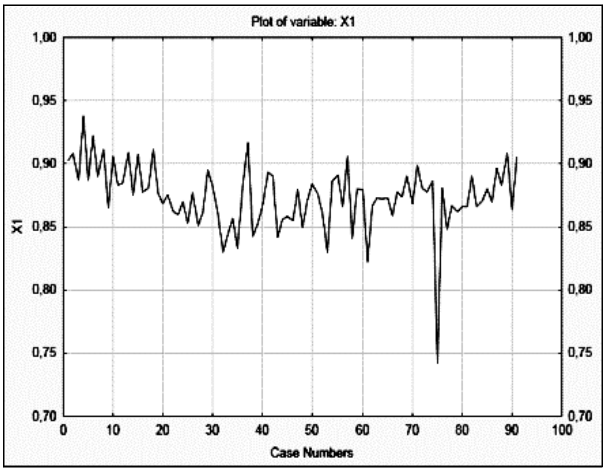

Рис. 1. Динамика ряда X1

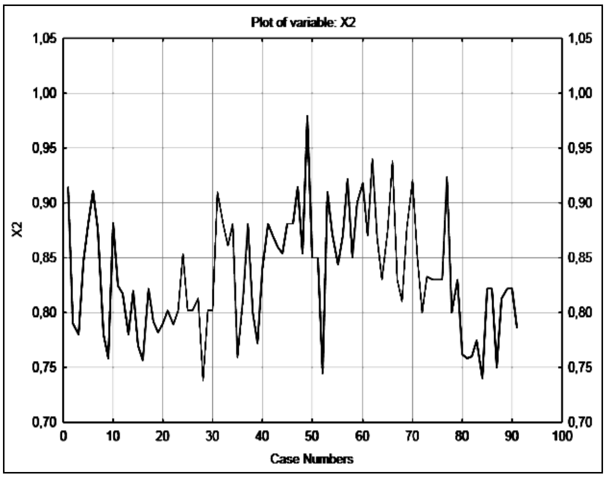

Рис. 2. Динамика ряда X2

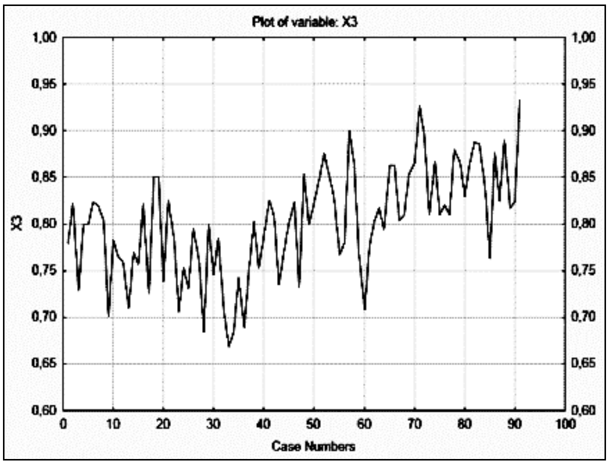

Рис. 3. Динамика ряда X3

Исследование возможности использования ARIMA-моделей для моделирования последовательностей коэффициентов выхода кондиционной продукции

Допустим, что каждый из трех рядов наблюдений X1, X2, X3 образует случайную выборку с генеральной совокупностью, которая имеет некоторую модельную функцию распределения $F_{mod}(X;\theta_1,\theta_2,...\theta_s)$ с известным общим видом функции F_{mod} (тип модели) и параметрами $\theta_1,\theta_2,...\theta_s$, значения которых могут быть как известными, так и неизвестными. Тогда задание закона распределения наблюдаемых значений рядов заключается в проверке статистической гипотезы

$$H_0: F(x) = F_{mod}(x;\theta_1,\theta_2,...\theta_s). \qquad (1)$$

Для этого могут быть использованы разные так называемые критерии согласия, которые основаны на использовании разных мер расстояний между анализируемой эмпирической функцией распределения $\bar{F}^{(n)}(x)$ (которая определяется по выборке) и гипотетической модельной $F_{mod}(x;\theta_1,\theta_2,...\theta_s)$. Рассмотрим наиболее используемые критерии согласия [3].

Критерий χ^2–Пирсона позволяет выполнять проверку гипотезы (1) в условиях, когда параметры $\theta_1,\theta_2,...\theta_s$ модельной функции распределения неизвестны. Для измерения степени отклонения эмпирического распределения от модельного этот критерий использует статистику χ^2:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'}, \qquad (2)$$

где k – количество интервалов группировки выборочных значений; n_i – эмпирическая частота i-го интервала; n_i' – рассчитанная теоретическая частота i-го интервала.

Данная статистика имеет χ^2-распределение с k-s-1 степенями свободы.

Для вычисления n_i', i=1,2…k необходимо сначала получить выборочные оценки $\bar{\theta}_1,\bar{\theta}_2,...\bar{\theta}_S$ параметров $\theta_1,\theta_2,...\theta_s$ модельной функции распределения. Гипотеза (1) не отклоняется, если выполняется условие $\chi^2 < \chi^2_{\alpha/2}(k-s-1)$, где α – заданный уровень значимости.

Критерий Колмогорова-Смирнова позволяет выполнять проверку гипотезы (1) в условиях, когда модельная функция $F_{mod}(x) = F_0(x)$ известна полностью, то есть не зависит от неизвестных параметров. Статистики критерия Колмогорова-Смирнова определяются следующим образом:

$$D_n = \underbrace{\sup}_{x \in R^1} \left| \bar{F}^{(n)}(x) - F_0(x) \right|; \qquad (3)$$

$$D_n^+ = \underbrace{\sup}_{x \in R^1} \left(\bar{F}^{(n)}(x) - F_0(x) \right); \qquad (4)$$

$$D_n^- = \underbrace{\sup}_{x \in R^1} \left(F_0(x) - \bar{F}^{(n)}(x) \right). \qquad (5)$$

Статистики $\sqrt{n}D_n$ и $\sqrt{n}D_n^-$ являются статистиками критериев Колмогорова и Смирнова соответственно.

Для практического использования критерия Колмогорова-Смирнова статистики (3)–(5) представляются в виде:

$$D_n = \max\left(D_n^+, D_n^-\right);$$
$$D_n^+ = \underbrace{\max}_{1\le i\le n}\left(\frac{i}{n} - t_i\right); \qquad (6)$$
$$D_n^- = \underbrace{\max}_{1\le i\le n}\left(t_i - \frac{i-1}{n}\right),$$

где $t_i = F_0(x_i)$ – значение гипотетической функции распределения в i-й точке вариационного ряда.

Сначала проверим распределение значений исследуемых рядов по нормальному закону, то есть проверим нулевую гипотезу, в случае отклонения нулевой гипотезы можно будет перейти к проверке гипотез о других законах распределения.

Проверка базируется на сравнении данных наблюдений и теоретических частот. Обычно они отличаются, поэтому необходимо исследовать, является ли случайной эта разница. Если разница не значимая, то гипотезу о том, что ряд распределен по нормальному закону отклоняют.

Были построены с помощью программы STATISTICA гистограммы частот для рассматриваемых рядов, которые оказались близки к кривой нормального распределения. Проведено тестирование гипотезы о нормальном распределении с использованием критериев Пирсона и Колмогорова-Смирнова. Вычисленные значения статистик с использованием критерия Колмогорова-Смирнова представлены в таблице 1.

Поскольку значения P-Value для всех рядов больше 0.1, то гипотеза о том, что эти выборки взяты из генеральных совокупностей, распределенных по нормальному закону, соответствует действительности с вероятностью 0,95.

Таблица 1

Вычисленные значения статистик критерия Колмогорова-Смирнова

Статистика	X1	X2	X3
D_n^+	0.058	0.078	0.054
D_n^-	0.047	0.039	0.076
D_n	0.058	0.078	0.076
P-Value	0.916	0.637	0.677

Таким образом, установлено, что значения коэффициентов выхода кондиционной продукции для трех марок продукции распределены по нормальному закону, и следовательно, можно перейти к синтезу ARIMA-модели.

Идентификация и оценка параметров модели ARIMA (p, d, q)

Методика идентификации хорошо описана в [4]. На первом шаге необходимо выяснить, является ли исследуемый временной ряд стационарным. Для этого анализируются выборочные автокорреляционные функции.

Если их поведение доказывает нестационарность процесса, то предпринимается попытка приведения данного ряда к стационарному путем взятия последовательных разностей и исследования поведения тех же функций для новых рядов.

Если автокорреляционная функция ряда быстро затухает, значит достигнут необходимый порядок разницы d.

Затем, анализируя для стационарного ряда поведение выборочных автокорреляционной и частной автокорреляционной функций, принимается решение о количестве параметров оператора авторегрессии p и оператора скользящего среднего q.

Значения выборочных оценок автокорреляций $r(k)$ и частных автокорреляций $\Phi(k,k)$ этих рядов, вычислены с помощью соотношений (7 – 9) и в графическом виде представлены для ряда X3 на рис. 4 и 5.

$$r(k) = \gamma(k)/\gamma(0); \qquad (7)$$

$$\gamma(k) = \frac{1}{N}\sum_{n=1}^{N-k}\left(X(n)-\overline{X}\right)\left(X(n+k)-\overline{X}\right),, \qquad (8)$$

где k – лаг, а $\overline{X}$ – средний уровень ряда$

$$\Phi(k,k) = \begin{cases} \dfrac{r(k) - \sum\limits_{j=1}^{k-1}\Phi(k-1,j)r(k-j)}{1-\sum\limits_{j=1}^{k-1}\Phi(k-1,j)r(j)}, & k>1 \\ r(k), & k=1 \end{cases}, \qquad (9)$$

где $\Phi(k,j) = \Phi(k-1,j) - \Phi(k,k)\Phi(k-1,k-j),$
$j = 1, 2..., k-1.$

Рассмотрим этапы идентификации модели для ряда X3.

Аналогичные исследования проведем для рядов X1, X2 и результаты занесем в табл. 2.

Необходимо отметить, что оценки автокорреляций, которые лежат в основе идентификации, могут иметь достаточно большие дисперсии и быть автокоррелированными.

По этой причине отсутствует строгое соответствие между теоретической и выборочной автокорреляционными функциями. Это обуславливает существенные сложности при определении параметров p, d и q.

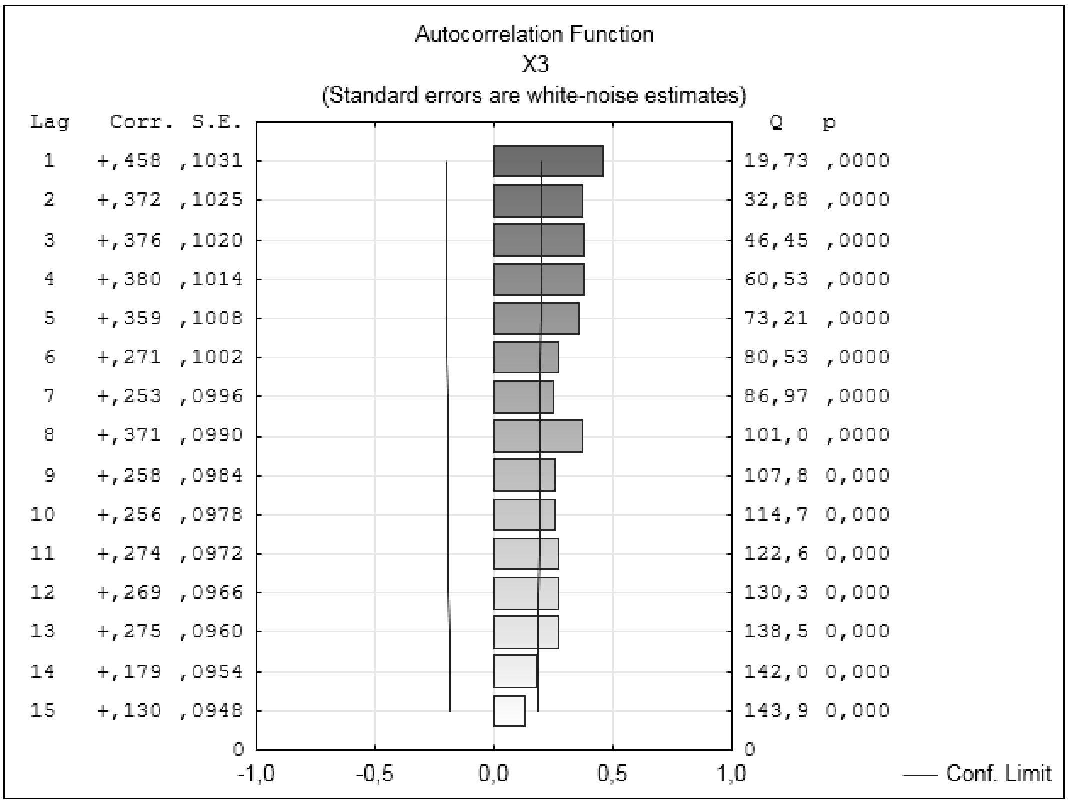

Рис. 4. Автокорреляционная функция для ряда X3

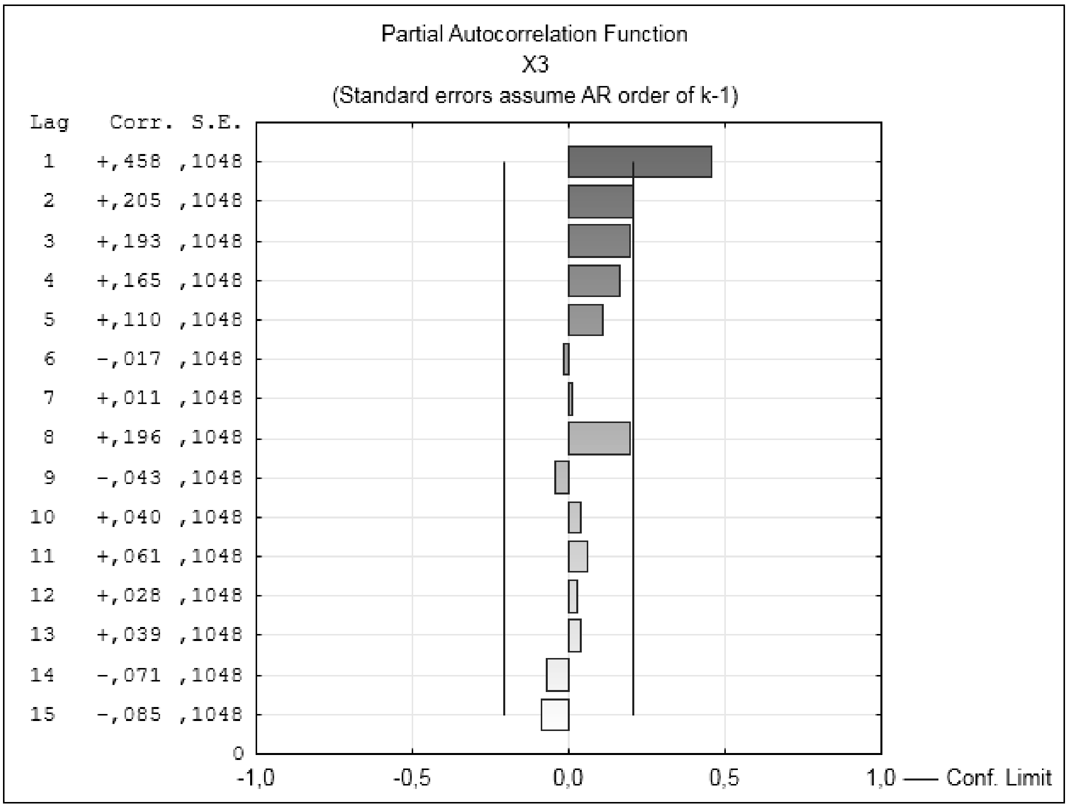

Рис. 5. Частная автокорреляционная функция для ряда X3

Поэтому на данном этапе целесообразно выбрать несколько подходящих моделей, а решение о «наилучшей» модели можно будет принять после получения некоторой количественной информации о пригодности выбранных моделей для прогнозирования. Кроме того, наличие нескольких альтернативных моделей дает возможность построения селективных или комбинированных моделей прогнозирования, которые позволяют повысить точность прогнозов.

В связи с этим, для каждого исследуемого ряда было выбрано по две структуры модели, которые не противоречат результатам проведенного анализа исходных данных.

Существует много разных методов оценивания моделей временных рядов. Наибольшее распространение принял подход к оцениванию ARIMA-модели, который заключается в получении начальных оценок параметров с помощью общего метода, основанного на использовании уравнений Юла-Уоккера для оценивания параметров оператора авторегрессии, и алгоритма Ньютона-Рафсона для оценивания параметров оператора скользящего среднего [4], с последующим уточнением оценок с помощью метода Марквардта, с помощью которого и было выполнено оценивание рассматриваемых в данной работе моделей.

Результаты оценивания приведены в табл. 2.

Таблица 2

Оценки параметров модели

Ряд	X1				X2				X3			
Модель ARIMA (p,d,q)	(0,0,2)	σ	(2,0,0)	σ	(0,1,1)	σ	(3,1,0)	σ	(0,1,1)	σ	(2,1,0)	σ
Константа	0,874		0,874		-0,009		-0,001		0,001		0,001	
p(1)			0,067	0,1047			-0,566	0,1032			-0,543	0,1046
p(2)	-		0,237	0,1050	-		-0,446	0,1091	-		-0,320	0,1043
p(3)	-				-		-0,313	0,1035	-			
q(1)	-0,039	0,1060	-		0,740	0,0701	-		0,832	0,0772	-	
q(2)	-0,223	0,1004	-		-		-		-		-	

Чтобы убедиться в том, что разработанные модели достоверно отображают свойства исходных данных, то есть, что они являются адекватными, нами использовалась проверка, которая основана на использовании так называемого совокупного критерия согласия (критерий Бокса-Пирса) [4], который базируется на анализе остатков. Этот критерий определяется следующим соотношением:

$$Q = N\sum_{l=1}^{L} r^2_{\bar{a}_t}(l), \qquad (10)$$

где $r_{\bar{a}_t}(l)$, $l = 1,2,...,L$ – оценки автокорреляций остатков; N – объем выборки остатков; L – максимальная задержка (на практике достаточно взять L=20).

Анализ графиков остатков для рассматриваемых моделей и их автокорреляций дает возможность сделать вывод, что последовательности остатков образуют процессы типа «белый шум» с нулевым математическим ожиданием и ограниченной дисперсией. Это служит подтверждением гипотезы об адекватности моделей. Рассчитанные значения критерия и его критические значения для уровня значимости 0,05 приведены в табл. 3.

Таблица 3

Вычисленные и критические значения статистики Бокса-Пирса

Модель	f	Q	Qкр(f, 0.05)
X1(2,0,0)	18	12.674	28.9
X2(0,1,1)	18	7.959	28.9
X3(0,1,1)	18	8.159	28.9

Поскольку для всех моделей справедливо отношение $Q < Q_{кр}$, то с вероятностью 0,95 можно утверждать, что рассматриваемые модели адекватны.

Разработка адаптивного метода комбинированного прогнозирования

Пусть имеется k прогнозных моделей одного и того же показателя Y. Эти модели различаются тем, что с разной степенью адекватности описывают некоторые стороны моделируемого показателя. Эти модели дополняют одна другую, поэтому получается спектр прогнозных оценок, что позволяет построить комбинированный прогноз, который будет превосходить лучший из начальных прогнозов.

При этом следует ввести некоторую меру P_j, которая позволит измерять степень точности прогнозов. Меру выберем так, чтоб ее значение уменьшалось с возрастанием точности прогнозов.

В качестве $P_j(t)$ можно взять сглаженный средний квадрат ошибки прогноза:

$$P_j(t) = \bar{\varepsilon}_j^2(t) = \\ = \sum_{i=0}^{t} \gamma(1-\gamma)^i \varepsilon_j^2(t-i) = (1-\lambda)\bar{\varepsilon}_j^2(t-1) + \gamma\varepsilon_j^2(t), \qquad (11)$$

где $\varepsilon_j(t) = Y(t) - \hat{Y}_j(t)$, $\gamma = \overline{0.3 - 0.7}$.

Представим комбинированный прогноз в виде

$$\hat{Y}_0(t) = B_1(t)\hat{Y}_1(t) + B_2(t)\hat{Y}_2(t) + ... + B_k(t)\hat{Y}_k(t), \qquad (12)$$

где $\hat{Y}_i(t)\left(i = \overline{1,k}\right)$ – частные прогнозы, $B_i\left(i = \overline{1,k}\right)$ – параметры обобщения, k — количество частных моделей прогнозирования. Тогда оценки параметров $B_i\left(i = \overline{1,k}\right)$ из соотношения (12) найдем путем решения системы уравнений:

$$B_i(t+1) = p(t)/P_i(t), \quad i = \overline{1,k}; \qquad (13)$$

$$\sum_{i=1}^{k} B_i(t+1) = 1. \qquad (14)$$

Определим параметр p(t), для чего подставим в (14) уравнение (13). Получим:

$$p(t) = \frac{\prod_{i=1}^{k} P_i(t)}{\prod_{i=2}^{k} P_i(t) + \prod_{\substack{i=1 \\ i\neq 2}}^{k} P_i(t) + \ldots + \prod_{i=1}^{k-1} P_i(t)} = \frac{\prod_{i=1}^{k} P_i(t)}{\sum_{n=1}^{k} \prod_{\substack{i=1 \\ i\neq n}}^{k} P_i(t)}. \qquad (15)$$

Затем получим:

$$B_i(t+1) = \prod_{j=1,\, j\neq i}^{k} P_j(t) \Bigg/ \sum_{n=1}^{k} \prod_{j=1,\, j\neq n}^{k} P_j(t), \quad i = \overline{1,k}. \qquad (16)$$

Сравнение качества прогнозов оценивалось с помощью среднего квадрата ошибки. Приведем результаты использования разработанного метода для прогнозирования ряда X3. Как частные модели были использованы две модели, разработанные для этого ряда раньше (0,1,1), (2,1,0). Прогнозирование выполнялось на проверочной последовательности, которая является продолжением во времени последовательности X3 (четвертый месяц) и которая не использовалась при разработке моделей.

Таблица 4

Анализ качества частных и комбинированного прогнозов

№ п/п	Тип прогноза	СКО
1	По модели (0,0,1)	0.001481
2	По модели (2,1,0)	0.001515
3	Комбинированный	0.001368

Как видно из приведенной таблицы, комбинированный прогноз имеет определенное преимущество над частными прогнозами.

Выводы

1. Синтезированы ARIMA-модели последовательностей значений коэффициентов выхода кондиционной продукции, которые позволят получить прогнозные значения этих коэффициентов, необходимых для реализации алгоритмов оптимальной загрузки производственного оборудования.

2. Получил дальнейшее развитие метод построения комбинированного прогноза, что даст возможность повысить точность прогнозирования по сравнению с применением частных моделей прогнозирования.

3. Полученные результаты будут использованы при имитационном моделировании процесса производства полупроводниковых материалов.

Список литературы

1. Оксанич И.Г. Математические методы в оперативном управлении производством полупроводниковых материалов / И.Г. Оксанич, И.В. Шевченко, Н. . Рилова // Вісник Кременчуцького національного університету імені Михайла Остроградського. Збірник наукових праць. – Кременчук, 2013. – № 6 (83). – С. 28-33.

2. Оксанич И.Г. Разработка критерия эффективности оперативных решений по управлению производством кремниевых структур / И.Г. Оксанич, Н.В. Рилова. – Вісник Кременчуцького національного університету імені Михайла Остроградського. Збірник наукових праць. – Кременчук, 2014. – № 5 (88). – С. 9-13.

3. Айвазян С.А. Прикладная статистика. Основы моделирования и первичная обработка данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М. : Финансы и статистика, 1983. – 471 с.

4. Бокс Дж. Анализ временных рядов. Прогноз и управление. Вып. 1. / Дж. Бокс, Г. Дженкинс; пер. с англ. А.Л. Левшина. – М. : Мир, 1974. – 406 с.

Поступила в редколлегию 19.03.2015

Рецензент: д-р техн. наук, проф. Д.И. Левинзон, Кременчугский национальный университет им. М. Остроградского, Кременчуг.

СИНТЕЗ ARIMA-МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ КОЕФІЦІЄНТІВ ВИХОДУ КОНДИЦІЙНИХ НАПІВПРОВІДНИКОВИХ МАТЕРІАЛІВ

Н.В. Рилова, І.Г. Оксанич

В статті запропоновано математичну модель для моделювання послідовностей коефіцієнтів виходу кондиційної продукції при виробництві напівпровідникових матеріалів, проведена ідентифікація структури моделі, оцінка параметрів і перевірена адекватність моделі. А також розроблено метод побудови комбінованого прогнозу, що дасть можливість підвищити точність прогнозування у порівнянні із застосуванням часткових моделей прогнозування. Розроблена модель та метод використовуються для оперативного управління виробництвом при прогнозуванні значень коефіцієнтів виходу кондиційної продукції.

Ключові слова: *ARIMA-модель, ідентифікація моделі, часова послідовність, гіпотеза, прогнозування.*

SYNTHESIS ARIMA-MODEL TO PREDICT THE COEFFICIENTS OF CONDITIONAL SEMICONDUCTOR MATERIAL OUTPUT

N.V. Rylova, I.G. Oksanych

The paper proposes a mathematical model to simulate the sequence of coefficients of conditional product output in the production of semiconductor materials. The model structure, parameter estimation were identified and the adequacy of the model was verified. And the method for combined prognosis enabling to get higher exact prognostication as compared to applied specific models of prognostication is developed. The developed model and method is used for operative management for predicting the coefficients of conditional products.

Keywords: *ARIMA-model, model identification, time sequence, hypothesis, prediction.*