

## АВТОМАТИЗАЦИЯ ПРОЦЕССОВ ИСПЫТАНИЯ, ОТЛАДКИ И ВЕРИФИКАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ В ЗАДАЧАХ УПРАВЛЕНИЯ ПРОМЫШЛЕННЫМИ ОБЪЕКТАМИ

Рассмотрена задача построения программно-алгоритмического комплекса (ПАК), предназначенной для испытаний, отладки и верификации математических моделей (ММ), используемых в различных задачах наблюдения и управления. ПАК включает в себя базу знаний ММ, программный генератор тестовых сигналов и анализатор качества функционирования моделей. Реализация системы осуществлена на основе web-технологий.

Program complex construction problem intended for mathematical models debugging and verifications used in various observation and control tasks is considered. The complex includes base of knowledge of mathematical models, the program generator of test signals and the analyzer of quality of models functioning. System realization is carried out on the basis of web-technologies.

### Введение и постановка задачи

Современные системы поддержки принятия решений и связанные с ними *аналитические информационные технологии* (АИТ) [3, 4, 7, 10] в своей основе базируются на *динамическом подходе* к анализу ситуаций. Данное замечание требует осознания достаточно очевидного факта – любое управленческое решение носит прогностический характер (по отношению ко времени исполнения и получения результата соответствующего управления). Для того чтобы оценить качество принимаемого управленческого решения, необходимо построить некоторый прогностический виртуальный сценарий развития многопараметрической ситуации, базирующийся на адекватной математической модели (или системы моделей) управления.

Прикладная математика предоставляет разработчикам систем управления обширный арсенал эффективных *математических моделей* (ММ), используемых для решения задач оценивания, идентификации, распознавания, прогнозирования и т.п. В частности, широко применяются методы вероятностно-статистического моделирования и связанные с ними алгоритмы статистической обработки данных (параметрического оценивания, проверки гипотез, многомерного статистического анализа и т.п.). Среди современных средств компьютерной математики следует указать технологии искусственных нейронных сетей, генетические алгоритмы, эволюционные методы и другие. Указанные подходы относятся к общей методологии *Data Mining* (DM, интеллектуальный анализ данных) и широко ис-

пользуются при решении задач поддержки принятия решения и оптимизации процессов управления [2, 4, 7].

Однако непосредственное применение существующих ММ и отвечающих им программных комплексов сопряжено с существенными трудностями. Проблема состоит в несоответствии статистической и динамической структур исходных данных (полученных в результате мониторинга состояния объекта управления и среды взаимодействия) совокупности ограничений, гарантирующих сохранение функциональной эффективности используемых ММ.

Как правило, результаты мониторинга образуют нестационарные нелинейные временные ряды, содержащие как систематические, так и случайные ошибки. При этом статистические характеристики погрешностей измерений заранее неизвестны и эволюционируют во времени. Возникает неадекватность между изменяющейся реальной ситуацией и описывающей ее моделью. В результате этого прямое применение априори заданных ММ в конкретных задачах управления часто приводит к существенному снижению точности оценки состояния объектов управления, достоверности прогноза и, в конечном счете, к существенным потерям эффективности формируемых решений.

Естественным выходом из создавшегося положения является периодическая перенастройка и отладка алгоритмов управления и соответствующего программного обеспечения. При этом динамика изменения протекающих процессов может быть столь значительной, что возникает необходимость в автоматической настройке и адаптации используемых моделей. Таким образом, возникает проблема создания специализированного *программно-алгоритмического комплекса* (ПАК), ориентированного на процессы автоматизированной *испытаний, отладки и верификации* (ИОВ) ММ, используемых в качестве основы для разработки современных средств управления сложными динамическими системами.

Современные подходы к задачам данного класса допускают возможность не только параметрической и структурной адаптации модели, но и выбора класса используемых ММ. При этом предполагается, что разрабатываемый программно-алгоритмический комплекс будет содержать *базы знаний* (БЗ) ММ, используемых в задачах управления сложными динамическими объектами.

Прикладная (и демонстрационная) часть настоящей статьи посвящена частному вопросу совершенствования процессов управления - решению задач последовательной оптимизации управления технологическими процессами промышленного предприятия.

## 2. Структура и функционирование программно-алгоритмического комплекса

Функциональная структура программно-алгоритмического комплекса ИОВ ММ представлена на рис. 1.

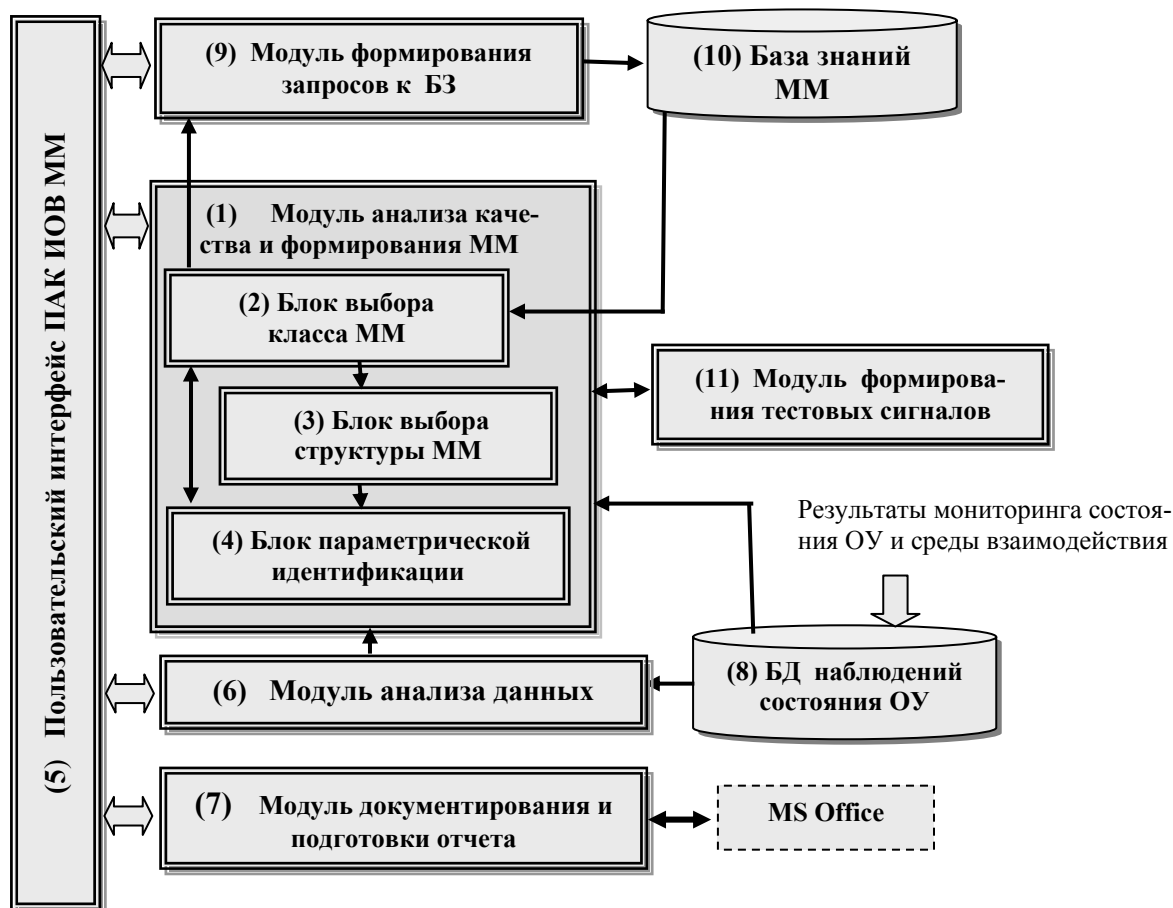


Рис 1. Функциональная структура ПАК ИОВ ММ

Центральным элементом ПАК является модуль анализа качества и формирования ММ (1), включающий в себя блок выбора класса ММ (2), блок выбора структуры ММ (3) и блок параметрической идентификации (4).

Исходными данными для функционирования модуля являются результаты анализа данных, полученные модулем (6), а также оперативные и ретроспективные данные, полученные в процессе мониторинга системы управления и хранящиеся в БД наблюдений (8). В процессе выбора и идентификации ММ могут использоваться тестовые последовательности (сигналы), формируемые модулем (11). В процессе выбора класса и структуры ММ используются запросы к базе знаний ММ (10), формируемых блоком (2) и модулем (9). Формирование задания на создание и адаптацию ММ, выбор опций и визуализация результата (качества моделирования) осуществляется с помощью пользовательского интер-

фейса (5). Результаты моделирования и последовательной адаптации используемых ММ автоматически документируются средствами модуля (7), связанного с программами MS Office.

Функционирование ПАК ИОВ ММ происходит следующим образом. Результаты оперативного мониторинга состояния объекта управления и среды взаимодействия поступают БД (8). Оперативная информация и заданный объем ретроспективных данных направляются в модуль анализа данных (6). В модуле (6) анализируются динамические и статистические свойства временных рядов наблюдений, используемые в качестве исходных сведений для предварительного выбора типа и структуры ММ, глубины ретроспективных данных, используемых для корректирующего обучения (параметрической и структурной идентификации ММ), выбора наиболее значимых предикторов (факторов влияния) и при решении других задач адаптации.

В частности, модуль анализа осуществляет дескриптивный статистический анализ, исследование динамических свойств, проверку гауссовости и независимости невязок наблюдений относительно аппроксимирующей модели, анализ корреляционных связей, обнаружение аномальных наблюдений, обнаружение мультиколлинеарности многомерных данных и т.п.

Непосредственно выбор и последовательная коррекция ММ осуществляются в модуле анализа качества и формирования ММ (1). Качество ММ определяется ее адекватностью текущей ситуации и оценивается с помощью критериев, приведенных, например, в [1, 6, 8, 11]. Для этого обучающую выборку ретроспективных данных целесообразно разделить на собственно обучающую, используемую в процессе структурной и параметрической идентификации, и тестовую, применяемую для оценки качества ММ.

Выбор класса и структуры ММ осуществляется итерационно, систематически (например, перебором возможных вариантов) или методом случайного поиска. Основные классы ММ, их базовые структуры, описания, алгоритмы и программы формирования, а также ограничения на их применение составляют содержание базы знаний ММ (10). Обращение к БЗ ММ осуществляется модулем формирования запросов (9). Инициация и формирование параметров запросов к БЗ реализуется в блоке выбора класса ММ (2) модуля (1). Основными классами ММ, хранящимися в БЗ, являются полиномиальные модели, представления функций в различных ортогональных базисах, модели с аддитивными стохастическими компонентами, временные ряды (авторегрессия, скользящее среднее, модель Бокса-Дженкинса), нейросетевые модели и т.п.

Параметрическая идентификация осуществляется на множестве обучающих (точнее, собственно обучающих) данных, обычно, путем использования традиционных мето-

дов статистического оценивания (например, методом наименьших квадратов (МНК), методом максимального правдоподобия, методом наименьших модулей и т.п.).

Верификация модели производится путем ее применения при решении базовых функциональных задач на множестве тестовых ретроспективных данных. Например, используя ММ для формирования прогноза, можно сравнить полученные оценки с результатами реального управления и, тем самым, оценить точность прогнозирования. При этом точность формируемых оценок является естественным критерием качества ММ, позволяющим выбрать наиболее адекватную модель управления.

В случаях, когда отсутствуют ретроспективные данные (например, на этапе проектирования), качество ММ может быть оценено путем подачи на ее вход тестовых последовательностей, и сравнения выхода модели с ожидаемым результатом. В качестве источника таких последовательностей используется модуль формирования тестовых сигналов (11), позволяющий поучить реализации детерминированных и стохастических временных рядов с заданными свойствами.

Результаты испытаний, верификации и отладки ММ регистрируются в форме текстового файла средствами модуля документирования и подготовки отчета (7).

### **Пример применения схемы последовательной верификации и коррекции ММ в задаче прогнозирования качества выходной продукции**

В качестве примера реализации прогнозирования с использованием комплекса ИОВ ММ рассмотрим вариант решения задачи оценки ожидаемого качества выходной продукции  $Y$  технологического процесса (ТП) в зависимости от значений выбранных управлений  $X_j$ ,  $j = 1, \dots, m$ . Пусть в качестве исходного класса ММ выбрана совокупность простейших линейных моделей, параметрическая идентификация которых осуществляется на основе классической методологии многомерной линейной регрессии. Многомерная регрессия (multiple regression, MR) представляют собой обобщение одномерной линейной регрессии на случай, когда имеется несколько взаимосвязанных предикатов  $X$ , определяющих структуру базовой модели в виде:

$$\bar{Y}_i = c_0 + c_1 X_{1i} + c_2 X_{2i} + \dots + c_m X_{mi} + \varepsilon_i, \quad i=1, \dots, n.$$

С точки зрения формирования линейного прогноза применение MR означает возможность оценивания влияния совокупности управляющих воздействий на свойства выходных потоков. При этом предполагается, что все  $X_i$  не являются случайными величинами и, следовательно, для  $\forall i=1, \dots, n$  формируемые управления не коррелированы с шумами системы, т.е.  $\text{cov}(\varepsilon_i, X_i)=0$ .

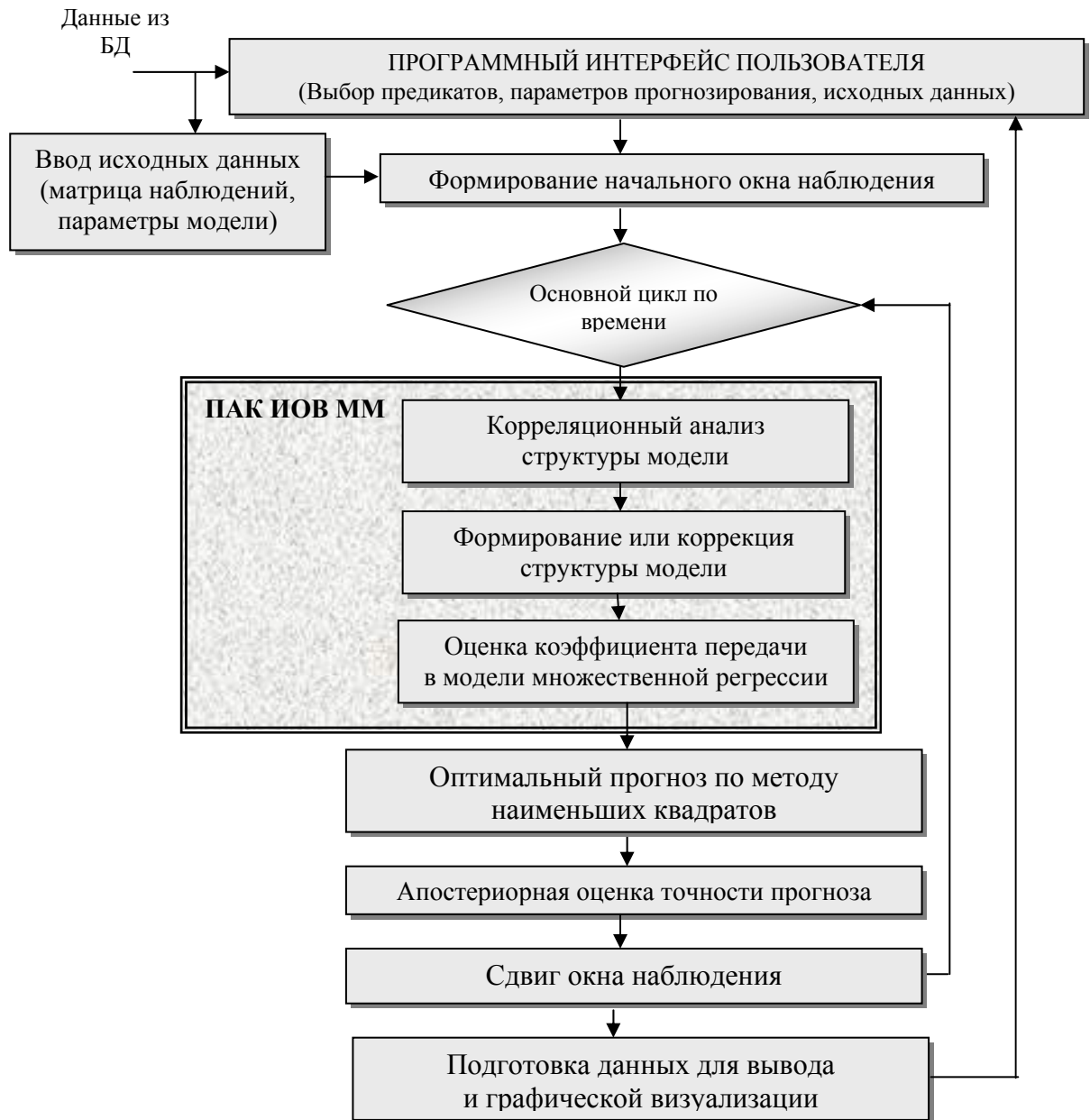


Рис. 3. Блок-схема программы прогнозирования с использованием ПАК ИОВ ММ на основе вычислительной схемы множественной регрессии

Кроме того, предполагается, что  $\forall Y_i, i=1, \dots, n$ , - независимые случайные величины, и, следовательно,  $\text{cov}(Y_i, Y_j) = E\{\varepsilon_i \varepsilon_j\} = 0$  для  $\forall i \neq j$ .

Используя матричную запись, введем вектор оценок коэффициентов регрессионной модели  $\tilde{c} = (\tilde{c}_0, \tilde{c}_1, \dots, \tilde{c}_m)$  и матрицу наблюдений

$$X = \begin{bmatrix} 1 & X_{11} & \dots & X_{m1} \\ 1 & X_{12} & \dots & X_{m2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{1n} & \dots & X_{mn} \end{bmatrix}$$

Тогда, в соответствии с МНК [5, 9], искомая оценка параметров модели, минимизирующая сумму квадратов невязок наблюдений  $\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\tilde{c})^T (Y - X\tilde{c})$  будет иметь вид  $\tilde{c} = (X'X)^{-1} X'Y$ .

Переходя к построению модуля прогноза с использованием ПАК ИОВ ММ (в рамках методологии MR), отметим, что реализация MR-прогнозатора на множестве нестационарных и нелинейных процессов предполагает использование скользящего окна наблюдения. Скользящее окно используется в качестве совокупности исходных данных, применяемых для построения текущих оценок коэффициента передачи МНК-фильтра  $C = (X'X)^{-1} X'$  и последовательной коррекции структуры базовой модели.

Блок-схема программной реализации модуля MR-прогноза с ИОВ ММ представлена на рис. 3.

Формирование задания и выбор имени загружаемого массива данных осуществляется средствами программного интерфейса пользователя, либо, в случае его отсутствия, процедурой загрузки и формирования исходных данных.

Далее управление передается основной процедуре прогнозирования. Прогнозирование выходных параметров осуществляется в цикле по времени на скользящем интервале (окне) наблюдения. При этом на каждом шаге скольжения осуществляется коррекция значений соответствующих ковариационных структур и средних значений.

Выбор перечня предикатов из общего списка параметров, подлежащих мониторингу, осуществляется одним из трех способов:

вручную, без использования ПАК ИОВ ММ на основе опыта оператора;

автоматически, средствами ПАК ИОВ ММ, на основе корреляционного анализа данных на скользящем окне наблюдения (выбирается заданное число параметров управления и состояния, наиболее коррелированных к прогнозируемому параметру);

автоматически, средствами ПАК ИОВ ММ, как и в предыдущем случае, однако множество параметров управления выбирается исходя из условия превышения коэффициентами корреляции заданного уровня.

Прогнозируемое значение качества выходной продукции определяется путем векторного перемножения данных текущего мониторинга на коэффициент передачи МНК-фильтра. Значения прогноза, результаты мониторинга выходных параметров и соответствующих им ошибки накапливаются в процессе скользящего наблюдения и служат для формирования усредненных оценок качества прогноза. В роли показателей качества

обычно используются среднее квадратическое отклонение (ско) и среднее значение относительной ошибки прогноза.

Внешний вид пользовательского интерфейса прогнозатора представлен на рис. 4.

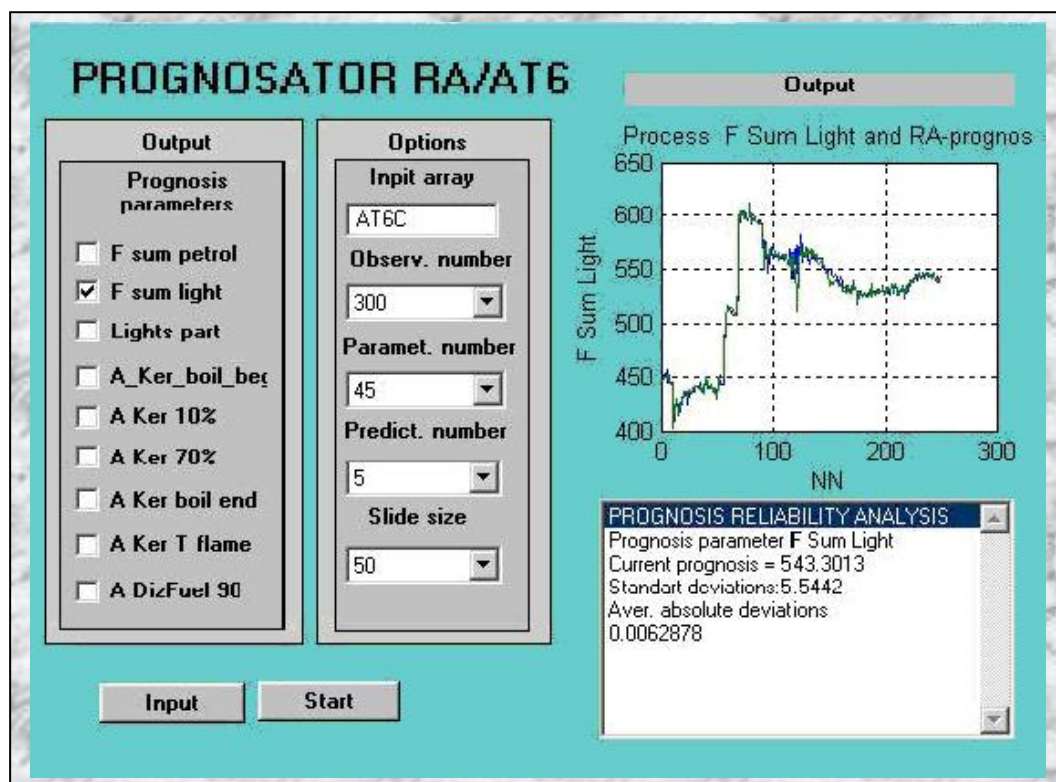


Рис. 4. Внешний вид интерфейса MR-прогнозатора

Средствами панелей Prognosis parameters выбираются выходные параметры технологического процесса, подлежащие прогнозированию. При этом число прогнозируемых параметров должно быть меньше или равно числа предикатов модели. Настройка процесса ввода данных и скользящего окна наблюдения производится средствами панели Options. Здесь осуществляется выбор:

имени массива входных данных;

число ретроспективных наблюдений, используемых при решении задачи прогноза;

общее число параметров управления, состояния и выхода;

размер скользящего окна наблюдения;

среднюю задержку реакции свойств выходных потоков на изменения управляющих воздействий.



Задание на прогнозирование, выбранные параметры и имя файла с массивами исходных данных вводятся в программу кнопкой Input, запускающей процедуру ввода данных.

Таблица 1. Результаты прогнозирования параметра A\_BegBoilKer при ручном выборе предикторов

```

*****
КОНТРОЛИРУЕМЫЙ ПАРАМЕТР A_BegBoilKer
Количество выбранных предикторов NF = 8
Перечень выбранных предикторов:
'F Input oil #14 corr = -0.53672'
'P down K2 #24 corr = -0.61301'
'T from K2-K6 #25 corr = 0.51929'
'T From K2-K7 #26 corr = -0.48635'
'T down K2 #27 corr = -0.45539'
'T down K6 #28 corr = 0.15579'
'T down K7 #29 corr = -0.43831'
'T CO1 T3.1 #30 corr = -0.37856'
*****
Средняя относительная погрешность прогнозирования = 0.010177
Средне квадратическая погрешность прогнозирования = 3.125

```

В случае если введенные данные соответствуют выбору пользователя, кнопкой Start осуществляется запуск основной программы прогнозатора, осуществляющей решение поставленной задачи. Результаты расчетов выводятся в окнах Output интерфейса. При этом верхнее окно обеспечивает графический, а нижнее – текстовый вывод результатов прогноза.

В качестве примера рассмотрим задачу прогнозирования такой характеристики выходной продукции (товарного керосина) как значе-

нии начала кипения бензиновой фракции (A\_BegBoilKer). Выходная продукция формируется на установке атмосферной перегонки нефти. В качестве предикторов могут быть использованы 45 параметров ТП (температуры, давления, потоки и т.п.), полученные в процессе мониторинга его состояния. Очевидно, что многие из этих параметров не влияют или не существенно влияют на значение прогнозируемого параметра. Кроме того, в зависимости от условий проведения ТП, те

или иные параметры оказываются более или менее значимыми с точки зрения точности формируемой прогнозстической оценки.

Наличие модуля ИОВ ММ позволяет на каждом шагу прогнозирования (и управления) оценить степень значимости каждого из 45 предикатов и выбрать те из них, которые позволяют получить наиболее высокую точность прогноза.

Сравним рассматриваемые варианты. В первом случае используется ручной выбор структуры модели с 8 предикторами (на основе опыта работы операторов), в каче-

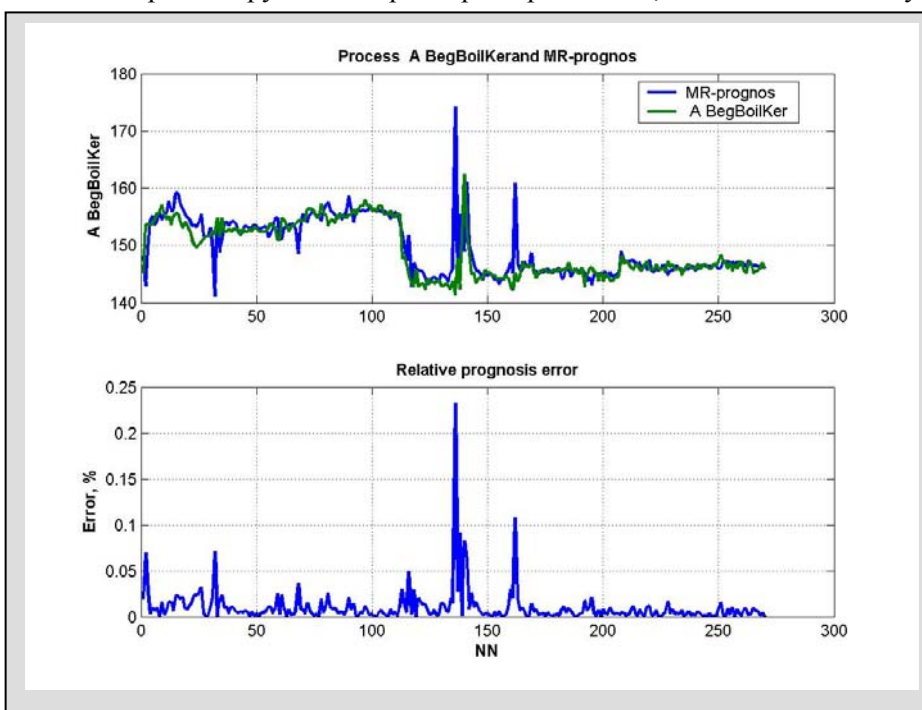


Рис. 5. Изменение выходного параметра, результат его прогнозирования и относительная ошибка прогноза при ручном выборе параметров модели

стве которых использовались потоки острого орошения (применяются для охлаждения верха ректификационной колонны), поток горячей струи (разогрев низа колонны), давление внутри колонны и т.д. Параметры и результаты управления приведены ниже, в таблице 1, а графики изменения выходного параметра, результат его прогнозирования и относительная ошибка прогноза при ручном выборе параметров модели – на рис. 5.

Второй и третий варианты прогнозирования качества выходной продукции приведены, соответственно, в таблицах 2 и 3. При этом, данные в табл. 2 получены для случая, когда селекция предикатов осуществляется по критерию выбора заданного числа наиболее значимых параметров, а в табл. 3 - без ограничения на численность предикатов, по критерию превышения значимости корреляции между предикатами и прогнозируемым параметром некоторого априорно выбранного уровня.

Таблица 2. Основные параметры и результаты при автоматической настройке ММ и заданном числе предикатов

```

*****
Исследуется ряд наблюдений за параметром/анализом A BegBoiKer
*****
КОНТРОЛИРУЕМЫЙ ПАРАМЕТР A BegBoiKer
Количество выбранных предикторов NF = 8
Перечень выбранных предикторов:
'T 1COroshK2 corr=0.56826 #31'
'F CO2 outK2 corr=-0.57188 #3'
'F мазута K2 corr=-0.57524 #16'
'P down K2 corr=-0.61301 #24'
'F180-240 K6 corr=-0.67819 #17'
'T up K2 corr=0.68563 #8'
'F CO1 outK2 corr=-0.69794 #2'
'F OOrosh K2 corr=-0.72754 #10'
*****
Средняя относительная погрешность прогнозирования = 0.0085657
Средне квадратическая погрешность прогнозирования = 2.1273

```

Таблица 3. Основные параметры и результаты при автоматической настройке ММ и заданном уровне корреляции предикатов и прогнозируемого параметра

```

Исследуется ряд наблюдений за параметром/анализом A BegBoiKer
*****
КОНТРОЛИРУЕМЫЙ ПАРАМЕТР A BegBoiKer
Уровень отсечки корреляций CorLevel = 0.5
Количество выбранных предикторов NF = 13
Перечень выбранных предикторов:
'T COrosh2 K2 corr=-0.51842 #33'
'T fromK2-K6 corr=0.51929 #25'
'F Input oil corr=-0.53672 #14'
'F oil-benzK1 corr=-0.54324 #19'
'P press K2 corr=-0.56307 #23'
'T 1COroshK2 corr=0.56826 #31'
'F CO2 outK2 corr=-0.57188 #3'
'F мазута K2 corr=-0.57524 #16'
'P down K2 corr=-0.61301 #24'
'F180-240 K6 corr=-0.67819 #17'
'T up K2 corr=0.68563 #8'
'F CO1 outK2 corr=-0.69794 #2'
'F OOrosh K2 corr=-0.72754 #10'
*****
Средняя относительная погрешность прогнозирования = 0.0093335
Средне квадратическая погрешность прогнозирования = 2.3254

```

Оценка самой матрицы корреляций осуществляется на основе данных из скользящего окна ретроспективных данных, полученных на L предыдущих шагах управления.

Графики изменения выходного параметра (зеленая линия), результаты его прогнозирования (синяя линия) и относительная ошибка прогноза для обоих вариантов ИОВ модели (нижний график) приведены, соответственно, на рис. 6 и 7.

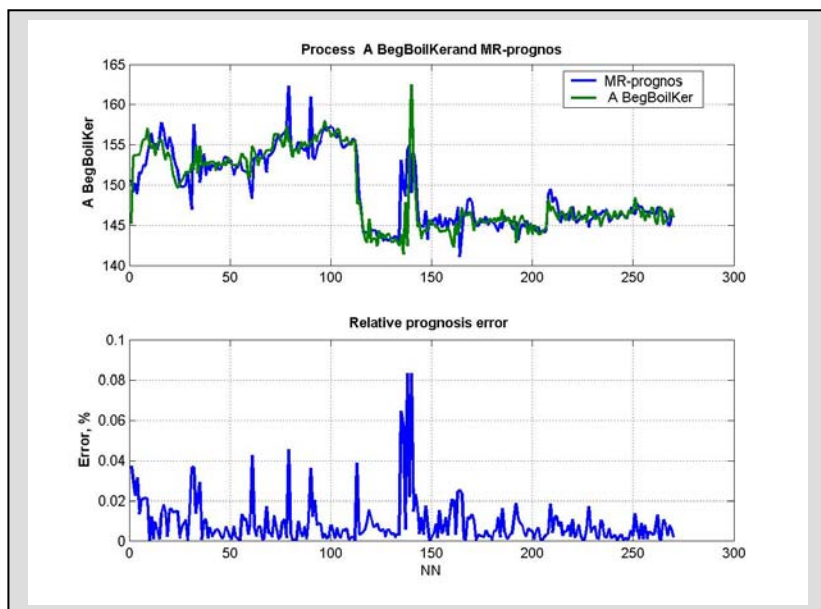


Рис. 6. Изменение выходного параметра, результат его прогнозирования и относительная ошибка прогноза при ИОВ модели с выбором заданного числа параметров

Сравнение полученных результатов, приведенных в таблицах 1-3, позволяет сделать вывод о том, что последовательная верификация и коррекция ММ позволяет существенно улучшить качество прогноза. Так, например, применение модуля ИОВ с коррекцией ММ первого типа позволяет снизить относительную погрешность прогноза на 20%, а с коррекцией ММ второго типа – на 10%.

Повышение качества прогноза за счет снижения ско для первого и второго типа коррекции с использованием модуля ИОВ ММ составляет, соответ-

ственно, 32% и 26%. Более низкий уровень выигрыша для коррекции второго типа связан со слишком низким значением выбранного порога корреляции ( $r=0,5$ ). В этом случае в структуру ММ попали параметры, которые несут в себе существенную стохастическую компоненту и относительно

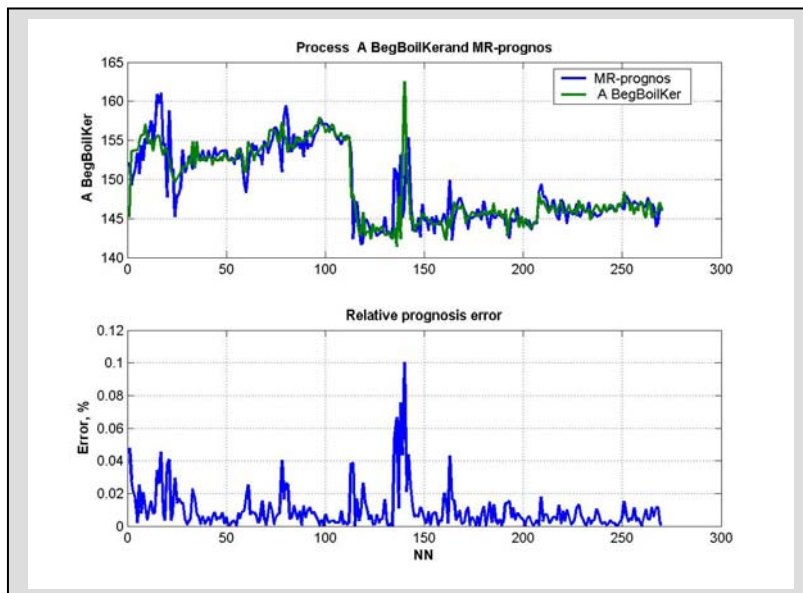


Рис. 7. Изменение выходного параметра, результат его прогнозирования и относительная ошибка прогноза с автоматическим выбором числа параметров с уровнем коэффициента корреляции, превышающим заданный порог

ско, определяемое стохастической природой наблюдений.

## Заключение

Программный комплекс ИОВ ММ является унифицированным изделием и пригоден для использования в любых предметных областях, связанных с разработкой и использованием математических моделей сложных динамических систем.

Элементы данного комплекса апробированы при построении математических моделей, используемых в задачах оптимизации управления технологическими процессами. Дальнейшее направление работ в рамках указанной проблемы связано с решением таких вопросов, как создание вариантов реализации ПАК ИОВ М, основанных на принципах самоорганизации и искусственного интеллекта.

*Работа выполнена в ОАО специализированная инжиниринговая компания "Севзап-монтажавтоматика" (ОАО СПИК СЗМА) и в Санкт-Петербургском институте информатики и автоматизации РАН.*

*Контактный телефон +7(812)3505885.*

*E-mail: Vladimir\_Nikitin@szma.com, musayev@yandex.ru.*

## Список литературы

1. Вознюк М. А., Мусаев А. А., Елшин А. В. Теоретические основы квалиметрии информационных систем. - СПб.: МО РФ, 1999. - 108с.
2. Дюк В., Самойленко А. Data Mining: Учебный курс. – СПб.: Питер, 2001. – 366с.
3. Карпов Е. А., Мусаев А. А., Шерстюк Ю. М. Многоцелевая аналитическая информационная система. Методология создания и основные проектные решения. – СПб.: МО РФ, 2000. – 143с.
4. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах // Открытые системы, 1997, №4. – С. 41–44.

небольшой объем полезных сведений об изменении прогнозируемого параметра. Данный недостаток может быть устранен путем включение значения порога селекции (или численности сохраняемых предикторов модели) в число параметров, подлежащих верификации и коррекции.

Не менее важно в указанный список корректируемых параметров включить размер скользящего окна  $L$ , на котором осуществляется верификация и уточнение модели. Проблема состоит в выборе размера обучающей выборки, обеспечивающего минимум суммарной ошибки, включающей в себя смещение, обусловленное динамическими погрешностями наблюдения, и

5. Линник Ю. В. Метод наименьших квадратов и основы теории обработки наблюдений М.: Физматгиз, 1958.- 349с.
6. Морозов Л. М., Петухов Г. Б., Сидоров В. Н. Методологические основы теории эффективности. - Л.: ВИКИ им. А. Ф. Можайского, 1982. –236с.
7. Мусаев А.А. Алгоритмы аналитического управления производственными процессами. - «Автоматизация в промышленности», 2004, №1, с. 30-35.
8. Охтилев М. Ю., Соколов Б. В., Юсупов Р. М. Интеллектуальные технологии мониторинга состояния и управления структурной динамикой сложных технических объектов. – М: Наука, 2006. - 410с.
9. Рао С. Р. Линейные статистические методы и их применение /Пер. с англ. под ред. Ю. В. Линника. - М.: Наука, 1972. - 591с.
10. Сахаров А. А. Концепции построения и реализации информационных систем, ориентированных на анализ данных // Системы управления базами данных, 1996, №4. – С. 55–70.
11. Элементы теории испытаний и контроля технических систем /Городецкий В. И., Дмитриев А. К., Марков В.М. и др. Под ред. Р. М. Юсупова. – М.: Энергия, 1978. –191с.