

Министерство образования Республики Беларусь
Министерство науки и высшего образования Российской Федерации
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет»

А. И. Якимов, Е. А. Якимов, Е. М. Борчик

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ ДЛЯ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ПРОИЗВОДСТВЕННЫХ СИСТЕМ

Могилев
«Белорусско-Российский университет»
2021

УДК 004.94
ББК 32.973.26-02
Я45

Рекомендовано к изданию ученым советом Белорусско-Российского университета
«30» декабря 2020 г., протокол № 6

Р е ц е н з е н т ы :

д-р техн. наук, доц. *А. В. Аверченков*;
д-р техн. наук, доц. *А. Ю. Влагова*

Якимов, А. И.

Я45 Интеллектуальный анализ данных для имитационного моделирования
производственных систем: [монография] / А. И. Якимов, Е. А. Якимов, Е. М.
Борчик. – Могилев: Белорус.-Рос. ун-т, 2021. – 184 с.
ISBN 978-985-492-256-0.

В монографии представлены результаты исследования методов анализа данных для имитационного моделирования производственных систем. Разработан метод извлечения полезной информации из числовых последовательностей данных, экспериментально исследованы метод сингулярного спектрального анализа, методы кластерного анализа, статистические методы на основе обобщенных законов распределения. Показано применение методов интеллектуального анализа данных в оптимизационной задаче распределения ресурсов в производстве готовой продукции на примере текстильного производства.

Предназначена для научных и инженерно-технических работников, решающих вопросы системного анализа, моделирования и автоматизации деятельности предприятий, а также для аспирантов, слушателей курсов повышения квалификации, студентов соответствующих специальностей.

УДК 004.94
ББК 32.973.26-02

ISBN 978-985-492-256-0

© Якимов А. И., Якимов Е. А.,
Борчик Е. М., 2021
© Межгосударственное образовательное
учреждение высшего образования
«Белорусско-Российский
университет», 2021

Перечень сокращений и условных обозначений

- АО – алгоритм оптимизации.
 ГА – генетический алгоритм.
 ДЭ – дифференциальная эволюция.
 ИМ – имитационная модель.
 ИО – имитация отжига.
 ИС – информационная система.
 ИЭ – имитационный эксперимент.
 КИ – контрольная инструкция.
 КИС – корпоративная информационная система.
 КЛ – колористическая лаборатория.
 КП – карта процесса.
 ЛВС – локальная вычислительная сеть.
 МИ – методологическая инструкция.
 ОП – отделочное производство.
 ОТК – отдел технического контроля.
 ОФ – отделочная фабрика.
 ПМ – программный модуль.
 ПО – программное обеспечение.
 ПП КТНС – производственный процесс крашения тканей набивным способом.
 ПП – производственный процесс (процесс производства).
 ПрО – производственная операция.
 ПТКИ – программно-технологический комплекс имитации.
 РК – руководство по качеству.
 РЧ – рой частиц.
 СМ – система моделирования.
 СМК – система менеджмента качества.
 СПЭС – сложная производственно-экономическая система.
 ССК – служба системы качества.
 СУЦ – складально-уборочный цех.
 СЭС – социально-экономическая система.
 ТБЦ – товаро-браковочный цех.
 ТНПА – технические, нормативные, правовые акты в области технического нормирования и стандартизации.
 ТП – технологический процесс.
 ТР – технологический режим.
 ТР ПП – технологический режим производственного процесса.
 ТФ – ткацкая фабрика.
 УПМ – управляющая программа моделирования.
 ЦТЛ – центральная технологическая лаборатория.

Введение

Внутрифирменное управление промышленным предприятием является одним из важнейших объектов исследования в теории управления социально-экономическими системами (СЭС). Повышение эффективности управления в СЭС достигается, в частности, использованием информационных систем, построенных на современных компьютерных технологиях. Корпоративная информационная система (КИС) промышленного предприятия, являясь средой информационной поддержки целенаправленной коллективной деятельности сотрудников предприятия, решает задачу эффективного управления материально-техническими, финансовыми, технологическими, трудовыми и интеллектуальными ресурсами предприятия для достижения заданных показателей производственно-экономической деятельности [26]. Она включает в себя различные программно-аппаратные платформы, универсальные и предметно-ориентированные программные приложения, интегрированные в единую информационно-однородную систему.

Существенными проблемами для повышения эффективности механизмов управления и качества принимаемых решений является обеспечение оперативности и достоверности информации в условиях структурных и параметрических изменений процессов функционирования предприятия и его окружающей среды.

При решении задач внутрифирменного управления для каждого подразделения фирмы определяются показатели, позволяющие оценивать экономические результаты его хозяйственной деятельности. В основном это стоимостные или объемные показатели выполняемых работ и, соответственно, затраты на исполнение работ (на оплату труда сотрудников подразделения, материальные и энергетические затраты, накладные расходы).

Эффективность управленческих решений значительно повышается, если используется имитационное моделирование производственной деятельности, позволяющее, с одной стороны, оценить на модели качество принятого решения, а с другой, является средством обучения сотрудников [64].

Оптимизация процессов функционирования производственной фирмы (предприятия) осуществляется путем декомпозиции из этих процессов отдельных компонентов и связанных с ними функций структурных подразделений, позволяющих выполнить формальное описание и построить модели оптимизации, именуемые оптимизационными моделями внутрифирменного управления. Задачи математического моделирования и оптимизации процессов деятельности предприятий в условиях структурных и параметрических изменений производственной и внешней среды, дальнейшего

совершенствования методов имитационного моделирования процессов функционирования промышленного предприятия для их системного представления с учетом принятия решений весьма актуальны при внедрении корпоративных информационных систем и дальнейшем развитии информационных технологий.

Интеллектуальный анализ данных [85], являясь современной концепцией анализа данных, позволяет обнаружить в «сырых» данных на входе и/или выходе имитационных моделей ранее неизвестные, нетривиальные, практически полезные, доступные интерпретации знаний, необходимые для принятия решений.

1 Краткий обзор методов интеллектуального анализа данных в имитационном моделировании

1.1 Анализ эксплуатации программно-технологического комплекса имитации сложных систем BelSim

Программно-технологический комплекс имитации сложных систем (ПТКИ) BelSim позволяет реализовать метод имитационного моделирования производственно-экономической деятельности промышленных предприятий [67]. В состав комплекса входит следующее программное обеспечение: построения функциональной модели системы на основе технологии IDEF0; интегрированной среды разработки приложений на языке C++; системы имитационного моделирования PSTL; планирования, проведения и обработки результатов имитационных экспериментов; решения оптимизационных задач; анализа и отображения данных; интеграции имитационных моделей в комплексную информационную систему; организации распределенных вычислений.

В подсистеме планирования, проведения и обработки результатов имитационных экспериментов обмен данными между элементами подсистемы организован на основе иерархических структур, физически представленных в виде файлов в формате XML, доступ к которым реализован через стандартный интерфейс XML DOM Level 3 с использованием библиотеки MSXML 4.0. Блок ввода/вывода информации реализован в виде отдельного приложения Experiment Designer на основе Microsoft .NET Framework. Блок создания плана эксперимента, блок формирования данных для отдельных опытов и блок предварительного статистического анализа реализованы в виде макросов DesignOfExperiment и ExperimentData в специализированном пакете статистического анализа Statistica 6.0.

Опыт использования ПТКИ BelSim в условиях промышленных предприятий с комплексной информационной системой потребовал обратить внимание на исследование и анализ статистических данных, накапливающихся в хранилищах. Это в свою очередь требует разработки специализированного программного модуля в составе ПТКИ для анализа статистических данных, используемых для имитационной модели (рисунок 1.1).

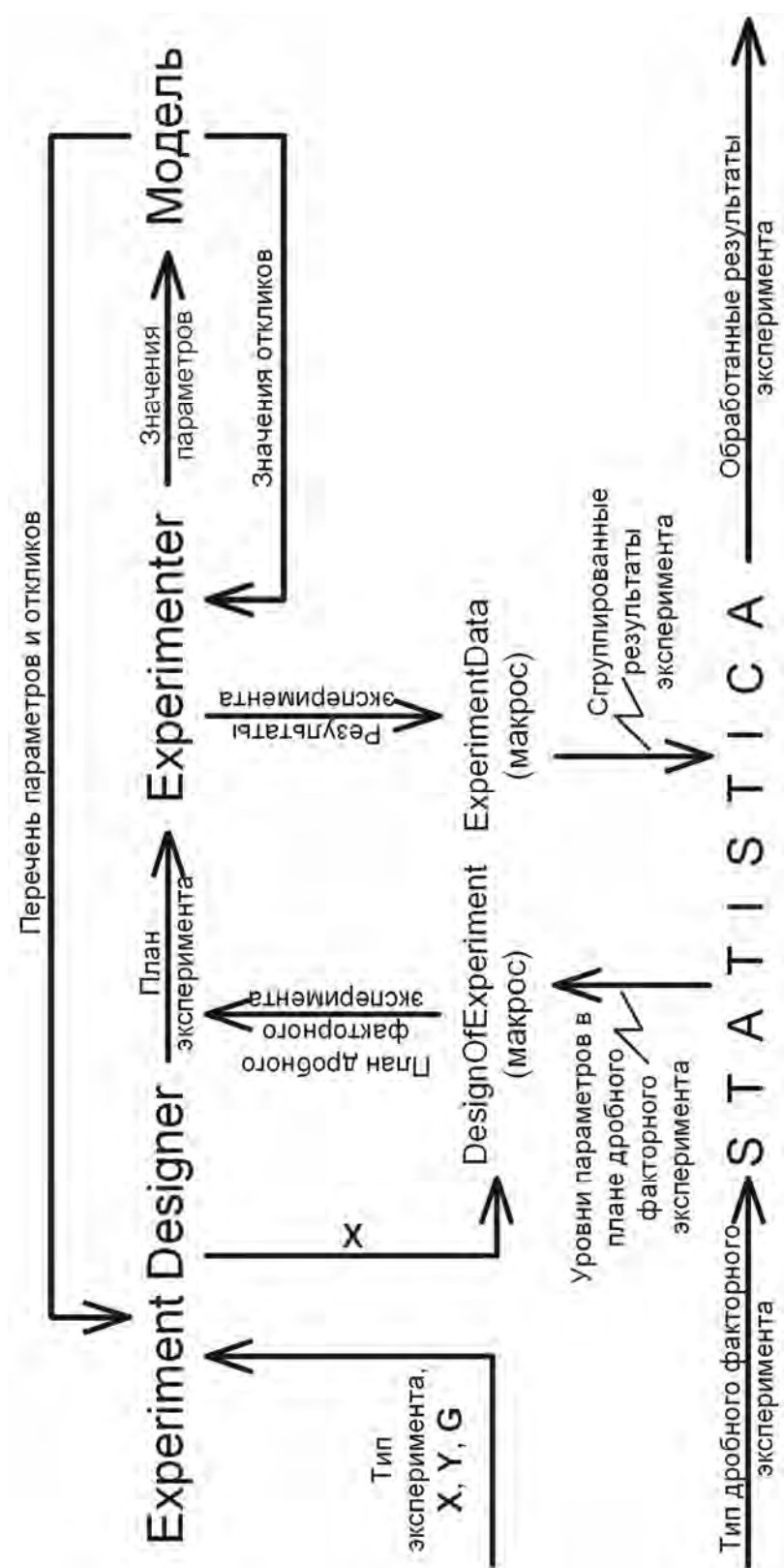


Рисунок 1.1 – Схема постановки имитационных экспериментов

Проведенные исследования обрывности нитей при их вытягивании на заводе полиэфирных нитей дают возможность утверждать, что анализ временных последовательностей статистических данных обрывности (рисунок 1.2) позволяет объективно оценить состояние технологических потоков и своевременно предпринять необходимые действия для их корректировки.

В то же время результаты анализа позволяют принять решение о проведении технических мероприятий по повышению эффективности технологических процессов получения готовой продукции, по снижению затрат при их производстве. Для ведения целенаправленной работы крайне необходимо на каждом конкретном этапе устанавливать адресность причин, выявлять факторы, вызывающие снижение качества технологического процесса.

Как показали проведенные исследования, анализ временных последовательностей текущей обрывности нитей является эффективным инструментом при обеспечении требуемого качества технологии.

Проведены исследования по выявлению свойств ИМ контрактов продажи, купли, на дилерское обслуживание для оценки влияния закона распределения исходных данных модели на отклики модели.

Эксперименты проведены с применением ПТКИ BelSim на основе базовой имитационной модели промышленного предприятия [67]. В ходе экспериментов все варианты файлов данных модели контракта продажи имеют одинаковые значения параметров, кроме одного, изменение которого в процессе эксперимента осуществляется случайным образом по заданному закону распределения.

В качестве изменяемого в ходе эксперимента выбран параметр «Реализация: Объем отгрузки» со средним значением, равным 500, стандартным отклонением, равным 100, для каждого варианта модели. Вид функции плотности распределения задан для первого варианта модели – 0 (равномерный), для второго – 1 (нормальный). В качестве откликов выбрана «Рентабельность продукции» (Y_3). Тип эксперимента – полный факторный. При планировании эксперимента для каждого варианта задано по 100 прогонов модели. Вывод по результатам исследований: закон изменения величин выходных откликов значительно зависит от закона изменения входных параметров.

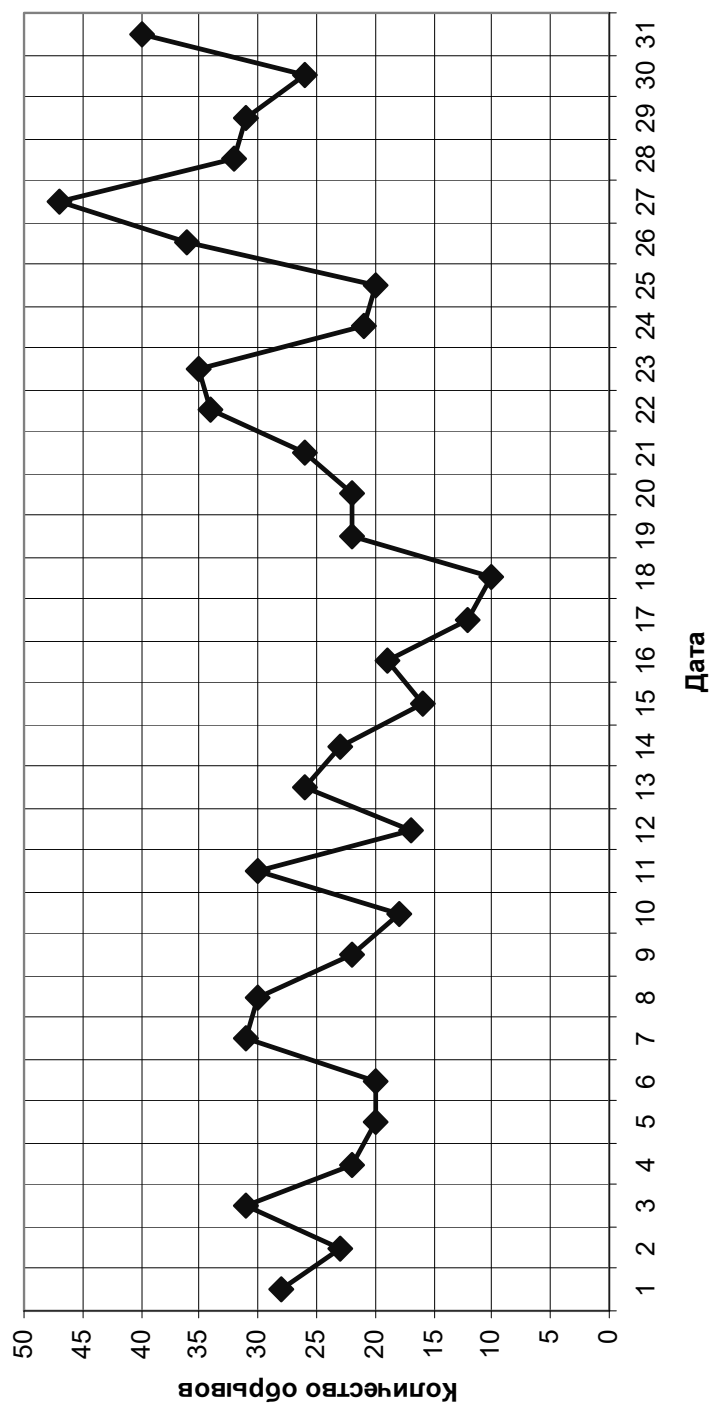


Рисунок 1.2 – Средняя обрывность нити по дням

Следует отметить, что в обычной реляционной базе данных (БД) все таблицы имеют два измерения – столбцы и строки, которые хранят мгновенный снимок предметной области. Применительно к темпоральной БД можно говорить о добавлении третьего измерения в существующие таблицы. Это позволяет хранить не только мгновенную картину, но и всю историю изменений. Основной тезис темпоральных систем состоит в том, что для любого объекта данных, созданного в момент времени t_1 и уничтоженного в момент времени t_2 , в БД сохраняются (и доступны пользователям) все его состояния во временном интервале $[t_1, t_2]$. В темпоральных БД выделяют четыре типа таблиц:

- 1) регулярные – обычные реляционные таблицы без поддержки временных атрибутов;
- 2) таблицы с временем транзакции – фиксирующие время внесения информации в БД;
- 3) таблицы с временем актуальности – показывающие время, когда произошло событие в реальности;
- 4) битемпоральные – таблицы, хранящие оба вышеуказанных временных атрибута.

Преобразование нетемпоральной БД в темпоральную дает следующие преимущества: возможность получить сведения о состоянии информации в любой момент времени в прошлом, а не только о ее текущем состоянии; возможность производить анализ динамики данных.

1.2 Исследование входных последовательностей данных в имитационном моделировании

Математическое моделирование позволяет получить информацию о возможных состояниях объекта прогноза (модели) и путях достижения этих состояний. При этом на основе имеющейся последовательности данных выстраиваются модели прогнозирования поведения объекта. Например, применяют простейшее прогнозирование, при котором полагают, что ситуация в будущем эквивалентна текущей:

$$Y(t+1) = Y(t).$$

Для повышения достоверности прогноза применяют модифицированные методы, например

$$Y(t+1) = Y(t) + [Y(t) - Y(t-1)].$$

Другим повышением достоверности прогноза является применение методов среднего или скользящего среднего. При этом усреднение выполняется на всем временном интервале:

$$Y(t+1) = \frac{1}{t}(Y(t) + Y(t-1) + \dots + Y(1))$$

или пользователь выбирает самостоятельно временной интервал прогноза. Используют также экспоненциальные средние, методы Хольта, Винтера, регрессионные методы прогнозирования, методы Бокса–Дженкинса [59].

Цели изучения временных последовательностей данных, иначе именуемых временными рядами, могут быть различными. Можно, например, стремиться предсказать будущее на основании знания прошлого, управлять процессом, порождающим ряд, выяснить механизм, порождающий ряд, или кратко описать характерные особенности ряда. На основании ограниченного количества информации о временном ряде конечной длины статистические методы позволяют делать выводы о вероятностном механизме, порождающем этот ряд, анализировать структуру, лежащую в его основе [4]. При этом рассматриваются регрессионный анализ, тренды и сглаживание, циклические тренды, линейные вероятностные модели с конечным числом параметров, сериальная корреляция, выборочные средние, ковариации и спектральная плотность [92].

Одной из основных проблем при подготовке и проведении стохастических экспериментов является правильный выбор модели входных данных, т. е. закона распределения вероятностей, который точно соответствует поведению случайных входных процессов. При этом стандартные законы распределения могут не соответствовать вероятностному поведению реальных входных процессов, особенно на границах распределения. Подстройка или редактирование формы подобранного распределения является трудным процессом из-за ограниченного числа параметров, доступных для управления формой распределения, и из-за отсутствия эффективного механизма для непосредственного управления формой подобранного распределения, одновременно изменяя соответствующий параметр распределения. Для решения этой проблемы используется обобщенное распределение Джонсона [100] или кривые Безье [92].

Для интерактивного моделирования входных данных с использованием семейства распределения вероятностей, основанных на кривых Безье, разработана графическая технология [101].

При построении модели входных данных используется байесовский подход с предварительной информацией о законе распределения

вероятностей. Предложено дисперсию случайных величин представлять двумя компонентами: одна дисперсия создается контролируемыми параметрами; вторая дисперсия связана с неопределенностью знания закона распределения контролируемых величин [102].

Адекватное моделирование входных процессов часто требует учета корреляции подобно аналитической модели марковских цепей [79, 81].

Для оценки эффективности деятельности стохастических систем в имитационном моделировании исследуют выходные данные в переходном процессе за конечный период времени и в установившемся решении за длительный период [95].

При исследовании поведения системы в установившемся режиме за длительный период времени по гистограмме распределения откликов моделирования оцениваются отклонения, которые являются мерой риска при принятии решения [96].

Проекты с дискретно-обобщенным моделированием в большей степени полагаются на высококачественные входные данные. Предлагается структурированная методология управления входными данными, охватывающая все процессы, включая идентификацию, сбор и подготовку входных данных для имитационной модели [99].

А. М. Лоу в [94] обращает внимание на проведение статистического анализа выходных данных для определения окончания имитационного моделирования с оцениванием периода разгона модели.

Кроме вышеизложенных вопросов, изучаемых при исследовании входных данных, в имитационном моделировании рассматриваются требования к точности генераторов случайных чисел для формирования последовательности входных данных с заданным законом распределения [91].

1.3 Поиск сложных неперiodических шаблонов в последовательностях данных

Продуктивным направлением анализа временных рядов на сегодняшний день является подход, связанный с преобразованием временного ряда в матрицу с помощью однопараметрической сдвиговой процедуры. Этот подход независимо разрабатывался в России (Санкт-Петербург, Москва) и США (там его аналог получил название SSA – Singular Spectrum Analysis) и показал себя мощным средством исследования временных рядов (в основном в метеорологии, гидрологии, климатологии) [29].

Теоретические вопросы развития SSA-метода, а также примеры применения опубликованы в работах Д. Л. Данилова, В. Н. Солнцева,

М. М. Кислицына, А. А. Жиглявского, В. В. Некруткина, Н. Э. Голяндиной, Ф. И. Александрова и др. [2, 86, 87].

Цель SSA состоит в том, чтобы превратить путем декомпозиции исходную последовательность данных в сумму небольшого количества независимых и поддающихся толкованию компонентов, таких как слабо выраженный тренд, периодическая и шумовая составляющие.

В [2] отмечается, что необходимым для применимости SSA-метода является возможность разделения детерминированного сигнала и шума, которая опирается на два асимптотических результата: о разделимости стационарного сигнала определенного вида и стационарного шума; о разделимости любого сигнала и гауссовского шума.

При исследовании SSA трендом полагают медленно меняющуюся составляющую, описывающую глобальное поведение ряда. Под шумовой составляющей ряда понимают реализацию случайной последовательности чисел [2].

Свободный от модели метод SSA предназначен для исследования структуры временных рядов и совмещает в себе достоинства многих других методов, включающих элементы классического анализа временных рядов, многомерной статистики, многомерной геометрии, динамических систем и обработки сигналов. Возможные прикладные области SSA разнообразны: от математики и физики до экономики и финансовой математики, от метеорологии и океанологии до социологии и изучения рыночной конъюнктуры. Любая сложная последовательность данных с потенциальной структурой может представляться как пример успешного применения SSA.

Суть метода заключается в следующем: одномерный ряд преобразуется в матрицу развертки с помощью однопараметрической сдвиговой процедуры, затем осуществляется сингулярное разложение этой матрицы. Алгоритм преобразования временного ряда в матрицу данных состоит в том, что анализу подвергается временной ряд $\{x_i\}_{i=1}^N$, образованный последовательностью N равноотстоящих значений некоторой (возможно, случайной) функции $f(t)$:

$$x_i = f((i-1), \Delta t), \quad i = 1, 2, \dots, N.$$

Выбирают некоторое число $L < N$, называемое длиной сдвига, и первые L значений последовательности f представляют в качестве первой строки матрицы X . В качестве второй строки матрицы берут значения последовательности с x_2 по x_{L+1} . Последнюю строку с номером $k = N - L + 1$ составляют последние L элементов последовательности.

После этого из полученного набора главных компонент выбираются те компоненты, по которым нужно восстановить ряд. Таким образом, результатом применения метода является разложение временного ряда на компоненты, которые могут быть интерпретированы как медленные тренды, циклические составляющие и шум [86].

SSA является полезным инструментом, который может использоваться для того, чтобы решить следующие задачи: обнаружение тенденций в различных областях; сглаживание; извлечение (обнаружение) компонентов сезонности; одновременное извлечение (распознавание) циклов с маленькими и большими периодами; извлечение периодичности с переменной амплитудой; одновременное извлечение (распознавание) сложных тенденций и периодичностей; обнаружение структуры в коротких временных последовательностях; определение точки изменения тенденции. Решение всех этих задач соответствует основным возможностям SSA. Чтобы достигнуть вышеупомянутых возможностей SSA, нет необходимости знать параметрическую модель рассматриваемых временных рядов [88].

1.4 Data Mining в исследовании временных последовательностей данных

Для анализа закономерностей временного ряда все чаще применяют методы Data Mining [7], предназначенные для обнаружения различных шаблонов (паттернов) во временном ряде [98]. При этом особую ценность в обнаружении таких шаблонов имеют логические методы. Эти методы позволяют находить логические *if-then* правила, характерные для строк m -мерной временной матрицы и нехарактерные для случайно сгенерированной m -мерной матрицы данных. Они пригодны для анализа как числовых, так и символьных последовательностей, и их результаты имеют прозрачную интерпретацию.

К методам Data Mining относят искусственные нейронные сети (ИНС), нечеткую логику, кластерный анализ и др.

Используемые для анализа данных искусственные нейронные сети – математические модели, а также их программные или аппаратные реализации, построенные по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток живого организма. ИНС представляют собой систему соединенных и взаимодействующих между собой простых пороговых логических элементов (искусственных нейронов). Нейронные сети не программируются в привычном смысле этого слова, они обучаются. Возможность обучения – одно из главных преимуществ нейронных сетей перед традиционными алгоритмами. Технически обучение

заключается в нахождении коэффициентов связей между нейронами. С математической точки зрения обучение нейронных сетей – это многопараметрическая задача нелинейной оптимизации. В процессе обучения нейронная сеть способна выявлять сложные зависимости между входными данными и выходными, а также выполнять обобщение. Это значит, что в случае успешного обучения сеть сможет вернуть верный результат на основании данных, которые отсутствовали в обучающей выборке, а также неполных и/или «зашумленных», частично искаженных данных [41, 77].

Использованию понятий нечетких множеств и нечеткой логики в искусственных нейронных сетях посвящены работы [11, 50].

Кластерный анализ – задача разбиения заданной выборки объектов на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя. Кластерный анализ – это многомерная статистическая процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в сравнительно однородные группы (кластеры). Однако универсальность применения привела к появлению большого количества несовместимых терминов, методов и подходов, затрудняющих однозначное использование и непротиворечивую интерпретацию кластерного анализа [54, 89].

Быстрая разработка и эффективная эксплуатация имитационных моделей часто сдерживается проблемами сбора и ввода данных. Основные проблемы – плохая пригодность имеющихся данных, детализация модели высокого уровня, трудности в определении доступных источников данных, сложность исследуемой системы [97].

1.5 Особенности подготовки данных, накопленных в информационных системах

Эксплуатация имитационных моделей в производственных условиях затруднена отсутствием методов обработки накопленных данных и подготовки их для использования в моделях [97].

Эффективная подготовка входных данных для имитационного моделирования становится возможной при использовании данных, накопленных в информационных системах организаций. Для этого разработан программный инструментальный, который позволяет подключаться к базе данных и извлекать их, представляя в CMSD – формате (CMSD – Core Manufacturing Simulation Data), пригодном для дискретно-событийных моделей [80].

Имитационные модели и инструменты их построения используются в тех случаях, когда становятся невозможными аналитические подходы. Обычно исследуют поведение сложной системы при полной неопределенности, состоящей из двух компонентов:

- 1) изменчивость, определенная случайными процессами в системе, именуемая стохастической неопределенностью, случайной неопределенностью или непреодолимой неопределенностью;
- 2) неопределенность, обусловленная нехваткой знаний или достаточной информации о системе.

Предлагается учитывать оба компонента при формировании последовательности входных данных для имитационной модели, например

$$Exp([3; 10]) \rightarrow Norm([2; 5], 8) \rightarrow Norm([4; 7], [2; 8]),$$

где $Exp([3; 10])$ – источник поступления заявок;

$Norm([2; 5], 8)$ – время обслуживания заявки узла 1;

$Norm([4; 7], [2; 8])$ – время обслуживания заявки узла 2 [78].

Проведены исследования для экспоненциального и нормального закона распределения, где параметры в интервалах также распределены по нормальному и экспоненциальному закону распределения.

В [82] представлен анализ программных средств для определения закона распределения случайных величин, используемых для ввода в имитационную модель, например, программа StatFit имеет 32 различных распределения, BestFit – 37 распределений вероятности, ExpertFit – 40 распределений.

Задача установления типа выравнивающей кривой распределения и нахождения оценок параметров последовательностей данных, хранящихся в информационной системе предприятия, к настоящему времени до конца не решена. Для описания статистических распределений параметров производственной системы В. В. Нешиной предлагает использовать не отдельные распределения (нормальный закон, Вейбулла, Рэлея, Максвелла и др.), а обобщенные четырехпараметрические распределения, включающие как частные случаи большое количество известных распределений, образующих единую систему. Три основные системы непрерывных распределений В. В. Нешиной представлены следующими плотностями распределения [49]:

$$p(x) = Ne^{\gamma x} \left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1};$$

$$p(t) = Nt^{\gamma-1} \left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}-1};$$

$$p(y) = \frac{N(\ln y)^{\gamma-1}}{y} \left[1 - \alpha u (\ln y)^{\beta} \right]^{\frac{1}{u}-1}$$

с нормирующим множителем N и четырьмя параметрами α , β , γ , u .

С применением методов, разработанных в рамках теории обобщенных распределений, исследованы возможности, например, оценки параметров эмпирических распределений биомедицинских данных. Использование теории обобщенных распределений позволяет: вычислять выравнивающее теоретическое распределение, а не «подбирать» его путем перебора известных распределений; устранить извечную проблему биологических и медицинских исследований по определению «нормы» при интерпретации биомедицинских данных – среднее значение и стандартное отклонение не могут быть использованы для этих целей, т. к. большая часть количественных переменных, характеризующих различные морфофизиологические признаки организма человека, не имеет нормального распределения [65].

Методики определения типа распределения, предложенные В. В. Неши-тым, представляются перспективными для использования в имитационном моделировании производственно-экономических систем для задания исходных данных, получаемых из комплексной информационной системы предприятия. Этим достигается новый, значительно более высокий уровень точности решения прикладных задач, связанных с выравниванием статистических распределений.

Перспективным направлением представляется использование имитационной модели (ИМ) в системе принятия решений и параметрическая адаптация ее на основе алгоритма случайного поиска (АСП) для выявления закономерностей в технологических процессах и технических системах посредством анализа временных последовательностей статистических данных. Программные системы с использованием ИМ для принятия решений и АСП для параметрической адаптации модели к временным последовательностям в настоящее время отсутствуют. В случае, когда анализируются временные последовательности, АСП позволяет производить параметрическую адаптацию ИМ для согласования с временной последовательностью и, таким образом, определять параметры исследуемой системы, используемые затем для принятия решения по улучшению ее качества. Основная идея построения такого программного комплекса состоит в выделении классов характеристик, определяющих особые состояния исследуемого объекта.

2 Моделирование производственного процесса

2.1 Содержательное описание и концептуальное моделирование производственного процесса крашения тканей набивным способом

Процесс производства готовых текстильных полотен (рисунок 2.1) осуществляется в соответствии со следующими входными параметрами, управляющими параметрами и исполняющими механизмами.

Входные параметры процесса: суровые ткани и трикотажные полотна с соответствующей сопроводительной документацией; химикаты и красители, прошедшие входной контроль; вода; электроэнергия; пар; газ.

Цель процесса – производство готовых тканей и трикотажных полотен в соответствии с планом производства и планом по качеству, требованиями ТНПА и удовлетворяющих требованиям потребителей.

Выходные параметры процесса: готовые ткани и трикотажные полотна с соответствующей сопроводительной документацией, а также потери ресурсов (химикатов, красителей, воды) в процессе производства.

Регламентируют процесс производства готовых текстильных полотен, а также содержат установленные нормы и требования к процессу производства следующие потоки: заявки на выполнение заказов потребителей, график очередности выполнения заказов; контракты на переработку продукции потребителя; документация системы менеджмента качества (СМК); образцы-эталон; ТНПА на продукцию; технологическая документация и др.

Для поддержания в рабочем состоянии процесса производства готовых тканей и трикотажных полотен и повышения его результативности используются следующие ресурсы: персонал ОФ; инфраструктура (здания, помещения, технологическое оборудование, транспорт, связь); производственная среда; средства измерения.

Процесс производства включает работы по планированию, функционированию, мониторингу (рисунок 2.2). Стадия обеспечения функционирования процесса производства готовых тканей и трикотажных полотен включает этапы (рисунок 2.3): обеспечение ресурсами процесса производства готовых тканей; производство готовых тканей и трикотажных полотен, контроль в процессе производства; контроль качества готовых тканей и трикотажных полотен.

Стадия обеспечения ресурсами процесса производства готовых текстильных полотен включает следующее: подготовка воды; подготовка химикатов, красителей, печатных красок; подготовка текстильных полотен.

Подготовка воды происходит согласно документации, включает этапы: подготовка воды для технических нужд; контроль качества воды.

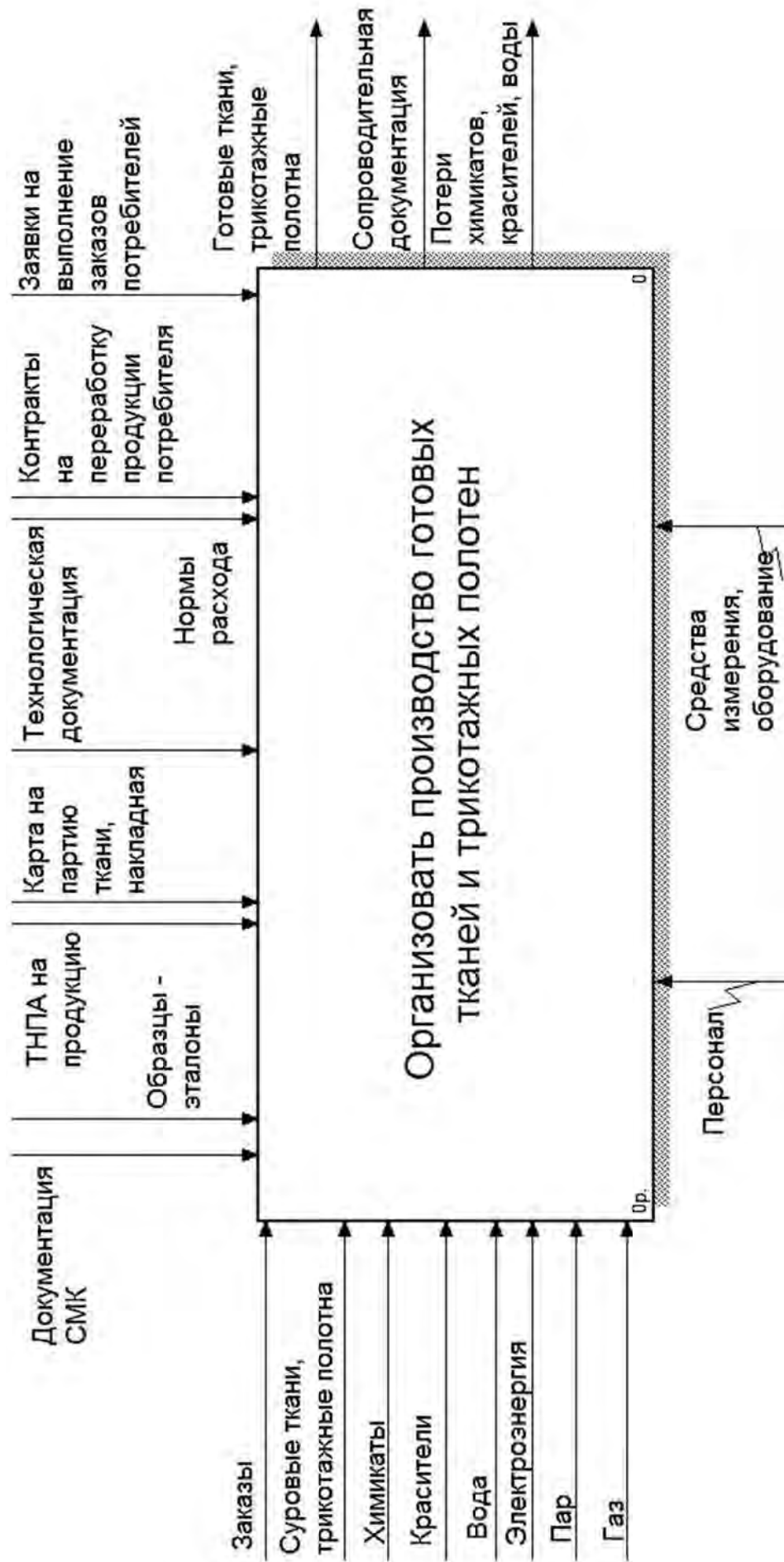


Рисунок 2.1 – Организация производства готовых тканей и трикотажных полотен

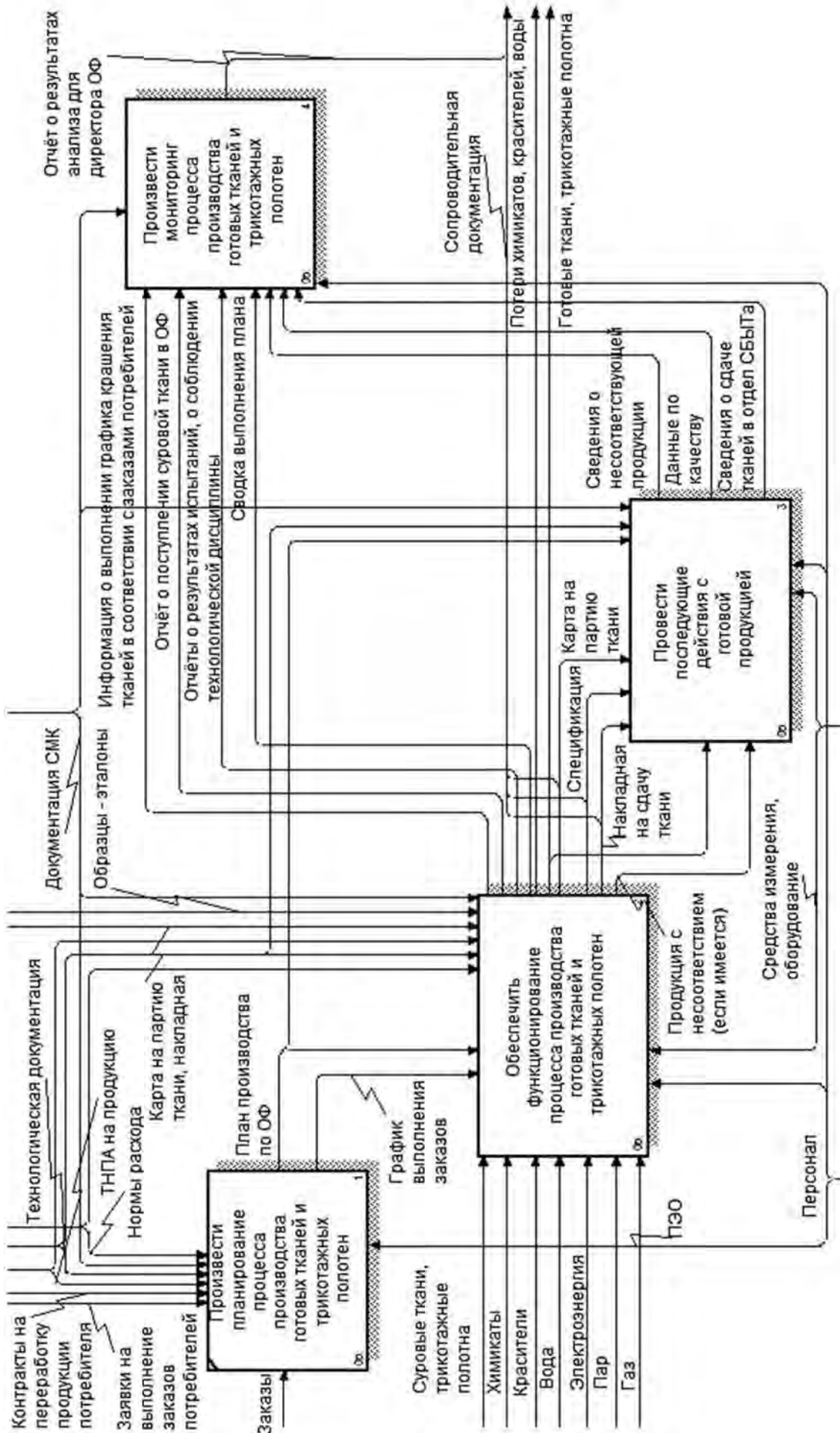


Рисунок 2.2 – Организация производства готовых тканей и трикотажных полотен. Декомпозиция

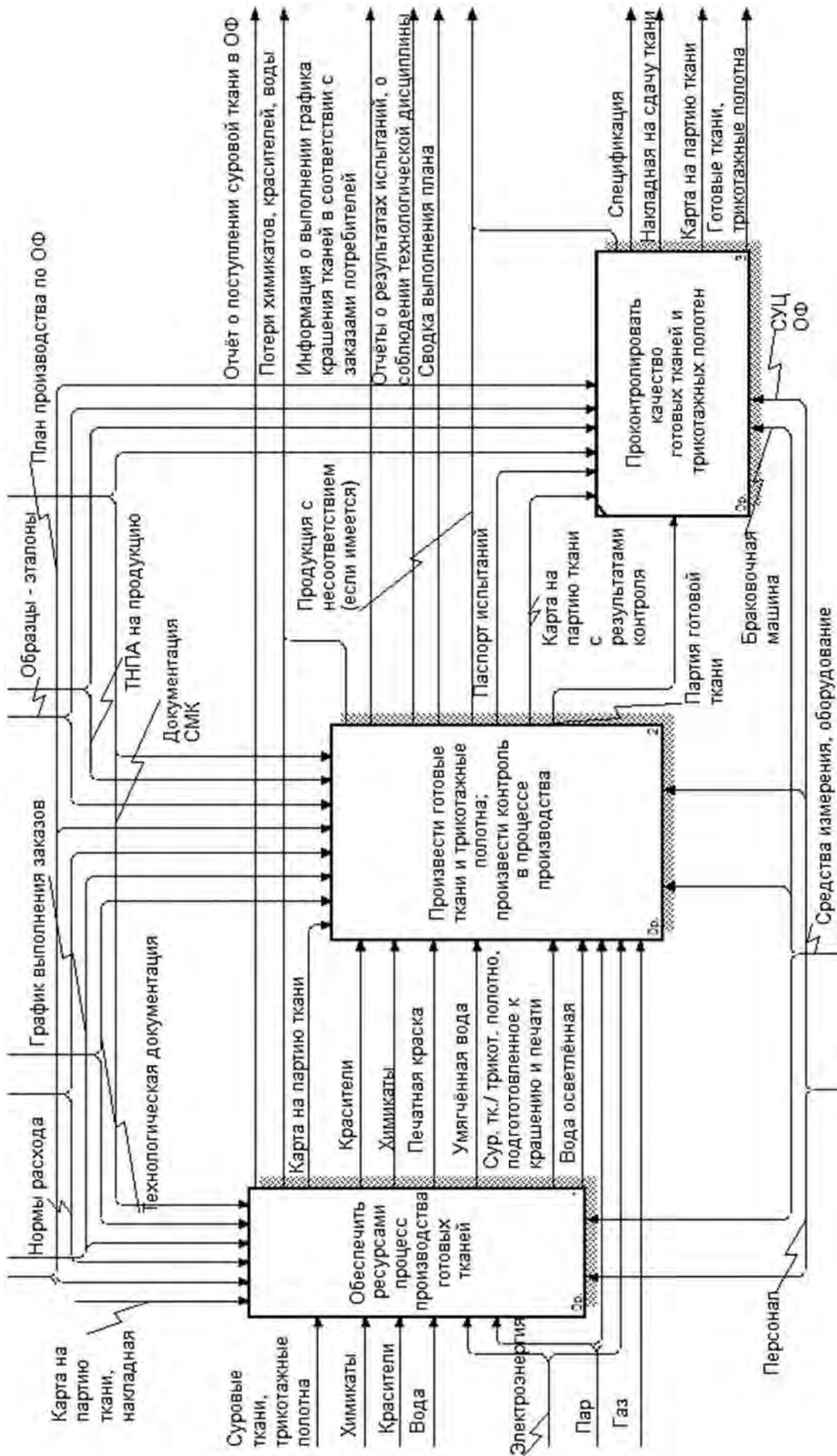


Рисунок 2.3 – Обеспечение функционирования процесса производства готовых тканей и трикотажных полотен. Декомпозиция

Процесс подготовки химматериалов, печатных красок включает этапы: контроль качества закупаемых химикатов и красителей; подготовка растворов химикатов, красителей, печатных красок; контроль концентрации химических растворов. Подготовка растворов химикатов, красителей [6, 47], печатных красок осуществляется в автоматической химстанции с учетом технологической документации предприятия и специально разработанных рецептов.

Необходимое условие предотвращения пороков печатания – химическая и механическая подготовка тканей надлежащего качества. Ткани, предназначенные для печатания, должны иметь определенную капиллярность и белизну, на поверхности не должно быть складок, засечек, заминов и механических загрязнений. Нити в основе и утке должны быть расположены соответственно раппорту ткацкого рисунка. До печатания ткани равномерно высушиваются и доводятся до заданной ширины. Поэтому подготовка текстильного полотна [39] включает стадии: приемка из ТФ партий суровых тканей и трикотажных полотен; производство химической и механической подготовок ткани; контроль качества на всех стадиях.

Процесс химической подготовки ткани включает стадии, или операции: отварка; стабилизация ткани; другие возможные виды подготовки [40, 47], а также стадию контроля качества подготовленного текстильного полотна (рисунок 2.4).

Операция отварки текстильных полотен (стадия химической подготовки) заключается в варке текстильных полотен в щелочных растворах при повышенном или атмосферном давлении для отделения примесей, а также с целью обеспечения возможности дальнейшего качественного беления, крашения, набивки полотна. Процесс отварки ткани проводится согласно технологической документации на одном из имеющихся типов оборудования [8]. При этом вначале осуществляется настройка соответствующего оборудования, затем – загрузка ткани в оборудование, далее – процесс отварки ткани. Конечный этап – выгрузка ткани из оборудования с соответствующими отметками в карте на партию ткани, передача ткани на следующую стадию обработки.

Оборудование отварки включает части [8]: промывные коробки; ванна для замочки; отварочная ванна и др. Все части оборудования настраиваются в соответствии с параметрами ТР. Ткань в процессе отварки последовательно проходит стадии: замочка; отварка; промывка в соответствующих частях оборудования. При этом ванны для оборудования заполняются химическим раствором. Химический раствор готовится в химстанции согласно данным ТР, учитывающего состав и концентрации химикатов.

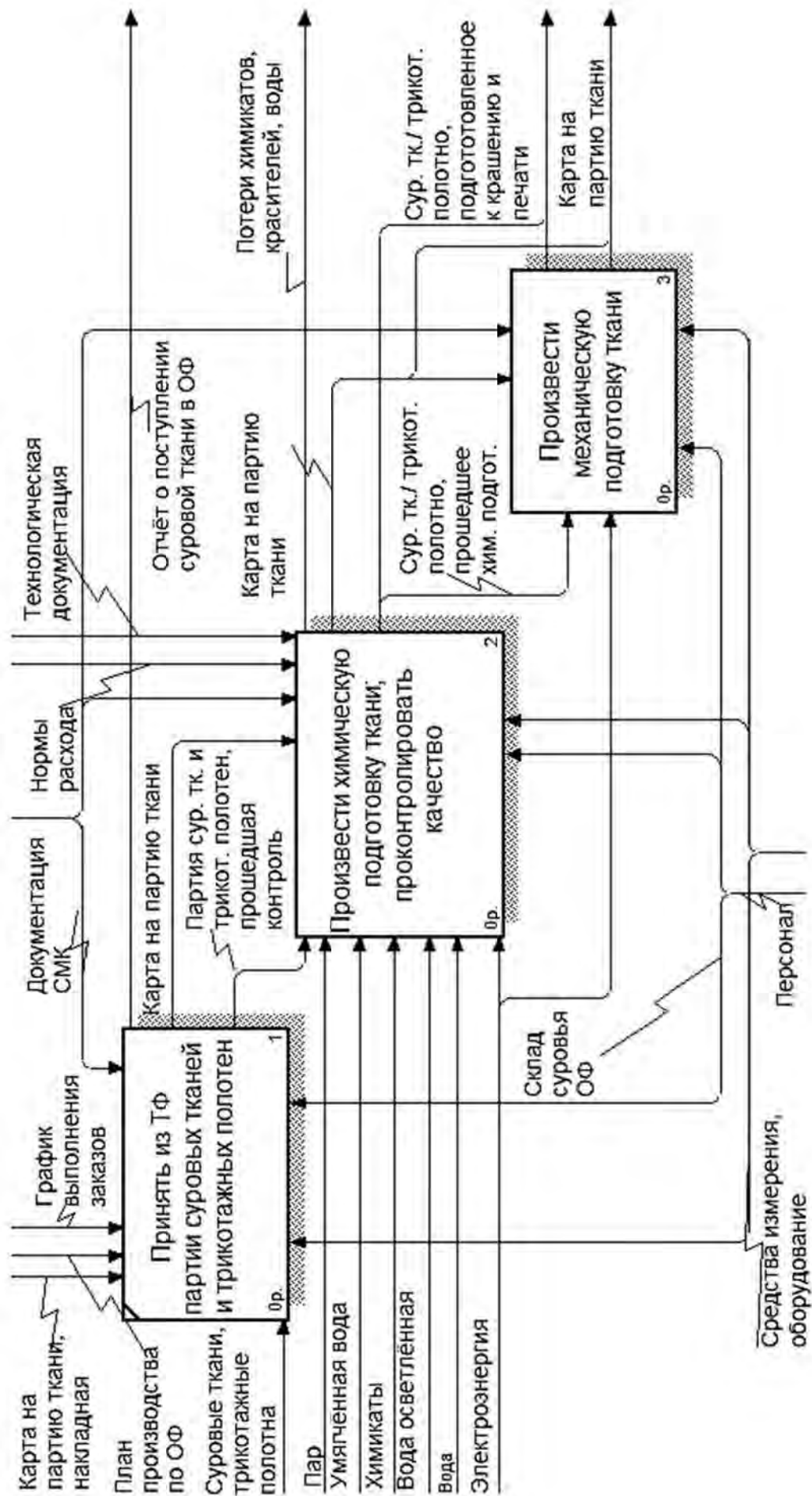


Рисунок 2.4 – Проведение химической подготовки ткани. Контроль качества. Декомпозиция

Концентрации химикатов варьируются в диапазоне $\langle \text{среднее} \rangle \pm \langle \text{модуль отклонения от среднего} \rangle$, имеют стохастический характер. В качестве моющих веществ, согласно примечаниям ТР, могут применяться в одинаковой концентрации как основной химикат, так и указанные возможные для основного химиката химикаты-заменители.

После прохождения тканью стадий замочки и отварки в ваннах оборудования производится ее отжим согласно ТР. При этом определенное количество химикатов уносится из раствора вместе с тканью, определенное – возвращается в химвраствор. Согласно ТР по истечении заданных периодов времени (один раз в неделю, один раз в смену и т. д.) производится слив химвраствора из ванн и промывных коробок оборудования.

Стадии процесса отварки текстильного полотна выполняются в соответствии с технологической документацией с применением указанных в ней значений временных, температурных и других параметров.

Операция стабилизации ткани. Ткань в процессе стабилизации последовательно проходит стадии: пропитка; сушка; термофиксация. Для каждой из частей оборудования стабилизации согласно ТР и в зависимости от веса обрабатываемой ткани устанавливаются температурные настройки для термофиксации. В качестве химвраствора применяется вода умягченная.

Стадия механической подготовки текстильного полотна включает этапы: стрижка; чистка лицевой стороны полотна; разглаживание полотна; исправление всевозможных перекосов утка и другие с соответствующими пометками в карте на партию ткани о проведении механической подготовки [6].

Стадия производства готовых текстильных полотен. Процесс производства готовых текстильных полотен (рисунок 2.5) включает этапы: стадия процесса печатания на ткани; стадия заключительной отделки; контроль в процессе производства [6, 30, 52].

Производственный процесс печати на ткани – процесс нанесения на ткань печатного рисунка с помощью одной или нескольких красок. При этом ткань окрашивается лишь частично – там, где нанесено изображение. Изображение на полотно переносится с использованием машин для печати на ткани. Качество печати характеризуется устойчивостью окраски к различным видам внешних воздействий, а также четкостью контуров изображения, ровностью печати, глубиной проникновения печатной краски в ткань и др.

В ПП КТНС применяются специальные краски для печати. В отличие от красильных растворов печатные краски обладают повышенной вязкостью. При нанесении красителя на ткань в загущенном вязком виде форма контуров печатного рисунка остается неизменной.

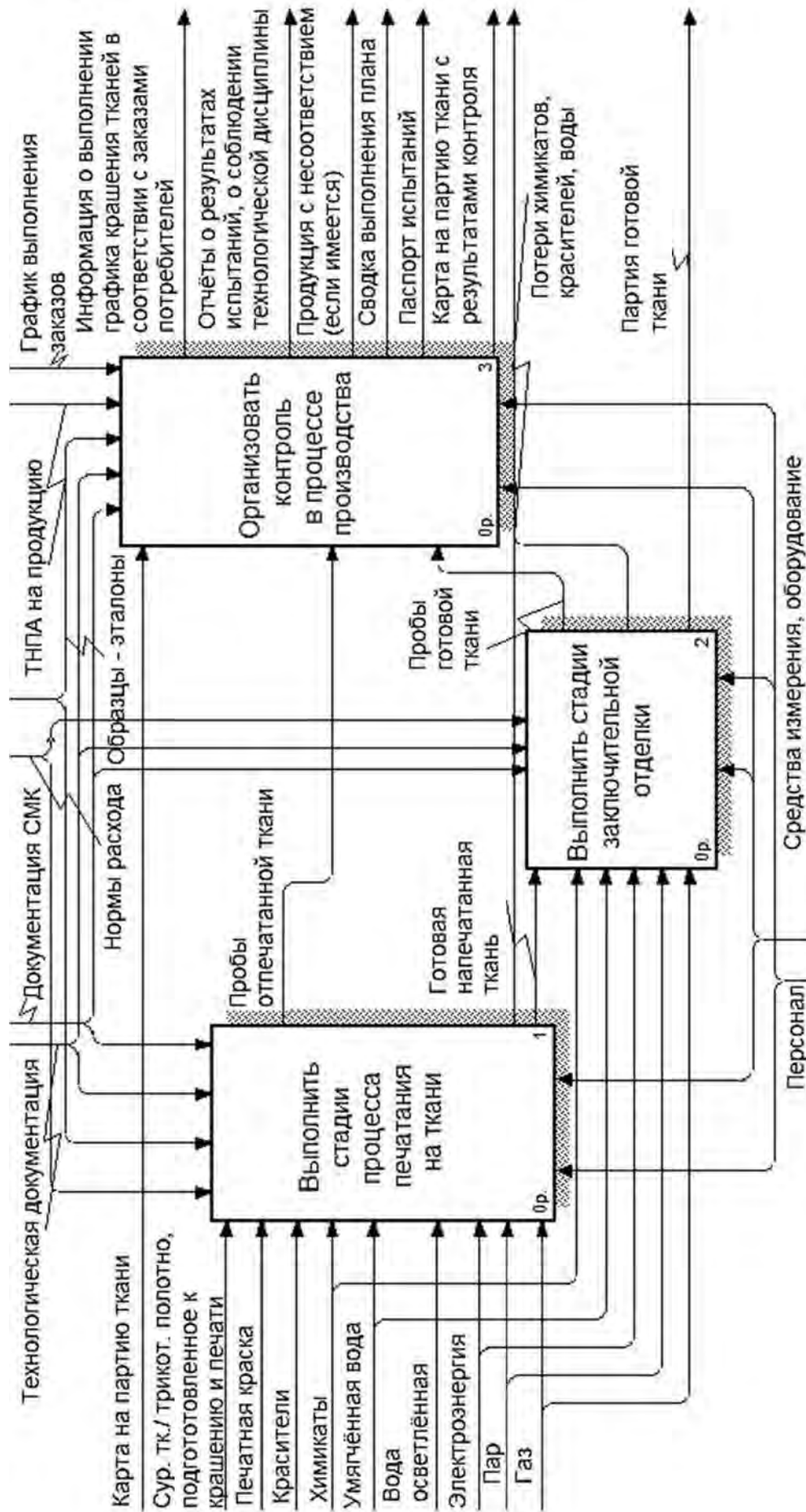


Рисунок 2.5 – Производство готовых тканей и трикотажных полотен. Контроль в процессе производства

В ПП КТНС используются различные виды красителей, такие как активные, пигментные, кубовые и др. [5]. Загустителями выступают водорастворимые вещества (крахмал, трагант, декстрин, камедь и др.), которые при разваривании или растворении в воде образуют вязкие системы-загустки.

Состав красок для печати может включать также растворители, окислители, диспергаторы, кислоты, восстановители, соли, щелочи, пеногасители, пластификаторы. Качество печати зависит как от красителей, так и от используемых загустителей, условий нанесения печатной краски на ткань и условий обработки ткани после печати. В отличие от крашения при печати в качестве красильной ванны используется вязкий раствор загустителя. Наличие загустителя в печатной краске может оказывать как замедляющее, так и ускоряющее влияние на процесс перехода красителя на волокно. Характер и интенсивность такого влияния зависит от особенностей строения загустителя [6].

Печатные краски готовятся в автоматической химстанции согласно данным ТР, учитывающим состав и концентрации химикатов. Концентрации химикатов варьируются в диапазоне $\langle \text{среднее} \rangle \pm \langle \text{модуль отклонения от среднего} \rangle$, имеют стохастический характер.

Печатная краска на ткань может быть перенесена различными способами. Далее рассмотрен способ печати с использованием печатных машин с цилиндрическими (ротационными) сетчатыми шаблонами. Цилиндрический сетчатый печатный шаблон – перфорированный никелевый цилиндр заданных размеров. Изготавливается гальваническим способом. Варианты размещения цилиндрических шаблонов в печатных машинах: плоский стол; грузовик. Есть различие конструкций rakelного узла печатных машин [6].

Перенос рисунка на шаблон осуществляется фотохимическим методом. Рисунок для печати разрабатывается и редактируется художниками по рисунку, уже на стадии редактирования рисунка известны проценты заполнения поверхности каждым из используемых цветов.

В процессе крашения тканей (рисунок 2.6) может использоваться от четырех до шестнадцати цветов, для каждого из которых применяется свой ротационный шаблон с выгравированным для данного цвета рисунком. При этом в печатной машине с плоским печатным столом шаблоны для отдельных цветов располагаются последовательно вдоль печатного стола, который представляет собой непрерывное полотно-конвейер. Шаблоны прижимаются к печатному столу за счет давления rakel, расположенной внутри него, и имеют двусторонний привод. Раكلي на машинах различных фирм имеют вид резиновой пластины или прутка.

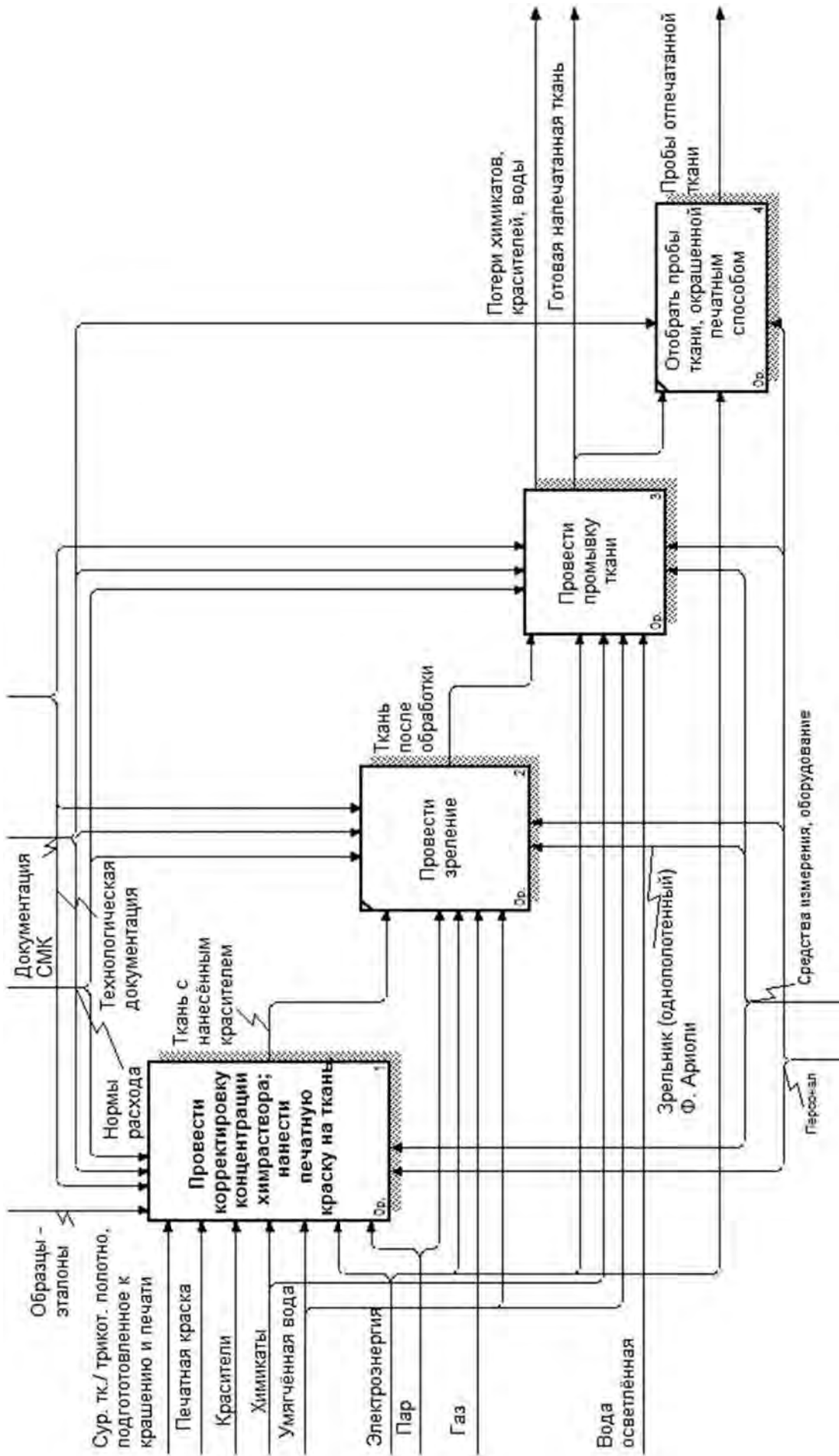


Рисунок 2.6 – Выполнение стадии процесса печатания на ткани

Краска в шаблон подается автоматически через боковую щель трубы ракельного устройства. При постоянном натяжении ткань с больших рулонов вводным устройством подается на конвейер и приклеивается к нему. Напечатанная ткань поступает в сушилку [53].

Плановая норма затрат печатной краски в ПП КТНС и потери ресурсов составляют фактическую норму затрат химических материалов в ПП КТНС. Потери делятся на: постоянные – количество красителя, остающееся в насосе, шлангах и шаблоне; переменные – избыточный объем красителя, сливаемый в канализацию. Независимо от процента заполнения ткани цветом постоянные потери для каждого цвета одинаковы [8].

Операция зреления ткани (обработка ткани после печатания). После печати рисунка на полотне и последующей сушки краситель находится в пленке из печатной краски на поверхности волокон. Термообработка напечатанного полотна во влажной среде специального оборудования (зрельника) обеспечивает переход красителя из пленки в глубь волокна. В зрельнике волокно набухает в процессе конденсации влаги на холодной ткани. При этом компоненты печатной краски (краситель и текстильно-вспомогательные вещества) растворяются в воде, сорбированной на волокне. Вследствие выделения тепла при растворении вспомогательных веществ набухает загустка.

Краситель перемещается из печатной краски в волокна ткани и там фиксируется. На этой стадии происходит набухание пленки печатной краски за счет поглощения влаги из паровой среды зрельника [8]. Кроме того, образуется высококонцентрированный раствор загустителя – красильная ванна, в которой растворяются составляющие печатной краски (краситель, вспомогательные вещества); происходит процесс прокрашивания ткани. В зависимости от специфических свойств применяемого красителя, состава печатной краски и специфических свойств ткани устанавливаются временные и температурные настройки зреления. От выбора зрельника и качества его работы зависит степень фиксации красителей на ткани, яркость и устойчивость окраски [6].

Операция промывки ткани. Последующая после зреления промывка текстильного полотна позволяет убрать с его поверхности излишки загустителя и красителя. На оборудовании промывки может осуществляться также специальная обработка ткани для завершающего закрепления красителя.

Промывка может производиться на оборудовании различных фирм [8]. Вначале осуществляется загрузка ткани на одну из линий промывки, затем – процесс промывки на соответствующей линии, далее – выгрузка ткани из оборудования с соответствующими отметками в карте на партию ткани.

Оборудование для промывки включает части: пропиточные и промывные коробки; ванны для замочки и релаксации; сушильная камера и др. Для каждой из частей оборудования согласно технологическому режиму устанавливаются соответствующие температурные настройки. Также, в зависимости от веса обрабатываемой ткани, настраивается скорость.

Процесс проведения промывки включает стадии: замочка; промывка; сушка. Коробки и ванны оборудования заполняются химическим раствором, который готовится в химстанции согласно данным ТР, учитывающего состав и концентрации химикатов. Концентрации химикатов варьируются в диапазоне $\langle \text{среднее} \rangle \pm \langle \text{модуль отклонения от среднего} \rangle$, имеют стохастический характер. В качестве моющих веществ, согласно примечаниям ТР, могут применяться в одинаковой концентрации как основной химикат, так и указанные возможные химикаты-заменители для основного химиката.

После прохождения тканью промывки производится ее отжим. При этом определенное количество химикатов уносится из раствора вместе с тканью, определенное – возвращается в химраствор. По истечении заданных периодов времени производится слив химраствора из ванн оборудования.

Стадия заключительной отделки текстильного полотна. Под операциями отделки текстильного полотна, с одной стороны, понимают все возможные операции облагораживания тканей (отбелка, крашение, печать, виды заключительных отделок). С другой стороны, под понятиями отделки понимают только заключительную отделку, которая улучшает вид, а также качество полотна, повышающую его износоустойчивость и др. Последовательность операций при заключительной отделке текстильного полотна зависит от его особенностей и назначения. В отдельных случаях некоторые операции не применяются.

Выполнение заключительной отделки может осуществляться на оборудовании различных фирм. Включает стадии: стабилизация; антистатическая и грязеотталкивающая отделка ткани; придание наполненного грифа и др.; промывка ткани после заключительной отделки; отбор проб промытой ткани после заключительной отделки с целью проверки качества обработки [8].

Оборудование заключительной отделки включает части: плюсовка; камеры оборудования и др. Все части оборудования настраиваются согласно технологическим режимам. Ткань в процессе заключительной отделки последовательно проходит стадии: пропитка; сушка; фиксация в определенных камерах оборудования. Скорость работы оборудования устанавливается в зависимости от режима и веса обрабатываемой ткани. В качестве химматериалов применяются вода умягченная и химикат в зависимости от режима. Концентрации химикатов варьируются в диапазоне

$\langle \text{среднее} \rangle \pm \langle \text{модуль отклонения от среднего} \rangle$, имеют стохастический характер.

Контроль качества готовых текстильных полотен. После завершения всех стадий ПП КТНС проводятся испытания образцов ткани с соответствующими пометками в карте на партию ткани, составляются отчеты. Готовое текстильное полотно передается в складально-уборочный цех (СУЦ).

ПП готовой продукции, подобные ПП КТНС, исследуются методами потокового программирования, линейного программирования. На каждом из этапов ПП КТНС имеет место стохастический характер параметров данных ПП, что не позволяет в явном виде применить аналитические методы моделирования ПП и вызывает необходимость применения методов имитационного моделирования [34].

2.2 Формализованное описание производственного процесса крашения тканей набивным способом

Пусть технологический процесс (ТП) по производству готового текстильного полотна (продукции) $TexPr$ (рисунок 2.7) состоит из n определенных фиксированных технологических операций (этапов):

$$TexPr \stackrel{def}{=} \{Op_i / i = 1, \dots, n\}. \quad (2.1)$$

Операции Op_i , $i = 1, \dots, n$, производятся последовательно.

Все технологические операции Op_i предполагают выбор [38, 61] на этапах $TexPr$ (2.1) одного из $M_i \in N$ предполагаемых технологических режимов (ТР) $TR_{R(i,k)}$ обработки продукции, которые осуществляются с использованием соответствующего технологического оборудования [8] (рисунок 2.8):

$$Op_i \stackrel{def}{=} \{TR_{R(i,k)} / k = 1, \dots, M_i\}, \quad i = 1, \dots, n;$$

$$R(i,k) \stackrel{def}{=} (i, \xi_{ik}, k), \quad \xi_{ik} \in \{1, \dots, |\xi|\}, \quad k \in \{1, \dots, M_i\}, \quad (2.2)$$

где $R(i,k)$ – вектор, идентифицирует номер ТР;

i – номер операции Op_i технологического процесса (2.1) по порядку;

k – номер ТР в рамках нумерации режимов операции Op_i ;

ξ_{ik} – код технологического оборудования.



Рисунок 2.7 – Последовательность технологических операций (2.1)

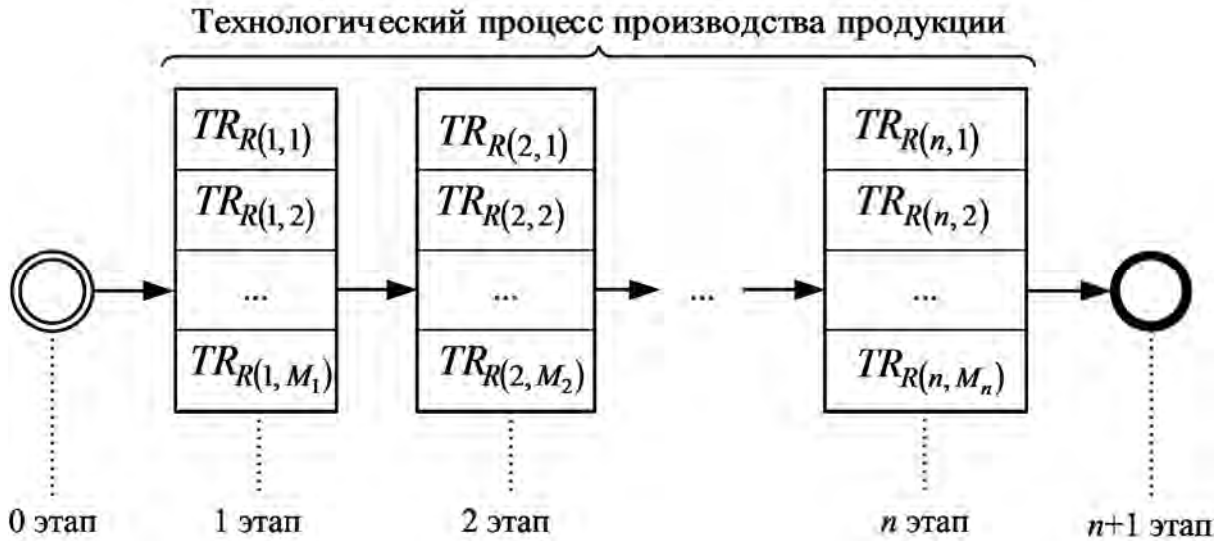


Рисунок 2.8 – Последовательность ТР операций ТП (2.1)

Выбор ТР и технологического оборудования на этапах ТП может осуществляться в соответствии с требованиями заказчика продукции (будет иметь место фиксированный выбор определенного ТР для операции ТП).

Себестоимости ТР – суммарные стоимости ресурсов, применяемых в рамках данных ТР. Расход времени, затрачиваемого на обработку заданного количества продукции в рамках рассматриваемых ТР, зависит от параметров оборудования, особенностей текстильного полотна, настроек ТР, а также прочих факторов.

Рассмотрим множества ресурсов ТР Res и соответствующих им цен на ресурсы Sen :

$$Res \stackrel{def}{=} \{Res_r / r = 1, \dots, |Res|\}; \quad (2.3)$$

$$\begin{aligned}
Cen &\stackrel{\text{def}}{=} \{Cen_r / Cen_r \in R, \quad r = 1, \dots, |Res|\}; \\
Cen_r &\stackrel{\text{def}}{=} CenVal_r \cdot CoursVal_r, \quad CenVal_r, CoursVal_r \in R,
\end{aligned} \tag{2.4}$$

где $CenVal_r$ – цена ресурса Res_r в валюте;

$CoursVal_r$ – курс валюты в белорусских рублях на момент закупки ресурса.

ТП производства готового текстильного полотна (2.1) возможно представить как нагруженный ориентированный ациклический граф [34], для которого вершинами являются технологические режимы (2.2). Представление ТП (2.1) в виде графа производится путем его разделения на n однотипных этапов, или операций. Нулевой этап ТП (2.1) – исток графа. Нумерация вершин графа осуществляется от истока к стоку последовательно. Исток – нулевой этап, сток – $(n+1)$ этап. Вершины соединяются дугами согласно последовательности операций Op_i (см. рисунок 2.8).

Пусть τ_{uv} – нагрузка на дугу графа (рисунок 2.9), которая исходит из вершины с номером $u \in M_B(i-1)$ и входит в вершину с номером $v \in M_B(i)$, где $M_B(i)$ – множество номеров вершин для i -го этапа графа (2.1):

$$M_B(i) \stackrel{\text{def}}{=} \left\{ \sum_{j=1}^{i-1} M_j + 1, \dots, \sum_{j=1}^i M_j \right\}, \quad i = 1, \dots, n;$$

$$M_B(i, k) = \sum_{j=1}^{i-1} M_j + k, \quad i = 1, \dots, n, \quad k \in \{1, \dots, M_i\}. \tag{2.5}$$

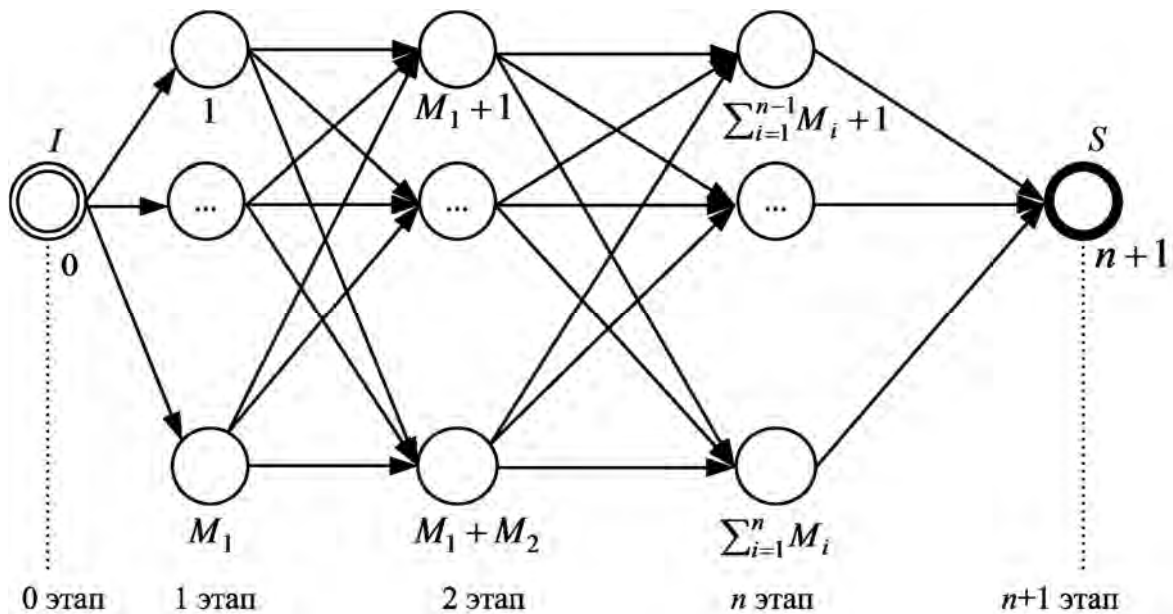


Рисунок 2.9 – Граф ТП производства готовой продукции

Нагрузки τ_{uv} интерпретируются таким образом – это временные затраты и/или стоимостные затраты ресурсов, расходуемых на выпуск определенного объема (1000 п. м) текстильного полотна согласно ТР $TR_{R(u,v)}$ (2.2) ТП (2.1). При этом T_i – норма расхода времени; S_i – общие стоимости расхода ресурсов, затрачиваемые на обработку данной продукции операцией Op_i ТП (2.1):

$$T_i \stackrel{def}{=} (t_{i1}, \dots, t_{iM_i}), t_{ij} \in R, i=1, \dots, n, j=1, \dots, M_i, n, M_i \in N; \quad (2.6)$$

$$S_i \stackrel{def}{=} (s_{i1}, \dots, s_{iM_i}), s_{ij} \in R, i=1, \dots, n, j=1, \dots, M_i, n, M_i \in N. \quad (2.7)$$

В общем случае нагрузки τ_{uv} имеют стохастический характер; для возможности применения аналитических методов моделирования ПП необходимо уточнение нагрузок методами имитационного моделирования и статистического анализа данных ТП.

2.3 Обзор законов распределения статистических данных в производственном процессе крашения тканей набивным способом

Как показано в формальном описании (см. подразд. 2.2), характер нагрузок на дуги графа производственного процесса (2.1) в общем случае стохастический, что вынуждает к применению методов статистического анализа и позволит сделать выводы об изучаемом ПП, уточнить нагрузки на дуги графа, построив статистическую модель (кривую) распределения выборочных данных

$$X = \{x_i / x_i \in R, i=1, \dots, m\}, \quad (2.8)$$

наилучшим образом описывающую рассматриваемую выборку. При этом для определения вида статистической модели, описывающей изучаемое реальное явление по выборочным данным-откликам испытаний, часто нет заслуживающих доверие теоретических обоснований. Иногда достаточно представления имеющейся информации в виде гистограммы. В других случаях желательно выразить данные через распределения.

Поставленную задачу построения статистической модели можно решить, например, с использованием существующих известных законов распределений [93]. Имеется множество способов определения видов распределений выборок экспериментальных данных, наиболее распространенным из

которых является применение нормального распределения [51]. Согласно центральной предельной теореме [27] адекватно описывает многие из явлений реальной действительности именно нормальное распределение [51]. Распределения гамма- и логарифмически нормальное обычно описывают случайные величины, ограниченные сверху и снизу. Несмотря на то, что данные модели конструируют распределения разнообразной формы, они часто не дают необходимой степени обобщения.

Более общим способом описания выборок эмпирических данных является применение распределений семейств универсальных статистических моделей [22]: Пирсона, Джонсона и другие, а также рядов специального вида: разложения Корниша–Фишера, разложения в ряд Грама–Шарлье, разложения в ряд Эджвота и т. д.

Обобщенные законы распределений Пирсона, Джонсона и другие чаще всего применяются для наиболее надежного построения статистической модели распределения выборочных данных.

Распределения К. Пирсона. Пирсон предложил построение семейства обобщенных распределений следующим способом:

$$df / dx = (x - a) / (b_0 + b_1x + b_2x^2) f \mid a, b_0, b_1, b_2 \in R; \quad (2.9)$$

$$f = C \cdot e^{\psi(x)}, \psi(x) = \int (x - a) / (b_0 + b_1x + b_2x^2) dx, C = \text{const}; \quad (2.10)$$

$$b_0 + b_1x + b_2x^2 = 0. \quad (2.11)$$

Дифференциальное уравнение (ДУ) Пирсона (2.9) – линейное однородное дифференциальное уравнение первого порядка. Его общий интеграл имеет вид (2.10) и представляет собой функцию плотности [27] распределения случайной величины (СВ) X . Решения ДУ (2.9) для различных значений корней квадратного уравнения (2.11) образуют семейство распределений Пирсона, состоящее из семи основных типов распределений и частных случаев (нормальное, равномерное, экспоненциальное).

Пусть дискриминант квадратного уравнения (2.11) имеет вид:

$$D = b_1^2(1 - 1/\aleph), \aleph = b_1^2 / (4b_0b_2), \quad (2.12)$$

где \aleph – критерий классификации типов распределений, введенный Пирсоном.

Тогда с учетом оценки классическим методом моментов коэффициентов a, b_0, b_1, b_2 ДУ (2.9) критерий \aleph принимает следующий вид:

$$\aleph = \beta_1(\beta_2 + 3)^2 / (4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)), \quad (2.13)$$

где β_1, β_2 – коэффициенты асимметрии и эксцесса, введенные Пирсоном:

$$\beta_1 = \mu_3^2 / \mu_2^3, \quad \beta_2 = \mu_4 / \mu_2^2; \quad (2.14)$$

μ_2, μ_3, μ_4 – центральные моменты выборки X (2.8) второго, третьего и четвертого порядка соответственно:

$$\mu_k = 1/m \cdot \sum_{i=1}^m (x_i - v_1)^k \quad | \quad k = 2, 3, 4, \quad m = |X|, \quad v_1 = 1/m \cdot \sum_{i=1}^m x_i. \quad (2.15)$$

Имеет место следующая классификация типов распределений Пирсона в зависимости от значений критерия \aleph (2.13) и показателей асимметрии и эксцесса β_1, β_2 вида (2.14). Значение критерия $\aleph < 0$ определяет I тип; $\aleph = 0$ соответствует II и VII типам. Условие $0 < \aleph < 1$ соответствует IV типу; $\aleph = 1$ – V типу; $\aleph > 1$ – VI типу; $\aleph = \pm\infty$ – III типу. Для II типа имеют место соотношения $\beta_1 = 0$ и $1 < \beta_2 < 3$; для VII типа – соотношения $\beta_1 = 0$ и $\beta_2 > 3$. Условие $(\beta_1, \beta_2) = (0; 1,8)$ определяет равномерное распределение; при $(\beta_1, \beta_2) = (0; 3)$ – нормальное; при $(\beta_1, \beta_2) = (4; 9)$ – экспоненциальное. При этом частным случаем I типа является бета-распределение первого рода. II тип (частный случай – равномерное распределение) является вариантом кривых Пирсона I типа. Частными случаями кривых III типа являются распределения: гамма- и хи-квадрат. V тип посредством преобразований сводится к III типу. Частными случаями VI типа являются бета-распределение второго рода, а также распределение F (Фишера–Снедекора). Частный случай VII типа – распределение T (Стьюдента). X тип распределений Пирсона – экспоненциальное распределение.

Распределения семейства Пирсона обладают следующими свойствами:

- 1) единственность моды, т. е. $df/dx = 0$ в некоторой точке $x = a$;
- 2) гладкое соприкосновение с осью x на концах интервала, на котором сосредоточено распределение, т. е. $df/dx = 0$ при $f = 0$;
- 3) существование не более двух точек перегиба, определяемых соотношением $x^2 = b_0/b_2$.

Таким образом, распределения семейства Пирсона унимодальны. Если у распределения количество точек перегиба равно двум, то они отстоят от моды на одинаковом расстоянии. Одна из точек может быть расположена вне зоны

сосредоточения распределения, т. е. распределения семейства (2.9) могут иметь также и J-, и U-образную форму [37].

Карта Пирсона построена в координатах β_1, β_2 (2.14) с целью возможности классификации типов распределений случайных величин. Типы различных распределений представлены на карте точками, кривыми либо областями. Нижняя граница всех распределений представлена прямой $\beta_2 - \beta_1 - 1 = 0$, т. к. не существует распределений, для которых выполняется условие $\beta_2 - \beta_1 - 1 < 0$. Уравнение $\beta_2 - \beta_1 - 1 = 0$ является условием вырождения в пределе случайной величины, имеющей U-образное распределение, на два значения, сосредоточенные по краям размаха распределения. Таким образом, существует возможность классификации типа распределения в соответствии с показателями (β_1, β_2) (2.14) либо \aleph (2.13). При этом вначале для исследуемой выборки X находятся оценки показателей β_1^*, β_2^* .

Графический метод классификации предполагает нанесение на карту Пирсона точки с координатами (β_1^*, β_2^*) . В том случае, если точка (β_1^*, β_2^*) находится на близком расстоянии по отношению к точке, кривой либо области, которая обозначает одну из вышеперечисленных моделей, то есть вероятности применимости данного типа распределения к описанию выборочных эмпирических данных. После определения типа кривой плотности распределения, соответствующего исследуемой выборке X , приступают к оценкам параметров распределения, используя соответствующие формулы. Оценки параметров кривых семейства Пирсона получены методом моментов Пирсона. При этом нормирующий множитель N для кривых $f(x)$ определен из условия нормирования $\int_a^b N \cdot f(x) dx = 1$, где a, b – границы сегмента (отрезка), в пределах которого производится построение, $a = \min X$, $b = \max X$, $\min X = \min\{x_i\}$, $\max X = \max\{x_i\}$, $i = 1, \dots, m$, $m = |X|$.

Имеют место следующие ограничения распределений К. Пирсона. Для любого множества данных X параметры β_1^*, β_2^* – только оценки показателей β_1, β_2 , склонны к колебаниям от выборки данных к выборке и очень чувствительны к небольшому числу крайних значений (особенно для выборок объемом $|X| < 200$). Форма распределения с помощью показателей асимметрии и эксцесса не определяется со стопроцентной точностью. Поэтому из определения вида распределения с целью описания заданного множества не следует адекватность выбранной модели; для контроля полученных результатов необходимо применение статистических методов

оценки адекватности выбранной модели, например, статистических критериев согласия [62].

2.4 Обзор методов кластерного анализа данных производственного процесса крашения тканей набивным способом

Для выявления особенностей поведения ПП КТНС требуется анализ многомерных данных. Так, необходимо решить задачу построения кластеров непересекающихся подмножеств множества $X \subset R^n$:

$$X = \{x_i / x_i \in R^n, i = 1, \dots, m\}, m \geq 3. \quad (2.16)$$

Задача решается с использованием методов кластеризации данных, через матрицы сходства, эвристические алгоритмы перебора, идеи математического программирования, через анализ свойств функции плотности эмпирического распределения выборки X (2.16) и др. [7].

Описание методов кластеризации данных достаточно широко отражено в различных источниках. Большинство алгоритмов, при всем их разнообразии, основаны на гипотезе компактности [7], состоящей в том, что объекты одного класса по результатам измерений близко расположены по отношению друг к другу, а объекты разных классов заметно различаются между собой. Вследствие анализа рассмотренных методов кластеризации составлена их классификация [15], представленная на рисунке 2.10. Все методы разбиения множеств на кластеры возможно разделить на иерархические, а также неиерархические. Для исследования отобраны методы кластеризации K-means [7, 84], Tree Clustering, Fuzzy Relation Clustering (FRC) [7, 17], которые являются представителями основных методов к разделению множества, которое рассматривается, на кластеры. Описание особенностей отобранных методов кластеризации приведено в [7].

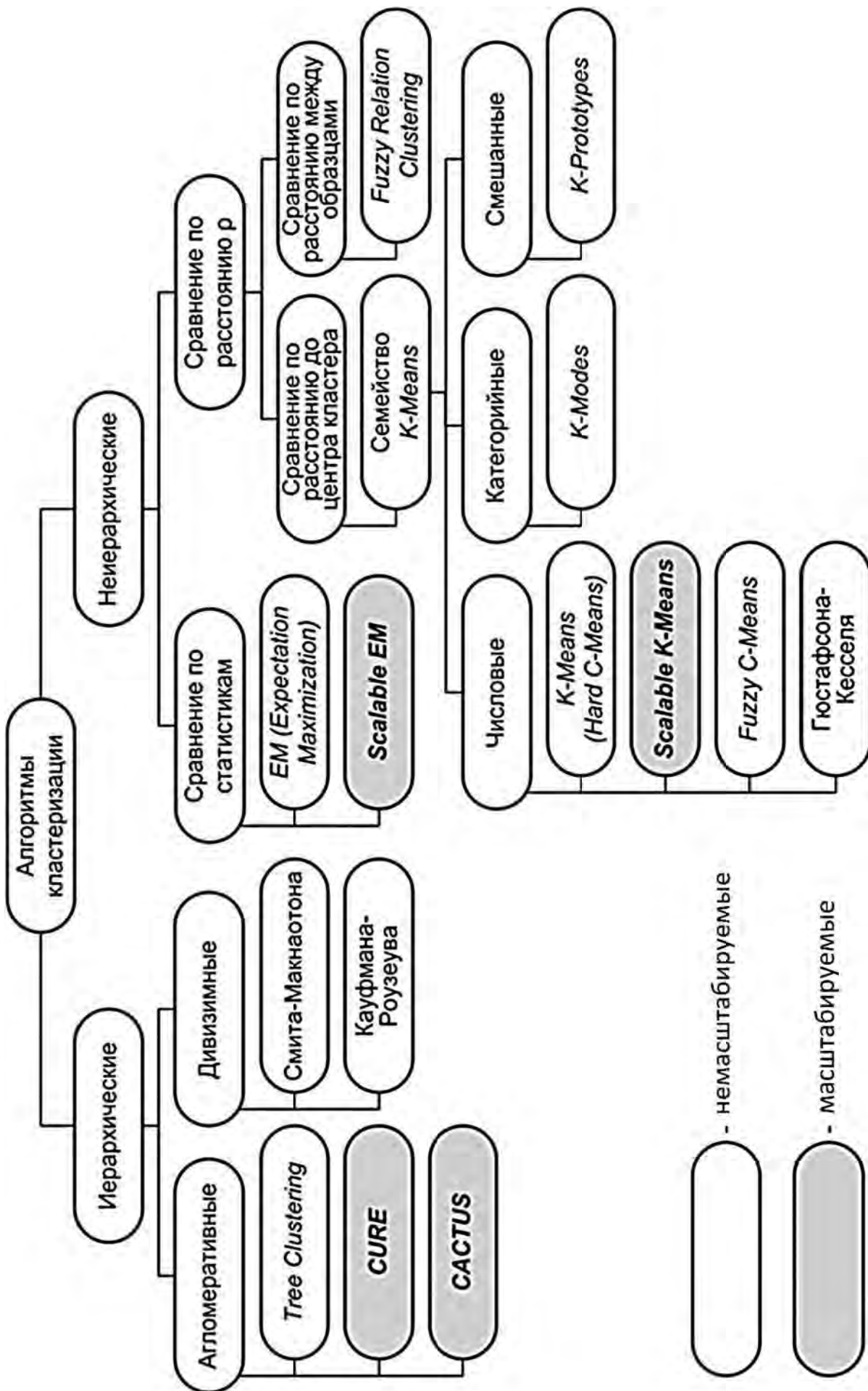


Рисунок 2.10 – Классификация алгоритмов кластерного анализа

3 Метод извлечения полезной информации из последовательности числовых данных

3.1 Идея метода извлечения полезной информации из последовательности числовых данных для применения в имитационном моделировании

В общем случае знания представляют собой обобщенное описание данных, которое отражает существенные закономерности, присущие исследуемым объектам. Эти закономерности могут принимать вид функциональных, логических или структурных связей. Знания, в отличие от данных, выполняют активную функцию: на основе знаний принимаются решения, вырабатываются стратегии, осуществляется планирование, проводится оптимизация и решаются другие задачи.

Интеллектуальный анализ данных (ИАД) – процесс управляемого извлечения зависимостей из больших баз данных. В этом процессе центральное место занимает автоматическое порождение моделей, правил и/или функциональных зависимостей, характеризующих анализируемые данные. В целом процесс извлечения знаний в ИАД условно делят на следующие этапы [60], которые в совокупности предложено использовать на этапе эксплуатации имитационной модели.

Шаг 1. Отбор данных: анализ задач пользователя; выбор целевого множества данных; определение переменных.

Шаг 2. Предобработка данных: устранение зашумленности; обработка пропущенных значений; итоговые показатели по группам данных.

Шаг 3. Редукция и проекция данных: ищутся полезные особенности данных для решения поставленных задач; сокращается пространство переменных.

Шаг 4. Поиск закономерностей: выбор метода поиска закономерностей с учетом объема и типа данных, их зашумленности и осуществление поиска закономерностей.

Шаг 5. Оценка и интерпретация найденных закономерностей: оценка и упорядочение закономерностей по их релевантности; проверка согласованности предыдущих и вновь найденных знаний. Возможно возвращение к любому шагу от 1 до 4 для дальнейших итераций.

Шаг 6. Использование найденных знаний: прямое использование; передача заинтересованным лицам; включение в интеллектуальные системы, основанные на знаниях.

Предлагается итерационный процесс P извлечения полезной информации, представленный композицией следующих операторов:

$$P = T_6 \circ T_5 \circ T_4 \circ T_3 \circ T_2 \circ T_1,$$

где T_1 – оператор, представляет отбор данных, накопленных в комплексных информационных системах предприятий, определение (с привлечением экспертов в исследуемой предметной области) переменных для решения поставленной задачи;

T_2 – предварительная обработка данных, устранение выбросов и пропущенных данных;

T_3 – выбор данных для исследования в соответствии с решаемой задачей, *отличается* формированием XML-файла для автоматизированной обработки;

T_4 – выбор метода поиска закономерностей, основанных главным образом на применении статистических исследований, *отличается* использованием метода сингулярного спектрального анализа и обобщенного закона распределения для оценки структуры распределения случайных величин;

T_5 – построение модели исходных данных для применения в имитационной модели на этапе ее эксплуатации;

T_6 – формирование XML-файла с моделью последовательности данных для последующего применения.

Для разработки технологии извлечения знаний из накопленных баз данных исследован сингулярный спектральный метод, включающий этапы: вложение; сингулярное разложение; группировка; диагональное усреднение.

В ранее проведенных исследованиях [1] основной задачей являлось выделение из известного ряда неизвестных трендовой и периодической составляющих с автоматизацией процедур визуального исследования. Дальнейшее развитие SSA-метода связано с исследованием шумовой составляющей и оценкой погрешности ее восстановления.

3.2 Содержательное описание SSA-метода

Данные в компьютерной информационной системе, представленные вещественными числами и определенные последовательно во времени, именуют временной последовательностью данных. В общем случае такую последовательность данных, обычно упорядоченных во времени, называют временным рядом. Поэтому в дальнейшем будем использовать понятие временного ряда, и для извлечения информации из последовательностей данных рассмотрим метод сингулярного спектрального анализа (SSA-метод), который применяется для анализа временных рядов и может быть

использован на этапе эксплуатации имитационной модели (ИМ). Он позволяет выделить составляющие временного ряда, которые описывают тренд и гармонические колебания, и те составляющие временного ряда, которые относят к шуму. При этом метод не требует стационарности временного ряда, знания модели тренда, а также сведений о наличии периодических составляющих и их периодах. Также с помощью данного метода можно определить модель тренда и использовать это знание для дальнейшей обработки временного ряда с уже известной моделью тренда, что важно, например, при автоматизации определения длительности переходного процесса в ИМ.

Математической основой SSA-метода является сингулярное разложение. Для успешного применения SSA-метода следует последовательно пройти несколько шагов.

Вложение. На этом шаге выбирается ширина окна L , от выбора которой зависят результаты применения SSA-метода. Из-за того, что нет общих рекомендаций по выбору ширины окна, параметр L зависит от решаемой задачи и предварительной информации, известной о временном ряде. Например, для выделения тренда рекомендуется выбирать ширину окна не слишком большой. С другой стороны, для выделения гармонических колебаний рекомендуется большая ширина окна. Поэтому в [45] рекомендуется выполнять два прохода с разной шириной окна. После выбора ширины окна в соответствии с L строится траекторная матрица A ряда, которая будет являться по условию ее построения ганкелевой [25].

Сингулярное разложение. Для матрицы $S = A \cdot A^T$ находятся собственные числа μ [42] и ортонормированные собственные векторы U . Упорядоченные по убыванию собственные числа, которые больше нуля, часто называются сингулярными числами, а соответствующие им собственные векторы – левыми сингулярными векторами U . После этого вычисляются векторы V , которые называются правыми сингулярными векторами, и находятся элементарные матрицы, на сумму которых раскладывается первоначальная траекторная матрица.

Группировка. На данном этапе элементарные матрицы группируются по принципу принадлежности к тренду, гармоническим колебаниям или к шуму. Этот этап является наиболее сложным при применении SSA-метода. Для нахождения тренда на диаграммах собственных векторов (по оси абсцисс откладывается порядковый номер координаты собственного вектора, а по оси ординат – значение координаты собственного вектора) выделяют медленно меняющиеся векторы. Сумма элементарных матриц, соответствующих этим векторам, будет являться траекторной матрицей тренда ряда. После этого

восстанавливают гармонические колебания ряда. Для отделения шума можно воспользоваться несколькими замечаниями: нерегулярное поведение сингулярных векторов может говорить о принадлежности их к набору, порожденному шумовой компонентой; также об этом может свидетельствовать медленное, практически без скачков, убывание собственных чисел с некоторого номера.

Диагональное усреднение. Если полученные сгруппированные матрицы ганкелевы, то они являются траекторными матрицами некоторого временного ряда, который может быть легко по ним восстановлен. Однако обычно сгруппированные матрицы редко получаются ганкелевыми, поэтому для восстановления временного ряда прибегают к диагональному усреднению. В соответствии с этим этапом каждый член восстановленного временного ряда будет являться средним арифметическим соответствующей ему побочной диагонали траекторной матрицы.

В результате проделанных шагов получается несколько временных рядов: один описывает тренд первоначального (исходного) временного ряда, другой – гармонические колебания, а третий – шумовые составляющие.

3.3 Методика исследования SSA-метода на основе информационных технологий

Для исследования SSA-метода применяется комплекс информационных технологий, представленный табличным процессором MS Excel, математическим пакетом Mathcad и пакетом статистической обработки данных Statistica (рисунок 3.1). При этом не требуется разработки специализированного программного обеспечения [71].

Этап вложения. Для экспериментальных исследований исходный ряд $G = (g_0, g_1, \dots, g_{N-1})$ может быть задан по известным функциям либо определен результатами функционирования исследуемой системы. По известным функциям ряд формируется на рабочем листе MS Excel и затем в пакете Mathcad формируется матрица A , которая по правилам построения – ганкелева. Процедура вложения является преобразованием исходного одномерного ряда $G = (g_0, g_1, \dots, g_{N-1})$ в последовательность L -мерных векторов, число которых $K = N - L + 1$:

$$A_i = (g_{i-1}, \dots, g_{i+L-2})^T, \quad 1 \leq i \leq K. \quad (3.1)$$



Рисунок 3.1 – Схема использования информационных технологий при исследовании SSA-метода

Эти векторы образуют траекторную матрицу $A = [A_1; \dots; A_K]$ ряда G , в которой $a_{ij} = g_{i+j-2}$, т. е. матрица A имеет одинаковые элементы на диагонали $i + j = \text{const}$.

Этап сингулярного разложения. Обозначим $S = A \cdot A^T \in R^{L \times L}$. Матрица $A \cdot A^T$ симметричная и неотрицательно определенная, а значит, ее собственные числа $\{\mu_k\}_{k=1}^L$ вещественны и неотрицательны. Представленные в виде $\mu_1 \geq \dots \geq \mu_L \geq 0$ собственные числа называют сингулярными значениями матрицы S . Пусть U_1, \dots, U_L – соответствующие им ортонормированные собственные векторы. Будем называть $p = \max\{k \mid \mu_k > 0\}$ порядком сингулярного разложения. Обозначим

$$V_k = \frac{1}{\sqrt{\mu_k}} A^T U_k, \quad k = 1, \dots, p. \quad (3.2)$$

Тогда сингулярным разложением матрицы A называется ее представление в виде суммы элементарных матриц:

$$A = A_1 + A_2 + \dots + A_p, \quad A_k = \sqrt{\mu_k} U_k V_k^T. \quad (3.3)$$

Каждая из матриц A_k имеет ранг, равный единице. Поэтому их можно назвать элементарными матрицами. Вектор U_k называют k -м левым сингулярным вектором или просто k -м собственным вектором, вектор V_k – правым сингулярным вектором. Набор $\langle \sqrt{\mu_k}, U_k, V_k \rangle$ называют k -й собственной тройкой. Обозначим корень собственного числа матрицы S через $\lambda_k = \sqrt{\mu_k}$ и будем использовать это обозначение в дальнейших исследованиях.

Сингулярные собственные числа $\{\lambda_k\}_{k=1}^L$ матрицы A в пакете Mathcad представлены вектором **d**. Вектор **d** сингулярных значений в Mathcad определяется с использованием функции svds() [33]:

$$\mathbf{d} := \text{svds}(A). \quad (3.4)$$

Диагональная матрица ds сингулярных значений матрицы A в пакете Mathcad определяется с использованием функции diag():

$$ds := \text{diag}(\mathbf{d}). \quad (3.5)$$

Объединенная матрица AS с левыми и правыми сингулярными векторами определяется с использованием функции $\text{svd}()$:

$$AS := \text{svd}(A). \quad (3.6)$$

Для разделения левых и правых сингулярных векторов из матрицы AS используется функция $\text{submatrix}()$ [33].

Этап группировки. Вид левых и правых сингулярных векторов, трактуемых в SSA как временные ряды, является очень важным для следующего шага метода – группировки. При этом для одномерного SSA левые и правые сингулярные векторы обладают определенной симметрией, т. к. в этих случаях сингулярные разложения траекторных матриц с длиной окна L и $K = N - L + 1$ эквивалентны.

Процедура группировки формально одинакова для всех разновидностей SSA. На основе разложения (3.3) процедура группировки делит все множество индексов $\{1, \dots, p\}$ на m непересекающихся подмножеств I_1, \dots, I_m .

Пусть $I = \{i_1, \dots, i_p\}$. Тогда результирующая матрица A_I , соответствующая группе I , определяется как $A_I = A_{i_1} + \dots + A_{i_p}$. Такие матрицы вычисляются для $I = I_1, \dots, I_m$, тем самым разложение (3.3) может быть записано в сгруппированном виде:

$$A = A_{I_1} + \dots + A_{I_m}. \quad (3.7)$$

Процедура выбора множеств $I = I_1, \dots, I_m$ и называется группировкой собственных троек. Для определения $I = I_1, \dots, I_m$ в MS Excel используется лепестковая диаграмма, которая является аналогом графика в полярной системе координат, отображая распределение значений относительно начала координат. По особенностям представления сингулярных векторов на лепестковой диаграмме принимается решение о принадлежности их одной группе.

Этап диагонального усреднения. На последнем шаге базового алгоритма каждая матрица сгруппированного разложения переводится в новый ряд длины N . Для произвольной матрицы X процедуру приведения ее к ганкелевому виду и последующему преобразованию в ряд (обозначим его как G^B) выразим следующим образом. Пусть X – матрица размером $L \times K$ с элементами x_{ij} , $1 \leq i \leq L$, $1 \leq j \leq K$. Положим, $L^* = \min(L, K)$,

$K^* = \max(L, K)$ и $N = L + K - 1$. Пусть $z_{ij} = x_{ij}$, если $L < K$ и $z_{ij} = x_{ji}$ в остальных случаях. Тогда диагональное усреднение переводит матрицу X в ряд $(g_0^B, \dots, g_{N-1}^B)$ по формуле

$$g_k^B = \begin{cases} \frac{1}{k+1} \sum_{j=1}^{k+1} z_{j, k-j+2} & | 0 \leq k \leq L^* - 1; \\ \frac{1}{L^*} \sum_{j=1}^{L^*} z_{j, k-j+2} & | L^* - 1 \leq k \leq K^*; \\ \frac{1}{N-k} \sum_{j=k-K^*+2}^{N-K^*+1} z_{j, k-j+2} & | K^* \leq k \leq N. \end{cases} \quad (3.8)$$

Выражение (3.8) соответствует усреднению элементов матрицы вдоль побочных диагоналей $i + j = k + 2$: выбор $k = 0$ дает $g_0^B = x_{11}$, для $k = 1$ получаем $g_1^B = (x_{12} + x_{21})/2$ и т. д. Применив диагональное усреднение к матрицам, полученным на этапе группировки, приходим к разложению исходного ряда в сумму m рядов.

Процедуру диагонального усреднения просто и наглядно предложено выполнить в MS Excel. Для этого матрица, подлежащая диагонализации, размещается на рабочем листе. Затем блок матрицы, следующий за первой строкой, сдвигается вправо на одну позицию. В сдвинутом блоке также определяется блок, следующий за первой строкой, который сдвигается вправо на одну позицию. Процедура повторяется до тех пор, пока в очередном блоке не останется ни одной строки. Восстановленный ряд G^B определяется аналогично формуле (3.8) с использованием функции СРЗНАЧ() в MS Excel. Затем исследуется в пакете Statistica [71].

3.4 Технология исследования SSA-метода и пример его реализации

Для проведения предварительных исследований выбрана функция $F(x) = y_1(x) + y_2(x)$, где $y_1(x)$ – линейная функция, $y_1(x) = 0,1x + 2$; $y_2(x)$ – гармоническая функция, $y_2(x) = \sin(x)$. Переменная x принимает значения от 0 до 42 с шагом, равным единице. Таким образом, длина N ряда $G = (g_0, g_1, \dots, g_{N-1})$ равна 43. При этом длина окна $L = 22$, число L -мерных векторов $K = N - L + 1 = 22$.

Отдельно на рабочем листе MS Excel рассчитаны значения функций $y_1(x)$, $y_2(x)$ и $F(x)$. При этом ряду G_{y1} линейной функции, или тренда, соответствуют значения $G_{y1} = \langle g_{y1_0}, g_{y1_1}, \dots, g_{y1_{42}} \rangle = \langle 2,0; 2,1; \dots; 6,2 \rangle$. Ряд G_{y2} гармонической составляющей

$$G_{y2} = \langle g_{y2_0}, g_{y2_1}, \dots, g_{y2_{42}} \rangle = \langle 0; 0,841; \dots; -0,917 \rangle$$

и

$$G = \langle g_0, g_1, \dots, g_{42} \rangle = \langle 2,0; 2,94; \dots; 5,28 \rangle.$$

Элементы ряда G копируются в траекторную матрицу A на рабочем листе Mathcad 14.0. Сформированная матрица A является ганкелевой размером 22×22 . Следует отметить, что размеры матрицы при ручном вводе Insert Matrix ограничены возможностями Mathcad и не должны превышать 600 позиций.

Собственные числа при сингулярном разложении траекторной матрицы A с применением функции $\text{svds}()$ по формуле (3.4) принимают следующие значения: $\lambda_1 = 99,49$; $\lambda_2 = 10,99$; $\lambda_3 = 10,80$; $\lambda_4 = 2,10$; $\lambda_5 = 0,024$. Остальные значения $\lambda_6, \dots, \lambda_{22}$ представляют убывающую последовательность: $\lambda_6, \dots, \lambda_{22} = 0,022; 0,018; \dots; 3,06 \cdot 10^{-3}; 6,602 \cdot 10^{-4}$.

Сумма всех собственных чисел равна 116,577, причем сумма собственных чисел $\lambda_1 + \lambda_2 + \dots + \lambda_5 = 116,413$.

Сингулярное разложение в пакете Mathcad реализуется с помощью функции $\text{svd}()$ по формуле (3.6), формируется обобщенная матрица AS , включающая левые и правые сингулярные векторы. Левые и правые сингулярные векторы разделяются по формулам $U := \text{submatrix}(AS, 0, 21, 0, 21)$ и $V := \text{submatrix}(AS, 22, 43, 0, 21)$ соответственно.

Диагональная матрица ds сингулярных чисел определена по формуле (3.5). Для контроля правильности выполненных вычислений определяется равенство

$$U \cdot ds \cdot V^T = A. \quad (3.9)$$

Для последующих расчетов определяются скалярные значения собственных чисел по формулам с использованием функции $\text{submatrix}()$. Например, для λ_1 (обозначение $ds0$ в пакете Mathcad)

$$ds0 := \text{submatrix}(ds, 0, 0, 0, 0), \quad ds0 = 92,49; \quad (3.10)$$

для λ_2 (обозначение $ds1$ в пакете Mathcad)

$$ds1 := \text{submatrix}(ds, 1, 1, 1, 1), \quad ds1 = 10,99 \text{ и т. д.} \quad (3.11)$$

Левые сингулярные векторы определяются из матрицы U аналогично с использованием функции $\text{submatrix}()$:

$$U_0 := \text{submatrix}(U, 0, 21, 0, 0); \quad (3.12)$$

$$U_1 := \text{submatrix}(U, 0, 21, 1, 1) \text{ и т. д.} \quad (3.13)$$

Правые сингулярные векторы определяются в пакете Mathcad по формуле

$$V_i := \frac{1}{\sqrt{dsi}} \cdot A^T \cdot U_i, \quad i = 0, 1, \dots, 4. \quad (3.14)$$

Элементарные матрицы X_i определены по формуле

$$X_i := \sqrt{dsi} \cdot (U_i \cdot V_i^T), \quad i = 0, 1, \dots, 4. \quad (3.15)$$

Для определения групп элементарных матриц X_i построены лепестковые диаграммы для векторов V_i в табличном процессоре MS Excel. Например, при исследовании временного ряда, заданного функцией $F_1(x) = 2 + 0,1x + \sin(x) \mid x = 0, \dots, 42$, векторам V_0 (рисунок 3.2) и V_3 (рисунок 3.3) соответствуют расходящиеся кривые.

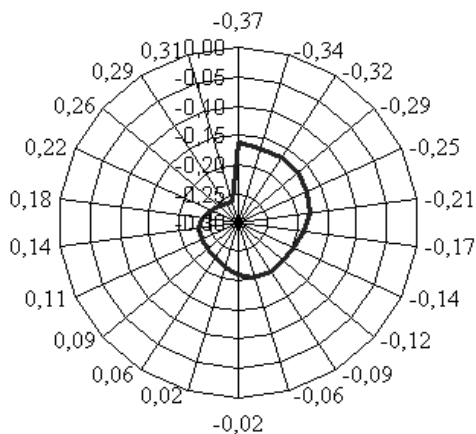


Рисунок 3.2 — Лепестковая диаграмма вектора V_0

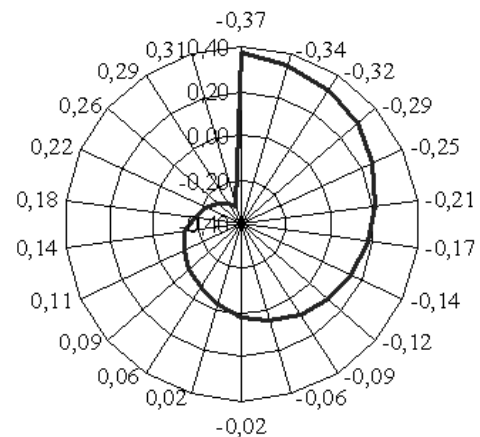


Рисунок 3.3 — Лепестковая диаграмма вектора V_3

Векторы V_1 , V_2 и V_4 , V_5 имеют вид, аналогичный изображению на рисунках 3.4–3.7.

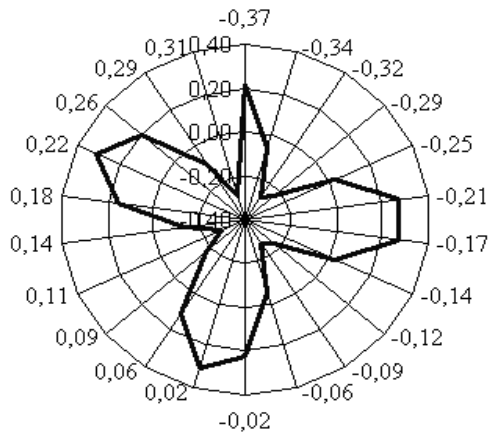


Рисунок 3.4 – Лепестковая диаграмма вектора V_1

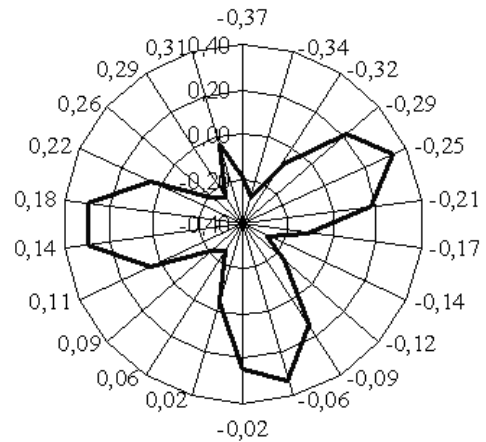


Рисунок 3.5 – Лепестковая диаграмма вектора V_2

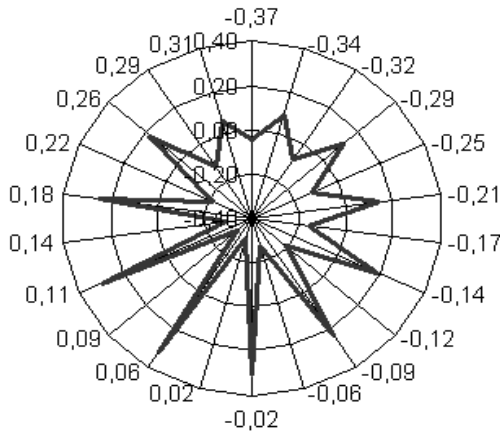


Рисунок 3.6 – Лепестковая диаграмма вектора V_4

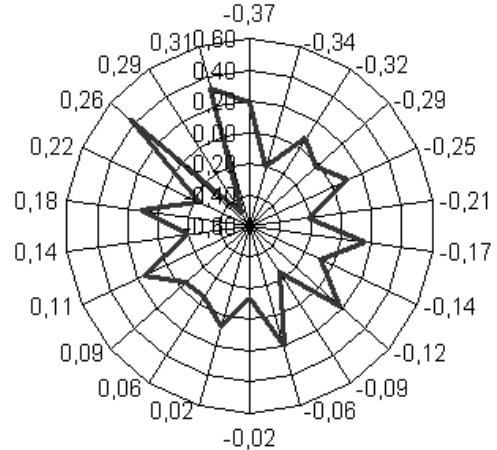


Рисунок 3.7 – Лепестковая диаграмма вектора V_5

Тогда группа $X1$ элементарных матриц соответствует тренду ряда: $X1 := X_0 + X_3$. Аналогично группа $X2 := X_1 + X_2 + X_4$ соответствует гармонической составляющей ряда.

На этапе диагонализации для выполнения расчетов по формуле (3.8) используется специальная процедура в MS Excel, которая иллюстрируется на примере матрицы $X1$ с левым верхним блоком

$$X1^{4 \times 4} = \begin{pmatrix} 2,141 & 2,242 & 2,276 & 2,303 \\ 2,242 & 2,345 & 2,377 & 2,402 \\ 2,276 & 2,377 & 2,407 & 2,431 \\ 2,303 & 2,402 & 2,432 & 2,453 \end{pmatrix}. \quad (3.16)$$

На рисунке 3.8 представлен фрагмент рабочего листа MS Excel с решением задачи диагонализации по предложенной процедуре.

Восстановленные значения тренда $g_{y1_i}^B$ определяются по столбцам с применением функции СРЗНАЧ(). Полученная таким образом матрица используется в пакете Statistica 6.0 для оценки характера распределения ошибки восстановления ряда.

Сдвиговая процедура блока матрицы X1 при диагонализации	2,141	2,242	2,276	2,303	$x1_{15}$	$x1_{16}$	$x1_{17}$
		2,242	2,345	2,377	2,402	$x1_{25}$	$x1_{26}$
			2,276	2,377	2,407	2,431	$x1_{35}$
				2,303	2,402	2,431	2,453
Восстановленные значения тренда $g_{y1_i}^B$	2,141	2,242	2,299	2,34
Значения тренда g_{y1_i}	2	2,1	2,2	2,3
Отклонения $g_{y1_i} - g_{y1_i}^B$	-0,141	-0,142	-0,099	-0,04	-0,0037	0,069	0,147

Рисунок 3.8 – Процедура диагонализации матрицы X1 в MS Excel

3.5 Предварительные результаты и их обсуждение

При исследовании временного детерминированного ряда G , образованного функцией вида $F(x) = ax + b\sin x + c = 0,1x + \sin x + 2$, построена ганкелева матрица A размером 22×22 . При сингулярном спектральном разложении матрицы A основными для восстановления ряда явились элементарные матрицы X_0 и X_3 для линейной составляющей (собственные числа $\lambda_1 = 99,49$ и $\lambda_4 = 2,10$, которым в пакете Mathcad соответствуют обозначения $ds0$ и $ds3$) и матрицы X_1 , X_2 , X_4 для гармонической составляющей (собственные числа $\lambda_2 = 10,99$, $\lambda_3 = 10,80$, $\lambda_5 = 0,024$, которым в пакете Mathcad соответствуют $ds1$, $ds2$ и $ds4$).

В ходе проведенных исследований временных детерминированных рядов, образованных функциями вида $F(x) = ax + b\sin x + c$, построена таблица 3.1.

Таблица 3.1 – Результаты исследования временных рядов

Временной ряд, заданный функцией $F_i(x)$	Трендовая составляющая		Гармоническая составляющая		
	$(V_{\text{тренд}}; \lambda_{\text{тренд}})$		$V_{\text{гарм}}; \lambda_{\text{гарм}}$		
$F_1(x) = 2 + 0,1x + \sin(x)$	$(V_0, 92,5)$	$(V_3, 2,10)$	$(V_1, 11,0)$	$(V_2, 10,8)$	$(V_4, 0,024)$
$F_2(x) = 2 + x + \sin(x)$	$(V_0, 542,1)$	$(V_1, 35,93)$	$(V_2, 10,99)$	$(V_3, 10,82)$	$(V_4, 0,024)$
$F_3(x) = 2 - x + \sin(x)$	$(V_0, 460,2)$	$(V_1, 42,35)$	$(V_2, 10,99)$	$(V_3, 10,82)$	$(V_4, 0,024)$
$F_4(x) = 2 + 12x + \sin(x)$	$(V_0, 6052,0)$	$(V_1, 463,54)$	$(V_2, 10,99)$	$(V_3, 10,82)$	$(V_4, 0,024)$
$F_5(x) = 2 - 12x + \sin(x)$	$(V_0, 5970,0)$	$(V_1, 469,92)$	$(V_2, 10,99)$	$(V_3, 10,82)$	$(V_4, 0,024)$
$F_6(x) = 2 + x + 10\sin(x)$	$(V_0, 543,9)$	$(V_3, 109,89)$	$(V_1, 107,84)$	$(V_2, 35,92)$	$(V_4, 0,021)$
$F_7(x) = 2 - 12x + 300\sin(x)$	$(V_0, 5951,0)$	$(V_3, 468,98)$	$(V_1, 3288,0)$	$(V_2, 3268,0)$	$(V_4, 0,025)$

Абсолютная ошибка восстановления ряда определена величиной отклонения

$$\Delta = \max |g_i - g_i^B|, \quad (3.17)$$

где g_i – элементы исходного ряда;

g_i^B – элементы восстановленного ряда, $i = 0, \dots, N - 1$.

Относительная ошибка восстановления ряда определяется по формуле

$$\delta = \frac{\max |g_i - g_i^B|}{\max |g_i|} \cdot 100 \%. \quad (3.18)$$

Абсолютная ошибка восстановления для тренда составляет 0,142, для гармонической составляющей – 0,143, для исходного ряда – 0,0055. Относительная ошибка восстановления тренда составляет 2,29 %, для гармонической составляющей – 14,3 %, для исходного ряда – 0,08 % [74].

Обнаружены интересные факты: абсолютная ошибка восстановления тренда и гармонической составляющей для начальных и конечных значений ряда имеет наибольшие значения (рисунок 3.9); распределения суммы квадратов отклонений значений побочных диагоналей матриц $X1$ и $X2$ имеют характерный вид (рисунок 3.10). Для тренда наибольшая ошибка ($g_i - g_i^B$) принимает отрицательные значения, для гармонической составляющей – положительные. Выявленные особенности оказывают существенное влияние на прогнозирующие свойства SSA-метода [35, 73].

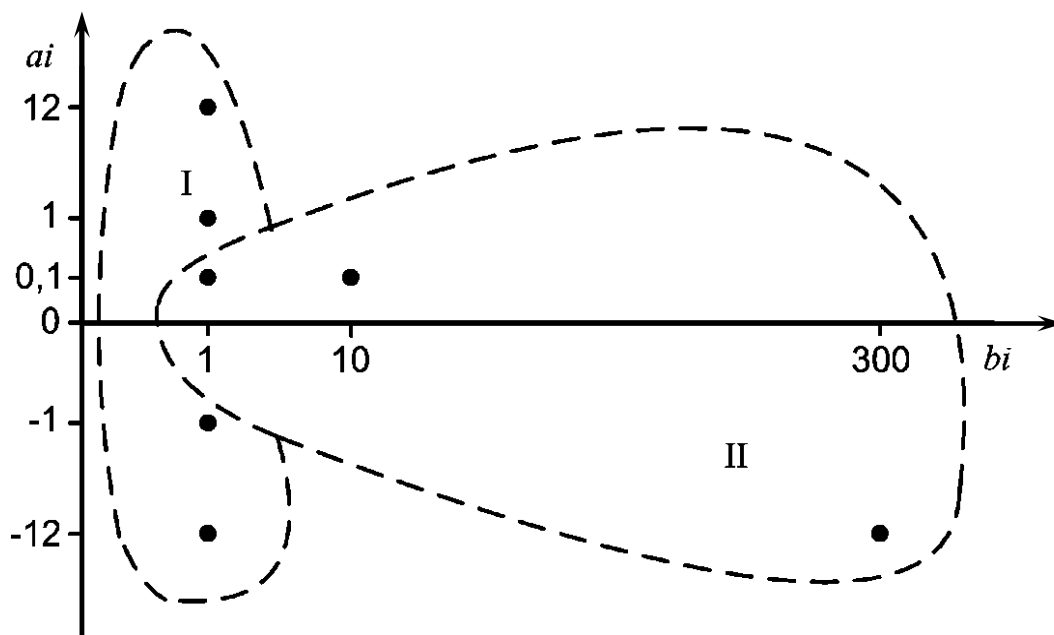


Рисунок 3.9 – Распределение абсолютной ошибки восстановления тренда



Рисунок 3.10 – Распределение суммы квадратов отклонений значений побочных диагоналей X_1

Особенность группировки составляющих ряда определяется перестановками сингулярных векторов, ответственных за трендовую и гармоническую составляющие (см. таблицу 3.1). По этому признаку образуются два кластера функций, представленных на рисунке 3.11.



I – кластер из функций $F_1(x)$, $F_6(x)$, $F_7(x)$; II – кластер из функций $F_2(x)$, $F_3(x)$, $F_4(x)$, $F_5(x)$

Рисунок 3.11 – Кластеризация исследуемых функций вида $F(x) = ax + b\sin x + c$

3.6 Исследование временных рядов с равномерным, нормальным и экспоненциальным шумом с помощью SSA-метода

3.6.1 Временной ряд с равномерным шумом.

Формула, задающая закон распределения случайной величины (функцию вероятности, плотность вероятности и т. п.), обязательно содержит хотя бы один параметр. Параметры закона распределения случайной величины принято делить на три вида: параметры положения μ , масштаба λ , формы θ . Распределения вероятностей случайных величин X и Y называют однотипными, если существуют постоянные $a > 0$ и b такие, что распределения случайных величин Y и $aX + b$ совпадают.

Пусть $X(0, 1)$ обозначает случайную величину с параметром положения μ , равным нулю, и параметром масштаба λ , равным единице, а $X(\mu, \lambda)$ – случайную величину того же самого типа с параметром положения μ и параметром масштаба λ . Тогда справедливы следующие соотношения:

$$X(\mu, \lambda) \sim \mu + \lambda X(0, 1); X(0, 1) \sim \frac{X(\mu, \lambda) - \mu}{\lambda}. \quad (3.19)$$

Запись $X \sim Y$ обозначает, что случайные величины X и Y имеют одно и то же распределение, т. е. имеют одинаковые функции вероятности, плотности вероятности и т. п. [23].

Равномерное (англ. uniform) распределение в научной литературе иногда называют прямоугольным. Случайная величина ξ называется равномерно распределенной на отрезке $[a, b]$, если ее плотность вероятности $f_{\xi}(x)$ постоянна на этом отрезке и равна нулю вне его, т. е.

$$f_{\xi}(x) = \begin{cases} \frac{1}{b-a} & | a \leq x \leq b; \\ 0 & | x < a \vee x > b. \end{cases} \quad (3.20)$$

Равномерное распределение иногда используется в качестве «нулевого приближения» в описании априорного распределения анализируемых параметров в так называемом байесовском подходе в условиях полного отсутствия априорной информации об этом распределении.

Основные числовые характеристики закона распределения случайной величины:

– характеристики положения: математическое ожидание, оцениваемое средним (mean); медиана (med) [23];

– характеристики рассеяния: среднеквадратическое отклонение, оцениваемое стандартным отклонением s ; максимум (max); минимум (min); диапазон (range); коэффициенты асимметрии распределения γ_3 и эксцесса γ_4 [23].

Цель исследования: изучение особенностей восстановления SSA-методом равномерной (нормальной, экспоненциальной) шумовой составляющей временного ряда.

Для исследования принята последовательность данных (ряд) G_1 с равномерным шумом

$$G_1 = Rnd(0; 1)_n, \quad (3.21)$$

где n – количество элементов ряда, $n = 43$,

что обусловлено ограничениями математического пакета Mathcad в режиме Insert Matrix.

Временной ряд G_1 сформирован в пакете Minitab: Calc\Random Data\Uniform...\ с параметрами, в соответствии с (3.20), $a = 0$, $b = 1$. После преобразования временного ряда G_1 SSA-методом получен восстановленный для шумовой составляющей ряд G_1^B , представленный на рисунке 3.12. При SSA-преобразовании получены собственные числа,

представленные на рисунке 3.13, где λ_1 соответствует трендовой составляющей G_{1T}^B .

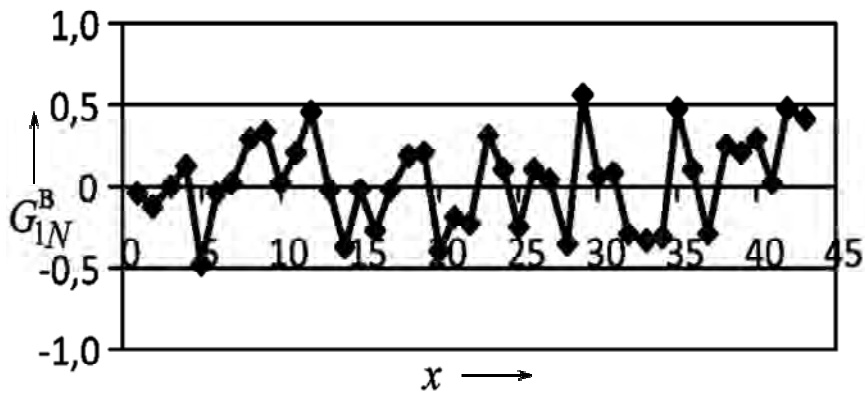


Рисунок 3.12 – Восстановленный шум ряда G_1^B

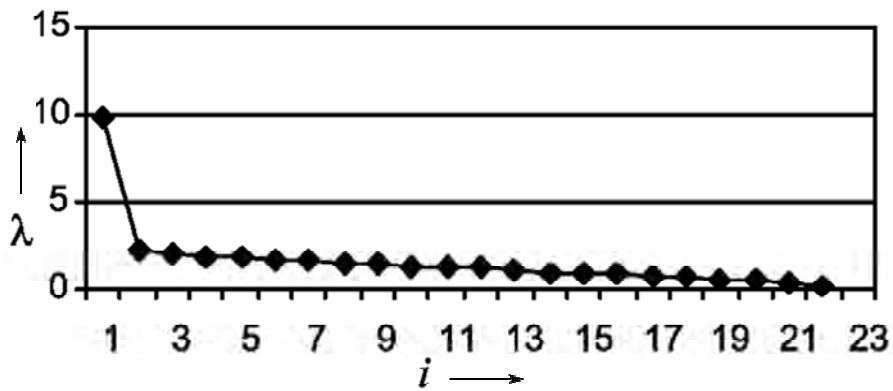


Рисунок 3.13 – Распределение собственных чисел

Основные числовые характеристики временного ряда G_1 с шумом, распределенным по равномерному закону с параметрами $a = 0$, $b = 1$, и характеристики восстановленной трендовой G_{1T}^B и шумовой G_{1N}^B составляющих представлены в таблице 3.2.

Таблица 3.2 – Основные характеристики временного ряда с равномерным шумом ($a = 0$, $b = 1$)

Ряд	mean	med	max	min	range	s	γ_3	γ_4
G_1	0,485	0,479	0,986	0,029	0,954	0,2903	-0,039	-0,938
G_{1N}^B	0,031	0,025	0,563	-0,484	1,047	0,2705	0,011	-0,784
G_{1T}^B	0,456	0,436	0,562	0,359	0,203	0,0662	0,332	-1,221

При этом трендовая составляющая G_{1T}^B восстановленного ряда равна математическому ожиданию (МО) исходного шума $Rnd(0; 1)$ в соответствии с обозначениями (3.19): $\mu = 0,5$.

Тогда исходная последовательность (3.21), в соответствии с (3.19), представлена следующим образом:

$$Rnd(0; 1) \sim Rnd^B(-0,5; 0,5) + 0,5 \mid \alpha = 1, \quad (3.22)$$

где $Rnd^B(-0,5; 0,5)$ – распределение восстановленного шума.

3.6.2 Временной ряд с нормальным шумом.

Нормальным называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$f(x) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left[-\frac{(x-\gamma)^2}{2\beta^2}\right], \quad (3.23)$$

где γ и β – параметры распределения [58].

Временные ряды G_2 и G_3 сформированы в пакете Minitab: Calc\Random Data\Normal...\ с параметрами $\gamma = 1$, $\beta = 1$ и $\gamma = 1$, $\beta = 5$ в соответствии с обозначениями (3.23).

Результаты исследования временного ряда с шумом по нормальному закону, проведенного аналогично временному ряду (3.21), представлены в таблицах 3.3 и 3.4 для разных значений параметров γ и β .

Таблица 3.3 – Основные характеристики временного ряда с нормальным шумом ($\gamma = 1$, $\beta = 1$)

Ряд	mean	med	max	min	range	s	γ_3	γ_4
G_2	0,994	0,944	2,918	-2,078	4,997	1,158	-0,323	-0,081
G_{2N}^B	0,004	-0,040	2,019	-2,937	4,956	1,075	-0,246	0,160
G_{2T}^B	0,985	0,941	1,486	0,864	0,622	0,148	2,141	3,827

При этом следует отметить для последующего обсуждения результатов, что

$$MO_H = \gamma \quad (3.24)$$

для нормального закона распределения случайной величины [93].

Таблица 3.4 – Основные характеристики временного ряда с нормальным шумом ($\gamma = 1, \beta = 5$)

Ряд	mean	med	max	min	range	s	γ_3	γ_4
G_3	1,184	1,160	12,048	-7,087	19,134	5,139	0,361	-0,753
G_{3N}^B	-0,375	-0,497	10,436	-8,420	18,856	4,763	0,330	-0,605
G_{3T}^B	1,539	1,194	3,966	0,370	3,596	0,937	1,349	0,745

3.6.3 Временной ряд с экспоненциальным шумом.

Часто на практике экспоненциальному закону распределения подчиняется длительность безотказной работы оборудования. Экспоненциальным называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$f(x) = \begin{cases} \frac{1}{\beta} \exp\left[\frac{-(x-\gamma)}{\beta}\right] & | x > \gamma, \\ 0 & | x \leq \gamma, \end{cases} \quad (3.25)$$

где γ и β – параметры распределения [93].

Для предварительного исследования принята последовательность данных (временной ряд) G с экспоненциальным шумом

$$G = \text{Exp}(\gamma; \beta)_n, n = 43, \quad (3.26)$$

где $\text{Exp}(\gamma; \beta)_n$ – функция, возвращающая n случайных чисел, распределенных по экспоненциальному закону с параметром смещения $\gamma = 0$ и параметром масштаба $\beta = 1$.

Последовательность данных G получена в пакете Minitab:

$G =$ 1,36; 0,87; 0,35; 0,12; 3,84; 0,20; 0,96; 2,02; 0,62; 0,68; 2,20;
1,16; 0,49; 0,35; 1,38; 0,10; 0,44; 3,59; 1,23; 1,30; 0,45; 0,71; 2,50;
0,51; 0,22; 0,80; 0,42; 0,68; 1,15; 0,14; 2,26; 0,56; 1,07; 0,85; 0,07;
1,09; 0,56; 0,20; 2,09; 0,45; 0,65; 0,55; 1,93.

После преобразования временного ряда G SSA-методом получены восстановленный для шумовой составляющей ряд \hat{G}_N и трендовая составляющая \hat{G}_T :

$$G \sim \hat{G} = \hat{G}_N + \hat{G}_T.$$

Проверка по критерию Колмогорова–Смирнова ($p = 0,01$) [48] соответствия полученных случайных чисел теоретическому экспоненциальному

распределению показала, что шумовая составляющая \hat{G}_N распределена по экспоненциальному закону. При этом трендовая составляющая \hat{G}_T восстановленного временного ряда \hat{G} совпадает с параметром масштаба β исходного временного ряда G .

Следует обратить внимание на гистограммы [90] распределения случайных чисел исходного временного ряда (рисунок 3.14) и восстановленной шумовой составляющей для этого ряда (рисунок 3.15).

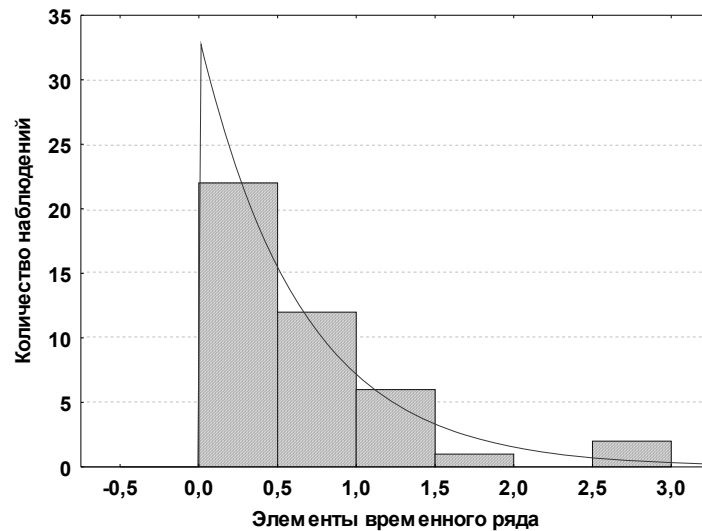


Рисунок 3.14 – Распределение случайных чисел исходного временного ряда G

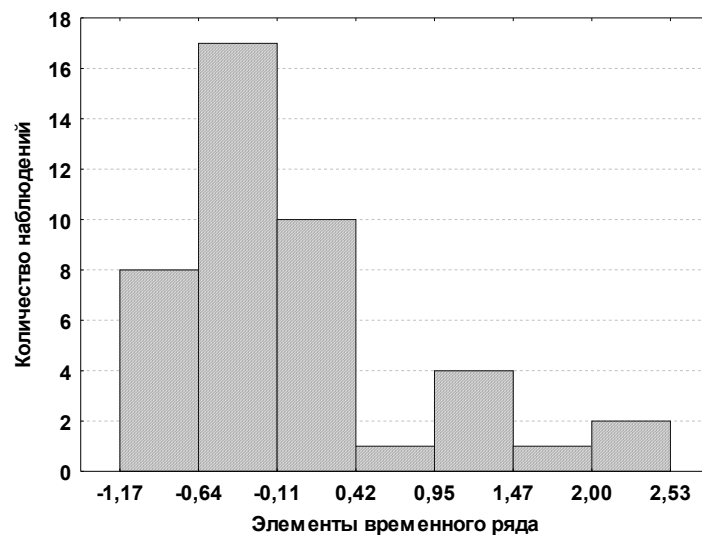


Рисунок 3.15 – Распределение случайных чисел восстановленной шумовой составляющей ряда G

На рисунке 3.14 представлен явно выраженный экспоненциальный закон распределения случайных чисел. Восстановленная шумовая составляющая

(см. рисунок 3.15) отличается визуально от исходного экспоненциального типа распределения.

В связи с явными отличиями представленных гистограмм на рисунках 3.14 и 3.15 ставится задача исследования восстановленной шумовой составляющей для поиска алгоритма определения типа распределения. Так как шумовая составляющая расположена в отрицательной и положительной областях вещественных чисел, рассмотрим восстановленный шум ряда G в следующем виде:

$$\hat{G}_{Nj} = \hat{G}_{Nj}^- + \hat{G}_{Nj}^+, \quad (3.27)$$

где \hat{G}_{Nj}^- – случайные числа в отрицательной области;

\hat{G}_{Nj}^+ – случайные числа в положительной области (см. рисунок 3.15).

Исследуем соответствие восстановленных случайных величин \hat{G}_{Nj}^- и \hat{G}_{Nj}^+ теоретическому экспоненциальному распределению по критерию Колмогорова–Смирнова [58]. Исследуем также и сумму этих временных рядов по формуле

$$\hat{G}'_{Nj} = |\hat{G}_{Nj}^-| + \hat{G}_{Nj}^+, \quad (3.28)$$

где $|\hat{G}_{Nj}^-|$ – модули случайных чисел в отрицательной области.

Временной ряд \hat{G}'_{Nj} представлен на рисунке 3.16.

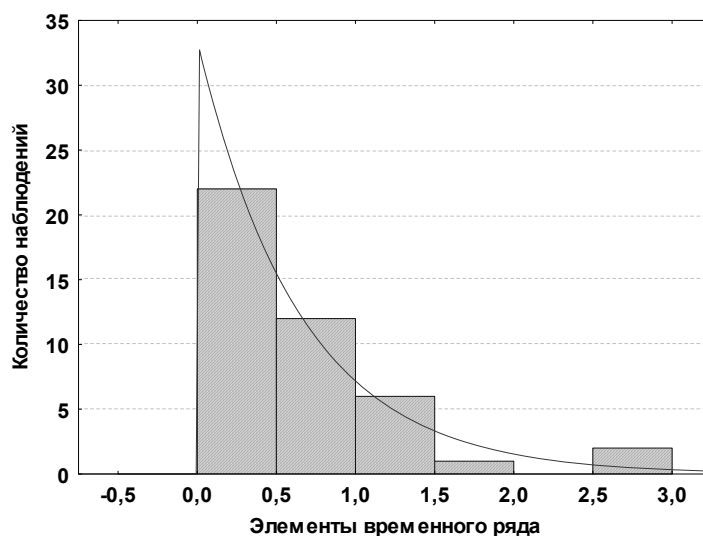


Рисунок 3.16 – Восстановленный ряд с данными, перенесенными из отрицательной области значений в положительную область значения ($n = 43$)

Критическое значение Δ_p для наибольшего отклонения эмпирического распределения от теоретического по [10, таблица 6.2] при $p = 0,01$ и, например, $n = 43$ равно 0,24332. Если наблюдаемое значение Δ больше или равно критическому значению, нулевая гипотеза H_0 отклоняется. Результаты исследований представлены в таблице 3.5.

Таблица 3.5 – Показатели соответствия восстановленных случайных величин \hat{G}_{Nj}^- и \hat{G}_{Nj}^+ теоретическому экспоненциальному распределению

Оценка	n	$\Delta_p (p = 0,01)$	Δ
\hat{G}_{N1}^+	15	0,40420	0,031
\hat{G}_{N1}^-	28	0,29971	0,149
\hat{G}_{N1}'	43	0,24332	0,161
\hat{G}_{N2}^+	16	0,39201	0,038
\hat{G}_{N2}^-	27	0,30502	0,142
\hat{G}_{N2}'	43	0,24332	0,141
\hat{G}_{N3}^+	16	0,39201	0,033
\hat{G}_{N3}^-	27	0,30502	0,142
\hat{G}_{N3}'	43	0,24332	0,161
\hat{G}_{N4}^+	18	0,37062	0,067
\hat{G}_{N4}^-	25	0,31657	0,151
\hat{G}_{N4}'	43	0,24332	0,161
\hat{G}_{N5}^+	17	0,38086	0,035
\hat{G}_{N5}^-	26	0,31064	0,158
\hat{G}_{N5}'	43	0,24332	0,184
\hat{G}_{N6}^+	16	0,39201	0,038
\hat{G}_{N6}^-	27	0,30502	0,142
\hat{G}_{N6}'	43	0,24332	0,161
\hat{G}_{N7}^+	17	0,38086	0,035
\hat{G}_{N7}^-	26	0,31064	0,158
\hat{G}_{N7}'	43	0,24332	0,184
\hat{G}_{N8}^+	17	0,38086	0,054
\hat{G}_{N8}^-	26	0,31064	0,158
\hat{G}_{N8}'	43	0,24332	0,184

Поскольку наблюдаемое значение Δ во всех случаях меньше критического значения Δ_p , то нулевая гипотеза H_0 не отклоняется, т. е. восстановленная шумовая составляющая соответствует теоретическому экспоненциальному закону распределения.

Для продолжения исследований временные ряды G_4 и G_5 получены в пакете Minitab: Calc\Random Data\Exponential...\ с параметрами $\gamma = 0, \beta = 1$ и $\gamma = 1, \beta = 1$ в соответствии с обозначениями (3.25).

Результаты исследования временного ряда с шумом, распределенным по экспоненциальному закону, представлены в таблицах 3.6 и 3.7 для разных значений параметров γ и β .

Таблица 3.6 – Основные показатели временного ряда с экспоненциальным шумом ($\gamma = 0, \beta = 1$)

Ряд	mean	med	max	min	range	s	γ_3	γ_4
G_4	0,9541	0,753	2,731	0,0098	2,721	0,757	0,785	–0,397
G_{4N}^B	–0,0013	–0,209	1,388	–1,028	2,416	0,698	0,648	–0,979
G_{4T}^B	0,955	0,914	1,549	0,606	0,943	0,254	0,655	–0,352

Таблица 3.7 – Основные показатели временного ряда с экспоненциальным шумом ($\gamma = 1, \beta = 1$)

Ряд	mean	med	max	min	range	s	γ_3	γ_4
G_5	2,172	1,692	7,911	1,009	6,903	1,420	2,825	8,609
G_{5N}^B	–0,004	–0,281	4,851	–1,759	6,610	1,296	2,004	5,273
G_{5T}^B	2,142	2,059	3,046	1,595	1,451	0,474	0,732	–0,790

При этом следует заметить, что

$$MO_3 = \gamma + \beta \quad (3.29)$$

для экспоненциального закона распределения случайной величины [93].

3.6.4 Обсуждение результатов исследования.

Исследуемый шум определяется типами непрерывных распределений, относящимися к различным классам. При нормальном распределении случайная величина принимает значения на положительной и отрицательной полуосях, экспоненциальное распределение относится к классу распре-

делений с возможными значениями на положительной полуоси, равномерное – к классу распределений с возможными значениями на ограниченном интервале.

В ходе анализа полученных результатов выявлены следующие факты.

Исходный ряд G_1 (см. таблицу 3.2) с равномерным распределением случайных величин имеет параметры, отличающиеся от теоретических с погрешностью 5,8 % для среднего mean, 0,6 % для среднеквадратического отклонения и 4,6 % для диапазона range. Восстановленный ряд G_{1N}^B имеет среднее 0,031, близкое к нулю. Стандартное отклонение отличается от исходного ряда на 6,8 %. SSA-метод выделяет трендовую составляющую G_{1T}^B со значением среднего 0,456, что с погрешностью 6,0 % совпадает со средним исходного ряда G_1 .

Для проверки согласия восстановленных случайных величин G_i^B , $i = 1, \dots, 5$, теоретическому распределению используется критерий Колмогорова–Смирнова. Критическое значение Δ для наибольшего отклонения эмпирического распределения от теоретического по [10, таблица 6.2] при $p = 0,01$ и $n = 43$ равно 0,24332. Если наблюдаемое значение Δ больше или равно критическому значению, нулевая гипотеза H_0 отклоняется. Например, для ряда G_1 наблюдаемое значение $\Delta = 0,091$, для восстановленного ряда G_{1N}^B (см. таблицу 3.1) $\Delta = 0,118$, что меньше критического, и нулевая гипотеза H_0 о равномерном распределении восстановленных случайных величин принимается.

Исходный ряд G_2 (см. таблицу 3.3) с нормальным распределением случайных величин имеет параметры, отличающиеся от теоретических с погрешностью 0,6 % для среднего mean и 15,6 % для среднеквадратического отклонения. Восстановленный ряд G_{2N}^B имеет среднее 0,004, близкое к нулю. Стандартное отклонение отличается от исходного ряда на 6,8 %. Диапазон случайных величин отличается от исходного ряда на 0,8 %. SSA-метод выделяет трендовую составляющую G_{2T}^B со значением среднего 0,985, что с погрешностью 0,9 % совпадает со средним исходного ряда G_2 .

Исходный ряд G_3 (см. таблицу 3.4) с нормальным распределением случайных величин имеет параметры, отличающиеся от теоретических с погрешностью 18,4 % для среднего mean и 2,8 % для среднеквадратического отклонения. Восстановленный ряд G_{3N}^B имеет среднее $-0,375$. Диапазон случайных величин отличается от исходного ряда на 1,5 %, а стандартное отклонение отличается от исходного ряда на 7,3 %. SSA-метод выделяет

трендовую составляющую G_{3T}^B со значением среднего 1,539, что с погрешностью 30,0 % совпадает со средним исходного ряда G_3 . Увеличение погрешности восстановления объясняется ростом дисперсии шума в 25 раз.

Исходный ряд G_4 (см. таблицу 3.6) с экспоненциальным распределением случайных величин имеет отклонение от математического ожидания с погрешностью 4,6 %. Восстановленный ряд G_{4N}^B имеет среднее $-0,0013$, практически равное нулю. Стандартное отклонение отличается от исходного ряда на 7,8 %. SSA-метод выделяет трендовую составляющую G_{4T}^B со значением среднего 0,955, что с погрешностью 0,1 % совпадает со средним исходного ряда G_4 .

Исходный ряд G_5 (см. таблицу 3.7) с экспоненциальным распределением случайных величин имеет отклонение от математического ожидания с погрешностью 8,6 %. Восстановленный ряд G_{5N}^B имеет среднее $-0,004$, близкое к нулю. Стандартное отклонение отличается от исходного ряда на 8,7 %. SSA-метод выделяет трендовую составляющую G_{5T}^B со значением среднего 2,142, что с погрешностью 1,4 % совпадает со средним исходного ряда G_5 [72].

4 Интеллектуальный анализ данных при оптимизации распределения ресурсов в процессе производства готовой продукции

4.1 Решение задачи многокритериального управления производственным процессом крашения тканей набивным способом

Динамическое программирование [28, 29] в общем случае есть один из методов [34, 40, 55] последовательного нахождения оптимальных решений задачи со сложной многоэтапной структурой [63]: задачи оптимального выбора технологических режимов вида (2.2) производственного процесса крашения тканей набивным способом (2.1).

Производственный процесс крашения ткани набивным способом – управляемая система [31], которая находится в одном из возможных состояний, трансформируемых на каждом из этапов ТП посредством управляющего воздействия (управления). Целевая функция (ЦФ), зависящая от многих параметров (состояния системы, управляющего воздействия), оценивает эффективность процесса управления.

Пусть в начальный момент времени – нулевой этап ТП (2.1) – система находится в исходном состоянии τ_0 (рисунок 4.1). На следующем шаге в результате управления y_1 система из состояния τ_0 переходит в состояние $\tau_1 = g_1(\tau_0, y_1)$ – переход с нулевого этапа ТП (2.1) на первый этап ТП (рисунок 4.1).

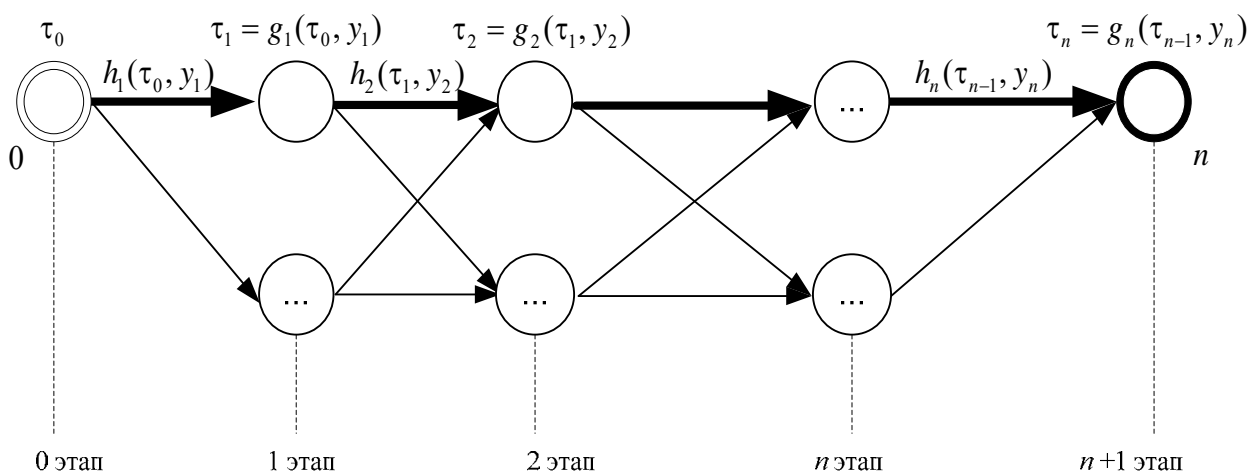


Рисунок 4.1 – Последовательность состояний и управлений

При этом достигается эффект $h_1(\tau_0, y_1)$. Далее управление y_2 переводит систему из состояния τ_1 в состояние $\tau_2 = g_2(\tau_1, y_2)$ (см. рисунок 4.1) и достигается эффект $h_2(\tau_1, y_2)$. Таким образом, за n этапов ТП строится последовательность состояний $\tau_0, \tau_1, \dots, \tau_n$ и последовательность управлений y_1, y_2, \dots, y_n , где $\tau_{i+1} = g_{i+1}(\tau_i, y_{i+1})$, $i = 1, \dots, n$.

Общий эффект по каждому этапу вычисляется по формуле

$$J_i(\tau_0, y_1, y_2, \dots, y_i) = h_1(\tau_0, y_1) + h_2(\tau_1, y_2) + \dots + h_i(\tau_{i-1}, y_i), \quad i = 1, \dots, n. \quad (4.1)$$

Необходимо отыскать такую последовательность управлений $y_1^*, y_2^*, \dots, y_i^*$, при которой ЦФ J_i (2.1) достигает \max либо $\min(\max(\min))$, т. е. выполняется условие

$$J_i(\tau_0, y_1^*, y_2^*, \dots, y_i^*) = \max_{y_1, y_2, \dots, y_i}(\min)(J_i(\tau_0, y_1, y_2, \dots, y_i)), \quad i = 1, \dots, n. \quad (4.2)$$

Согласно принципу оптимальности Беллмана при последовательности $y_1^*, y_2^*, \dots, y_i^*$, $i = 1, \dots, n$, оптимальных управлений должна достигать \max либо \min [57] и каждая из функций:

$$f_i(\tau_{i-1}, y_i, y_{i+1}, \dots, y_n) = h_i(\tau_{i-1}, y_i) + h_{i+1}(\tau_i, y_{i+1}) + \dots + h_n(\tau_{n-1}, y_n), \quad i = 1, \dots, n. \quad (4.3)$$

Введем следующее обозначение:

$$\varphi_i(\tau_{i-1}) \stackrel{\text{def}}{=} \max_{y_i, y_{i+1}, \dots, y_n}(\min)(f_i(\tau_{i-1}, y_i, y_{i+1}, \dots, y_n)), \quad i = 1, \dots, n. \quad (4.4)$$

Тогда с учетом принципа (4.3) и обозначения (4.4) имеют место функциональные уравнения Беллмана [46]

$$\varphi_i(\tau_{i-1}) = \max_{y_i}(\min)(\varphi_{i+1}(g_{i+1}(\tau_{i-1}, y_i)) + h_i(\tau_{i-1}, y_i)), \quad i = 1, \dots, n. \quad (4.5)$$

Решение функциональных уравнений Беллмана (4.5) дает возможность построить последовательность оптимальных управляющих воздействий и соответствующих им значений ЦФ. Уравнение (4.5) с учетом обозначений τ_{uv} нагрузок дуг графа ПП КТНС (см. рисунок 2.9) приобретает следующий вид:

$$\tau_v = \min_u (\max_u \{ \tau_u + \tau_{uv} \}), \tau_0 = 0, \quad (4.6)$$

где $v \in M_B(i)$ – номера вершин, входящих в дуги графа i -го этапа ТП;

$M_B(i)$ – множество, обозначающее номера вершин графа ТП вида (4.5) для i -го этапа ТП;

$u \in M_B(i-1)$ – номера вершин графа ПП КТНС, из которых исходят дуги $(i-1)$ -го этапа ПП.

Нагруженные дуги, которые исходят из одинаковых вершин графа, эквивалентны вследствие интерпретации характеристик нагрузок на дуги графа – это затраты времени и/или стоимости расхода ресурсов для выпуска определенного количества продукции для одного технологического режима ПП КТНС (2.1). Справедливы соотношения.

Соотношение 1. Для одномерных нагрузок $\tau_{uv} \in R$ на дуги графа ПП КТНС (2.1) справедливы равенства [18]

$$(\forall u \in M_B(i-1))(\forall v_1, v_2 \in M_B(i)) [\tau_{uv_1} = \tau_{uv_2}], i = 1, \dots, n. \quad (4.7)$$

Соотношение 2. Для двумерных нагрузок на дуги графа ПП КТНС $\tau_{uv} = (\tau_{uv}^{(1)}, \tau_{uv}^{(2)}) \in R^2$, $\tau_{uv_1} = (\tau_{uv_1}^{(1)}, \tau_{uv_1}^{(2)})$, $\tau_{uv_2} = (\tau_{uv_2}^{(1)}, \tau_{uv_2}^{(2)})$ истинны равенства

$$\begin{aligned} & (\forall u \in M_B(i-1))(\forall v_1, v_2 \in M_B(i)) \left[\left(\tau_{uv_1}^{(1)} = \tau_{uv_2}^{(1)} \right) \wedge \left(\tau_{uv_1}^{(2)} = \tau_{uv_2}^{(2)} \right) \Leftrightarrow \right. \\ & \left. \Leftrightarrow \left(\tau_{uv_1} = \tau_{uv_2} \right) \right], i = 1, \dots, n. \end{aligned} \quad (4.8)$$

Соотношение 3. Для m -мерных нагрузок на дуги графа ПП КТНС (2.1) $\tau_{uv} = (\tau_{uv}^{(1)}, \tau_{uv}^{(2)}, \dots, \tau_{uv}^{(m)}) \in R^m$ справедливы равенства [18]

$$\begin{aligned}
& (\forall u \in M_B(i-1))(\forall v_1, v_2 \in M_B(i)) \left[\left(\tau_{uv_1}^{(1)} = \tau_{uv_2}^{(1)} \right) \wedge \left(\tau_{uv_1}^{(2)} = \tau_{uv_2}^{(2)} \right) \wedge \dots \wedge \right. \\
& \left. \wedge \left(\tau_{uv_1}^{(m)} = \tau_{uv_2}^{(m)} \right) \Leftrightarrow \left(\tau_{uv_1} = \tau_{uv_2} \right) \right], \quad i = 1, \dots, n.
\end{aligned} \tag{4.9}$$

Утверждение 1. Для одномерных нагрузок $\tau_{uv} \in R$ на дуги графа ПП КТНС (2.1) из истинности равенств (4.7) следует истинность представления функционального уравнения Беллмана (4.6) в виде [18]

$$\tau_v = L_i = L_{i-1} + \min_{u \in M_B(i-1)} (\max) \{ \tau_{uv} \}, \quad i = 1, \dots, n, \tag{4.10}$$

где L_i – оптимальное состояние на i -м этапе, $L_i \in R$, $L_0 = \tau_0 = 0$;

(2.5)

$v \in M_B(i)$ – номера вершин, в которые входят дуги графа i -го этапа ТП;

(2.5)

$u \in M_B(i-1)$ – номера вершин, из которых исходят дуги графа $(i-1)$ -го этапа ТП.

Доказательство утверждения 1 проводится методом математической индукции.

$$\text{Этап 1. } u \in M_B(0) \equiv \{0\}, \quad v \in M_B(1) \equiv \{1, \dots, M_1\}. \tag{2.5}$$

$$\tau_v = \tau_0 + \tau_{0v} = [\tau_0 = 0] = \tau_{0v}, \quad v \in M_B(1); \tag{4.10}$$

$$\left(\tau_{01}^{(4.7)} = \tau_{02}^{(4.7)} = \dots = \tau_{0M_1}^{(4.7)} \right) \Leftrightarrow (\tau_1 = \tau_2 = \dots = \tau_{M_1} = L_1). \tag{4.12}$$

$$\text{Этап 2. } u \in M_B(1) \equiv \{1, \dots, M_1\}, \quad v \in M_B(2) \equiv \{M_1 + 1, \dots, M_1 + M_2\}. \tag{2.5}$$

$$\tau_v = \min_{u \in M_B(1)} (\max) \{ \tau_u + \tau_{uv} \} \stackrel{(4.12)}{=} \min_{u \in M_B(1)} (\max) \{ L_1 + \tau_{uv} \} = L_1 + \min_{u \in M_B(1)} (\max) \{ \tau_{uv} \}; \tag{4.13}$$

$$\begin{aligned}
& \left(\tau_1(\min(M_B(2))) \stackrel{(4.7)}{=} \dots \stackrel{(4.7)}{=} \tau_1(\max(M_B(2))) \right) \wedge \left(\tau_2(\min(M_B(2))) \stackrel{(4.7)}{=} \dots \stackrel{(4.7)}{=} \tau_2(\max(M_B(2))) \right) \wedge \dots \wedge \\
& \wedge \left(\tau_{M_1}(\min(M_B(2))) \stackrel{(4.7)}{=} \dots \stackrel{(4.7)}{=} \tau_{M_1}(\max(M_B(2))) \right) \Rightarrow \\
& \Rightarrow \left(\begin{aligned} & \min(\max)_{u \in M_B(1)} \{ \tau_u(\min(M_B(2))) \} = \min(\max)_{u \in M_B(1)} \{ \tau_u(\min(M_B(2))+1) \} = \\ & = \dots = \min(\max)_{u \in M_B(1)} \{ \tau_u(\max(M_B(2))) \} \end{aligned} \right) \Rightarrow \\
& \Rightarrow \left(\tau_{\min(M_B(2))} = \tau_{\min(M_B(2))+1} = \dots = \tau_{\max(M_B(2))} = L_2 \right); \\
& \tau_v = L_1 + \min(\max)_{u \in M_B(1)} \{ \tau_{uv} \} = L_2, \quad v \in M_B(2).
\end{aligned}$$

Этап i . Пусть для любого $i \in \{1, \dots, n\}$ выполняется условие

$$\tau_v = L_i = L_{i-1} + \min(\max)_{u \in M_B(i-1)} \{ \tau_{uv} \}, \quad v \in M_B(i) \stackrel{(2.5)}{\equiv} \left\{ \sum_{j=0}^{i-1} M_j + 1, \dots, \sum_{j=0}^i M_j \right\}. \quad (4.14)$$

$$\text{Этап } i+1. \quad u \stackrel{(4.14)}{\in} M_B(i), \quad v \in M_B(i+1) \stackrel{(2.5)}{\equiv} \left\{ \sum_{j=0}^i M_j + 1, \dots, \sum_{j=0}^{i+1} M_j \right\}.$$

Докажем равенство $\tau_{\min(M_B(i))} = \tau_{\min(M_B(i))+1} = \dots = \tau_{\max(M_B(i))}$.

$$\tau_v = \min(\max)_{u \in M_B(i)} \{ \tau_u + \tau_{uv} \} \stackrel{(4.14)}{=} L_i + \min(\max)_{u \in M_B(i)} \{ \tau_{uv} \},$$

$$\begin{aligned}
& \left(\overset{def}{\tau_{(\min(M_B(i)))}(\min(M_B(i+1)))} = \dots = \overset{def}{\tau_{(\min(M_B(i)))}(\max(M_B(i+1)))} \right) \wedge \\
& \wedge \left(\overset{def}{\tau_{(\min(M_B(i))+1)}(\min(M_B(i+1)))} = \dots = \overset{def}{\tau_{(\min(M_B(i))+1)}(\max(M_B(i+1)))} \right) \wedge \dots \wedge \\
& \wedge \left(\overset{def}{\tau_{(\max(M_B(i)))}(\min(M_B(i+1)))} = \dots = \overset{def}{\tau_{(\max(M_B(i)))}(\max(M_B(i+1)))} \right) \Rightarrow \\
& \Rightarrow \left(\min_{u \in M_B(i)}(\max)\{\tau_u(\min(M_B(i+1)))\} = \min_{u \in M_B(i)}(\max)\{\tau_u(\min(M_B(i+1))+1)\} = \right. \\
& \quad \left. = \dots = \min_{u \in M_B(i)}(\max)\{\tau_u(\max(M_B(i+1)))\} \right) \Rightarrow \\
& \Rightarrow (\tau_{\min(M_B(i))} = \dots = \tau_{\max(M_B(i))} = L_{i+1}). \tag{4.15}
\end{aligned}$$

Равенство доказано, из чего следует соотношение

$$\tau_v = L_{i+1} = L_i + \min_{u \in M_B(i)}(\max)\{\tau_{uv}\}, \quad u \in M_B(i), \quad v \in M_B(i+1). \tag{4.16}$$

Следовательно, соотношение (4.10) имеет место для любого $i \in \{1, \dots, n\}$.

Доказательство закончено.

Замечание к утверждению 1. Пусть $isFixOp_i \in \{True, False\}$ – управляющий параметр фиксации операции $(i-1)$ -го этапа; $isFixTR_{ij_0} \in True$ – параметр фиксации ТР данного этапа (вершины графа ТП с номером $u_{j_0} \in M_B(i-1)$). Тогда с учетом введенных в рассмотрение управляющих параметров фиксации [13, 18] функциональное уравнение (4.10) принимает следующий вид:

$$\tau_v = L_i = L_{i-1} + \left(\left\{ \begin{array}{l} \min_{u \in M_B(i-1)}(\max)\{\tau_{uv}\} \Big| isFixOp_i = False, \\ \tau_{u_{j_0}v} \Big| isFixOp_i = True \end{array} \right. \right),$$

$$L_i, \tau_{uv} \in R, \quad L_0 = \tau_0 = 0, \quad v \in M_B(i), \quad u \in M_B(i-1), \quad i = 1, \dots, n. \tag{4.17}$$

Утверждение 2. Для двумерных нагрузок $\tau_{uv} = (\tau_{uv}^{(1)}, \tau_{uv}^{(2)}) \in R^2$ на дуги графа ПП КТНС (2.1) из истинности равенств (4.8) следует истинность представления функционального уравнения Беллмана (4.6) в виде [18]

$$L_i = \left\{ \begin{array}{l} \left(\begin{array}{l} L_{(i-1)1} + \underset{u \in M_B(i-1)}{Fm_1} \left(\{\tau_{uv}^{(1)}\} \right), \\ L_{(i-1)2} + Fm_2 \left\{ \tau_{u,v}^{(2)} \middle| \tau_{u,v}^{(1)} = \underset{u \in M_B(i-1)}{Fm_1} \left(\{\tau_{uv}^{(1)}\} \right) \right\}, \end{array} \right) \Bigg| pOpt = (1, 2); \\ \left(\begin{array}{l} L_{(i-1)1} + Fm_1 \left\{ \tau_{u,v}^{(1)} \middle| \tau_{u,v}^{(2)} = \underset{u \in M_B(i-1)}{Fm_2} \left(\{\tau_{uv}^{(2)}\} \right) \right\}, \\ L_{(i-1)2} + \underset{u \in M_B(i-1)}{Fm_2} \left(\{\tau_{uv}^{(2)}\} \right) \end{array} \right) \Bigg| pOpt = (2, 1), \end{array} \right. \quad (4.18)$$

где $L_i = (L_{i1} \quad L_{i2})^T$, $L_0 = (0 \quad 0)^T$;

$i = 1, \dots, n$ – номера этапов ТП;

$u \in M_B(i-1)$ – номера вершин графа ПП КТНС, из которых исходят дуги $(i-1)$ -го этапа ПП;

$v \in M_B(i)$ – номера вершин графа, в которые входят дуги i -го этапа ПП;

$pOpt \in R^2$ (где $pOpt_k \stackrel{def}{\in} \{1, 2\}$, $k \in 1, \dots, 2$) – параметр приоритета критериев оптимизации ПП КТНС (2.1) по первой (времени), затем по второй (стоимости ресурсов) координате $\tau_{uv} \in R^2$, или наоборот;

$Fm_1, Fm_2 \in \{\min, \max\}$ – критерии оптимизации ПП КТНС для первой и второй координат векторов состояний на этапах ПП КТНС, которые лексикографически упорядочил параметр $pOpt$, соответственно.

Замечание 1 к утверждению 2. Для доказательства утверждения 2, по аналогии с доказательством утверждения 1, применяется метод математической индукции. При этом учитываются значения координат векторов $\tau_{uv}^{(pOpt_k)} \in R$, $k \in 1, \dots, 2$. Оптимизация ПП (2.1) с использованием метода динамического программирования (в рассматриваемой задаче выбора оптимальных ТР ПП КТНС с учетом многих критериев оптимизации) вначале осуществляется по более значимому критерию (координате $pOpt_1 \in \{1, 2\}$ для векторов $\tau_{uv}^{(pOpt)} \in R$). Если оптимальные значения нескольких ТР этапа ПП КТНС по более значимому критерию совпадают, производится оптимизация по менее значимой координате-критерию ($pOpt_2 \in \{1, 2\}$ для векторов

$\tau_{uv}^{(pOpt_k)} \in R$, $k \in 1, \dots, 2$) и по ней определяются предпочитаемые технологические режимы.

Замечание 2 к утверждению 2. Пусть $isFixOp_i \in \{True, False\}$ – управляющий параметр фиксации операции $(i-1)$ -го этапа; $isFixTR_{ij_0} \in True$ – параметр фиксации ТР данного этапа (вершины графа ТП с номером $u_{j_0} \in M_B(i-1)$). Тогда с учетом введенных в рассмотрение управляющих параметров фиксации функциональное уравнение (4.18) принимает следующий вид [13, 18]:

$$\tau_v = \begin{cases} \left(\begin{array}{l} L_{(i-1)1} + Fm_1 \left(\{\tau_{uv}^{(1)}\} \right), \\ L_{(i-1)2} + Fm_2 \left\{ \tau_{u,v}^{(2)} \middle| \tau_{u,v}^{(1)} = Fm_1 \left(\{\tau_{uv}^{(1)}\} \right) \right\}, \end{array} \right)_{pOpt=(1,2)} \\ \left(\begin{array}{l} L_{(i-1)1} + Fm_1 \left\{ \tau_{u,v}^{(1)} \middle| \tau_{u,v}^{(2)} = Fm_2 \left(\{\tau_{uv}^{(2)}\} \right) \right\}, \\ L_{(i-1)2} + Fm_2 \left(\{\tau_{uv}^{(2)}\} \right), \end{array} \right)_{pOpt=(2,1)} \end{cases} \left| \begin{array}{l} isFixOp_i = False, \\ isFixOp_i = True, u_{j_0} \in M_B(i-1) \end{array} \right. \quad (4.19)$$

где $\tau_{uv} \in R^2$, $\tau_v = L_i = (L_{i1} \ L_{i2})^T$, $L_0 = \tau_0 = (0 \ 0)^T$;

$u \in M_B(i-1)$ – номера вершин $(i-1)$ -го этапа ТП;

$v \in M_B(i)$ – номера вершин i -го этапа ТП;

$i = 1, \dots, n$ – номера этапов ТП;

$pOpt \in R^2$ – параметр приоритета критериев оптимизации ТП:

$pOpt_k \stackrel{def}{\in} \{1, 2\}$, $k \in 1, \dots, 2$;

$Fm_1, Fm_2 \in \{\min, \max\}$ – критерии оптимизации ПП КТНС для первой и второй координат векторов состояний на этапах ПП КТНС, которые лексикографически упорядочил параметр $pOpt$, соответственно.

Утверждение 3. Для m -мерных нагрузок $\tau_{uv} = (\tau_{uv}^{(1)}, \tau_{uv}^{(2)}, \dots, \tau_{uv}^{(m)}) \in R^m$ на дуги графа ПП КТНС (2.1) из истинности равенств (4.8) следует истинность представления функционального уравнения Беллмана (4.6) в виде [18]

$$L_i = \begin{cases} G_1(L_{(i-1)}, \{\tau_{uv}\}) \big| pOpt = I_1 = (1, 2, 3, \dots, m), \\ G_2(L_{(i-1)}, \{\tau_{uv}\}) \big| pOpt = I_2 = (2, 1, 3, \dots, m), \\ \dots \\ G_{m!}(L_{(i-1)}, \{\tau_{uv}\}) \big| pOpt = I_{m!}, \end{cases} \quad (4.20)$$

где $L_i = (L_{i1} \ L_{i2} \ \dots \ L_{im})^T$, $L_0 = (0 \ 0 \ \dots \ 0)^T$, $\dim(L_i) = m \times 1$;

$i = 1, \dots, n$ – номера этапов ТП;

$u \in M_B(i-1)$ – номера вершин $(i-1)$ -го этапа ТП, из которых исходят дуги графа;

$v \in M_B(i)$ – номера вершин, в которые входят дуги графа i -го этапа ТП;

$pOpt \in R^m$, $pOpt_k \stackrel{def}{\in} \{1, 2, 3, \dots, m\}$, $k \in 1, \dots, m$ – параметр приоритета критериев оптимизации ПП КТНС (2.1) по первой (времени), затем по второй (стоимости первой группы ресурсов ПП) и т. д., по m -й (стоимости $(m-1)$ -й группы ресурсов ПП) координате $\tau_{uv} \in R^m$;

$I_1, I_2, \dots, I_{m!}$ – перестановки приоритета критериев параметра $pOpt$.

$$G_1(L_{(i-1)}, \{\tau_{uv}\}) = \begin{pmatrix} L_{(i-1)1} + \underset{u \in M_B(i-1)}{Fm_1} \left(\{\tau_{uv}^{(1)}\} \right), \\ L_{(i-1)2} + Fm_2 \left\{ \tau_{u_1v}^{(2)} \big| \tau_{u_1v}^{(1)} = \underset{u \in M_B(i-1)}{Fm_1} \left(\{\tau_{uv}^{(1)}\} \right) \right\}, \\ \dots \\ L_{(i-1)n} + Fm_n \left\{ \tau_{u_{(n-1)}v}^{(n)} \big| \tau_{u_{(n-1)}v}^{(n-1)} = Fm_{n-1} \left\{ \tau_{u_{(n-2)}v}^{(n-1)} \right\} \right\} \end{pmatrix}; \quad (4.21)$$

$$G_2(L_{(i-1)}, \{\tau_{uv}\}) = \begin{pmatrix} L_{(i-1)1} + Fm_1 \left\{ \tau_{u_1v}^{(1)} \big| \tau_{u_1v}^{(2)} = \underset{u \in M_B(i-1)}{Fm_2} \left(\{\tau_{uv}^{(2)}\} \right) \right\}, \\ L_{(i-1)2} + \underset{u \in M_B(i-1)}{Fm_2} \left(\{\tau_{uv}^{(2)}\} \right), \\ \dots \\ L_{(i-1)n} + Fm_n \left\{ \tau_{u_{(n-1)}v}^{(n)} \big| \tau_{u_{(n-1)}v}^{(n-1)} = Fm_{n-1} \left\{ \tau_{u_{(n-2)}v}^{(n-1)} \right\} \right\} \end{pmatrix} \quad (4.22)$$

и т. д.,

где $Fm_1, Fm_2, \dots, Fm_m \in \{\min, \max\}$ – критерии оптимизации ПП КТНС для первой, второй, ..., m -й координат векторов состояний на этапах ПП КТНС, которые лексикографически упорядочил параметр $pOpt$.

Далее приведены **шаги методики оптимизации ПП КТНС** (2.1), в основу которой легли утверждения 1, 2, 3.

Шаг 1. На этапах ПП КТНС строится последовательность оптимальных состояний ПП вида

$$L_i \stackrel{def}{=} \varphi_i(L_{i-1}, S_i, T_i, Params_i), L_i, L_{i-1} \in \{R, R^2\}, L_0 \in \{0, (0 \ 0)^T\}, \quad (4.23)$$

где $L_i, L_{i-1} \in R$ – итоговые временные и/или стоимостные накопления относительно затрачиваемых ресурсов для $i, i \in 1, \dots, N$, этапа ПП (2.1). Так, при $L_i, L_{i-1} \in R^2$, $L_i = (L_{i1} \ L_{i2})^T$ имеем L_{i1} – оптимальное состояние ПП по времени и L_{i2} – оптимальное состояние ПП по стоимости затрат ресурсов;

φ_i – целевые функции многих измерений (одно-, двумерные), зависящие от многих критериев (оптимизация по времени и/или оптимизация по стоимости затрат ресурсов);

$Params_i$ – параметры, посредством которых осуществляется управление

ПП КТНС (2.1). Например, параметры $pOpt \stackrel{def}{\in} \{1, 2\}$ и $Fm_1, Fm_2 \in \{\min, \max\}$.

Шаг 2. На этапах ПП КТНС строится оптимальный маршрут по графу путем выбора оптимальных технологических режимов ПП КТНС – последовательность оптимальных управлений ПП КТНС – предикатов вида

$$Fl \stackrel{def}{=} \{fl_{ij} \mid fl_{ij} \in \{True, False\}, i = 1, \dots, n, j = 1, \dots, M_i\}. \quad (4.24)$$

Замечание 1 к методике оптимизации. Пусть $\tau_{uv}, L_i, L_{i-1} \in R$. Тогда с учетом введенных ранее управляющих параметров $isFixOp_i, isFixTR_{ij_0} \in \{True, False\}$ фиксации [13, 18] операции $(i-1)$ -го этапа параметр фиксации ТР данного этапа – функция (4.24), соответствующая функциональному уравнению (4.17), имеет вид:

$$Fl_1(\tau_{uv}, i, k_0, j_0, isFixOp_i) = \begin{cases} \tau_{u_{k_0}v} = \min_{u \in M_B(i-1)}(\max)\{\tau_{uv}\} & | isFixOp_i = False; \\ \begin{cases} True & | j_0 = k_0, \\ False & | j_0 \neq k_0 \end{cases} & | isFixOp_i = True, \end{cases} \quad (4.25)$$

где $\tau_0 = 0$;

$i = 1, \dots, n$ – номера этапов ТП;

$u \in M_B(i-1)$ – номера вершин $(i-1)$ -го этапа ТП;

$v \in M_B(i)$ – номера вершин i -го этапа ТП;

$u_{k_0}, u_{j_0} \in M_B(i-1)$ – фиксированные номера вершин $(i-1)$ -го этапа ТП,

где $k_0, j_0 \in \{1, \dots, M_i\}$.

Замечание 2 к методике оптимизации. Пусть $\tau_{uv}, L_i, L_{i-1} \in R^2$. Тогда с учетом введенных ранее управляющих параметров $isFixOp_i, isFixTR_{ij_0} \in \{True, False\}$ функция (4.24), соответствующая функциональному уравнению (4.19), имеет вид (2.26) [18]:

$$Fl_1(\tau_{uv}, i, k_0, j_0, isFixOp_i, pOpt) = \begin{cases} \left(\left(\tau_{u_{k_0}v}^{(1)} = Fm_1 \left(\tau_{uv}^{(1)} \right) \right) \wedge \right. \\ \left. \wedge \left(\tau_{u_{k_0}v}^{(2)} = Fm_2 \left\{ \tau_{u_i v}^{(2)} \mid \tau_{u_i v}^{(1)} = Fm_1 \left(\tau_{uv}^{(1)} \right) \right\} \right) \right) & \left. \begin{matrix} pOpt = (1, 2), \\ isFixOp_i = False, \end{matrix} \right. \\ = \begin{cases} \left(\tau_{u_{k_0}v}^{(1)} = Fm_1 \left\{ \tau_{u_i v}^{(1)} \mid \tau_{u_i v}^{(2)} = Fm_2 \left(\tau_{uv}^{(2)} \right) \right\} \right) \wedge \\ \left(\tau_{u_{k_0}v}^{(2)} = Fm_2 \left(\tau_{uv}^{(2)} \right) \right) & \left. \begin{matrix} pOpt = (2, 1), \\ isFixOp_i = True \end{matrix} \right. \\ \left. \begin{cases} True & | j_0 = k_0, \\ False & | j_0 \neq k_0 \end{cases} \right. \end{cases} \quad (4.26)$$

где $\tau_0 = 0$;

$i = 1, \dots, n$ – номера этапов ТП;

$u \in M_B(i-1)$ – номера вершин $(i-1)$ -го этапа ТП;

$v \in M_B(i)$ – номера вершин i -го этапа ТП;

$u_{k_0}, u_{j_0} \in M_B(i-1)$ – фиксированные номера вершин $(i-1)$ -го этапа ТП,

где $k_0, j_0 \in \{1, \dots, M_i\}$;

$pOpt \in R^2$ – параметр приоритета критериев оптимизации ПП КТНС:

$pOpt_k \stackrel{def}{\in} \{1, 2\}, k \in 1, \dots, 2$;

$Fm_1, Fm_2 \in \{\min, \max\}$ – критерии оптимизации ПП КТНС для первой и второй координат векторов состояний на этапах ПП КТНС, которые лексикографически упорядочил параметр $pOpt$, соответственно.

Замечание 3 к методике оптимизации. Пусть $\tau_{uv}, L_i, L_{i-1} \in R^m$. Тогда с учетом введенных ранее управляющих параметров $isFixOp_i, isFixTR_{ij_0} \in \{True, False\}$ функция (4.24), соответствующая функциональному уравнению (4.20), имеет вид:

$$Fl_1(\tau_{uv}, i, k_0, j_0, isFixOp_i, pOpt) =$$

$$= \begin{cases} Fl_G_1(\tau_{u_{k_0}v}, \{\tau_{uv}\}) \Big| pOpt = I_1 = (1, 2, 3, \dots, m); \\ Fl_G_2(\tau_{u_{k_0}v}, \{\tau_{uv}\}) \Big| pOpt = I_2 = (2, 1, 3, \dots, m); \\ \dots \\ Fl_G_{m!}(\tau_{u_{k_0}v}, \{\tau_{uv}\}) \Big| pOpt = I_{m!}, \end{cases} \quad (4.27)$$

где $pOpt \in R^m$, $pOpt_k \stackrel{def}{\in} \{1, 2, 3, \dots, m\}, k \in 1, \dots, m$ – параметр приоритета критериев оптимизации ПП КТНС (2.1) по первой (времени), затем по второй (стоимости первой группы ресурсов ТП) и т. д., по m -й (стоимости $(m-1)$ -й группы ресурсов ТП) координате $\tau_{uv} \in R^m$;

$$Fl_G_1(\tau_{u_{k0}v}, \{\tau_{uv}\}) = \left(\begin{aligned} & \left(\tau_{u_{k0}v}^{(1)} = Fm_1 \left(\{\tau_{uv}^{(1)}\} \right) \right) \wedge \\ & \wedge \left(\tau_{u_{k0}v}^{(2)} = Fm_2 \left\{ \tau_{u_1v}^{(2)} \middle| \tau_{u_1v}^{(1)} = Fm_1 \left(\{\tau_{uv}^{(1)}\} \right) \right\} \right) \wedge \\ & \wedge \dots \wedge \\ & \wedge \left(\tau_{u_{k0}v}^{(n)} = Fm_n \left\{ \tau_{u_{(n-1)}v}^{(n)} \middle| \tau_{u_{(n-1)}v}^{(n-1)} = Fm_{n-1} \left\{ \tau_{u_{(n-2)}v}^{(n-1)} \right\} \right\} \right) \end{aligned} \right); \quad (4.28)$$

$$Fl_G_2(\tau_{u_{k0}v}, \{\tau_{uv}\}) = \left(\begin{aligned} & \left(\tau_{u_{k0}v}^{(1)} = Fm_1 \left\{ \tau_{u_1v}^{(1)} \middle| \tau_{u_1v}^{(2)} = Fm_2 \left(\{\tau_{uv}^{(2)}\} \right) \right\} \right) \wedge \\ & \wedge \left(\tau_{u_{k0}v}^{(2)} = Fm_2 \left(\{\tau_{uv}^{(2)}\} \right) \right) \wedge \\ & \wedge \dots \wedge \\ & \wedge \left(\tau_{u_{k0}v}^{(n)} = Fm_n \left\{ \tau_{u_{(n-1)}v}^{(n)} \middle| \tau_{u_{(n-1)}v}^{(n-1)} = Fm_{n-1} \left\{ \tau_{u_{(n-2)}v}^{(n-1)} \right\} \right\} \right) \end{aligned} \right) \quad (4.29)$$

и т. д.;

$I_1, I_2, \dots, I_{m!}$ – перестановки приоритета критериев параметра $pOpt$;

$i = 1, \dots, n$ – номера этапов ТП;

$u \in M_B(i-1)$ – номера вершин $(i-1)$ -го этапа ТП;

$v \in M_B(i)$ – номера вершин i -го этапа ТП;

$Fm_1, Fm_2, \dots, Fm_m \in \{\min, \max\}$ – критерии оптимизации ПП КТНС для первой, второй, m -й координат векторов состояний на этапах ПП КТНС, которые лексикографически упорядочил параметр $pOpt$.

4.2 Построение кластеров состояний/управлений производственным процессом крашения тканей набивным способом по итогам решения задачи многокритериального управления производственным процессом

Технологический процесс (ТП) производства готового текстильного полотна представлен в виде взвешенного ориентированного ациклического графа (см. подразд. 2.2), вершинами которого являются технологические режимы (ТР). Нагрузки (стоимости ресурсов, временные затраты) на дуги графа в общем случае стохастические и уточнены методом Монте-Карло посредством применения моделей, разработанных с целью их имитации [13, 14], и/или посредством анализа статистических данных ТП [57]. При этом

для каждой технологической операции ТП (2.1) определены нормы расхода времени, затрачиваемого на обработку заданного количества продукции:

$T_i = (t_{i1}, \dots, t_{iM_i})$, $t_{ij} \in R$; общие стоимости затрат ресурсов, которые расходуются на обработку необходимого заданного объема продукции (1000 п. м

тканей): $S_i = (s_{i1}, \dots, s_{iM_i})$, $s_{ij} \in R$, $i = 1, \dots, n$, $j = 1, \dots, M_i$, $n, M_i \in N$.

Стоимости $s_{ij} \in S_i$ расхода ресурсов ТР $R(i, j)$ определены с учетом значений параметров управления потоками ресурсов при имитации ТП [13, 14] ($isRes_r \in \{0, 1\}$ – наличия/отсутствия r -го ресурса ТП), стоимостей затрат r -го ресурса для ТР $R(i, j)$ операции Op_i , а также норм расхода r -го ресурса ТП (2.1) для ТР $R(i, j)$. Таким образом, s_{ij} могут рассматриваться как общие стоимости расхода ресурсов ТР $R(i, j)$ либо как стоимости расхода заданных групп ресурсов, например химических материалов (химикаты, красители), энергоресурсов (пар, газ, электроэнергия).

Для устранения неоднозначности выбора технологических режимов на каждом из этапов ТП поставлена задача многокритериальной оптимизации с использованием принципа оптимальности Беллмана с лексикографическим упорядочением критериев оптимизации (по стоимостным и/или временным затратам ресурсов). Требуется определить состояния ТП, близкие в соответствии с заданной мерой близости к *оптимальному* состоянию.

Для решения поставленной задачи предлагается методика анализа выборок данных возможных состояний ТП. При этом общее количество всех возможных маршрутов по графу ТП составляет [27]:

$$\eta = C_{M_1}^1 \cdot C_{M_2}^1 \cdot \dots \cdot C_{M_n}^1 = M_1 \cdot M_2 \cdot \dots \cdot M_n = \prod_{i=1}^n M_i. \quad (4.30)$$

Пусть расход ресурсов (2.6), (2.7) ТП представлен матрицами вида

$$X = \|x_{ij}\|, \quad x_{ij} \in R, \quad i = 1, \dots, n, \quad j = 1, \dots, \max\{M_1, \dots, M_n\}; \quad (4.31)$$

$$X_T = \psi\left(\{T_i | i = 1, \dots, n\}\right) = \|x_{ij}\|, \quad x_{ij} = \begin{cases} t_{ij} & | j = 1, \dots, M_i, \\ 0 & | j = (M_i + 1), \dots, \max\{M_1, \dots, M_n\}; \end{cases} \quad (4.32)$$

$$X_S^{(2.7)} = \psi\left(\{S_i | i=1, \dots, n\}\right) = \|x_{ij}\|, x_{ij} = \begin{cases} s_{ij} & | j_i = 1, \dots, M_i, \\ 0 & | j_i = (M_i + 1), \dots, \max\{M_1, \dots, M_n\}. \end{cases} \quad (4.33)$$

Этапы решения поставленной оптимизационной задачи

Этап 1. Построение многомерных выборок данных состояний ТП.

Введем в рассмотрение функции $Z_i : X \rightarrow R, i \in \{1, \dots, n\}$ построения возможных состояний ТП (2.1) на соответствующих i -х этапах:

$$Z_1(X, j_1) \stackrel{(4.31)}{=} x_{1 j_1}, j_1 \in \{1, \dots, M_1\}; \quad (4.34)$$

$$Z_2(X, j_1, j_2) \stackrel{(4.31)}{=} x_{1 j_1} + x_{2 j_2}, j_k \in \{1, \dots, M_k\}, k = 1, 2; \quad (4.35)$$

...

$$Z_i(X, j_1, j_2, \dots, j_i) \stackrel{(4.31)}{=} \sum_{k=1}^i x_{k j_k}, j_k \in \{1, \dots, M_k\}, k = 1, \dots, i, i \in \{1, \dots, n\}; \quad (4.36)$$

...

$$Z_n(X, j_1, j_2, \dots, j_n) \stackrel{(4.31)}{=} \sum_{k=1}^n x_{k j_k}, j_k \in \{1, \dots, M_k\}, k = 1, \dots, n. \quad (4.37)$$

Тогда временные ряды возможных состояний ТП (2.1) на соответствующих этапах определяются соотношением вида

$$y_{li} \stackrel{(4.31)}{=} (i, Z_i(X, j_1, j_2, \dots, j_i)) \in R^2, i = 1, \dots, n, j_k = 1, \dots, M_k,$$

$$k = 1, \dots, i, l = 1, \dots, \eta, \eta \stackrel{(4.30)}{=} \prod_{i=1}^n M_i. \quad (4.38)$$

Виды возможных выборок состояний ТП:

1) вид одномерных выборок возможных на n -м этапе состояний ТП

$$Y = \{y_1, y_2, \dots, y_\eta\}, y_l \in R, l = 1, \dots, \eta, \eta \stackrel{(4.30)}{=} \prod_{i=1}^n M_i; \quad (4.39)$$

$$y_l^{(4.33), (4.37)} = Z_n(X_S, j_1, j_2, \dots, j_n), \quad j_k = 1, \dots, M_k, \quad k = 1, \dots, n, \quad l \in 1, \dots, \eta; \quad (4.40)$$

$$y_l^{(4.32), (4.37)} = Z_n(X_T, j_1, j_2, \dots, j_n), \quad j_k = 1, \dots, M_k, \quad k = 1, \dots, n, \quad l \in 1, \dots, \eta; \quad (4.41)$$

2) вид двумерных выборок возможных на n -м этапе состояний ТП

$$Y = \{(y_1^{(1)}, y_1^{(2)}), (y_2^{(1)}, y_2^{(2)}), \dots, (y_\eta^{(1)}, y_\eta^{(2)})\}, \quad (y_l^{(1)}, y_l^{(2)}) \in R^2,$$

$$l = 1, \dots, \eta, \quad \eta^{(4.30)} = \prod_{i=1}^n M_i; \quad (4.42)$$

$$(y_l^{(1)}, y_l^{(2)})^{(4.32), (4.33), (4.37)} = (Z_n(X_T, j_1, j_2, \dots, j_n), Z_n(X_S, j_1, j_2, \dots, j_n)),$$

$$j_k = 1, \dots, M_k, \quad k = 1, \dots, n, \quad l \in 1, \dots, \eta; \quad (4.43)$$

3) вид m -мерных, $m \in \{1, \dots, |Res|\}$, выборок всех возможных на n -м этапе состояний ТП (2.1) определяется с учетом следующего равенства:

$$S = S_{Gr_1} + S_{Gr_2} + \dots + S_{Gr_{(m-1)}}, \quad (4.44)$$

где $S_{Gr_k} \in R$ – стоимость затрат некоторой k -й, $k = 1, \dots, (m-1)$, группы ресурсов (например, химикатов либо энергоресурсов (электроэнергии, пара, газа)), входящей в состав общей стоимости затрат ресурсов $S \in R$ ТП (2.1).

$$Y = \{(y_1^{(1)}, y_1^{(2)}, \dots, y_1^{(m)}), (y_2^{(1)}, y_2^{(2)}, \dots, y_2^{(m)}), \dots, (y_\eta^{(1)}, y_\eta^{(2)}, \dots, y_\eta^{(m)})\},$$

$$(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)}) \in R^m, \quad l = 1, \dots, \eta, \quad \eta^{(4.30)} = \prod_{i=1}^n M_i; \quad (4.45)$$

$$(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)})^{(4.32), (4.33), (4.37)} = (Z_n(X_T, j_1, j_2, \dots, j_n), Z_n(X_{S_{Gr_1}}, j_1, j_2, \dots, j_n),$$

$$Z_n(X_{S_{Gr_2}}, j_1, j_2, \dots, j_n), \dots, Z_n(X_{S_{Gr_{(m-1)}}}, j_1, j_2, \dots, j_n)) \in R^m,$$

$$j_k = 1, \dots, M_k, \quad k = 1, \dots, n, \quad l \in 1, \dots, \eta. \quad (4.46)$$

Этап 2. Кластерный анализ многомерных данных.

Требуется проанализировать множество многомерных данных вида (4.45) ((4.39), (4.42) соответственно); разделить (4.45) на непересекающиеся подмножества (кластеры); определить при этом состояния ТП, близкие наиболее оптимальному $(y_{\alpha}^{(1)}, y_{\alpha}^{(2)}, \dots, y_{\alpha}^{(m)}) \in R^m$ и наименее оптимальному $(y_{\omega}^{(1)}, y_{\omega}^{(2)}, \dots, y_{\omega}^{(m)}) \in R^m$, $\alpha, \omega \in \{1, \dots, \eta\}$ (в соответствии с заданными критериями оптимизации), состояниям ТП. Для решения поставленной задачи применяются методы кластеризации данных. Пусть в результате разбиения (4.45) на кластеры элементам $(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)}) \in R^m$, $l = 1, \dots, \eta$, поставлены в соответствие кластеры, номера которых обозначим через K_{ζ} , $\zeta = 1, \dots, |\zeta|$, где обозначение $|\zeta|$ – итоговое количество кластеров, которые были построены.

Искомыми являются кластеры K_1 , K_2 , в состав которых входят элементы $(y_{\alpha}^{(1)}, y_{\alpha}^{(2)}, \dots, y_{\alpha}^{(m)}) \in K_1 \subset R^m$, $(y_{\omega}^{(1)}, y_{\omega}^{(2)}, \dots, y_{\omega}^{(m)}) \in K_2 \subset R^m$ соответственно. В случае $|\zeta| > 2$ предлагается объединение элементов, не входящих в состав кластеров K_1 , K_2 , в некоторый обобщающий кластер K_3 , как прочие элементы (рисунок 4.2).

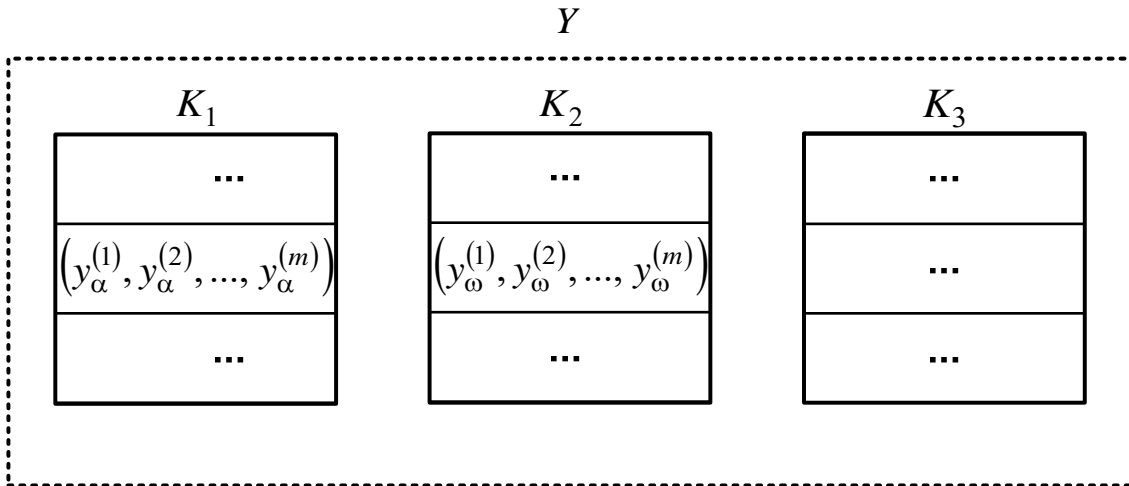


Рисунок 4.2 – Разделение исследуемой выборки Y на кластеры K_1 – K_3

Этап 3. Дополнительный анализ структур построенных кластеров.

Этап 3.1. Выявление факторов, оказывающих наиболее существенное влияние на попадание объектов в соответствующие кластеры. Каждому элементу множества (4.45) соответствует информация о маршруте по графу ТП – последовательность порядковых номеров вершин (рисунок 4.3).

$$(M_B(1, j_1), M_B(2, j_2), \dots, M_B(n, j_n)), j_i \in \{1, \dots, M_i\}, i = 1, \dots, n,$$

$$M_B(i, k) = \sum_{j=1}^{i-1} M_j + k, i = 1, \dots, n, k \in \{1, \dots, M_i\}. \quad (4.47)$$

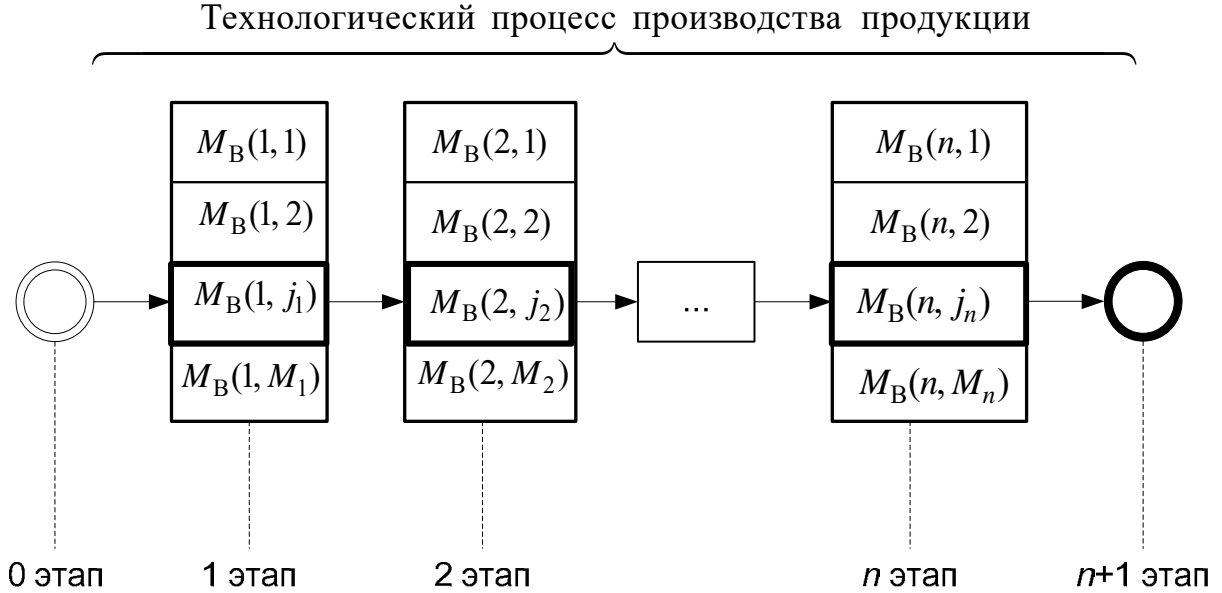


Рисунок 4.3 – Информация о маршруте по графу ТП

Предлагается способ анализа структур кластеров по информации (4.47), включающий следующие шаги.

Шаг 1. Выбор объектов исследования. В качестве объектов исследования предлагается рассмотрение одиночных независимых ТР, или вершин графа ТП с номерами u , $u \in M_B(i)$, $i = 1, \dots, n$. Поскольку граф ТП является связным, допустимо построение любой n -ки исследуемых ТР вида (4.47).

Шаг 2. Определение элементов матрицы вероятностей принадлежности объектов $u \in M_B(i)$, $i = 1, \dots, n$, построенным кластерам через отношение количества случаев попадания объекта в кластер к общему количеству измерений, произведенных над объектом.

Шаг 3. Определение классов объектов. Принадлежность объекта кластеру определяется согласно критерию: объект u , $u \in M_B(i)$, $i = 1, \dots, n$, принадлежит кластеру K_{j_0} , $j_0 \in 1, \dots, \zeta'$, $\zeta' \in \{2, 3\}$, тогда и только тогда, когда вероятность принадлежности объекта кластеру в u -й строке матрицы вероятностей принадлежности объектов кластерам максимальна.

При этом интерес представляют номера вершин графа ТП (номера ТР, тип технологического оборудования соответственно), с наибольшей долей

вероятности обеспечивающие попадание итоговых состояний ТП в кластер состояний ТП, наиболее близких к оптимальному $(y_{\alpha}^{(1)}, y_{\alpha}^{(2)}, \dots, y_{\alpha}^{(m)}) \in K_1$.

Этап 3.2. Определение оценочных характеристик для кластеров.

Шаг 1. Определение оценочных характеристик кластеров. Для выражения в сжатой форме наиболее существенных особенностей (и возможности последующего *качественного* анализа) распределений данных в пределах построенных кластеров K_{ζ} , $\zeta \in 1, \dots, \zeta'$, $\zeta' \in \{2, 3\}$, достаточно указания лишь некоторых их характерных черт, или числовых характеристик, в частности, предлагается определение:

1) математических ожиданий $\Theta_1(K_{\zeta})$ выборок значений по каждой k -й компоненте, $k = 1, \dots, m$, векторов состояний ТП $(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)})$, составляющих кластер K_{ζ} :

$$\Theta_1(K_{\zeta}) = (\Theta m_{\zeta}^{(1)}, \Theta m_{\zeta}^{(2)}, \dots, \Theta m_{\zeta}^{(m)}) \in R^m, \quad \zeta \in 1, \dots, \zeta', \quad \zeta' \in \{2, 3\}; \quad (4.48)$$

2) средних квадратических отклонений $\Theta_2(K_{\zeta})$ выборок значений по каждой k -й компоненте, $k = 1, \dots, m$, векторов состояний ТП $(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)})$, составляющих кластер K_{ζ} :

$$\Theta_2(K_{\zeta}) = (\Theta \sigma_{\zeta}^{(1)}, \Theta \sigma_{\zeta}^{(2)}, \dots, \Theta \sigma_{\zeta}^{(m)}) \in R^m, \quad \zeta \in 1, \dots, \zeta', \quad \zeta' \in \{2, 3\}; \quad (4.49)$$

3) в соответствии с евклидовой метрикой обобщенных характеристик $\Theta_3(K_{\zeta})$ кластеров K_{ζ} , $\zeta \in 1, \dots, \zeta'$, $\zeta' \in \{2, 3\}$, в зависимости от соответствующих оценок математических ожиданий выборок значений по каждой из k -х компонент, $k = 1, \dots, m$, векторов состояний ТП $(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)})$, составляющих кластер K_{ζ} :

$$\Theta_3(K_{\zeta}) = \sqrt{(\Theta m_{\zeta}^{(1)})^2 + (\Theta m_{\zeta}^{(2)})^2 + \dots + (\Theta m_{\zeta}^{(m)})^2} \in R, \quad \zeta \in \{1, \dots, 3\}. \quad (4.50)$$

Шаг 2. Покомпонентное построение и анализ статистических моделей распределения внутри кластеров данных об итоговых состояниях ТП. Для каждого из построенных на этапе 2 кластеров K_{ζ} , $\zeta \in 1, \dots, \zeta'$, $\zeta' \in \{2, 3\}$, требуется проанализировать многомерные множества вида (4.45) ((4.39), (4.42) соответственно), составляющие кластеры K_{ζ} . При этом для выборок

данных по каждой из k -х компонент, $k=1, \dots, m$, векторов состояний $(y_l^{(1)}, y_l^{(2)}, \dots, y_l^{(m)})$ ТП, образующих рассматриваемый кластер, необходимо построить статистическую модель распределения выборочных данных (кривую), наилучшим образом описывающую исследуемую выборку [12].

Поставленную задачу, в частности, можно решить с использованием семейств универсальных статистических моделей распределений (Пирсона, Джонсона и др. [10, 37, 83]), особенность которых состоит в возможности аппроксимации лишь одномодальных и U-образных распределений. При этом для распределений, существенно отличных от нормального, семейства универсальных распределений, по сравнению с множеством «известных» законов распределений, обеспечивают большее разнообразие необходимых форм при меньшем общем количестве функций плотностей распределений, составляющих данные семейства.

Пусть для каждой из k -х компонент $y_l^{(k)}$, $k=1, \dots, m$, векторов состояний ТП построены статистические модели распределения выборочных данных (функции плотности распределения Пирсона [12, 36]):

$$f_k(y_l^{(k)}) = \begin{cases} (|K_1|/\eta) \cdot f_{K_1}^{(k)}(y_l^{(k)}) & | (y_l^{(1)}, \dots, y_l^{(k)}, \dots, y_l^{(m)}) \in K_1; \\ \dots & \\ (|K_\zeta|/\eta) \cdot f_{K_\zeta}^{(k)}(y_l^{(k)}) & | (y_l^{(1)}, \dots, y_l^{(k)}, \dots, y_l^{(m)}) \in K_\zeta, \end{cases} \quad (4.51)$$

$$k=1, \dots, m, \quad \zeta \in 1, \dots, \zeta', \quad \zeta' \in \{2, 3\}, \quad l \in \{1, \dots, \eta\}, \quad \eta \stackrel{(4.27)}{=} \prod_{i=1}^n M_i,$$

где $(|K_\zeta|/\eta)$ – нормирующие множители для функций $f_{K_\zeta}^{(k)}(y_l^{(k)})$ плотностей распределения данных каждого из кластеров K_ζ , при этом $|K_\zeta| \geq 30$.

Тогда вероятность достижения некоторого *прогнозируемого* итогового состояния ТП по соответствующей k -й компоненте $y_l^{(k)}$ векторов состояний ТП, $k=1, \dots, m$ (при $m=2$ $y^{(1)}$ – время; $y^{(2)}$ – общая стоимость ресурсов ТП), – попадания $y_l^{(k)}$ в заданный промежуток $[\alpha, \beta]$ – равняется площади криволинейной трапеции, ограниченной сверху графиком функции $f_k(y_l^{(k)})$ вида (4.51), справа и слева – прямыми $x=\alpha$, $x=\beta$ ($\alpha, \beta \in R$):

$$P(\alpha \leq y_l^{(k)} \leq \beta) \stackrel{(4.48)}{=} \int_{\alpha}^{\beta} f_k(y_l^{(k)}) d(y_l^{(k)}), \quad l \in \{1, \dots, \eta\}, \quad \eta \stackrel{(4.27)}{=} \prod_{i=1}^n M_i. \quad (4.52)$$

4.3 Кластерный анализ многомерных данных на этапах производственного процесса крашения тканей набивным способом

Пусть на этапах производственного процесса крашения тканей набивным способом собраны наблюдения $X = \{x_i / x_i \in R^n, i = 1, \dots, m\}$. При этом перед *Экспертом* – специалистом предметной области, формулирующим запросы к анализу данных, – стоит задача разделения выборки X на непересекающиеся подмножества (кластеры).

Для решения поставленной задачи применяются методы кластеризации. Вследствие анализа методов кластеризации [7, 54] составлена их классификация (см. подразд. 2.4). Для исследования отобраны методы кластеризации K-means, Tree Clustering, FRC (Fuzzy Relation Clustering), которые являются представителями основных методов к разделению множества, которое рассматривается, на кластеры [12].

Для разбиения множества X на кластеры предполагается использование нескольких методов кластеризации для проверки и уточнения результатов. Пусть для проверки и уточнения результатов кластеризации данных в качестве дополнительных начальных условий *Эксперт* определяет методы кластеризации данных, предполагаемое количество кластеров, а также контрольные точки (КТ) для кластеров – элементы исходного исследуемого множества, по заключению *Эксперта* с вероятностью 100 % принадлежащие соответствующим кластерам.

Для принятия решения о разделении множества наблюдений X на кластеры *Эксперт* формулирует запрос к точности принятия решений. Задаёт пороговые нижние границы для необходимого и достаточного количества методов кластеризации, в соответствии с которыми принимается экспертное решение (оценка): голосование большинством голосов, принятие решений максимально возможным количеством голосов, принятие решений по крайней мере заданным количеством голосов и т. д. При этом вначале разделение производится двумя методами. Если результаты разбиений не совпадают, то применяют третий и так далее методы. При этом необходим поиск алгоритма обобщения результатов кластеризации исследуемого множества X несколькими методами по запросам *Эксперта*.

Кластерный анализ проводится на многомерных данных:

$$X = \{x_i / x_i \in R^n, i = 1, \dots, m\}. \quad (4.53)$$

Пусть элементы $x_i \in X$, $i = 1, \dots, m$, представляют собой измерения n параметров объектов b_r из множества

$$B = \{b_r / r = 1, \dots, |B|\}, |B| < m. \quad (4.54)$$

Тогда каждому $x_i \in X$ соответствует некоторый объект b_r и идентифицирующая информация об этом объекте [19].

Пусть для разделения множества (4.53) на кластеры используются методы кластеризации M_l , $l = 1, \dots, L$, $L \geq 3$ (например, K-means, Tree Clustering, FRC и т. д.). В результате разбиения множества X на кластеры каждый из L применяемых методов кластеризации M_l , $l = 1, \dots, L$, ставит в соответствие номерам элементов $x_i \in X$ номера $j = 1, \dots, k_l$ (кластеров K_j), где k_l – количество кластеров, построенных методом M_l .

Обобщение результатов кластеризации данных несколькими методами. Введем необходимые определения и утверждения, позволяющие интерпретировать результаты кластеризации множества X методами M_l , $l = 1, \dots, L$, произвести их обобщение и последующее выделение кластеров объектов $b_r \in B$.

Утверждение 4. Результат кластеризации множества X вида (4.53) для каждого из методов кластеризации M_l , $l = 1, \dots, L$, может быть представлен в виде матрицы вероятностей принадлежности объектов $b_r \in B$ определенным кластерам

$$P_l = \|p_{lrj}\|, p_{lrj} \in [0, 1], r = 1, \dots, |B|, j \in 1, \dots, k_l, \quad (4.55)$$

где r – номер объекта из множества B ;

j – номер кластера K_j ;

k_l – количество кластеров, построенных методом M_l ;

p_{lrj} – вероятность принадлежности r -го объекта $b_r \in B$ кластеру K_j в соответствии с методом M_l ;

l – номер примененного метода кластеризации M_l , $l = 1, \dots, L$.

Вероятности p_{lrj} в (4.55) рассчитываются на основе классического определения вероятности как отношение количества случаев попадания объекта b_r в кластер K_j к общему количеству измерений, выполненных над объектом b_r .

Замечание к утверждению 4. Сумма строчных элементов матриц P_l вида (4.55) постоянна и равна единице.

Вероятности отнесения объектов к определенным кластерам одновременно L выбранными методами малы и вместе с этим вероятности принадлежности объектов определенным кластерам по крайней мере одним из заданных методов очень велики [19]. Поэтому рассматривается случай отнесения объекта к данному кластеру L методами из L выбранных методов кластеризации, где $L_E < L^* \leq L$, L_E – точная нижняя граница, определяемая *Экспертом* в зависимости от уровня значимости задачи. Например, в случае аналогии голосования большинством голосов $L_E = L/2$.

Определение 1. Объект $b_r \in B$, $r = 1, \dots, |B|$, является элементом кластера K_j , $j = 1, \dots, k$, тогда и только тогда, когда он отнесен к данному кластеру по крайней мере L методами из L выбранных методов кластеризации, причем $L_E < L^* \leq L$, $L \geq 3$ [12].

Утверждение 5. Пусть P_1, P_2, \dots, P_L – матрицы вида (4.55) вероятностей принадлежности объектов b_r , $r = 1, \dots, |B|$, определенным кластерам K_j , $j = 1, \dots, k$, согласно методам кластеризации M_1, M_2, \dots, M_L соответственно. Тогда значения элементов $p_{rj} \in [0, 1]$ обобщенной (в смысле определения 1) матрицы P могут быть найдены суммированием коэффициентов $P_{v,L}$, $v = L^*, \dots, L$, производящей функции [16, 18, 24]

$$\varphi(z) = \prod_{l=1}^L (q_{lrj} + p_{lrj} \cdot z) = \sum_{v=0}^L P_{v,L} \cdot z^v, \quad (4.56)$$

где p_{lrj} – элементы матрицы P_l , $l = 1, \dots, L$, $p_{lrj} \in [0, 1]$, $q_{lrj} = 1 - p_{lrj}$ [16];

$r = 1, \dots, |B|$ – номера объектов b_r ;

$j = 1, \dots, k$ – номера кластеров K_j ,

соответственно,

$$p_{rj} = \sum_{v=L^*}^L P_{v,L}. \quad (4.57)$$

Доказательство утверждения 5. Пусть производится L независимых [25] испытаний. Под испытанием будем понимать проведение кластеризации одним из L методов M_1, M_2, \dots, M_L .

Кластерный анализ разными методами проводится в одинаковых условиях, используются одни и те же исходные данные. Результаты обработки данных одним из методов не зависят от результатов анализа другими методами кластеризации [18, 19]. Испытания независимы.

Введем в рассмотрение событие A – попадание объекта в результате испытания в определенный кластер. В каждом испытании событие A появляется со своими вероятностями.

Обозначим: p_1, p_2, \dots, p_L – вероятности попадания объекта в заданный кластер в первом, втором, ..., L -м испытании соответственно; $q_1 = 1 - p_1$, $q_2 = 1 - p_2$, ..., $q_L = 1 - p_L$ – вероятности непопадания объекта в определенный кластер в первом, втором, ..., L -м испытании соответственно.

Вероятность того, что в этих L испытаниях событие A произойдет 0, 1, 2, ..., L раз [19], – коэффициент при соответствующей степени z производящей функции

$$\begin{aligned} \varphi(z) &= \prod_{l=1}^L (q_l + p_l \cdot z) = (q_1 + p_1 \cdot z)(q_2 + p_2 \cdot z) \cdot \dots \cdot (q_L + p_L \cdot z) = \\ &= q_1 q_2 \cdot \dots \cdot q_L + \dots + p_1 p_2 \cdot \dots \cdot p_L \cdot z^L = \sum_{v=0}^L P_{v,L} \cdot z^v, P_{v,L} \in [0, 1]. \end{aligned} \quad (4.58)$$

Таким образом, вероятность [19] в этих L испытаниях событию A не произойти ни разу $P_{0,L} = q_1 q_2 \cdot \dots \cdot q_L$; ...; ровно L раз из L – $P_{L,L} = p_1 p_2 \cdot \dots \cdot p_L$.

Тогда вероятность события Y – попадания объекта в определенный кластер по крайней мере в L испытаниях из L [19] –

$$p(Y) = \sum_{v=L}^L P_{v,L}. \quad (4.59)$$

В результате кластеризации множества данных X вида (4.53) L методами кластеризации получены матрицы вероятностей P_1, P_2, \dots, P_L вида (4.55) принадлежности объектов определенным кластерам [18]

$$P_l = \|p_{lrj}\|, p_{lrj} \in [0, 1], l = 1, \dots, L, r = 1, \dots, /B/, j = 1, \dots, k. \quad (4.60)$$

Для оценки обобщенного результата кластеризации L методами, применяя к элементам [19] p_{lrj} , $l = 1, \dots, L$, матриц P_1, P_2, \dots, P_L формулу (4.59), строим обобщенную матрицу

$$P = \|p_{rj}\|, p_{rj} \in [0, 1], r = 1, \dots, /B/, j = 1, \dots, k. \quad (4.61)$$

В общем виде элементы матрицы P могут быть вычислены по формуле

$$p_{rj} = \sum_{v=L^*}^L P_{v,L}, r=1, \dots, |B|, j=1, \dots, k, \quad (4.62)$$

где $P_{v,L}$ – коэффициенты производящей функции

$$\varphi(z) = \prod_{l=1}^L ((1 - p_{lrj}) + p_{lrj} \cdot z) = \sum_{v=0}^L P_{v,L} \cdot z^v, r=1, \dots, |B|, j=1, \dots, k. \quad (4.63)$$

Доказательство закончено.

Следствие из утверждения 5. Пусть P_1, P_2, P_3 – матрицы вида (4.55) вероятностей объектов $b_r, r=1, \dots, |B|$, определенным кластерам $K_j, j=1, \dots, k$, согласно методам кластеризации M_1, M_2, M_3 соответственно. Тогда значения элементов $p_{rj} \in [0,1]$ обобщенной (в смысле определения 1) матрицы P определяются по формуле [16]

$$\varphi(z) = \prod_{l=1}^3 (q_{lrj} + p_{lrj} \cdot z) = \sum_{v=0}^3 P_{v,3} \cdot z^v; \quad (4.64)$$

$$p_{rj} = (1 - p_{1rj})p_{2rj}p_{3rj} + p_{1rj}(1 - p_{2rj})p_{3rj} + p_{1rj}p_{2rj}(1 - p_{3rj}) + p_{1rj}p_{2rj}p_{3rj}, \quad (4.65)$$

где $q_{lij} = 1 - p_{lij}, p_{lij} \in [0,1], l=1, \dots, 3, r=1, \dots, |B|, j=1, \dots, k$ – элементы матриц P_1, P_2, P_3 соответственно.

Вывод обобщающей формулы для $L=3$ приведен в [20].

Замечание 1 к обобщению результатов кластеризации. Для каждого из методов кластеризации свойственна собственная система нумерации кластеров K_j [7, 54]. Номера $j=1, \dots, k_l$ присваиваются кластерам K_j в соответствии с порядком их построения, который зависит от особенностей алгоритмов методов кластеризации $M_l, l=1, \dots, L$, основанных на использовании матриц сходства, эвристических алгоритмов перебора, идей математического программирования, на оценивании функций плотности статистического распределения и др.

Замечание 2 к обобщению результатов кластеризации. Для возможности обобщения результатов кластеризации размерность обобщаемых матриц P_1, P_2, \dots, P_L должна быть одинаковой. В случае разбиения множества X на k_1 кластеров методом M_1, k_2 кластеров методом M_2, \dots, k_L клас-

теров методом M_L необходимо предварительно привести матрицы P_1, P_2, \dots, P_L к одной размерности $k = \max \{k_1, k_2, \dots, k_L\}$. Преобразование матрицы к необходимой размерности возможно посредством дополнения ее столбцами с нулевыми вероятностями попадания объекта в добавленные кластеры [16].

Замечание 3 к обобщению результатов кластеризации. Для обеспечения возможности обобщения результатов кластеризации необходимо упорядочивание столбцов обобщаемых матриц P_l вида (4.55) размерности $|B| \times k$ таким образом, что для любого $j_0 \in 1, \dots, k$ кластеры $K_{j_0}^{(l)}$, построенные методами M_l , $l = 1, \dots, L$, будут соответствовать *эталону*, определенному *Экспертом*, т. е. иметь максимально возможное количество общих элементов $x_i \in X$ с заданным *эталон*ом.

Введем понятие: **контрольные точки (КТ) кластеров**. Пусть *Экспертом* заданы следующие дополнительные начальные условия кластеризации $X \subset R^n$:

- 1) ψ – предполагаемое количество кластеров K_j , $j = 1, \dots, \psi$;
- 2) $Y \subset X$, $|Y| = \gamma \geq \psi$ – множество контрольных точек кластеров:

$$Y = \{y_j / y_j \in R^n, j = 1, \dots, \gamma\}. \quad (4.66)$$

При этом для контрольных точек кластеров $y_i \in Y$ выполняется условие

$$(\forall y_i \in Y)(\exists ! K_j)[y_i \in K_j], i \in 1, \dots, \gamma, j \in 1, \dots, \psi. \quad (4.67)$$

Предлагается следующая **методика определения кластеров объектов**.

Этап 1. Кластеризация множества X методами M_l , $l = 1, \dots, L$; интерпретация результатов кластеризации; подготовка к этапу обобщения.

Этап 2. Обобщение результатов кластеризации. Определение кластеров объектов.

Первый этап методики определения кластеров объектов

Определение 2. Матрицы $P_\xi = \|p_{\xi rj}\|$, $P_\eta = \|p_{\eta rj}\|$ вида (4.55) размерности $|B| \times k$ эквивалентны тогда и только тогда, когда выполняется условие

$$(P_\xi = P_\eta) \Leftrightarrow (\forall r \in 1, \dots, |B|)(\forall j \in 1, \dots, k)[p_{\xi rj} = p_{\eta rj}]. \quad (4.68)$$

Определение 3. Приведение матриц $P_l = \|p_{lrj}\|$, $l = 1, \dots, w$, $1 < w \leq L$ размерностей $|B| \times k_l$ (множество X разделено методом M_l на k_l кластеров) к одной размерности $|B| \times k_l$, где $k = \max\{k_1, \dots, k_w\}$, равносильно построению матриц $P'_l = \|p'_{lrj}\|$ за счет дополнения матриц P_l столбцами с нулевыми вероятностями попадания объектов в добавленные кластеры:

$$p'_{lrj} = \begin{cases} p_{lrj} & | \quad j \leq k_l, \\ 0 & | \quad k_l < j \leq k. \end{cases} \quad (4.69)$$

Введем необходимые определения и утверждения, позволяющие провести упорядочивание столбцов матрицы P_T (тестируемая матрица) по отношению к матрице P_E (эталонная, согласно оценке *Эксперта*, матрица).

Определение 4. Обозначим через P_E – (эталонную) матрицу вероятностей принадлежности контрольных точек (КТ) – начальное условие кластеризации, определенное *Экспертом*, – кластерам; P_T – (тестируемую) матрицу вероятностей принадлежности элементов множества X , соответствующих КТ, построенную по результатам кластеризации X методом M_l :

$$P_E = \|p_{Eij}\|, \quad p_{Eij} = \begin{cases} 1, & y_i \in K_j, \\ 0, & y_i \notin K_j; \end{cases} \quad (4.70)$$

$$P_T = \|p_{Tij}\|, \quad p_{Tij} = \begin{cases} 1, & (\exists x_k \in X) \left[(x_k = y_i) \wedge (x_k \in K_j) \right], \\ 0, & (\exists x_k \in X) \left[(x_k = y_i) \wedge (x_k \notin K_j) \right]. \end{cases} \quad (4.71)$$

Утверждение 6. Пусть в результате кластеризации X методом M_l построено K_j , $j = 1, \dots, \psi'$, кластеров. Пусть выполняется условие $\psi' \geq \psi$. Возможны следующие случаи.

Случай 1. $\gamma = \psi = \psi'$: предполагаемое количество кластеров совпадает с количеством построенных, для каждого из кластеров задана единственная КТ; выполняется условие

$$\left(|Y| = |K_j| \right) \wedge (\forall y_i \in Y) (\exists ! K_j) [y_i \in K_j], \quad i \in 1, \dots, \gamma, \quad j \in 1, \dots, \psi; \quad \gamma = \psi = \psi'. \quad (4.72)$$

Тогда матрица P_E вероятностей принадлежности КТ кластерам может быть приведена к единичной матрице: $P_E = E$ (квадратная симметричная матрица).

Случай 2. $\gamma > \psi = \psi'$: предполагаемое количество кластеров совпадает с количеством построенных, для каждого из кластеров задано не менее одной КТ; выполняется условие

$$(|Y| > |K_j|) \wedge (\forall y_i \in Y)(\exists K_j)[y_i \in K_j], i \in 1, \dots, \gamma, j \in 1, \dots, \psi, \gamma > \psi = \psi'. \quad (4.73)$$

Матрица P_E отличается от единичной: $P_E = E$ (прямоугольная матрица).

Случай 3. $\gamma = \psi < \psi'$: количество построенных кластеров превышает предполагаемое количество кластеров, таким образом, не для всех построенных кластеров заданы КТ; выполняется условие

$$(|Y| < |K_j|) \wedge (\forall y_i \in Y)(\exists K_j)[y_i \in K_j], i \in 1, \dots, \gamma, j \in 1, \dots, \psi, \gamma = \psi < \psi'. \quad (4.74)$$

Возможны следующие случаи.

Случай 3.1. Количество построенных кластеров на один превышает количество контрольных точек (КТ):

$$|Y| + 1 = |K_j|, |K_j| - |Y| = 1. \quad (4.75)$$

Случай 3.2. Количество построенных кластеров превышает количество КТ более чем на один (более одного кластера не обозначено контрольной точкой):

$$|Y| + 1 < |K_j|, |K_j| - |Y| = \delta > 1. \quad (4.76)$$

Замечание к утверждению 6. В силу прикладной специфики поставленной задачи предлагается приведение случая 3.2 к случаю 3.1 за счет объединения δ кластеров, не отмеченных контрольными точками, в один общий кластер, как прочие элементы.

Определение 5. Перестановкой из элементов конечного множества I называется всякое упорядочивание элементов этого множества.

Определение 6. Обозначим через $P_T(I_t)$ матрицы, образованные из матрицы P_T перестановками I_t ее столбцов $j = 1, \dots, k$:

$$P_T(I_0) = P_T, P_T(I_1), \dots, P_T(I_t), t = 0, \dots, k!-1, \quad (4.77)$$

где t – номер выполненной перестановки I_t столбцов матрицы P_T с целью задания порядка их следования:

$$I_0 = (1, 2, \dots, k), I_1 = (2, 1, \dots, k), \dots, I_t, t = 0, \dots, k!-1. \quad (4.78)$$

Замечание к определению 6. Общее количество возможных перестановок k столбцов матрицы P_T составляет $k!$. Могут рассматриваться перестановки только тех столбцов матрицы P_T , которые не совпадают со столбцами P_E .

Определение 7. Для матричных пар $(P_T(I_t), P_E)$, $t = 0, \dots, (k!-1)$ определяется метрика $\rho: (P_T(I_t), P_E) \rightarrow R_+ \cup \{0\}$ вида

$$\rho(P_T(I_t), P_E) = \sum_{r=1}^{|B|} \sum_{j=1}^k \min(p_{Trj}(I_t), p_{Ej}), t = 0, \dots, k!-1, \quad (4.79)$$

где $p_{Trj}(I_t)$ – элементы матрицы $P_T(I_t)$ со столбцами j , следующими в порядке, соответствующем произведенной перестановке I_t столбцов матрицы P_T ;

p_{Ej} – элементы матрицы P_E .

Замечание к определению 7. Метрика $\rho(P_T(I_t), P_E) \geq 0$ позволяет определить меру совпадения элементов соответствующих матриц. Значение ρ прямо пропорционально количеству общих элементов $x_i \in X$ в кластерах матричной пары.

Утверждение 7. Перестановка I^* столбцов матрицы P_T по отношению к матрице P_E оптимальна тогда и только тогда, когда выполняется условие

$$\rho(P_T(I^*), P_E) = \max\{\rho(P_T(I_t), P_E) \mid t = 0, \dots, k!-1\}, \quad (4.80)$$

где ρ – метрика вида (4.79).

Замечание 1 к утверждению 7. Возвращаясь к прежним обозначениям, результат упорядочивания столбцов P_T относительно эталонной матрицы P_E [16]

$$P_T = P_T(I^*). \quad (4.81)$$

Замечание 2 к утверждению 7. Возможны варианты равновероятных случаев упорядочивания столбцов P_T относительно P_E . Если перестановка I^* в соответствии с (4.80) не единственна, имеет место неоднозначность относительно упорядочивания столбцов P_T . В этом случае метод кластеризации не применим к анализируемым данным, т. к. неадекватно описывает реальные данные в соответствии с оценкой (начальным условием) *Эксперта*. Результаты работы метода далее не рассматриваются.

Утверждение 8. Объект $x_k \in X$, соответствующий контрольной точке $y_i \in Y$, $x_k = y_i$, принадлежит кластеру K_{j_0} , $j_0 \in 1, \dots, \psi'$, тогда и только тогда, когда вероятность принадлежности $x_k = y_i$ кластеру в i -й строке упорядоченной матрицы P_T ненулевая и максимальна.

Утверждение 8 следует из определения 1.

Замечание к утверждению 8. В соответствии с принадлежностью элементов $x_k \in X$, соответствующих КТ (начальное условие), кластерам K_{j_0} , $j_0 \in 1, \dots, \psi'$, настраивается нумерация кластеров, на которые разделяется исследуемое множество X методом M_l .

Вышеописанная методика упорядочивания столбцов тестируемой матрицы P_T по отношению к столбцам эталонной матрицы P_E применима в случае $\dim(P_T) = \dim(P_E)$. В случае $\dim(P_T) \neq \dim(P_E)$ предлагается следующая адаптированная **методика упорядочивания**.

Шаг 1. На основании P_T строится P'_T – матрица вероятностей принадлежности элементов кластерам, помеченным контрольными точками, – таким образом, что $\dim(P'_T) = \dim(P_E) < \dim(P_T)$.

Шаг 2. Применяется методика упорядочивания P'_T относительно P_E .

Шаг 3. С учетом замечания к утверждению 6 производится объединение кластеров, не отмеченных контрольными точками, в один общий кластер, как прочие элементы. Размерность упорядоченной матрицы P'_T дополняется до исходной размерности $\dim(P_T)$. Для этого в матрицу P'_T добавляется строка и столбец с вероятностями попадания КТ (с учетом дополнительной КТ) в соответствующие кластеры. Матрица P_T упорядочена.

Утверждение 9. Методика кластеризации множества многомерных данных X методами M_l , $l = 1, \dots, L$, интерпретации результатов кластеризации и их подготовки к последующему этапу обобщения выполняема за L итераций, $L_E < L^* \leq L$, состоящих из шагов 1–5.

Шаг 1. Вводится в рассмотрение параметр номера итерации w с начальным значением $w = 0$.

Шаг 2. Проводится кластеризация элементов X методом M_{w+1} .

Шаг 3.1. По результатам кластеризации (см. шаг 2) в соответствии с определением 4 по КТ кластеров строятся тестируемая (P_T) и эталонная (P_E) матрицы. Согласно утверждению 4 определяется оптимальная перестановка столбцов P_T относительно P_E . Согласно утверждению 5 производится настройка нумерации построенных (см. шаг 1) кластеров.

Шаг 3.2. По результатам кластеризации (см. шаг 2) проводится построение матрицы P_{w+1} вида (4.55) размерности $|B| \times k_{w+1}$ вероятностей принадлежности объектов $b_r \in B$, $r = 1, \dots, |B|$, кластерам K_j , $j = 1, \dots, k_{w+1}$.

Шаг 4. При $w > 0$ проводится приведение матриц $P_l^{(w-1)}$, $l = 1, \dots, w$, размерности $|B| \times k^{(w-1)}$ и матрицы P_{w+1} одной размерности $|B| \times k^{(w)}$, где

$$k^{(w)} = \max \{ k^{(w-1)}, k_{w+1} \}. \quad (4.82)$$

Построение матриц $P_{w+1}^{(w)} = \|p_{w+1rj}^{(w)}\|$, $P_l^{(w)} = \|p_{lrj}^{(w)}\|$, $l = 1, \dots, w$, размерности $|B| \times k^{(w)}$.

Шаг 5. Определяется количество $S^{(w)}$ эквивалентных матриц $P_l^{(w)}$, $l = 1, \dots, w+1$. Если выполняется условие

$$(w < L-1) \wedge (S^{(w)} \leq L_E), \quad (4.83)$$

то $w := w+1$, шаги 2–5 повторяются.

Доказательство утверждения 9. Проводится с применением метода математической индукции. Пусть в результате выполнения шагов итерационной методики на последней итерации $w^* \geq 1$ ($(L > 3, L^* \geq 2) \Rightarrow w^* \geq 1$) получены приведенные к одной размерности $|B| \times k^{(w^*)}$ упорядоченные матрицы

$$P_1^{(w^*)}, \dots, P_{w^*+1}^{(w^*)}, w^* = \min \{ S^{(w^*)}, L-1 \}. \quad (4.84)$$

Поскольку итерация с номером $w^* \geq 1$ – последняя итерация шагов 1–5 методики, то условие продолжения шагов методики (4.83) не выполняется, т. е. выполняется условие

$$(w^* \geq L-1) \vee (S^{(w^*)} > L_E). \quad (4.85)$$

По построению w^* имеет место ограничение для значений w^*

$$w^* \leq L-1. \quad (4.86)$$

Из (4.85) и (4.86) следует

$$(w^* \geq L-1) \wedge (w^* \leq L-1) \Leftrightarrow (\max(w^*) = L-1). \quad (4.87)$$

По построению имеет место ограничение для значений $S^{(w^*)}$

$$S^{(w^*)} \leq w^* + 1. \quad (4.88)$$

Из формул (4.85), (4.87) и (4.88) следует

$$\stackrel{(4.85)}{L_E} < \stackrel{(4.88)}{S^{(w^*)}} \leq \stackrel{(4.87)}{w^* + 1} \leq L-1+1 = L, \quad (4.89)$$

или

$$L_E < w^* + 1 \leq L. \quad (4.90)$$

Поскольку начальное значение параметра w равно нулю, из (4.90) получаем общее количество выполненных итераций шагов методики

$$w = w^* + 1 = L^*, \quad L^* \in (L_E, L]. \quad (4.91)$$

Возвращаясь к прежним обозначениям, получим матрицы $P_l = P_l^{(w^*)}$, $l = 1, \dots, L$, где $L := L^* = w^* + 1$, размерности $|B| \times k$, где $k = k^{(w^*)}$.

Доказательство закончено.

Замечание 1 к утверждению 9. Построение на шаге 3.2 методики матриц P_{w+1} вида (4.52) проводится в соответствии с утверждением 4 по формуле (4.57).

Замечание 2 к утверждению 9. Приведение матриц P_l , $l = 1, \dots, w$, размерности $|B| \times k^{(w-1)}$ и P_{w+1} к одной размерности $k^{(w)}$ (4.82) на шаге 4

методики, равносильное построению матриц $P_{w+1}^{(w)} = \|p_{w+1rj}^{(w)}\|$, $P_l^{(w)} = \|p_{lrj}^{(w)}\|$ размерности $|B| \times k^{(w)}$, проводится по формуле (4.69):

$$p_{w+1rj}^{(w)} = \begin{cases} p_{w+1rj}^{(w)} & | j \leq k_{w+1}, \\ 0 & | k_{w+1} < j \leq k^{(w)}; \end{cases} \quad p_{lrj}^{(w)} = \begin{cases} p_{lrj}^{(w-1)} & | j \leq k^{(w-1)}, \\ 0 & | k^{(w-1)} < j \leq k^{(w)}, \end{cases} \quad l = 1, \dots, w. \quad (4.92)$$

Замечание 3 к утверждению 9. Определение на шаге 5 методики количества $S^{(w)}$ эквивалентных матриц $P_l^{(w)}$, $l = 1, \dots, w+1$, возможно посредством применения следующей формулы:

$$S^{(w)} = \max \left\{ \left(\sum_{j=1}^{w+1} d_{ij} - 1 \right) \mid d_{ij} = 1 - \Delta_{ij}, i = 1, \dots, w+1 \right\}, \quad (4.93)$$

где

$$\Delta_{ij} = \begin{cases} 0 & | P_i^{(w)} = P_j^{(w)}, \\ 1 & | P_i^{(w)} \neq P_j^{(w)}. \end{cases} \quad (4.94)$$

При расчете $S^{(w)}$ элементы главной диагонали d_{ii} , $i = 1, \dots, w+1$, не учитываются.

Рассмотрим *второй этап методики определения кластеров объектов*, представленный следующими шагами.

Шаг 1. Обобщение результатов кластеризации P_1, \dots, P_L в соответствии с утверждением 5.

Шаг 2. Определение кластеров объектов.

В результате кластеризации множества X вида (4.53) L методами получена обобщенная матрица $P = \|p_{rj}\|$, $r = 1, \dots, |B|$, $j = 1, \dots, k$ [18], на основе анализа значений элементов которой определяется принадлежность объекта к определенному кластеру.

Утверждение 10. Объект $b_r \in B$, $r \in 1, \dots, |B|$, принадлежит кластеру K_{j_0} , $j_0 \in 1, \dots, k$, тогда и только тогда, когда вероятность принадлежности объекта кластеру в r -й строке обобщенной матрицы P максимальна:

$$(b_r \in K_{j_0}) \Leftrightarrow (p_{rj_0} = \max \{p_{rj} \mid j = 1, \dots, k\}). \quad (4.95)$$

Утверждение 10 следует из определения 1.

4.4 Статистические модели распределения данных многомодальных выборок при моделировании нагрузок на дуги графа производственного процесса крашения тканей набивным способом

При моделировании нагрузок на дуги графа производственного процесса крашения тканей вида (2.1) получена выборка

$$X = \{x_i / x_i \in R, i = 1, \dots, m\}. \quad (4.96)$$

Необходимо построить статистическую модель распределения данных (кривую) для выборки X , наилучшим образом описывающую выборку на исследуемом отрезке $[a, b]$, $a = \min\{x_i\}$, $b = \max\{x_i\}$, $x_i \in X$.^(4.96)

Для решения поставленной задачи применимы законы распределения экспериментальных данных [93], ряды специального вида, семейства универсальных моделей распределений (Пирсона и др.) [10, 23, 93].

Каждый из подходов имеет свои достоинства и недостатки. Так, для распределений, существенно отличных от нормального [56] семейства универсальных распределений, по сравнению с множеством «известных» законов распределений, обеспечивают большее разнообразие необходимых форм при меньшем общем количестве функций плотностей распределений, составляющих данные семейства. К недостаткам подхода с применением специальных рядов относят большое количество коэффициентов, отсутствие физического смысла коэффициентов [9].

Посредством построения универсальных моделей распределений данных возможно аппроксимировать одномодальные и U-образные распределения [41]. Из многомодальности распределения следует неоднородность рассматриваемых данных (4.96) и предлагается с использованием методов кластеризации [19] разделить (4.96) на однородные выборки, или кластеры, а далее – построение для каждого из кластеров соответствующей ему функции плотности [21, 57].

Для построения функций семейства Пирсона – статистических моделей распределения данных X (4.96) разработан программный модуль BelSim2#.random [57]. В программно-технологическом комплексе имитации сложных систем BelSim2 [69] реализована методика разделения X (4.96) на кластеры.

Этапы построения функции плотности семейства Пирсона

Этап 1. Производится расчет точечных оценок выборки, необходимых для построения функций семейства Пирсона в BelSim2#.random. Так,

определяются моменты: начальный первого порядка ν_1 ; центральные μ_0, \dots, μ_4 ; рассчитываются коэффициенты β_1 (асимметрии) и β_2 (эксцесса) по Пирсону:

$$\beta_1 = \mu_3^2 / \mu_2^3, \quad \beta_2 = \mu_4 / \mu_2^2. \quad (4.97)$$

Построение функции $f(x)$, согласно классификации Пирсона, производится в следующем порядке.

Этап 2. Строится кривая согласно классификации Пирсона.

Шаг 2.1. Классифицируется тип кривой $f(x)$, зависящий от значений коэффициентов β_1, β_2 (4.97) и, соответственно, показателя Пирсона \aleph классификации $f(x)$ [37]:

$$\aleph = (\beta_1(\beta_2 + 3)^2) / (4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)). \quad (4.98)$$

Шаг 2.2. Классическим методом моментов, согласно методике, по Пирсону [21], оцениваются значения параметров функции $f(x)$, тип которой определен на шаге 2.1.

Шаг 2.3. Комплексом критериев согласия [10] (Колмогорова–Смирнова (λ), Пирсона (χ^2), Мизеса (ω^2)) выполняется проверка гипотезы о распределении выборки X по закону распределения с плотностью $f(x)$ [32]. Первыми применяются χ^2 и λ . При эквивалентности логических значений итогов работы критериев проверка гипотезы окончена. Иначе – применяется третий критерий: ω^2 и результат его работы принимается как итоговый при проверке гипотезы. Сохраняемые результаты работы критериев $\chi^2, \lambda, \omega^2$: наблюдаемые значения (Δ_k), критические значения ($\Delta_k^{\text{кр}}$), отношение наблюдаемых значений критериев к критическим значениям (dL_k), итоги работы критериев логические (bL_k):

$$\Delta_k, \quad k = \chi^2, \lambda, \omega^2; \quad (4.99)$$

$$\Delta_k^{\text{кр}}, \quad k = \chi^2, \lambda, \omega^2; \quad (4.100)$$

$$dL_k = \Delta_k / \Delta_k^{\text{кр}}, \quad k = \chi^2, \lambda, \omega^2; \quad (4.101)$$

$$bL_k = \begin{cases} True & | dL_k < 1; \\ False & | dL_k \geq 1, \end{cases} \quad k = \chi^2, \lambda, \omega^2. \quad (4.102)$$

Этап 3. В случае, если на этапе 2 комплекс статистических критериев χ^2 , λ , ω^2 отклонил гипотезу, проверяемую на шаге 2.3 либо по запросу пользователя, производится построение множества функций семейства Пирсона (все семь основных типов, за исключением типа функции $f(x)$, построенной на этапе 2) и последующая их проверка комплексом статистических критериев аналогично шагам 2.2 и 2.3.

Этап 4. По запросу пользователя производится построение и проверка комплексом статистических критериев χ^2 , λ , ω^2 равномерного, нормального, экспоненциального распределений.

Этап 5. На множестве кривых $\{f_i(x)\}$, построенных и не отклоненных комплексом статистических критериев на этапе 3, производится выбор статистической модели $f^*(x)$, которая наилучшим образом описывает выборку X (4.96).

Шаг 5.1. Если есть функции плотности $f_i(x)$, $i = 1, \dots, |\{f_i(x)\}|$, для которых проверено меньшее количество критериев $nk \in \{2, 3\}$, чем для остальных функций, для определения $f^*(x)$ для них дополнительно производится проверка критерием ω^2 .

Шаг 5.2. Функция $f^*(x) = f_{i_0}(x) \in \{f_i(x)\}$ наилучшим образом описывает X (4.96) на отрезке $[a, b]$, если справедливо следующее:

$$\rho(X, f_{i_0}(x)) = \min\{\rho(X, f_i(x)) \mid i = 1, \dots, |\{f_i(x)\}|\}, \quad (4.103)$$

где $\rho(X, f_i(x))$ – евклидова метрика отношений dL_k :

$$\rho(X, f_i(x)) = \sqrt{\sum_{k=1}^{nk} (dL_k(X, f_i(x)))^2}; \quad (4.104)$$

$nk \in \{2, 3\}$ – количество критериев, которыми проверена гипотеза относительно функции $f_i(x)$.

Если комплекс статистических критериев отклонил все построенные функции, то на рассматриваемом отрезке $[a, b]$ невозможно построить одномодальную либо U-образную функцию плотности распределения Пирсона, описывающую выборку X .

Этап 6. Кластеризация данных – один из этапов нахождения законченного аналитического решения поставленной задачи. Создание одной общей модели для всех данных не редко сложнее определения групп схожих объектов, изучения их особенностей и построения для каждой из групп отдельной модели.

Шаг 6.1. Выборка X (4.96) разделяется на кластеры K_j , $j = 1, \dots, |K_j|$, $X = K_1 \cup \dots \cup K_{|K_j|}$. Количество кластеров $|K_j|$ определяет *Эксперт* в соответствии с количеством участков отдельных «пиков-максимумов» в полигоне частот [25] выборки X (4.96). Для кластеризации применяются методы [19, подразд. 1.4]: Tree Clustering, K-means, FRC. Обобщение итогов кластеризации производится согласно утверждению 1.

Утверждение 11. Точка x_i из множества X принадлежит кластеру K_j , $j \in 1, \dots, |K_j|$, если x_i отнесена к рассматриваемому кластеру не менее чем двумя из трех примененными методами кластеризации данных.

Шаг 6.2. Если обнаружены выбросы, осуществляется их сглаживание, затем шаг 6.1 повторяется.

Шаг 6.3. Для всех кластеров K_j , $j \in 1, \dots, |K_j|$, строятся (этапы 1–5) соответствующие распределению их элементов функции плотности распределения $f_j^*(x)$ [9]. Если для кластера K_{j_0} , $j_0 \in 1, \dots, |K_j|$, комплекс статистических критериев отклонил все построенные кривые, для K_{j_0} невозможно построение $f_{j_0}^*(x)$ и рекомендуется дальнейшее разделение выборки K_{j_0} на несколько выборок-кластеров меньшего размера (см. шаг 6.1), повторение шагов этапов 1–5.

Построенная в результате выполнения этапов методики модель распределения выборки X [21]

$$f^*(x) = \sum_{j=1}^{|K_j|} k_j \cdot f_j^*(x), \quad k_j = \text{iif}(x \in K_j, 1, 0). \quad (4.105)$$

В выражении (4.105) функция *iif* является логическим оператором, принимающим значение 1 при выполнении условия $x \in K_j$, иначе – 0.

5 Исследование методом сингулярного спектрального анализа числовых последовательностей данных

5.1 Исследование методом сингулярного спектрального анализа последовательностей числовых данных с равномерным шумом

5.1.1 Исходные данные для исследования последовательностей числовых данных с равномерным шумом.

Данные в компьютерной информационной системе, представленные вещественными числами и определенные последовательно во времени, именуют временной последовательностью данных. В общем случае такую последовательность данных, обычно упорядоченных во времени, называют временным рядом. Поэтому в дальнейшем будем использовать понятие временного ряда.

Методика проведения экспериментальных исследований включает выбор исходного временного ряда G , который задан по известным функциям

$$G = F(x) = F_T(x) + F_H(x) + F_N(x) \mid x = 0, \dots, 42; \quad (5.1)$$

$$G_T = F_T(x) = ax + b \mid x = 0, \dots, 42; \quad (5.2)$$

$$G_H = F_H(x) = c \sin x \mid x = 0, \dots, 42; \quad (5.3)$$

$$G_N = F_N(x) = Rnd(n_{\inf}; n_{\sup}), \quad (5.4)$$

где a, b, c – параметры;

$Rnd(n_{\inf}; n_{\sup})$ – функция, возвращающая случайные равномерные числа в интервале $[n_{\inf}; n_{\sup}]$ (n_{\inf} – нижняя граница значений случайных чисел; n_{\sup} – верхняя граница значений случайных чисел);

T, H, N – индексы трендовой, гармонической и шумовой составляющих соответственно.

Для предварительного исследования принят временной ряд G_N с равномерным шумом

$$G_N = Rnd(0; 1)_n, n = 43, \quad (5.5)$$

где $Rnd(0; 1)_n$ – функция, возвращающая n случайных чисел, равномерно распределенных в интервале $[0; 1]$.

$G_N = 0,007; 0,059; 0,948; 0,536; 0,426; 0,756; 0,288; 0,939; 0,200; 0,588; 0,586; 0,397; 0,306; 0,139; 0,906; 0,756; 0,564; 0,333; 0,471; 0,000; 0,459; 0,449; 0,290; 0,997; 0,402; 0,181; 0,640; 0,256; 0,538; 0,747; 0,022; 0,603; 0,406; 0,684; 0,965; 0,518; 0,203; 0,077; 0,509; 1,000; 0,806; 0,198; 0,893.$

После предварительного преобразования временного ряда G_N SSA-методом получен восстановленный временной ряд \hat{G} , определяемый двумя составляющими,

$$\hat{G} = \hat{G}_N + \hat{G}_T, \quad (5.6)$$

где \hat{G}_N , \hat{G}_T – соответственно шумовая и трендовая составляющие.

При SSA-преобразовании получены собственные числа, где λ_1 соответствует трендовой составляющей \hat{G}_T , а $\lambda_i, i = 2, \dots, 43$, определяют шумовую составляющую \hat{G}_N .

Проверка по критерию Колмогорова–Смирнова ($p = 0,01$) соответствия полученных случайных чисел теоретическому равномерному распределению показала, что шумовая составляющая \hat{G}_N равномерно распределена на интервале $[-0,5; 0,5]$. При этом трендовая составляющая \hat{G}_T восстановленного временного ряда \hat{G} совпадает с математическим ожиданием исходного временного ряда G . В исходном временном ряде шумовая составляющая (5.5) представлена следующим образом:

$$G_N = Rnd(0; 1) \sim Rnd(-0,5; 0,5) + 0,5 / \alpha = 1. \quad (5.7)$$

5.1.2 Оценки качества восстановления составляющих исходного временного ряда с равномерным шумом.

Для оценивания результатов исследования каждой из составляющих используется сумма модулей значений элементов временного ряда, которую будем именовать Ф-оценкой соответственно для трендовой, гармонической и шумовой составляющих [75],

$$\Phi_T = \sum_{i=1}^n |F_T(x_i)| = \sum_{i=1}^n |a \cdot x_i + b|; \quad (5.8)$$

$$\Phi_H = \sum_{i=1}^n |F_H(x_i)| = \sum_{i=1}^n |c \cdot \sin x_i|; \quad (5.9)$$

$$\Phi_N = \sum_{i=1}^n |F_N(x_i)| = \sum_{i=1}^n |Rnd(n_{inf}; n_{sup})|, \quad (5.10)$$

где n – число элементов временного ряда, $n = 43$.

Исходные данные для проведения исследований представлены в таблице 5.1.

Таблица 5.1 – Исходные данные для исследования временных рядов с равномерным шумом

$F_j(x)$	$F_{Tj}(x)$	$F_{Hj}(x)$	$F_{Nj}(x)$	Φ_{Tj}	Φ_{Hj}	Φ_{Nj}
$F_1(x)$	$1 \cdot x + 2$	$1 \cdot \sin x$	$Rnd(0; 1)$	989	27,15	20,86
$F_2(x)$	$1 \cdot x + 2$	$1 \cdot \sin x$	$Rnd(0; 10)$	989	27,15	210,5
$F_3(x)$	$1 \cdot x + 2$	$10 \cdot \sin x$	$Rnd(0; 1)$	989	271,5	20,86
$F_4(x)$	$1 \cdot x + 2$	$10 \cdot \sin x$	$Rnd(0; 10)$	989	271,5	210,5
$F_5(x)$	$10 \cdot x + 2$	$1 \cdot \sin x$	$Rnd(0; 1)$	9030	27,15	20,86
$F_6(x)$	$10 \cdot x + 2$	$1 \cdot \sin x$	$Rnd(0; 10)$	9030	27,15	210,5
$F_7(x)$	$10 \cdot x + 2$	$10 \cdot \sin x$	$Rnd(0; 1)$	9030	271,5	20,86
$F_8(x)$	$10 \cdot x + 2$	$10 \cdot \sin x$	$Rnd(0; 10)$	9030	271,5	210,5

Для анализа временных рядов в практике имитационного моделирования и сравнения результатов SSA-преобразований будем использовать относительные ф-оценки, определяемые соответственно для трендовой, гармонической и шумовой составляющих,

$$\varphi_T = \frac{\Phi_T}{\min(\Phi_T, \Phi_H, \Phi_N)}; \quad (5.11)$$

$$\varphi_H = \frac{\Phi_H}{\min(\Phi_T, \Phi_H, \Phi_N)}; \quad (5.12)$$

$$\varphi_N = \frac{\Phi_N}{\min(\Phi_T, \Phi_H, \Phi_N)}. \quad (5.13)$$

Принятые ф-оценки характеризуют соотношение между значениями составляющих временного ряда. Результаты расчета относительных ф-оценок представлены в таблице 5.2.

Таблица 5.2 – Относительные ϕ -оценки исследуемых временных рядов $F_j(x)$

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
$\phi_{Tj}, j = 1, \dots, 8$	47,4	36,4	47,4	4,7	432,9	332,6	432,9	42,9
$\phi_{Hj}, j = 1, \dots, 8$	1,3	1,0	13,0	1,3	1,3	1,0	13,0	1,3
$\phi_{Nj}, j = 1, \dots, 8$	1,0	7,8	1,0	1,0	1,0	7,8	1,0	1,0

5.1.3 Обсуждение результатов исследования методом сингулярного спектрального анализа временных рядов с равномерным шумом.

При сингулярном спектральном анализе получено распределение собственных чисел λ_i (таблица 5.3), являющееся одной из характеристик для оценки качества восстановления составляющих исходного временного ряда.

Таблица 5.3 – Собственные числа при сингулярном спектральном анализе временных рядов $F_j(x)$

λ_i	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
λ_1	550,97	641,68	552,76	643,51	5059	5148	5060	5149
λ_2	33,98	28,41	108,81	117,84	383,7	372,7	383,6	372,6
λ_3	10,04	24,66	106,84	114,04	10,04	25,77	108,77	117,71
λ_4	9,94	22,43	33,94	25,68	9,94	24,65	107,14	114,42
λ_5	1,88	22,05	1,88	23,87	1,88	22,07	1,88	23,86
λ_6	1,77	21,17	1,77	21,26	1,77	21,26	1,77	21,87
λ_7	1,60	15,34	1,59	15,50	1,60	15,33	1,59	15,49
λ_8	1,58	14,63	1,56	14,83	1,58	14,74	1,56	15,13
λ_9	1,53	14,49	1,53	14,57	1,54	14,54	1,54	14,57
λ_{10}	1,51	14,41	1,526	14,47	1,51	14,42	1,53	14,47
λ_{11}	1,40	12,62	1,402	12,61	1,40	12,76	1,40	12,75
λ_{12}	1,25	11,67	1,253	11,93	1,25	11,67	1,25	11,93
λ_{13}	1,19	11,14	1,184	11,20	1,19	11,14	1,19	11,21
λ_{14}	1,06	9,70	1,051	9,87	1,06	9,69	1,05	9,86
λ_{15}	0,98	9,44	0,981	9,43	0,98	9,44	0,98	9,42
λ_{16}	0,89	8,59	0,898	9,19	0,89	8,62	0,90	9,27
λ_{17}	0,81	5,81	0,805	5,88	0,81	5,88	0,81	5,96

Окончание таблицы 5.3

λ_i	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
λ_{18}	0,77	4,70	0,764	4,78	0,77	4,65	0,76	4,72
λ_{19}	0,65	2,423	0,641	2,43	0,65	2,42	0,64	2,43
λ_{20}	0,53	2,417	0,521	2,42	0,53	2,418	0,52	2,42
λ_{21}	0,29	1,464	0,302	1,49	0,29	1,48	0,30	1,51
λ_{22}	0,25	1,456	0,238	1,43	0,25	1,42	0,24	1,39

Значения собственных чисел в таблице 5.3, которым соответствуют векторы, восстанавливающие трендовую составляющую, выделены полужирным шрифтом. Собственные числа, соответствующие гармонической составляющей, выделены полужирным шрифтом с подчеркиванием. Остальные λ_i определяют шумовую составляющую.

Аналогично Φ -оценкам и относительным φ -оценкам составляющих временного ряда приняты оценки для сингулярной последовательности собственных чисел

$$\Phi_{\lambda M} = \sum_{i=1}^k \lambda_i, | M = T, H, N; \quad (5.14)$$

$$\varphi_{\lambda M} = \frac{\Phi_{\lambda M}}{\min(\Phi_{\lambda T}, \Phi_{\lambda H}, \Phi_{\lambda N})} | M = T, H, N, \quad (5.15)$$

где k – количество собственных чисел.

Расчеты относительных φ_{λ} -оценок представлены в таблице 5.4.

Таблица 5.4 – Относительные φ_{λ} -оценки сингулярной последовательности собственных чисел

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
$\varphi_{\lambda T j}, j = 1, \dots, 8$	29,3	28,6	29,5	3,0	272,8	640,5	273,4	29,3
$\varphi_{\lambda H j}, j = 1, \dots, 8$	1,0	1,0	10,8	1,1	1,0	1,0	10,8	1,2
$\varphi_{\lambda N j}, j = 1, \dots, 8$	1,0	10,5	1,0	1,0	1,0	26,2	1,0	1,0

Качество восстановления составляющих временного ряда определяется следующими показателями:

$$\Delta F_{Tj}(x) = F_{Tj}(x) - \hat{G}_{Tj}, j = 1, \dots, 8; \quad (5.16)$$

$$\Delta F_{Hj}(x) = F_{Hj}(x) - \hat{G}_{Hj}, j = 1, \dots, 8, \quad (5.17)$$

где $F_{Tj}(x)$, $F_{Hj}(x)$ – трендовая и гармоническая составляющие исходного временного ряда;

\hat{G}_{Tj} , \hat{G}_{Hj} – восстановленные трендовая и гармоническая составляющие временного ряда.

Восстановленная шумовая составляющая \hat{G}_{Nj} , а также $\Delta F_{Tj}(x)$ и $\Delta F_{Hj}(x)$, $j = 1, \dots, 8$, определяются как характеристиками положения: среднее (mean), медиана (med), так и характеристиками рассеяния: стандартное отклонение s , максимум (max), минимум (min), диапазон (range), коэффициенты асимметрии распределения γ_3 и эксцесса γ_4 [23].

Для проверки соответствия восстановленных случайных величин \hat{G}_{Nj} теоретическому равномерному распределению используется критерий Колмогорова–Смирнова при $p = 0,01$. Критическое значение Δ_p для наибольшего отклонения эмпирического распределения от теоретического по [10, таблица 6.2] при $p = 0,01$ и $n = 43$ равно 0,24332. Поскольку наблюдаемое значение

$$\Delta = 0,071 - 0,187 \mid \forall \hat{G}_{Nj}, j = 1, \dots, 8,$$

меньше критического, гипотеза H_0 о равномерном распределении восстановленных случайных величин принимается.

Основные показатели качества восстановления составляющих исходного временного ряда SSA-методом представлены в таблице 5.5.

Таблица 5.5 – Основные показатели качества восстановления составляющих исходного временного ряда SSA-методом

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{T1}(x)$	–0,50	–0,48	–0,36	–0,76	0,40	0,123	–0,49	–1,03
$\Delta F_{H1}(x)$	0,02	0,01	0,23	–0,22	0,45	0,123	–0,08	–1,16
\hat{G}_{N1}	0	–0,01	0,46	–0,52	0,98	0,231	–0,01	–0,72
$\Delta F_{T2}(x)$	–5,73	–5,44	–3,46	–9,29	5,83	1,747	–0,49	–0,97
$\Delta F_{H2}(x)$	0,28	0,27	2,16	–1,19	3,35	0,811	0,23	–0,55
\hat{G}_{N2}	–0,55	–0,33	5,13	–8,68	13,81	3,205	–0,35	–0,10
$\Delta F_{T3}(x)$	–0,70	–0,54	–0,13	–2,12	1,98	0,497	–1,51	1,54

Окончание таблицы 5.5

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{H3}(x)$	0,21	0,10	1,37	-0,36	1,73	0,462	1,19	0,82
\hat{G}_{N3}	0	0	0,45	-0,52	0,98	0,231	-0,01	-0,77
$\Delta F_{T4}(x)$	-5,92	-5,52	-3,47	-10,32	6,85	2,023	-0,66	-0,69
$\Delta F_{H4}(x)$	0,26	0,28	3,56	-1,57	1,99	1,152	0,36	-0,13
\hat{G}_{N4}	-0,76	-0,97	6,03	-7,60	13,63	2,891	-0,12	0,13
$\Delta F_{T5}(x)$	-0,50	-0,48	-0,36	-0,75	0,39	0,122	-0,48	-1,11
$\Delta F_{H5}(x)$	0,02	0,01	0,23	-0,22	0,45	0,124	-0,07	-1,18
\hat{G}_{N5}	0	-0,01	0,45	-0,53	0,98	0,232	-0,01	-0,72
$\Delta F_{T6}(x)$	-4,97	-4,87	-3,86	-6,23	2,37	0,522	-0,69	0,43
$\Delta F_{H6}(x)$	0,04	0,01	1,23	-1,26	2,49	0,750	-0,07	-1,24
\hat{G}_{N6}	-0,03	-0,18	5,05	-5,48	10,52	2,926	0,15	-0,81
$\Delta F_{T7}(x)$	-0,67	-0,51	-0,10	-1,94	1,84	0,434	-1,54	1,67
$\Delta F_{H7}(x)$	0,18	0,13	1,44	-0,70	2,14	0,431	1,07	1,42
\hat{G}_{N7}	0	0	0,45	-0,52	0,97	0,232	-0,01	-0,77
$\Delta F_{T8}(x)$	-5,13	-4,89	-4,36	-7,07	2,71	0,692	-1,25	1,19
$\Delta F_{H8}(x)$	0,21	0,24	2,74	-1,66	1,08	1,085	0,15	-0,80
\hat{G}_{N8}	-0,03	-0,28	5,19	-4,00	9,19	2,420	0,44	-0,62

При исследовании методом сингулярного спектрального анализа временных рядов с равномерным шумом выявлены следующие факты.

1 Шумовая составляющая восстанавливается со средним значением, близким к нулю. Например, для временного ряда $F_8(x)$ среднее значение восстановленной шумовой составляющей равно -0,03 (см. таблицу 5.5), а для исходной шумовой составляющей, определяемой функцией $Rnd(0; 10)$, равно 4,89.

2 Постоянная составляющая шума исключается при восстановлении и суммируется с трендовой составляющей исходного временного ряда.

3 Шумовая составляющая восстанавливается с погрешностью стандартного отклонения, не превышающей 32 %. Например, для временного ряда $F_8(x)$ с соотношением φ_λ -оценок

$$\varphi_{T8} : \varphi_{H8} : \varphi_{N8} = 29,3 : 1,2 : 1,0$$

стандартное отклонение для исходной составляющей равно 2,945, а для восстановленной $s = 2,420$. При этом теоретическое значение стандартного отклонения для $Rnd(0; 10)$ равно 2,887. Отклонение от теоретического значения для исходной шумовой составляющей равно 2,0 %, а для восстановленной – 16,2 %.

4 Наихудшим случаем восстановления гармонической составляющей является $F_2(x)$ с соотношением φ_λ -оценок

$$\varphi_{T2} : \varphi_{H2} : \varphi_{N2} = 28,6 : 1,0 : 10,5,$$

т. е. шум превышает гармоническую составляющую. Полностью восстанавливается лишь количество периодов (рисунок 5.1). Максимальная ошибка превышает амплитуду гармонической составляющей в 2,16 раза (см. таблицу 5.5).

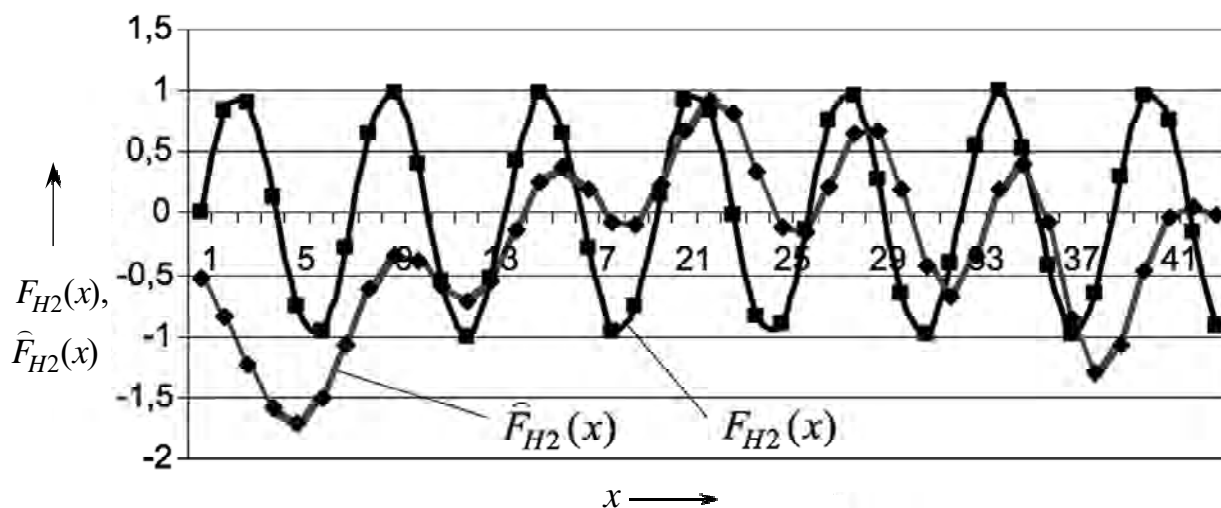


Рисунок 5.1 – Восстановление гармонической составляющей $F_{H2}(x)$

5 Ошибка восстановления трендовой составляющей распределяется в соотношении приблизительно 10 : 80 : 10. Например, ошибка $\Delta F_{T3}(x)$ в середине восстановленного тренда меньше, чем для крайних значений тренда, в 1,8–1,94 раза (рисунок 5.2).

6 Наихудшие результаты восстановления гармонических составляющих наблюдаются для рядов $F_2(x)$ и $F_6(x)$, характеризующихся значительной шумовой составляющей (см. таблицу 5.2). Этот факт проявляется и для φ_λ -оценок сингулярной последовательности собственных чисел (см. таблицу 5.4). При этом следует также обратить внимание на расположение собственных чисел, ответственных за восстановление гармонической составляющей (см. таблицу 5.3).

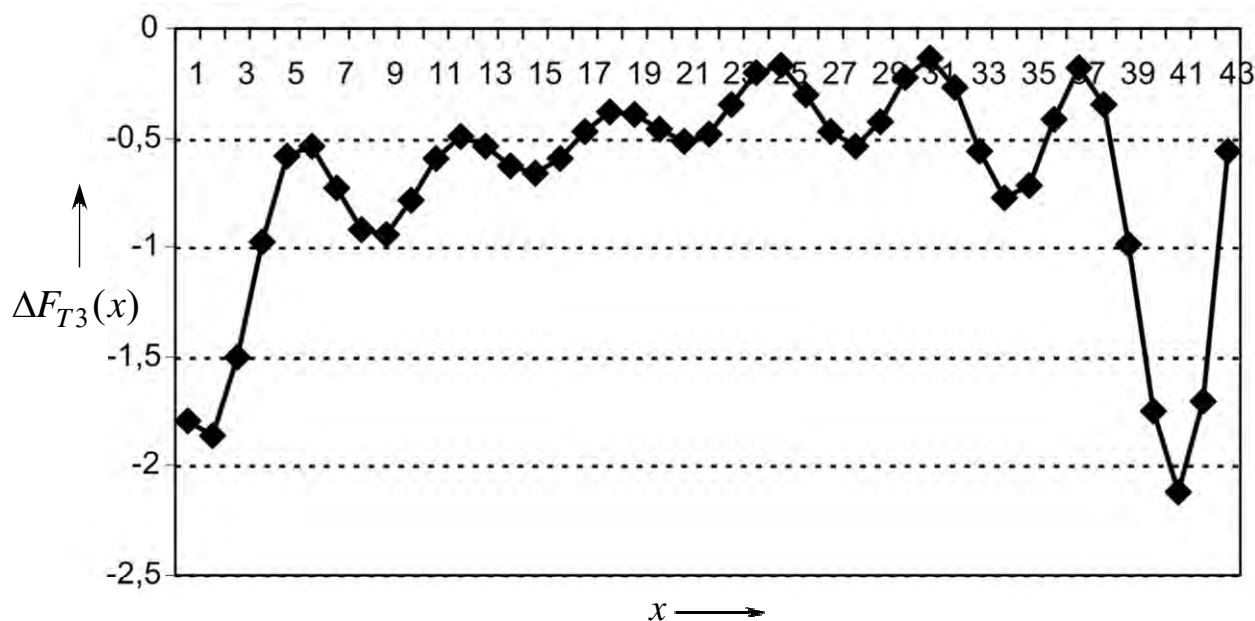


Рисунок 5.2 – Распределение абсолютной ошибки $\Delta F_{T3}(x)$ восстановления тренда $F_{T3}(x)$

Эти числа располагаются среди ниспадающей части собственных чисел, определяющих шумовую составляющую. Данный факт является косвенным предупреждением о возможных значительных ошибках восстановления гармонической составляющей [75].

Аналогичные результаты получены в ходе исследования методом SSA временного ряда с линейным трендом, гармонической сезонной компонентой и шумовой составляющей, имеющей равномерное дискретное распределение. Типичная интерпретация этого распределения – число, выбранное наудачу из n целых чисел, принадлежащих интервалу $[a, a+n-1]$. Генерирование случайных чисел производилось по формуле

$$x_i = [nr_i] + a,$$

где r_i – числа стандартной равномерной последовательности.

Проверка гипотезы о виде распределения осуществлялась при помощи критерия согласия Колмогорова–Смирнова при уровне значимости $p = 0,01$. При исследованиях применялся табличный процессор MS Excel, а также пакеты Minitab и Mathcad.

5.2 Исследование преобразования методом сингулярного спектрального анализа последовательностей данных с нормальным шумом

5.2.1 Исходные данные для исследования последовательностей данных с нормальным шумом.

Для предварительного исследования принята последовательность данных (временной ряд) G с нормальным шумом

$$G = \text{Norm}(1; 1)_n, n = 43, \quad (5.18)$$

где $\text{Norm}(1; 1)_n$ – функция, возвращающая n случайных чисел, распределенных по нормальному закону с математическим ожиданием, равным единице, и среднеквадратическим отклонением, равным единице.

$G = -0,118; 0,404; 1,022; 0,685; 1,650; 2,356; 0,103; 2,002; 1,122; 0,826; 0,949;$
 $1,501; 1,674; 0,890; 0,855; 0,795; -0,031; 0,157; 0,418; -0,578; 0,660; 1,533; 0,797;$
 $-0,234; 3,715; 0,783; 1,700; 0,947; 0,688; 2,117; 0,994; 1,768; 1,550; 1,964; 0,122;$
 $1,476; -0,956; -0,617; 1,549; 2,129; 2,375; 0,546; 1,452.$

После преобразования временного ряда G SSA-методом получен восстановленный временной ряд \hat{G} , определяемый двумя составляющими,

$$\hat{G} = \hat{G}_N + \hat{G}_T, \quad (5.19)$$

где \hat{G}_N , \hat{G}_T – шумовая (рисунок 5.3) и трендовая (рисунок 5.4) составляющие соответственно.

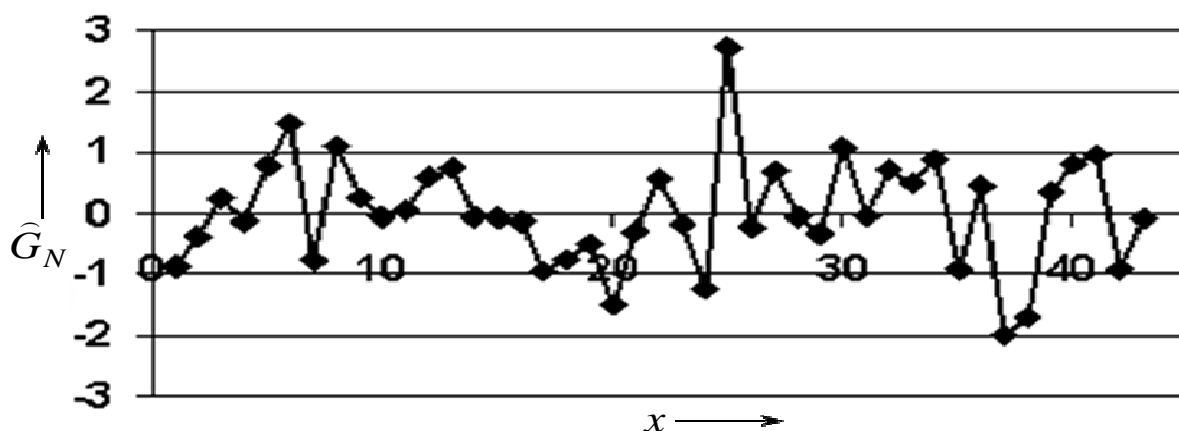


Рисунок 5.3 – Восстановленный шум временного ряда \hat{G}

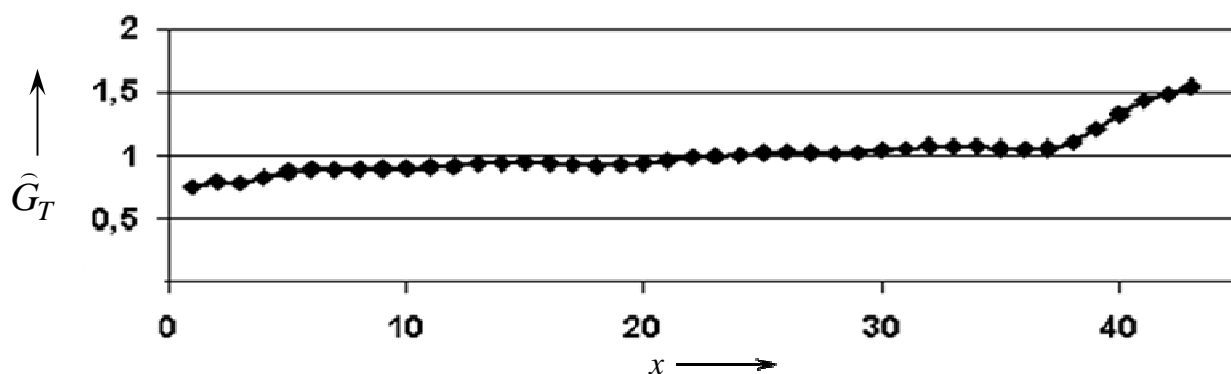


Рисунок 5.4 – Восстановленный тренд временного ряда \hat{G}

При SSA-преобразовании получены собственные числа (рисунок 5.5), где λ_1 соответствует трендовой составляющей \hat{G}_T , а λ_i , $i = 2, \dots, 43$, определяют шумовую составляющую \hat{G}_N . На лепестковой диаграмме (рисунок 5.6) представлены векторы V_1 (линия 1) и V_2 (линия 2), соответствующие собственным числам λ_1 и λ_2 на рисунке 5.6.

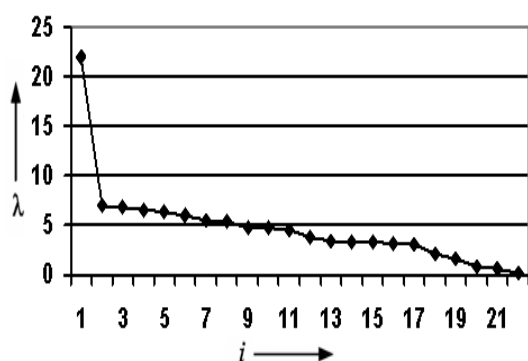


Рисунок 5.5 – Распределение собственных чисел

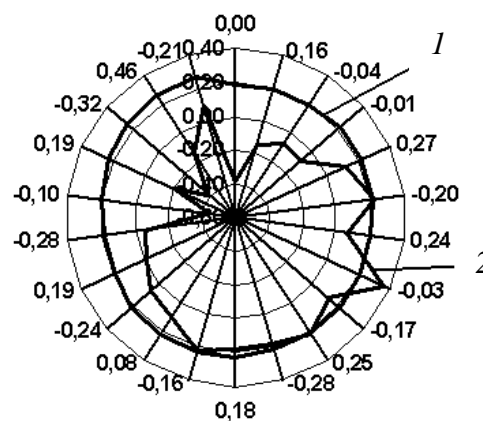


Рисунок 5.6 – Лепестковая диаграмма с векторами V_1 и V_2

Проверка по критерию Колмогорова–Смирнова ($p = 0,01$) соответствия полученных случайных чисел теоретическому нормальному распределению показала, что шумовая составляющая \hat{G}_N нормально распределена со смещением на единицу. При этом трендовая составляющая \hat{G}_T восстановленного временного ряда \hat{G} совпадает с математическим ожиданием исходного временного ряда G . В соответствии с обозначениями (5.19) $\mu = 1,0$ и исходный временной ряд (5.18), в соответствии с (5.19), представлен следующим образом:

$$G = \text{Norm}(1,0; 1,0) \sim \text{Norm}(0; 1,0) + 1,0 \mid \alpha = 1. \quad (5.20)$$

5.2.2 Методика исследования и анализа результатов преобразования методом сингулярного спектрального анализа временных рядов с нормальным шумом.

Методика проведения экспериментальных исследований включает выбор исходного временного ряда G (см. формулу (5.1)), который задан по известным функциям с трендовой G_T (см. формулу (5.2)) и гармонической G_H (см. формулу (5.3)) составляющими.

Шумовая составляющая G_N определена по формуле

$$G_N = F_N(x) = \text{Norm}(\gamma; \beta), \quad (5.21)$$

где $\text{Norm}(\gamma; \beta)$ – функция, возвращающая случайные нормально распределенные числа (γ – математическое ожидание случайных чисел; β – среднеквадратическое отклонение случайных чисел).

Для оценивания результатов исследования каждой из составляющих используются Φ -оценки (5.8) и (5.9) соответственно для трендовой и гармонической составляющих. Φ -оценка для шумовой составляющей в данном исследовании

$$\Phi_N = \sum_{i=1}^n |F_N(x_i)| = \sum_{i=1}^n |\text{Norm}(\gamma; \beta)|, \quad (5.22)$$

где n – число элементов временного ряда, $n = 43$.

Исходные данные для проведения исследований представлены в таблице 5.6.

Для анализа временных рядов и сравнения результатов SSA-преобразований используются относительные Φ -оценки, определяемые соответственно для трендовой, гармонической и шумовой составляющих по формулам (5.11)–(5.13).

Принятые Φ -оценки характеризуют соотношение между значениями составляющих временного ряда. Результаты расчета относительных Φ -оценок представлены в таблице 5.7.

Таблица 5.6 – Исходные данные для исследования временных рядов с нормальным шумом

$F_j(x)$	$F_{Tj}(x)$	$F_{Hj}(x)$	$F_{Nj}(x)$	Φ_{Tj}	Φ_{Hj}	Φ_{Nj}
$F_1(x)$	$0,042 \cdot x$	$1,4 \cdot \sin x$	$Norm(0; 1)$	37,9	38,1	38,3
$F_2(x)$	$0,042 \cdot x$	$1,4 \cdot \sin x$	$Norm(0; 10)$	37,9	38,1	386,9
$F_3(x)$	$0,042 \cdot x$	$14 \cdot \sin x$	$Norm(0; 1)$	37,9	381	38,3
$F_4(x)$	$0,042 \cdot x$	$14 \cdot \sin x$	$Norm(0; 10)$	37,9	381	386,9
$F_5(x)$	$0,42 \cdot x$	$1,4 \cdot \sin x$	$Norm(0; 1)$	379	38,1	38,3
$F_6(x)$	$0,42 \cdot x$	$1,4 \cdot \sin x$	$Norm(0; 10)$	379	38,1	386,9
$F_7(x)$	$0,42 \cdot x$	$14 \cdot \sin x$	$Norm(0; 1)$	379	381	38,3
$F_8(x)$	$0,42 \cdot x$	$14 \cdot \sin x$	$Norm(0; 10)$	379	381	386,9

Таблица 5.7 – Относительные ф-оценки исследуемых временных рядов $F_j(x)$

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
$\Phi_{Tj}, j = 1, \dots, 8$	1,0	1,0	1,0	1,0	9,94	9,94	9,89	1,0
$\Phi_{Hj}, j = 1, \dots, 8$	1,01	1,01	10,1	10,1	1,0	1,0	9,95	1,01
$\Phi_{Nj}, j = 1, \dots, 8$	1,01	10,2	1,0	10,2	1,01	1,02	1,0	1,02

5.2.3 Обсуждение результатов исследования методом сингулярного спектрального анализа временных рядов с нормальным шумом.

При сингулярном спектральном анализе временных рядов с трендовой, гармонической составляющими и нормальным шумом получено распределение собственных чисел λ_i , представленное в таблице 5.8 и являющееся одной из характеристик для оценки качества восстановления составляющих исходного временного ряда.

Значения собственных чисел в таблице 5.8, которым соответствуют векторы, восстанавливающие трендовую составляющую, выделены полужирным шрифтом. Собственные числа, соответствующие гармонической составляющей, выделены полужирным шрифтом с подчеркиванием. Остальные λ_i определяют шумовую составляющую.

Аналогично Ф-оценкам и относительным ф-оценкам составляющих временного ряда приняты оценки, определяемые по формулам (5.14) и (5.15), для сингулярной последовательности собственных чисел.

Расчеты относительных Φ_{λ} -оценок представлены в таблице 5.9.

Таблица 5.8 – Собственные числа при сингулярном спектральном анализе временных рядов $F_j(x)$

λ_i	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
λ_1	20,333	<u>97,225</u>	<u>152,823</u>	<u>231,99</u>	208,952	246,375	216,023	263,712
λ_2	<u>14,607</u>	<u>94,692</u>	<u>152,484</u>	<u>228,253</u>	18,422	<u>97,097</u>	<u>152,354</u>	<u>231,809</u>
λ_3	<u>14,029</u>	86,345	18,442	86,532	<u>14,595</u>	<u>92,826</u>	<u>145,128</u>	<u>208,92</u>
λ_4	9,933	83,912	9,954	84,132	<u>14,467</u>	85,806	18,403	86,342
λ_5	9,765	61,414	9,77	61,911	9,925	83,895	9,946	84,154
λ_6	7,429	59,189	7,448	58,983	9,767	61,238	9,772	61,057
λ_7	6,977	55,591	6,98	54,675	7,383	59,07	7,422	59,038
λ_8	6,793	48,206	6,776	48,288	6,938	51,854	6,962	51,175
λ_9	6,548	43,426	6,507	43,384	6,702	43,428	6,711	43,389
λ_{10}	5,939	42,534	5,874	42,708	6,522	42,693	6,502	42,855
λ_{11}	5,671	41,055	5,614	40,407	5,642	40,859	5,582	40,188
...
λ_{21}	0,875	5,476	0,854	5,573	0,891	5,414	0,86	5,515
λ_{22}	0,786	3,728	0,743	3,579	0,353	3,848	0,297	3,712

Таблица 5.9 – Относительные фл-оценки сингулярной последовательности собственных чисел

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
$\Phi_{\lambda T j}, j = 1, \dots, 8$	1,0	1,0	1,0	1,0	7,8	1,3	2,8	1,0
$\Phi_{\lambda H j}, j = 1, \dots, 8$	1,4	3,1	16,5	7,4	1,0	1,0	3,6	1,7
$\Phi_{\lambda N j}, j = 1, \dots, 8$	4,4	11,8	4,8	11,7	2,9	3,9	1,0	2,8

Для определения качества восстановления составляющих временного ряда приняты показатели ΔF_{Tj} , ΔF_{Hj} , определяемые по формулам (5.16) и (5.17) соответственно для трендовой и гармонической составляющих.

Восстановленная шумовая составляющая \hat{G}_{Nj} , а также $\Delta F_{Tj}(x)$ и $\Delta F_{Hj}(x)$, $j = 1, \dots, 8$, определяются характеристиками положения: среднее (mean), медиана (med); характеристиками рассеяния: стандартное отклонение s ,

максимум (max), минимум (min), диапазон (range), коэффициенты асимметрии распределения γ_3 и эксцесса γ_4 , представленными в таблице 5.10.

Таблица 5.10 – Основные показатели качества восстановления составляющих исходного ряда SSA-методом

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{T1}(x)$	0,03	0,09	0,35	-0,69	1,05	0,247	-1,39	1,47
$\Delta F_{H1}(x)$	0,02	0,05	0,58	-0,62	1,20	0,334	-0,23	-0,83
\hat{G}_{N1}	-0,10	-0,07	2,34	-1,89	4,23	1,024	0,09	-0,63
$\Delta F_{T2}(x)$	-1,14	-1,26	1,65	-2,83	4,48	1,131	0,79	0,24
$\Delta F_{H2}(x)$	-0,22	-0,05	9,32	-11,18	20,50	5,521	-0,05	-1,24
\hat{G}_{N2}	-0,63	-0,76	13,92	-21,37	35,29	8,352	-0,15	-0,50
$\Delta F_{T3}(x)$	0,09	0,10	0,22	-0,12	0,34	0,088	-0,63	-0,05
$\Delta F_{H3}(x)$	-0,05	-0,04	0,43	-0,55	0,98	0,282	0,01	-1,15
\hat{G}_{N3}	-0,10	-0,02	2,33	-1,83	4,16	1,025	0,12	-0,69
$\Delta F_{T4}(x)$	-1,31	-1,19	1,68	-4,44	6,12	1,349	-0,20	0,52
$\Delta F_{H4}(x)$	-0,14	-0,82	8,68	-9,21	17,89	5,424	-0,01	-1,37
\hat{G}_{N4}	-0,72	-1,39	14,25	-21,19	35,44	8,467	-0,08	-0,56
$\Delta F_{T5}(x)$	0,10	0,11	0,39	-0,31	0,69	0,157	-0,62	0,45
$\Delta F_{H5}(x)$	0,03	0,03	0,56	-0,48	1,04	0,274	-0,12	-1,02
\hat{G}_{N5}	-0,01	0,03	2,20	-1,80	4,01	0,981	-0,03	-0,64
$\Delta F_{T6}(x)$	-2,04	-1,59	2,69	-5,51	8,2	1,600	0,11	0,98
$\Delta F_{H6}(x)$	0,22	0,55	8,48	-8,68	17,16	5,158	0,03	-1,41
\hat{G}_{N6}	-1,09	-1,44	13,34	-23,30	36,64	8,410	-0,21	-0,32
$\Delta F_{T7}(x)$	-0,04	0,16	2,62	-4,88	7,51	1,710	-1,07	0,95
$\Delta F_{H7}(x)$	0,17	-0,06	4,98	-2,99	7,98	1,819	0,72	0,40
\hat{G}_{N7}	-0,01	0,06	2,20	-1,76	3,96	0,986	0,01	-0,70
$\Delta F_{T8}(x)$	-0,72	-0,32	8,32	-15,07	23,38	5,365	-0,68	0,37
$\Delta F_{H8}(x)$	-1,15	-1,79	12,14	-7,63	19,76	4,188	1,22	1,97
\hat{G}_{N8}	-1,14	-2,11	14,05	-23,62	37,67	8,676	-0,17	-0,36

Для проверки соответствия восстановленных случайных величин \hat{G}_{Nj} теоретическому нормальному распределению используется критерий Колмогорова–Смирнова. Критическое значение Δ_p для наибольшего отклонения эмпирического распределения от теоретического по [10, таблица 6.2] при $p = 0,01$ и $n = 43$ равно 0,24332. Если наблюдаемое значение Δ больше или равно критическому значению, нулевая гипотеза H_0 отклоняется. Поскольку наблюдаемое значение

$$\Delta = 0,027 \dots 0,074 \mid \forall \hat{G}_{Nj}, j = 1, \dots, 8,$$

меньше критического, гипотеза H_0 о нормальном распределении восстановленных случайных величин принимается.

Факты, обнаруженные при исследовании методом сингулярного спектрального анализа временных рядов с нормальным шумом.

1 Шумовая составляющая восстанавливается со средним значением, близким к нулю. Например, для ряда $F_8(x)$ среднее значение восстановленной шумовой составляющей равно $-1,14$ (см. таблицу 5.10), а для исходной шумовой составляющей, определяемой функцией $Norm(0; 10)$, среднее равно 4,89.

2 Постоянная составляющая шума исключается при восстановлении и суммируется с трендовой составляющей исходного временного ряда.

3 Наиболее сложными для восстановления составляющих исходного временного ряда являются $F_2(x)$ и $F_4(x)$ с соотношением ф-оценок соответственно $\varphi_{T2} : \varphi_{H2} : \varphi_{N2} = 1,0 : 1,01 : 10,2$ и $\varphi_{T4} : \varphi_{H4} : \varphi_{N4} = 1,0 : 10,1 : 10,2$. При этом соотношение относительных φ_λ -оценок сингулярной последовательности собственных чисел соответственно $\varphi_{\lambda T2} : \varphi_{\lambda H2} : \varphi_{\lambda N2} = 1,0 : 3,1 : 11,8$ и $\varphi_{\lambda T4} : \varphi_{\lambda H4} : \varphi_{\lambda N4} = 1,0 : 7,4 : 11,7$.

4 Трендовая составляющая у временных рядов $F_2(x)$ и $F_4(x)$ определена среди собственных чисел, соответствующих шумовой составляющей. При этом ошибка восстановления имеет характерный вид, представленный на рисунке 5.7.

5 При восстановлении временных рядов $F_1(x)$ и $F_8(x)$ трендовая составляющая определена векторами, соответствующими собственному числу λ_1 в обоих случаях. При этом ошибка восстановления для $F_{T1}(x)$ имеет вид, представленный на рисунке 5.8.

Основные результаты подтверждены также при экспериментальном исследовании методом SSA временного ряда с шумовой составляющей, распределенной по закону Коши. Подтверждена гипотеза о том, что выделенная по методу SSA шумовая составляющая, как и шумовая составляющая исходного ряда, распределена по закону Коши с

систематической ошибкой на величину параметра μ . На эту же величину изменена и трендовая составляющая восстановленного временного ряда.

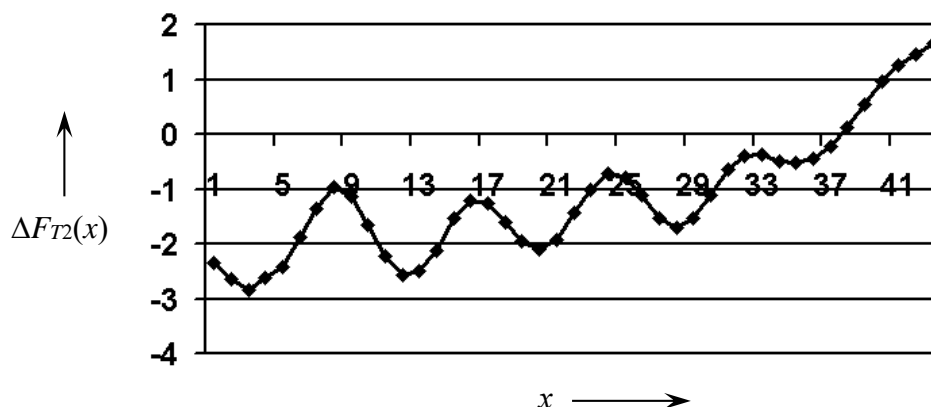


Рисунок 5.7 – Распределение абсолютной ошибки $\Delta F_{T2}(x)$ восстановления тренда $F_{T2}(x)$

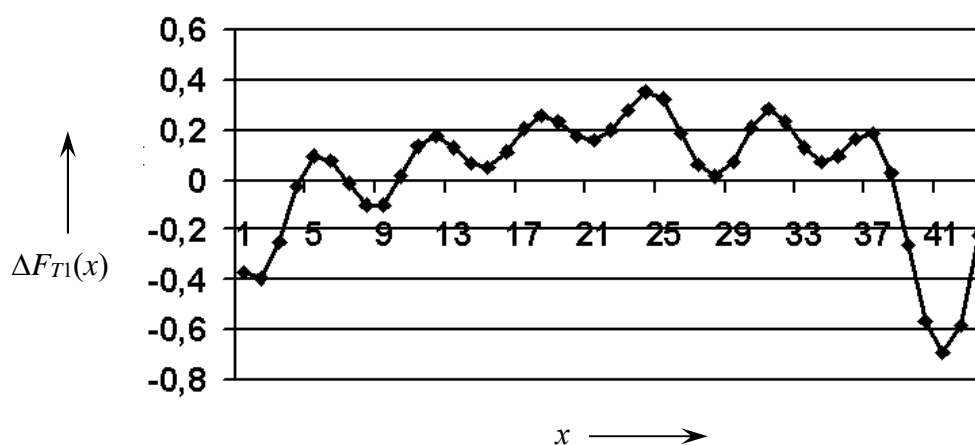


Рисунок 5.8 – Распределение абсолютной ошибки $\Delta F_{T1}(x)$ восстановления тренда $F_{T1}(x)$

5.3 Обоснование применения нормированного временного ряда

5.3.1 Влияние постоянной составляющей временного ряда на качество сингулярного спектрального анализа.

Задачей исследования является оценка влияния постоянной составляющей тренда на качество восстановления гармонической и шумовой составляющих временного ряда, определяемого моделью по формулам (5.1)–(5.4).

В [87] показано, что величина постоянной составляющей не оказывает влияния на качество восстановления временного ряда с шумовой составляющей. В отличие от [87] в настоящем исследовании временной ряд состоит

из постоянной и динамической составляющих тренда, гармонической и шумовой составляющих.

Представим гармоническую составляющую в общем случае с k периодами как

$$G_H = \sin\left(k \frac{2\pi(x-1)}{n-1}\right) \mid x=1, \dots, n, \quad k > 0. \quad (5.23)$$

В настоящем исследовании будем полагать $k = 2$.

Трендовую составляющую G_T определим по формуле

$$G_T = \frac{2(x-1)}{n-1} - 1 \mid x=1, \dots, n; \quad n = 43. \quad (5.24)$$

Представим G_T в виде двух составляющих:

$$G_T = G_{TC} + G_{TD},$$

где G_{TC} – постоянная составляющая;

G_{TD} – динамическая составляющая,

причем $G_{TC} = 0$.

Чтобы не вносить дополнительных постоянных составляющих, G_N определим по формуле

$$G_N = Rnd(-1; 1).$$

Исходные данные для проведения исследований представлены в таблице 5.11; результаты расчета относительных ф-оценок – в таблице 5.12.

Распределение собственных чисел λ_i представлено в таблице 5.13.

Таблица 5.11 – Исходные данные для исследования постоянной составляющей временных рядов $F_j(x)$

$F_j(x)$	G_{TCj}	$F_{TDj}(x)$	$F_{Hj}(x)$	$F_{Nj}(x)$	Φ_{TDj}	Φ_{Hj}	Φ_{Nj}
$F_1(x)$	-100	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89
$F_2(x)$	-10	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89
$F_3(x)$	-1	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89

Окончание таблицы 5.11

$F_j(x)$	G_{TCj}	$F_{TDj}(x)$	$F_{Hj}(x)$	$F_{Nj}(x)$	Φ_{TDj}	Φ_{Hj}	Φ_{Nj}
$F_4(x)$	0	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89
$F_5(x)$	1	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89
$F_6(x)$	10	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89
$F_7(x)$	100	$\frac{2x}{(n-1)} - 1$	$\sin\left(\frac{4\pi(x-1)}{n-1}\right)$	$Rnd(-1; 1)$	22,0	26,69	21,89

Таблица 5.12 – Относительные φ -оценки исследуемых временных рядов $F_j(x)$

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$
$\varphi_{Tj}, j = 1, \dots, 8$	196,5	19,65	1,96	1,01	1,96	19,65	196,5
$\varphi_{Hj}, j = 1, \dots, 8$	1,22	1,22	1,22	1,22	1,22	1,22	1,22
$\varphi_{Nj}, j = 1, \dots, 8$	1	1	1	1	1	1	1

Таблица 5.13 – Собственные числа при сингулярном спектральном анализе временных рядов $F_j(x)$

λ_i	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$
λ_1	2198,0	218,09	21,916	<u>10,953</u>	25,723	222,26	2202,0
λ_2	<u>10,158</u>	<u>10,122</u>	<u>9,631</u>	<u>10,927</u>	<u>10,421</u>	<u>10,197</u>	<u>10,165</u>
λ_3	<u>9,199</u>	<u>9,247</u>	<u>9,549</u>	5,274	<u>8,526</u>	<u>9,138</u>	<u>9,188</u>
λ_4	4,931	4,928	4,892	4,936	4,968	4,935	4,932
λ_5	4,886	4,886	4,889	4,486	4,882	4,885	4,886
λ_6	4,033	4,033	4,035	4,281	4,031	4,033	4,033
λ_7	3,869	3,863	3,815	3,974	3,949	3,877	3,870
λ_8	3,571	3,564	3,495	3,722	3,626	3,578	3,572
λ_9	3,308	3,308	3,309	3,302	3,307	3,308	3,308
λ_{10}	2,871	2,876	2,933	3,148	2,822	2,865	2,870

Окончание таблицы 5.13

λ_i	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$
λ_{11}	2,634	2,633	2,630	2,627	2,703	2,635	2,634
λ_{12}	2,484	2,464	2,284	2,577	2,618	2,508	2,489
λ_{13}	2,266	2,268	2,261	2,159	2,252	2,265	2,266
λ_{14}	1,755	1,757	1,782	1,821	1,738	1,753	1,755
λ_{15}	1,719	1,719	1,714	1,69	1,727	1,72	1,719
λ_{16}	1,571	1,573	1,591	1,686	1,552	1,569	1,571
λ_{17}	1,552	1,552	1,552	1,552	1,552	1,552	1,552
λ_{18}	1,437	1,425	1,307	1,433	1,544	1,451	1,44
λ_{19}	0,971	0,966	0,999	1,137	1,048	0,984	0,973
λ_{20}	0,964	0,959	0,756	0,854	0,94	0,96	0,963
λ_{21}	0,643	0,645	0,665	0,538	0,627	0,641	0,643
λ_{22}	0,052	0,052	0,059	0,034	0,048	0,051	0,052

Значения собственных чисел в таблице 5.13, которым соответствуют векторы, восстанавливающие трендовую составляющую, выделены полужирным шрифтом. Собственные числа, соответствующие гармонической составляющей, выделены полужирным шрифтом с подчеркиванием. Остальные λ_i определяют шумовую составляющую.

Аналогично относительным ф-оценкам составляющих временного ряда представлены расчеты оценок для сингулярной последовательности собственных чисел. Расчеты абсолютных Φ_λ -оценок сингулярной последовательности собственных чисел приведены в таблице 5.14.

Для определения качества восстановления составляющих временного ряда приняты следующие показатели:

$$\Delta F_{Tj}(x) = F_{Tj}(x) - \hat{G}_{Tj}, j = 1, \dots, 8;$$

$$\Delta F_{Hj}(x) = F_{Hj}(x) - \hat{G}_{Hj}, j = 1, \dots, 8,$$

где $F_{Tj}(x)$, $F_{Hj}(x)$ – трендовая и гармоническая составляющие исходного временного ряда;

\hat{G}_{Tj} , \hat{G}_{Hj} – восстановленные трендовая и гармоническая составляющие временного ряда.

Таблица 5.14 – Абсолютные Φ_λ -оценки сингулярной последовательности собственных чисел

Показатель	$F_1(x)$	$F_2(x)$	$F_3(x)$	$F_4(x)$	$F_5(x)$	$F_6(x)$	$F_7(x)$	$F_8(x)$
$\Phi_{\lambda T j}, j = 1, \dots, 8$	2200	218,1	21,92	–	25,72	222,3	2200	2200
$\Phi_{\lambda H j}, j = 1, \dots, 8$	19,357	19,369	19,18	21,88	18,947	19,335	19,353	19,357
$\Phi_{\lambda N j}, j = 1, \dots, 8$	45,517	45,471	44,968	51,231	45,934	45,57	45,528	45,517

Восстановленная шумовая составляющая \hat{G}_{Nj} , а также $\Delta F_{Tj}(x)$ и $\Delta F_{Hj}(x)$, $j = 1, \dots, 8$, определяются характеристиками положения: среднее (mean), медиана (med); характеристиками рассеяния: стандартное отклонение s , максимум (max), минимум (min), диапазон (range), коэффициенты асимметрии распределения γ_3 и эксцесса γ_4 . Показатели качества восстановления исходного временного ряда представлены в таблице 5.15.

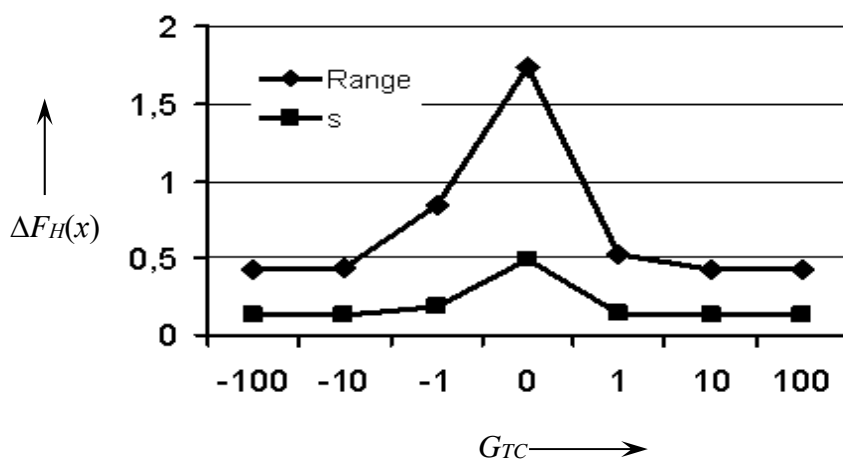
Таблица 5.15 – Основные показатели качества восстановления составляющих исходного ряда SSA-методом

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{T1}(x)$	–0,071	–0,096	0,071	–0,142	0,213	0,064	0,899	–0,410
$\Delta F_{H1}(x)$	0,029	0,047	0,245	–0,176	0,421	0,134	–0,031	–1,088
\hat{G}_{N1}	–0,029	–0,086	0,973	–1,163	2,136	0,557	–0,127	–0,796
$\Delta F_{T2}(x)$	–0,071	–0,095	0,088	–0,138	0,226	0,062	0,914	–0,291
$\Delta F_{H2}(x)$	0,032	0,053	0,260	–0,178	0,439	0,136	–0,056	–1,126
\hat{G}_{N2}	–0,027	–0,086	0,972	–1,160	2,132	0,556	–0,128	–0,792
$\Delta F_{T3}(x)$	–0,084	–0,089	0,124	–0,201	0,325	0,074	0,733	0,576
$\Delta F_{H3}(x)$	0,065	0,078	0,610	–0,229	0,839	0,187	0,535	0,222
\hat{G}_{N3}	–0,006	0,027	0,971	–1,139	2,111	0,550	–0,143	–0,674
$\Delta F_{T4}(x)$	–	–	–	–	–	–	–	–
$\Delta F_{H4}(x)$	–0,007	–0,059	0,873	–0,862	1,735	0,495	0,096	–0,699
\hat{G}_{N4}	0,006	–0,027	2,057	–2,142	4,199	0,755	–0,119	1,304
$\Delta F_{T5}(x)$	–0,061	–0,116	0,339	–0,206	0,545	0,136	1,863	2,676
$\Delta F_{H5}(x)$	–0,001	–0,028	0,261	–0,264	0,525	0,142	0,213	–0,752
\hat{G}_{N5}	–0,049	–0,083	0,988	–1,187	2,174	0,570	–0,161	–0,708

Окончание таблицы 5.15

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{T6}(x)$	−0,071	−0,096	0,085	−0,145	0,230	0,067	0,983	−0,196
$\Delta F_{H6}(x)$	0,027	0,041	0,248	−0,173	0,421	0,131	0,008	−1,032
\hat{G}_{N6}	−0,031	−0,086	0,974	−1,165	2,140	0,559	−0,127	−0,798
$\Delta F_{T7}(x)$	−0,071	−0,094	0,069	−0,142	0,212	0,065	0,909	−0,390
$\Delta F_{H7}(x)$	0,029	0,045	0,245	−0,175	0,421	0,133	−0,024	−1,078
\hat{G}_{N7}	−0,029	−0,086	0,973	−1,163	2,136	0,558	−0,127	−0,797

Основным результатом исследования является вывод о независимости качества восстановления составляющих временного ряда от величины постоянной составляющей исходного временного ряда (рисунки 5.9–5.11). Однако следует заметить, что при нулевой постоянной составляющей $G_{TC} = 0$ не удастся восстановить динамическую составляющую тренда исходного временного ряда (см. рисунок 5.10). При этом в 2–3 раза возрастает ошибка восстановления гармонической и шумовой составляющих (см. рисунки 5.9 и 5.11).

Рисунок 5.9 – Показатели ошибки $\Delta F_H(x)$ восстановления $F_{Hj}(x)$

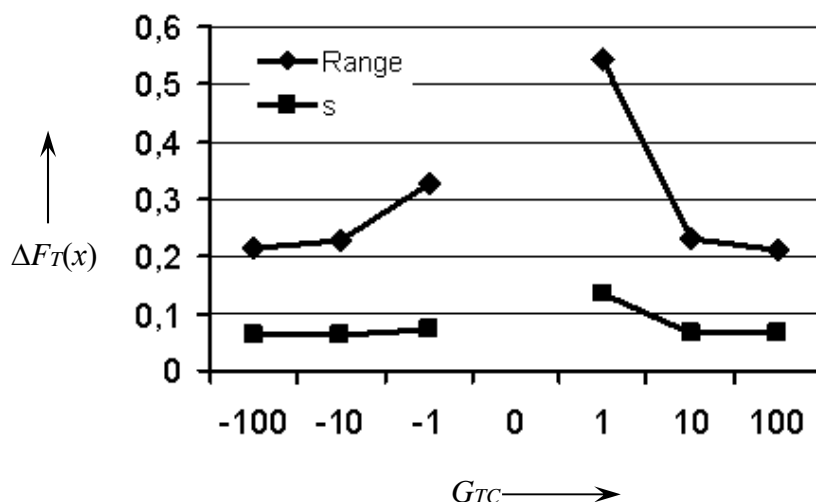


Рисунок 5.10 – Показатели ошибки $\Delta F_T(x)$ восстановления $F_{Tj}(x)$

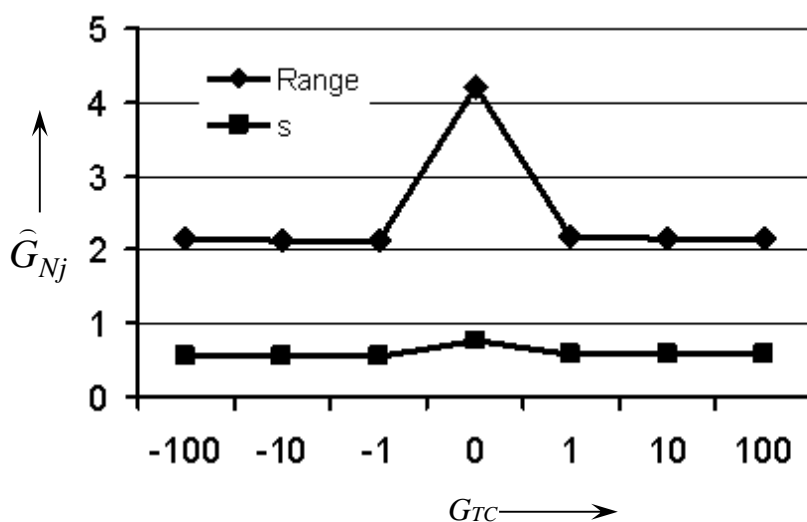


Рисунок 5.11 – Показатели восстановленной шумовой составляющей \hat{G}_{Nj}

5.3.2 Исследование сжатия/растяжения временного ряда.

Задачей исследования является оценка влияния процедуры сжатия/растяжения временного ряда на качество восстановления его составляющих методом сингулярного спектрального анализа.

Под сжатием/растяжением будем понимать пропорциональное уменьшение/увеличение значений элементов ряда путем умножения на коэффициент $1/\alpha$ ($0 < \alpha < 1$)/($\alpha > 1$).

Методика проведения экспериментальных исследований включает выбор исходного временного ряда G , который задан по известным функциям

$$G = F(x) = F_T(x) + F_H(x) + F_N(x);$$

$$G_T = F_T(x) = ax + b;$$

$$G_H = F_H(x) = c \sin x;$$

$$G_N = F_N(x) = \text{Norm}(\gamma; \beta),$$

где a, b, c – параметры;

$\text{Norm}(\gamma; \beta)$ – функция, возвращающая случайные нормально распределенные числа (γ – математическое ожидание случайных чисел; β – средне-квадратическое отклонение случайных чисел);

T, H, N – индексы трендовой, гармонической и шумовой составляющих соответственно.

Для оценивания результатов исследования каждой из составляющих используется сумма модулей значений элементов временного ряда, которая называется Ф-оценкой соответственно для трендовой, гармонической и шумовой составляющих. Исходные данные для проведения исследований представлены в таблице 5.16.

Таблица 5.16 – Исходные данные для исследования временных рядов с нормальным шумом и $1/\alpha = 5$, $1/\alpha = 10$

$F_j(x)$	$F_{Tj}(x)$	$F_{Hj}(x)$	$F_{Nj}(x)$	Φ_{Tj}	Φ_{Hj}	Φ_{Nj}
$F_1(x)$	$5 \left(\frac{2x}{(n-1)} - 1 \right)$	$5 \cdot \sin \left(\frac{4\pi(x-1)}{n-1} \right)$	$5 \cdot \text{Norm}(0; 1)$	2150	133,4	109,4
$F_2(x)$	$10 \left(\frac{2x}{(n-1)} - 1 \right)$	$10 \cdot \sin \left(\frac{4\pi(x-1)}{n-1} \right)$	$10 \cdot \text{Norm}(0; 1)$	4300	266,9	218,9

Для анализа временных рядов и сравнения результатов SSA-преобразований используются относительные ф-оценки, определяемые соответственно для трендовой, гармонической и шумовой составляющих. Принятые ф-оценки характеризуют соотношение между значениями составляющих временного ряда (таблица 5.17).

При сингулярном спектральном анализе получено распределение собственных чисел λ_i (таблица 5.18), являющееся одной из характеристик для оценки качества восстановления составляющих исходного временного ряда.

Значения собственных чисел в таблице 5.18, которым соответствуют векторы, восстанавливающие трендовую составляющую, выделены полужирным шрифтом, гармоническую составляющую – полужирным шрифтом с подчеркиванием. Остальные λ_i определяют шумовую составляющую. На ри-

сунках 5.12–5.16 векторы $U_1–U_5$ временного ряда $F_1(x)$ соответствуют собственным числам $\lambda_1–\lambda_5$. На рисунках 5.17–5.21 векторы $U_1–U_5$ временного ряда $F_2(x)$ соответствуют собственным числам $\lambda_1–\lambda_5$.

Таблица 5.17 – Относительные φ -оценки исследуемых временных рядов $F_1(x)$ и $F_2(x)$

Показатель	$F_1(x)$	$F_2(x)$
$\varphi_{Tj}, j = 1, 2$	19,65	19,64
$\varphi_{Hj}, j = 1, 2$	1,22	1,22
$\varphi_{Nj}, j = 1, 2$	1,00	1,00

Таблица 5.18 – Собственные числа при сингулярном спектральном анализе временных рядов $F_1(x)$ и $F_2(x)$

λ_i	$F_1(x)$	$F_2(x)$	λ_i	$F_1(x)$	$F_2(x)$
λ_1	1110	2220	λ_{12}	12,56	25,148
λ_2	<u>50,963</u>	<u>101,944</u>	λ_{13}	11,346	22,707
λ_3	<u>45,683</u>	<u>91,4</u>	λ_{14}	8,736	17,46
λ_4	24,652	49,34	λ_{15}	8,577	17,144
λ_5	24,416	48,869	λ_{16}	7,877	15,755
λ_6	20,175	40,349	λ_{17}	7,792	15,587
λ_7	19,389	38,778	λ_{18}	7,267	14,535
λ_8	17,896	35,796	λ_{19}	4,938	9,878
λ_9	16,552	33,116	λ_{20}	4,81	9,628
λ_{10}	14,334	28,684	λ_{21}	3,222	6,449
λ_{11}	13,214	26,437	λ_{22}	0,235	0,457

Анализ векторов $U_1–U_5$ для временных рядов $F_1(x)$ и $F_2(x)$ не выявил различий между ними. Различия не выявлены и для остальных векторов U_i , соответствующих собственным числам λ_i . Основные показатели восстановления временных рядов $F_1(x)$ и $F_2(x)$ представлены в таблице 5.19.

В соответствии с результатами таблицы 5.19 среднее ошибки восстановления трендовой составляющей $\Delta F_{T1}(x)$ находится в соотношении 1 : 2 к среднему ошибки восстановления трендовой составляющей $\Delta F_{T2}(x)$. В таком же соотношении находятся и другие показатели в таблице 5.19: med, max, min, range, s. При этом γ_3 и γ_4 не изменяются. Это же справедливо и для

ошибок восстановления гармонической составляющей, а также шумовой составляющей.

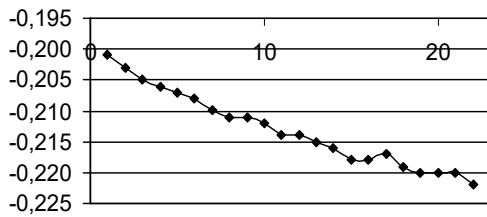


Рисунок 5.12 – Вектор U_1 ряда $F_1(x)$

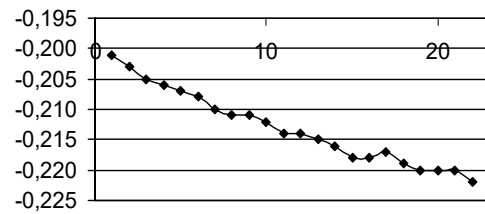


Рисунок 5.17 – Вектор U_1 ряда $F_2(x)$

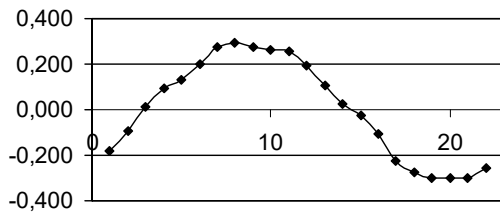


Рисунок 5.13 – Вектор U_2 ряда $F_1(x)$

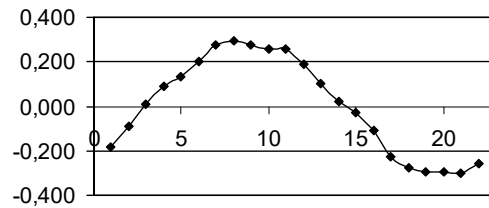


Рисунок 5.18 – Вектор U_2 ряда $F_2(x)$

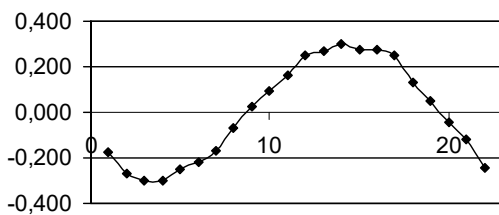


Рисунок 5.14 – Вектор U_3 ряда $F_1(x)$

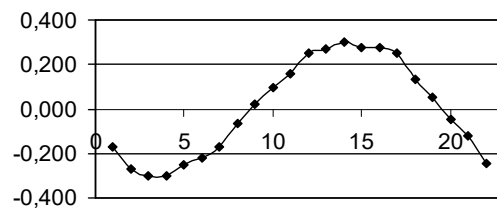


Рисунок 5.19 – Вектор U_3 ряда $F_2(x)$

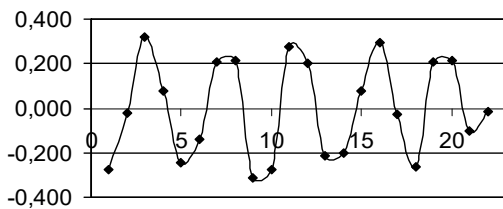


Рисунок 5.15 – Вектор U_4 ряда $F_1(x)$

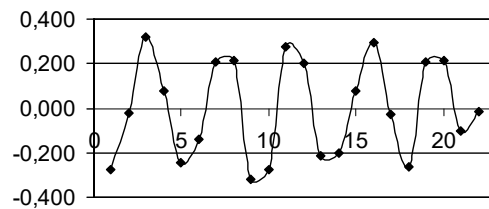


Рисунок 5.20 – Вектор U_4 ряда $F_2(x)$

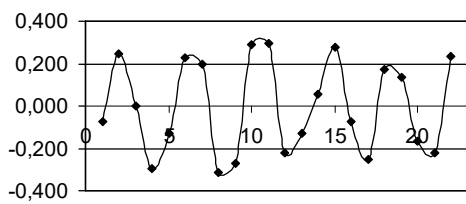


Рисунок 5.16 – Вектор U_5 ряда $F_1(x)$

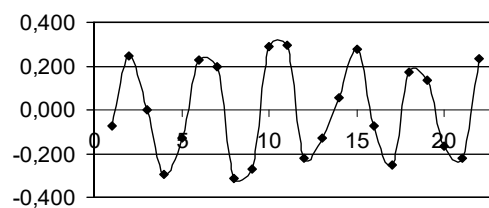


Рисунок 5.21 – Вектор U_5 ряда $F_2(x)$

Таблица 5.19 – Основные показатели качества восстановления составляющих исходного ряда SSA-методом

Оценка	mean	med	max	min	range	s	γ_3	γ_4
$\Delta F_{T1}(x)$	–0,348	–0,473	0,428	–0,716	1,144	0,334	0,981	–0,204
$\Delta F_{H1}(x)$	0,134	0,207	1,235	–0,870	2,106	0,658	–0,002	–1,039
\hat{G}_{N1}	–0,156	–0,440	4,884	–5,827	10,711	2,794	–0,125	–0,800
$\Delta F_{T2}(x)$	–0,694	–0,944	0,858	–1,430	2,288	0,669	0,980	–0,206
$\Delta F_{H2}(x)$	0,267	0,418	2,464	–1,740	4,204	1,314	–0,004	–1,040
\hat{G}_{N2}	–0,312	–0,870	9,757	–11,658	21,415	5,589	–0,125	–0,800

Утверждение. Изменение элементов временного ряда G пропорционально величине $\alpha \in R$ изменяет показатели (характеристики положения: среднее (mean), медиана (med); характеристики рассеяния: стандартное отклонение s , максимум (max), минимум (min), диапазон (range)) погрешности восстановления аддитивных трендовой G_T , гармонической G_H и шумовой G_N составляющих, $G = G_T + G_H + G_N$, методом сингулярного спектрального анализа также пропорционально величине α .

Доказательство утверждения. Пусть временной ряд $G = G_T + G_H + G_N$, причем G_T , G_H и G_N заданы известными функциями соответственно F_T , F_H и известен закон распределения G_N . В соответствии с сингулярным спектральным анализом по ряду G построена ганкелева матрица A заданного размера $L \times K$ (см. подразд. 3.3). Для матрицы $S = A \cdot A^T$ вычисляются решением характеристического уравнения $S - \mu \cdot E = 0$ собственные числа $\{\mu_k\}_{k=1}^L$, а также левые $\{U_k\}_{k=1}^L$ и правые $\{V_k\}_{k=1}^L$ векторы [1], формируя разложение $A = \sum_k A_k$, $A_k = \sqrt{\mu_k} \cdot U_k \cdot V_k^T$. Определены матрицы $A_T = \sum_{k \in T} A_k$, $A_H = \sum_{k \in H} A_k$, $A_N = \sum_{k \in N} A_k$. После диагонального усреднения матриц A_T , A_H , A_N получены восстановленные аддитивные составляющие \hat{G}_T , \hat{G}_H , \hat{G}_N , что позволяет оценить показатели ошибки восстановления $\Delta F_T = F_T - \hat{G}_T$, $\Delta F_H = F_H - \hat{G}_H$ и восстановленной шумовой составляющей \hat{G}_N .

Пусть далее $G_\alpha = \alpha \cdot G = \alpha \cdot (G_T + G_H + G_N)$. Тогда $A_\alpha = \alpha \cdot A$ и $S_\alpha = A_\alpha \cdot A_\alpha^T = (\alpha \cdot A) \cdot (\alpha \cdot A)^T = \alpha \cdot A \cdot \alpha \cdot A^T = \alpha^2 \cdot A \cdot A^T$ или $S_\alpha = \alpha^2 \cdot S$. Решением характеристического уравнения $S_\alpha - \mu_\alpha \cdot E = 0$ находят собственные числа $\{\mu_{\alpha k}\}_{k=1}^L$, причем $S_\alpha = \alpha^2 \cdot S$ и тогда $\alpha^2 \cdot S - \mu_\alpha \cdot E = 0 \rightarrow \mu_\alpha = \alpha^2 \cdot \mu$. При этом определяются левые $\{U_k\}_{k=1}^L$

и правые $\{\mathbf{V}_k\}_{k=1}^L$ векторы, формируя разложение $A_\alpha = \sum_k A_{\alpha k}$, $A_{\alpha k} = \sqrt{\alpha^2 \mu_k} \cdot \mathbf{U}_k \cdot \mathbf{V}_k^T = \alpha \cdot \sqrt{\mu_k} \cdot \mathbf{U}_k \cdot \mathbf{V}_k^T$. Определены матрицы $A_{\alpha T} = \sum_{k \in T} A_{\alpha k}$, $A_{\alpha H} = \sum_{k \in H} A_{\alpha k}$, $A_{\alpha N} = \sum_{k \in N} A_{\alpha k}$. После диагонального усреднения матриц $A_{\alpha T}$, $A_{\alpha H}$, $A_{\alpha N}$ получены восстановленные аддитивные составляющие $\hat{G}_{\alpha T}$, $\hat{G}_{\alpha H}$, $\hat{G}_{\alpha N}$, что позволяет оценить показатели ошибки восстановления $\Delta F_{\alpha T} = F_{\alpha T} - \hat{G}_{\alpha T}$, $\Delta F_{\alpha H} = F_{\alpha H} - \hat{G}_{\alpha H}$ и восстановленной шумовой составляющей $\hat{G}_{\alpha N}$.

Заметим, что $A_\alpha = \sum_k A_{\alpha k} = \sum_k \alpha \cdot A_k = \alpha \cdot \sum_k A_k$, т. е. $A_{\alpha T} = \alpha \cdot A_T$, $A_{\alpha H} = \alpha \cdot A_H$, $A_{\alpha N} = \alpha \cdot A_N$. Тогда становится очевидным, что $\Delta F_{\alpha T} = \alpha \cdot F_T - \alpha \cdot \hat{G}_T = \alpha \cdot (F_T - \hat{G}_T) = \alpha \cdot \Delta F_T$.

Аналогично $\Delta F_{\alpha H} = \alpha \cdot F_H - \alpha \cdot \hat{G}_H = \alpha \cdot F_H - \alpha \cdot \hat{G}_H = \alpha \cdot (F_H - \hat{G}_H)$. Соответственно, шумовая составляющая $\hat{G}_{\alpha N} = \alpha \cdot \hat{G}_N$. Поэтому по определению показателей погрешности восстановления, задаваемых характеристиками положения: математическое ожидание, оцениваемое средним (mean), медиана (med); характеристиками рассеяния: среднеквадратическое отклонение, оцениваемое стандартным отклонением s , максимум (max), минимум (min), диапазон (range), доказываем исходное утверждение.

Доказательство закончено.

Таким образом, можно утверждать об исследовании нормированного временного ряда, который рассчитывается по следующей формуле:

$$G_H = \beta \frac{G - \text{mean}(G)}{\alpha} + \gamma,$$

где $\text{mean}(G)$ – среднее элементов x_i , $i = 0, \dots, n-1$, временного ряда G :

$$\text{mean}(G) = \frac{1}{n} \sum_{i=0}^{n-1} x_i;$$

$1/\alpha$ – коэффициент сжатия/растяжения, при этом параметр α определяется наибольшим по модулю значением элемента ряда $G - \text{mean}(G)$:

$$\alpha = \max_{x_i \in (G - \text{mean}(G))} |G - \text{mean}(G)|;$$

β – нормирующий параметр масштаба, определяемый верхней границей исследуемых значений элементов нормированного временного ряда, например $\beta = 1$ или $\beta = 10$;

γ – параметр сдвига элементов временного ряда в область положительных вещественных чисел, $\gamma \geq \beta$.

Методика приведения исходного ряда G к нормированному виду состоит из следующих шагов.

Шаг 1. Находится величина $\text{mean}(G)$ временного ряда.

Шаг 2. Определяется разность между значениями элементов ряда G и его средним $\text{mean}(G)$.

Шаг 3. Находится параметр α как максимальное по модулю значение из элементов ряда $G - \text{mean}(G)$, полученных на шаге 2.

Шаг 4. Полученные на шаге 2 элементы ряда $G - \text{mean}(G)$ умножаются на коэффициент $1/\alpha$.

Шаг 5. Значения, полученные на шаге 4, умножаются на нормирующий параметр β , в данном примере принимается $\beta = 1$.

Шаг 6. Полученные значения на шаге 5 смещаются на величину параметра сдвига $\gamma = \beta$. Получен нормированный временной ряд (рисунок 5.22), предназначенный для сингулярного спектрального анализа [76].

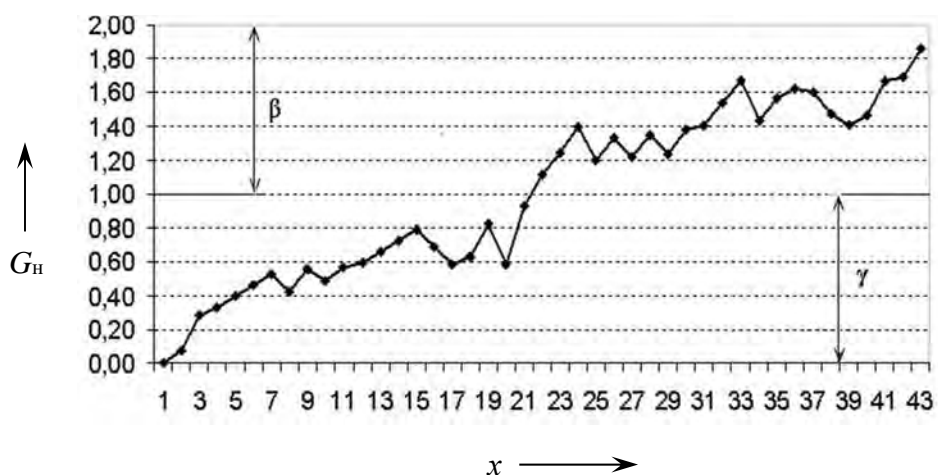


Рисунок 5.22 – Нормированный временной ряд

Таким образом, в результате проведенных исследований получено аналитическое выражение для нормированного временного ряда, включающее среднее элементов временного ряда, наибольшее по модулю значение элемента ряда, образованного разностью между элементами исходного ряда и средним его элементов, параметр масштаба, определяемый верхней границей нормированного временного ряда, параметр сдвига.

Аналитическое выражение для нормированного временного ряда позволяет независимым исследователям унифицировать результаты реальных исследований и создать базу знаний для выявления полезной информации с учетом особенностей применения метода сингулярного спектрального анализа.

6 Реализация методов интеллектуального анализа данных в программном комплексе BelSim

6.1 Программный модуль для оценки кривой плотности распределения шумовой составляющей

6.1.1 Структура и информационные потоки программного модуля.

Программный модуль (ПМ) расчета кривой плотности распределения случайной составляющей в последовательности данных BelSim2#.random предназначен для построения кривой плотности обобщенного распределения, описывающей представленную выборку. Предназначен для автономного режима работы и в составе программно-технологического комплекса имитации сложных систем BelSim2#.

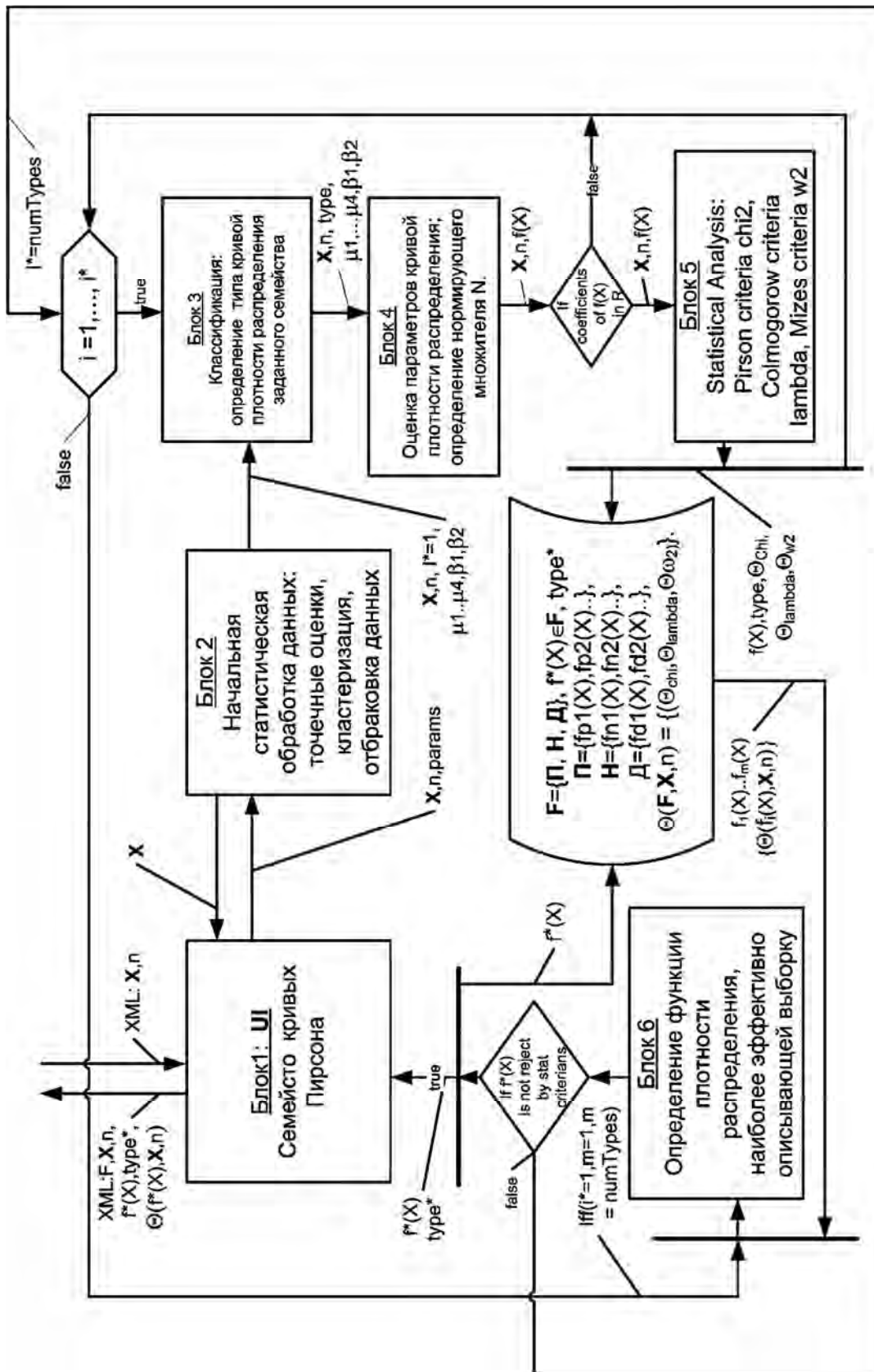
Соответствие теоретического распределения закону распределения с исследуемой плотностью проверяется по критериям Колмогорова–Смирнова, χ^2 , w^2 Мизеса. Семейство кривых Пирсона включает семь основных типов и три частных случая распределений: равномерное, нормальное, экспоненциальное [83].

Блок 1 на структурной схеме (рисунок 6.1) представляет интерфейс пользователя с возможностью расширения ПМ для использования иных обобщенных распределений, например, Джонсона, В. В. Нешистого. В блоке 2 реализуется начальная статистическая обработка данных: определяются точечные оценки, выполняется кластеризация и отбраковка данных. Блок 3 предназначен для определения типа кривой плотности распределения семейства Пирсона. В блоке 4 выполняется оценка параметров кривой плотности распределения, определяется нормирующий множитель. В блоке 5 кривая плотности распределения проверяется по критериям Колмогорова–Смирнова, χ^2 , w^2 Мизеса. В блоке 6 определяется функция плотности распределения, наиболее эффективно описывающая выборку. Результаты расчетов представлены в виде аналитической зависимости.

6.1.2 Основные компоненты программного модуля.

В состав ПМ входит программное обеспечение (ПО), включающее из следующие компоненты: kb; algorithm.pirson; criteria.kolmogorov; criteria.w2; criteria.xi2; criteria.normal.

Главным компонентом программного модуля является **kb**. В нем находится графический интерфейс пользователя (GUI), интерфейсы для остальных подключаемых компонентов, функции доступа к XML-файлам.



GUI программного модуля реализован стандартным компонентом Swing, входящим в JRE для GUI настольных приложений. Доступ к XML-файлам реализован с использованием стандартного компонента Streaming API for XML, входящего в JRE.

Algorithm.pirson реализует алгоритм Пирсона построения кривой плотности распределения для заданной выборки. Реализован общий интерфейс для всех алгоритмов построения кривой плотности распределения. После выбора кривой на основе входных параметров алгоритм использует по умолчанию критерий Колмогорова–Смирнова для проверки соответствия рассчитанной кривой заданной выборке и на основе этих проверок выбирает кривую, которая наиболее точно соответствует заданной выборке.

Criteria.kolmogorov реализует критерий Колмогорова–Смирнова для проверки полученной кривой на соответствие выборке, поддерживает общий интерфейс, находящийся в главном компоненте программного модуля, для всех критериев проверки кривой.

Criteria.w2 реализует критерий w^2 Мизеса для проверки полученной кривой на соответствие выборке, поддерживает общий интерфейс, находящийся в главном компоненте программного модуля, для всех критериев проверки кривой.

Criteria.xi2 реализует критерий χ^2 для проверки полученной кривой на соответствие выборке, поддерживает общий интерфейс, находящийся в главном компоненте программного модуля, для всех критериев проверки кривой.

Criteria.normal реализует критерии для проверки функции нормального распределения критериями Колмогорова–Смирнова, w^2 и χ^2 для проверки полученной кривой на соответствие выборке. Критерии реализуют общий интерфейс для всех критериев проверки кривой (таблица 6.1).

Таблица 6.1 – Перечень файлов программного модуля BelSim2#.random

Название файла	Размер, Кбайт	Примечание
run.bat	0,431	Файл для запуска программного модуля
kb-0.0.1-SNAPSHOT.jar	41,963	Файл, содержащий графический интерфейс пользователя (GUI), интерфейсы для подключаемых компонентов, функции доступа к XML-файлам
criteria.xi2-0.0.1-SNAPSHOT.jar	7,818	Файл, содержащий проверку по критерию χ^2

Окончание таблицы 6.1

Название файла	Размер, Кбайт	Примечание
criteria.w2-0.0.1-SNAPSHOT.jar	4,614	Файл, содержащий проверку по критерию ω^2
criteria.normal-0.0.1-SNAPSHOT.jar	16,740	Файл, содержащий реализацию проверок нормальной функции
criteria.kolmogorov-0.0.1-SNAPSHOT.jar	4,285	Файл с реализацией проверки по критерию Колмогорова–Смирнова
algorithm.pirson-0.0.1-SNAPSHOT.jar	24,909	Файл, содержащий реализацию алгоритма Пирсона
Model.xml	>1	Файл данных
spring-expression-3.0.5.RELEASE.jar	169,752	Библиотека Spring framework
spring-core-3.0.5.RELEASE.jar	382,442	Библиотека Spring framework
spring-context-3.0.5.RELEASE.jar	668,861	Библиотека Spring framework
spring-beans-3.0.5.RELEASE.jar	555,410	Библиотека Spring framework
spring-asm-3.0.5.RELEASE.jar	53,082	Библиотека Spring framework
commons-math-2.1.jar	832,410	Библиотека для расчета математических функций
commons-logging-1.1.1.jar	60,686	Библиотека для логирования данных

6.1.3 Основные характеристики программного модуля.

Основные характеристики ПМ представлены перечнем файлов, их размером и назначением в таблице 6.1.

Требования к операционной системе, оперативной памяти, интегрированной среде разработки представлены в таблице 6.2.

Таблица 6.2 – Технические требования ПМ BelSim2#.random

Операционная система	Windows 2000/XP и более поздние версии
Пространство на жестком диске	500 Мбайт
Оперативная память	512 Мбайт
Интегрированная среда разработки приложений на языке Java	IDE Eclipse 3.6.2
Общезыковая исполняющая среда и библиотека классов	JRE 6
Дата создания первой версии	2010 г.

Структура XML-файла Model.xml следующая:

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
  <model>
    <data>
      array>
        <element> $g_1$ </element>
        <element> $g_2$ </element>
        ...
        <element> $g_{|G|}$ </element>
      </array>
      <firstItem>1</firstItem>
      <length> $|G|$ </length>
    </data>
    <parameters>
      <isNearestTypesForcedVerify>false</isNearestTypesForcedVerify>
      <isNormVerify>true</isNormVerify>
      <isExpVerify>true</isExpVerify>
      <isRectVerify>true</isRectVerify>
      <ConfidenceLevel>0.05</ConfidenceLevel>
      <integralEps>0.01</integralEps>
    </parameters>
    <responses>
      <StatPointsDataEstimations />
      <FunctionByClussification />
      <NearestFunctions />
      <BestFunction />
    </responses>
  </model>
```

Файл с данными <model> содержит следующие теги: <data> для описания данных g_i , $i = 1, \dots, |G|$, временной последовательности G ; <parameters> для указания параметров модели данных; <responses> для представления откликов модели данных.

Тег данных модели <data> включает следующие теги:

<array> (тип <element>) – множество данных (выборка);

<firstItem> (тип Integer) – номер первого элемента, начиная с которого ведется отсчет элементов множества, участвующих в статистической обработке;

<length> (тип Integer) – количество элементов множества, участвующих в статистической обработке.

В теге данных модели <data> тег <array> состоит из тега <element>, описывающего элемент выборки, предназначенный для исследований.

Тег данных модели <parameters> включает управляющие параметры модели:

<isNearestTypesForcedVerify> (тип Boolean) – признак необходимости принудительной проверки всех ближайших типов кривых к заданной кривой (значение по умолчанию: False);

<isNormVerify> (тип Boolean) – признак необходимости принудительной проверки нормального распределения (значение по умолчанию: True);

<isExpVerify> (тип Boolean) – признак необходимости принудительной проверки экспоненциального распределения (значение по умолчанию: True);

<isRectVerify> (тип Boolean) – признак необходимости принудительной проверки равномерного распределения (значение по умолчанию: True);

<ConfidenceLevel> (тип Double) – уровень значимости применяемых статистических критериев (значение по умолчанию: $p = 0,05$);

<integralEps> (тип Double) – точность, с которой вычисляется интеграл при расчете нормирующего множителя кривой (значение по умолчанию: $\Delta = 0,05$).

Тег данных модели <responses> включает список тегов откликов, которые необходимо сохранить в файле модели (таблица 6.3).

Таблица 6.3 – Тег данных модели <responses>

Наименование тега	Комментарий
<StatPointsDataEstimations>	Признак необходимости сохранения в файл модели точечных оценок выборки
<FunctionByClussification>	Признак необходимости сохранения в файл модели функции f_0 плотности распределения выборки, построенной в соответствии с классификацией Пирсона по показателям (β_1, β_2)
<NearestFunctions>	Признак необходимости сохранения в файл модели всех построенных функций f_i плотностей распределения выборки ближайших типов по отношению к кривой функции плотности f_0 , построенной в соответствии с классификацией. Максимально возможное количество таких кривых f_i : шесть типов кривых Пирсона и три частных случая распределений (нормальное, экспоненциальное, равномерное)

Окончание таблицы 6.3

Наименование тега	Комментарий
<BestFunction>	Признак необходимости сохранения в файл модели функции плотности распределения f^* , наилучшим образом описывающей выборку

При наличии тегов <StatPointsDataEstimations>, <FunctionByClassification>, <NearestFunctions>, <BestFunction> в <responses> соответствующие им отклики сохраняются в файле модели:

```

<StatPointsDataEstimations>
<betta1>0.2174979780360239</betta1>
<betta2>4.359162966282781</betta2>
<dispersion>0.05298888888888884</dispersion>
<eps>0.1</eps>
<gamma3>0.4663667848764788</gamma3>
<gamma4>7.359162966282781</gamma4>
<maximum>6.2</maximum>
<minimum>5.0</minimum>
<mean>5.536666666666667</mean>
<mu2>0.05298888888888884</mu2>
<mu3>0.005688592592592536</mu3>
<mu4>0.01223975518518518</mu4>
<range>1.2000000000000002</range>
<sigma>0.2301931556082605</sigma>
<xi>0.08492929437067284</xi>
</StatPointsDataEstimations>

```

Отклики, выводимые в тег <StatPointsDataEstimations>, являются точечными оценками выборки, представленными в таблице 6.4.

Тег <function> имеет следующую структуру:

```

<function>
  <functionType>PirsonType4</functionType>
  <functionForm>y=-1.08220623958719152E17*(1+(x-7.247244663002339)^
2/0.39395264681423514)^(-5.562324028177368)*Exp(-30.31846086482367*
atan((x-7.247244663002339)/0.6276564719766977))</functionForm>
  <Kolmogorov>
  <NablValue>4.466884141908774</NablValue>

```



```

<KritValue>1.073</KritValue>
<DLevel>4.162986152757478</DLevel>
<BLevel>false</BLevel>
</Kolmogorov>
</function>

```

Таблица 6.4 – Отклики, выводимые в тег <StatPointsDataEstimations>

Наименование тега	Тип данных	Комментарий
<betta1>	Double	Показатель асимметрии распределения, введенный Пирсоном, $\beta_1 = \frac{\mu_3^2}{\mu_2^3}$
<betta2>	Double	Показатель эксцесса, введенный Пирсоном, $\beta_2 = \frac{\mu_4}{\mu_2^2}$
<dispersion>	Double	Дисперсия D
<eps>	Double	Точность, с которой проводится классификация типа кривой Пирсона по показателям (β_1, β_2)
<gamma3>	Double	Показатель асимметрии распределения в соответствии со Statistica $\gamma_3 = \sqrt{\beta_1}$
<gamma4>	Double	Показатель эксцесса в соответствии со Statistica $\gamma_4 = \beta_2 + 3$
<maximum>	Double	Минимальное значение выборки min
<minimum>	Double	Максимальное значение выборки max
<mean>	Double	Математическое ожидание mean
<mu2>	Double	Центральный момент выборки второго порядка μ_2
<mu3>	Double	Центральный момент выборки третьего порядка μ_3
<mu4>	Double	Центральный момент выборки четвертого порядка μ_4
<range>	Double	Размах range = max – min
<sigma>	Double	Среднеквадратическое отклонение $\sigma = \sqrt{D}$
<xi>	Double	Показатель, введенный Пирсоном для определения типа кривой плотности распределения, $\aleph = \frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 6)(4\beta_2 - 3\beta_1)}$

Тег `<function>` содержит отклики, выводимые в теги `<FunctionByClassification>`, `<NearestFunctions>`, `<BestFunction>`. Тег `<function>` содержит тег `<functionType>`, описывающий тип функции плотности распределения; `<functionForm>` описывает функцию плотности распределения; в теге `<Kolmogorov>` представлен результат проверки критерием Колмогорова–Смирнова гипотезы о распределении выборки по заданному закону распределения.

Тег `<Kolmogorov>` содержит сведения, представленные в таблице 6.5.

Таблица 6.5 – Сведения, содержащиеся в теге `<Kolmogorov>`

Наименование тега	Тип данных	Комментарий
<code><NablValue></code>	Double	Наблюдаемое значение критерия Колмогорова–Смирнова L_{nabl}
<code><KritValue></code>	Double	Критическое значение критерия Колмогорова–Смирнова L_{krit}
<code><DLevel></code>	Double	Действительный результат работы критерия – отношение наблюдаемого значения критерия к критическому: $D_{level} = \frac{L_{nabl}}{L_{krit}}$
<code><BLevel></code>	Boolean	Логический результат работы критерия $B_{level} = \text{iif}(D_{level} < 1; \text{true}; \text{false})$

Выбор кривой плотности распределения f^* , наилучшим образом описывающей выборку, выполняется на множестве тех кривых, которые не отклонены статистическим критерием ($B_{level} = \text{true}$). Для кривой плотности распределения, наилучшим образом описывающей выборку, значение D_{level} минимально.

6.2 Методика использования программного модуля

Методика использования ПМ для построения кривой семейства Пирсона, соответствующей эмпирическому распределению выборки исследуемых данных, состоит из этапов, описанных в подразд. 4.4.

Для указания алгоритма построения кривой плотности распределения и критериев проверки соответствия полученной кривой заданной выборке используется XML-файл. При этом используется Spring версии 3.0.5.RELEASE, который позволяет изменить реализацию алгоритма построения кривой плотности распределения и критерия проверки соответствия кривой заданной выборке без перекомпиляции программы.

Изменения выполняются редактированием параметров в XML-файле. По умолчанию заданы алгоритм Пирсона и критерий Колмогорова–Смирнова.

Программный модуль допускает расширение добавлением новых критериев или новых алгоритмов. Для этого достаточно добавить новый компонент, реализовать в нем интерфейс либо для алгоритма, либо для критерия и прописать в конфигурационном XML-файле использование созданного компонента в соответствующем качестве.

Пример – Технология применения SSA-метода и программного средства для определения типа шумовой составляющей.

Пусть числовая последовательность данных получена в результате имитационных экспериментов:

$G = \langle 95,0; 95,6; 97,1; 97,5; 98,0; 98,5; 99,0; 98,2; 99,2; 98,7; 99,3; 99,5; 100,0; 100,5; 101,0; 100,2; 99,4; 99,8; 101,2; 99,4; 102,0; 103,4; 104,4; 105,5; 104,0; 105,0; 104,2; 105,2; 104,3; 105,4; 105,6; 106,6; 107,6; 105,8; 106,8; 107,2; 107,1; 106,1; 105,6; 106,0; 107,6; 107,7; 109,0 \rangle$.

Задача построения модели последовательности числовых данных решается с применением программного модуля BelSim2#.random. Рассчитаны точечные оценки выборки G , в частности: $\min = 95,00$, $\max = 109,00$, $v_1 = 102,54$, $\mu_2 = 14,295$, $\mu_3 = -7,698$, $\mu_4 = 355,460$, $\beta_1 = 0,020$, $\beta_2 = 1,74$, $\aleph = -0,0064$ (с заданной точностью $\varepsilon = 0,1$ показатель $\aleph = 0$). Построены функции плотностей распределений $f_5(x)$, $f_7(x)$, $R(x)$, $E(x)$, $N(x)$.

Результат работы программы BelSim2#.random следующий:

$$f_5(x) = 6,244E30 * (x - 88,500)^{-20,020} * \exp(-281,084 / (x - 88,500));$$

$$f_7(x) = 0,114 * \exp(-(x - 102,540)^2 / 28,590);$$

$$R(x) = 0,071;$$

$$E(x) = 9,752E-3 * \exp(-1 * 9,752E-3 * x);$$

$$N(x) = 1 / (3,781 * (2 * \pi)^{(1/2)}) * \exp(-1 * (x - 102,54)^2 / 28,590).$$

Функции плотностей распределений $f_5(x)$, $N(x)$, $E(x)$ отклонены после проверки критериями χ^2 , λ , ω^2 , для отклонения гипотез о функциях $f_7(x)$, $R(x)$ нет оснований. Для функции $R(x)$ значение евклидовой метрики (4.104) отношений наблюдаемых значений критериев к критическим значениям минимально, поэтому принимается соотношение $f^*(x) = R(x)$.

Исследование внутренней структуры числовой последовательности G выполнено с помощью SSA-метода. При этом числовая последовательность G приводится к нормированному виду следующими шагами (методика реализована в программном модуле BelSim2#.SSA).

Шаг 1. Определяется среднее значение числовой последовательности G : $\text{mean} = 102,54$.

Шаг 2. Определяется разность между элементами числовой последовательности и ее средним значением: $G - \text{mean}$.

Шаг 3. Определяется максимальное по абсолютной величине значение из числовой последовательности $G - \text{mean}$: $\text{max} = 7,54$.

Шаг 4. Полученные на шаге 2 элементы числовой последовательности $G - \text{mean}$ делят на величину max (см. шаг 3).

Шаг 5. Элементы числовой последовательности $(G - \text{mean})/\text{max}$ (см. шаг 4) умножаются на нормирующий множитель β , равный единице или десяти (принимается $\beta = 1$).

Шаг 6. Элементы последовательности $\beta \cdot (G - \text{mean})/\text{max}$ (см. шаг 5) смещаются на значение параметра сдвига γ (в данном примере $\gamma=1$).

Шаг 7. Применяя SSA-метод к нормированной числовой последовательности G_n (см. шаг 6), строится ее модель разложением на аддитивные составляющие: трендовую, периодическую и шумовую.

Восстановленный тренд G_{T_B} для нормированной числовой последовательности G_n представлен на рисунке 6.2.

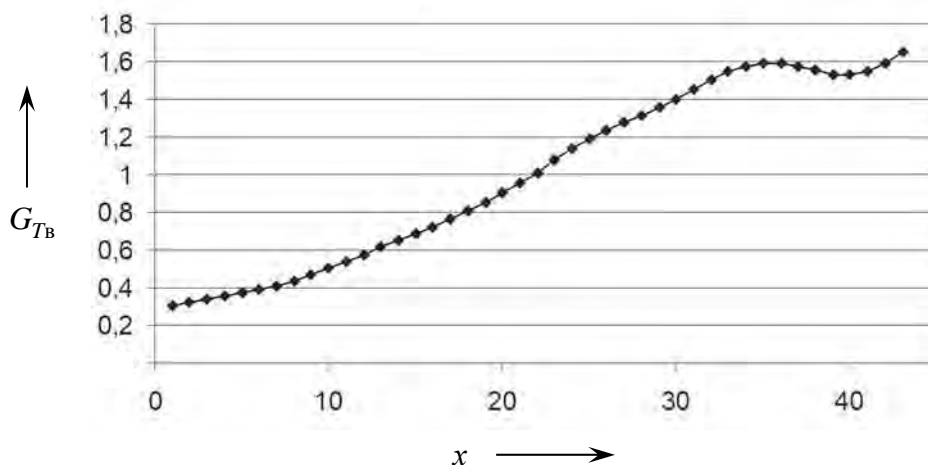


Рисунок 6.2 – Восстановленный тренд нормированной последовательности

Для трендовой составляющей G_{T_B} может быть получено аналитическое выражение в виде регрессионной модели с применением математического пакета Statistica.

Восстановленная периодическая составляющая $G_{HВ}$ для нормированной числовой последовательности G_H представлена на рисунке 6.3.

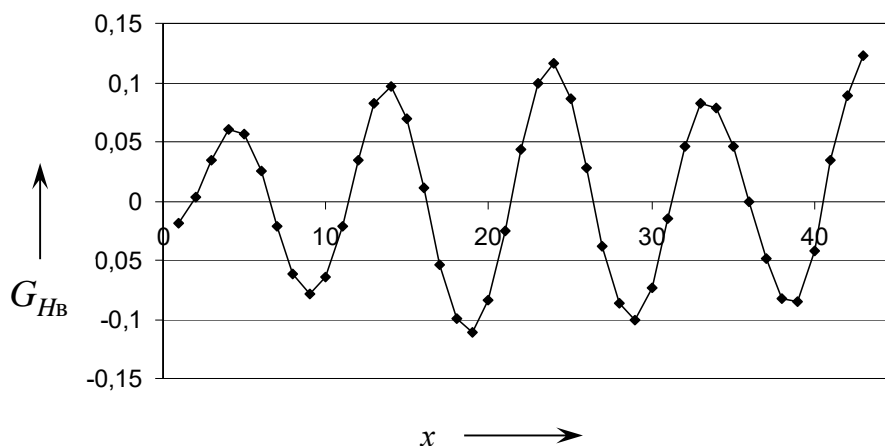


Рисунок 6.3 – Восстановленная периодическая составляющая

Аналитическая модель восстановленной периодической составляющей $G_{HВ}$ может быть построена в математическом пакете Mathcad с использованием дискретного преобразования Фурье.

Закон распределения восстановленной шумовой составляющей $G_{NВ}$ определяется программным модулем BelSim2#.random. При этом рассчитаны точечные оценки выборки $(\beta(G - \text{mean})/\text{max}) + \gamma$ (см. шаг 6): $\min = -0,194$, $\text{max} = 0,136$, $\nu_1 = -0,0104$, $\mu_2 = 0,0056$, $\mu_3 = -1,123$, $\mu_4 = 8,743$, $\beta_1 = 0,072$, $\beta_2 = 2,796$, $\aleph = -0,0884$ (с заданной точностью $\varepsilon = 0,1$ показатель $\aleph = 0$). Построены функции плотностей распределений $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_5(x)$, $f_7(x)$, $R(x)$, $E(x)$, $N(x)$.

Итоги работы программы BelSim2#.random следующие:

$$\begin{aligned}
 f_1(x) &= 5,077 * (1 + (x - 0,0104) / 0,321)^{7,141} * (1 - (x - 0,0104) / 0,333)^{7,426}; \\
 f_2(x) &= 5,346 * (1 - (x - 0,0104) / 0,153)^{12,206}; \\
 f_3(x) &= 5,524 * \exp(99,60 * (x - 0,0104)) * (1 + (x - 0,0104) / (-0,547))^{54,47}; \\
 f_5(x) &= 0,024 * (x - 0,289)^{-20,072} * \exp(-5,584 / (x - 0,289)); \\
 f_7(x) &= 5,512 * \exp(-(x - 0,0104)^2 / 0,0112); \\
 R(x) &= 3,032; \\
 E(x) &= -96,19 * \exp(96,19 * x); \\
 N(x) &= 1 / (0,075 * (2 * \text{Pi})^{(1/2)}) * \exp(-1 * (x - 0,0104)^2 / 0,0112).
 \end{aligned}$$

Функции $f_1(x)$, $E(x)$ отклонены критериями χ^2 , λ , ω^2 , для отклонения функций плотностей распределений $f_2(x)$, $f_3(x)$, $f_5(x)$, $f_7(x)$, $R(x)$, $N(x)$ нет оснований. Для функции $f_7(x)$ значение евклидовой метрики (4.104) отношений наблюдаемых значений критериев к критическим значениям минимально, поэтому является справедливым соотношение $f^*(x) = f_7(x)$.

Таким образом, определена внутренняя структура исследуемой последовательности числовых данных, включающая трендовую (см. рисунок 6.2), периодическую (см. рисунок 6.3) и шумовую составляющие. Полученный результат представляет исследователю имитационной модели (в сравнении со статистическими методами) значительно больше дополнительной информации, используемой для принятия решения. Тем самым достигается повышение эффективности применения имитационных моделей в реальных производственных условиях.

6.3 Кластеризация целевых функций

Идея кластеризации целевых функций. Применение генетического алгоритма (ГА) в BelSim Optimizer для поиска рациональных решений предполагает настройку алгоритма оптимизации (АО) в зависимости от типа используемой целевой функции (ЦФ). При этом для некоторых целевых функций параметры ГА, обеспечивающие минимальное время проведения экспериментов с моделью и высокую точность рациональных результатов, остаются неизменными [15]. Говоря иначе, АО для групп целевых функций ведут себя одинаково.

Пусть исследователь располагает множеством $\{CF_i\}, i = 1, \dots, n$, целевых функций и АО с m параметрами (например, вероятность кроссовера и мутации, доля родителей от размера популяции и др.) и r откликами (лучшее и худшее решение поколения, среднее значение приспособленности поколения и др.). Изменения откликов АО при варьировании значений параметров, т. е. чувствительность АО, для определенных ЦФ отличаются между собой. Требуется разбить совокупность ЦФ на непересекающиеся подмножества (классы), в пределах которых АО с характеристиками по чувствительности эквивалентны.

Для разбиения совокупности объектов на классы эквивалентности с определенным свойством применяют различные методы кластерного анализа, в основу которых могут быть положены эвристические методы перебора, методы математического программирования, использование матриц сходства, статистические распределения и др.

Предложена методика определения классов целевых функций с инвариантным поведением алгоритмов оптимизации. В соответствии с методикой применяют несколько разных методов кластерного анализа, которые отличаются методологическими подходами к разбиению множества объектов на кластеры (классы) для проверки и уточнения результатов кластеризации. Вначале разбиение выполняют двумя методами, например K-means и Tree Clustering. При несовпадении результатов разбиения дополнительно применяют третий метод, например Fuzzy Relation Clustering (FRC), для уточнения кластеризации.

В итоге разбиения кластеризуемого множества каждым из методов образуется в общем случае разное количество классов объектов и, следовательно, разное соответствие индексации объектов множества и полученных классов. Поэтому требуется обобщение итогов кластеризации для определения классов ЦФ, в пределах которых АО эквивалентны.

Методика определения классов целевых функций в ПТКИ BelSim

Последовательность шагов определения классов ЦФ следующая.

Шаг 1. Предварительный анализ АО.

Для каждой тестовой ЦФ выполняется исследование чувствительности откликов АО при варьировании значений параметров АО. В результате анализа чувствительности формируются матрицы

$$S = \|s_{ij}\|, s_{ij} \in R, i = 1, \dots, r, j = 1, \dots, m, \quad (6.1)$$

где s_{ij} – чувствительность i -го отклика АО при варьировании значений j -го параметра.

Для последующего кластерного анализа матрицы S вида (6.1) преобразуются к вектору:

$$S = (s_{11}, s_{12}, \dots, s_{1m}, \dots, s_{r1}, \dots, s_{rm}). \quad (6.2)$$

Анализ чувствительности проводится в общем случае t раз для каждой ЦФ. Поэтому в итоге получают $t \cdot n$ векторов вида (6.2), представленных множеством

$$\{S_i\}, i = 1, \dots, t \cdot n. \quad (6.3)$$

Шаг 2. Кластерный анализ.

При разбиении множества векторов (6.3) на классы используют ряд методов кластерного анализа для контроля и уточнения результатов.

В итоге множество (6.3) разбивается на k классов каждым из методов кластеризации. При этом индексам $i = 1, \dots, t \cdot n$ элементов множества (6.3) поставлены в соответствие номера классов K_j , $j = 1, \dots, k$.

Шаг 3. Интерпретация результатов кластерного анализа.

Векторы $S_i \in R^{m \cdot r}$, $i = 1, \dots, t \cdot n$, кластеризуемого множества (6.3) – результаты обработки статистических данных после экспериментов с n ЦФ. Поэтому результат кластеризации множества (6.3) может быть представлен матрицей относительных частот принадлежности ЦФ определенным классам

$$P = \|p_{ij}\|, \quad i = 1, \dots, n, \quad j = 1, \dots, k, \quad (6.4)$$

где i – номер ЦФ;

j – номер класса K_j ;

k – количество классов;

$p_{ij} \in [0, 1]$ – относительная частота принадлежности i -й ЦФ классу K_j .

Величины p_{ij} в (6.4) рассчитываются как отношение количества попаданий ЦФ CF_i в класс K_j к общему количеству экспериментов с исследуемой ЦФ.

Шаг 4. Обобщение итогов кластеризации разными методами.

Определение. ЦФ CF_i принадлежит классу K_j ($CF_i \in K_j$) тогда и только тогда, когда CF_i отнесена к данному классу K_j по меньшей мере двумя из трех используемых методов кластеризации.

Относительные частоты отнесения ЦФ к определенным классам одновременно тремя используемыми методами кластеризации ниже, чем относительные частоты отнесения ЦФ к определенным классам одним из используемых методов. Поэтому принята аналогия голосования большинством голосов – ЦФ относится к данному классу по меньшей мере двумя из трех используемых методов кластеризации.

Утверждение. Пусть P_1, P_2, P_3 – матрицы вида (6.4) относительных частот принадлежности ЦФ CF_i , $i = 1, \dots, n$, определенным классам в соответствии с тремя разными методами кластерного анализа. Тогда величины $p_{ij} \in [0, 1]$ обобщенной (по определению 1, см. подразд. 4.3)

матрицы P находятся суммированием коэффициентов $P_{2,3}$, $P_{3,3}$ производящей функции

$$\varphi(z) = \prod_{l=1}^3 (q_{lij} + p_{lij} \cdot z) = \sum_{v=0}^3 P_{v,3} \cdot z^v,$$

где $q_{lij} = 1 - p_{lij}$, $p_{lij} \in [0, 1]$, $l = 1, \dots, 3$, $i = 1, \dots, n$, $j = 1, \dots, k$ – элементы матриц P_1, P_2, P_3 , соответственно,

$$p_{ij} = (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}. \quad (6.5)$$

Доказательство утверждения. Пусть выполняются три независимых испытания, под которыми будем понимать кластерный анализ одним из трех разных методов.

Кластерный анализ проводится в одинаковых условиях с одними исходными данными и результаты разных методов кластеризации не зависят между собой.

Пусть событие A – попадание ЦФ в определенный класс в результате проведенного испытания. В каждом испытании событие A появляется со своей относительной частотой.

Введем обозначения:

p_1, p_2, p_3 – относительные частоты попадания ЦФ в некоторый класс в первом, втором, третьем испытании соответственно;

$q_1 = 1 - p_1$, $q_2 = 1 - p_2$, $q_3 = 1 - p_3$ – относительные частоты непопадания ЦФ в некоторый класс в первом, втором, третьем испытании соответственно.

Относительная частота появления события A ровно 0, 1, 2, 3 раза в указанных испытаниях равна коэффициенту при соответствующей степени z производящей функции

$$\varphi(z) = \prod_{l=1}^3 (q_l + p_l \cdot z);$$

$$\begin{aligned} \varphi(z) &= (q_1 + p_1 \cdot z)(q_2 + p_2 \cdot z)(q_3 + p_3 \cdot z) = q_1 q_2 q_3 + (p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3)z + \\ &+ (q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3)z^2 + p_1 p_2 p_3 z^3 = \sum_{v=0}^3 P_{v,3} \cdot z^v. \end{aligned}$$

Таким образом, относительная частота того, что в этих трех испытаниях событие A не появится ни разу, $P_{0,3} = q_1 q_2 q_3$; появится ровно один раз –

$$\begin{aligned}
P_{1,3} &= p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3; \text{ появится ровно два раза из трех} - \\
P_{2,3} &= q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3; \text{ появится ровно три раза из трех} - \\
P_{3,3} &= p_1 p_2 p_3.
\end{aligned}$$

Тогда относительная частота события Y – попадания ЦФ в некоторый класс по меньшей мере в двух испытаниях из трех – $p(Y) = P_{2,3} + P_{3,3}$, или

$$p(Y) = q_1 p_2 p_3 + p_1 q_2 p_3 + p_1 p_2 q_3 + p_1 p_2 p_3. \quad (6.6)$$

В итоге кластеризации множества (6.3) тремя методами получены матрицы относительных частот принадлежности ЦФ определенным классам P_1, P_2, P_3 вида (6.4) с элементами $p_{lij}, l = 1, \dots, 3$,

$$P_l = \|p_{lij}\|, p_{lij} \in [0, 1], i = 1, \dots, n, j = 1, \dots, k, l = 1, \dots, 3.$$

Для обобщения результатов кластеризации тремя разными методами, применяя к элементам $p_{lij}, l = 1, \dots, 3$, матриц P_1, P_2, P_3 формулу (6.6), получим следующую обобщенную матрицу:

$$P = \|p_{ij}\|, p_{ij} \in [0, 1], i = 1, \dots, n, j = 1, \dots, k,$$

элементы которой вычисляются по формуле

$$\begin{aligned}
p_{ij} &= (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}, \\
i &= 1, \dots, n, j = 1, \dots, k.
\end{aligned} \quad (6.7)$$

Доказательство закончено.

Замечание 1. ЦФ должна попадать хотя бы в один класс, поэтому сумма элементов строк обобщенной матрицы P должна быть равна единице. Тогда в общем случае элементы обобщенной матрицы P относительных частот принадлежности ЦФ $CF_i, i = 1, \dots, n$, определенным классам $K_j, j = 1, \dots, k$, вычисляются по формуле

$$p_{ij} = \begin{cases} (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}, & j < k; \\ 1 - \sum_{j=1}^{k-1} p_{ij}, & j = k. \end{cases} \quad (6.8)$$

Доказательство замечания 1. Пусть события C_1, C_2, \dots, C_k – попадание ЦФ $CF_i, i=1, \dots, n$, в класс K_1, K_2, \dots, K_k соответственно.

При отнесении методами кластерного анализа элементов множества (6.3) к определенному классу ЦФ попадет хотя бы в один из k классов (произойдет одно из событий $C_j, j=1, \dots, k$). В таком случае события $C_j, j=1, \dots, k$, образуют полную группу событий с выполнением условия

$$p(C_1) + p(C_2) + \dots + p(C_k) = 1. \quad (6.9)$$

Тогда элементы p_{ij} матрицы P при $j=1, \dots, k-1, i=1, \dots, n$ могут быть вычислены по формуле (6.7); при $j=k, i=1, \dots, n$, с учетом (6.9), – по формуле (6.8).

Доказательство закончено.

Замечание 2. Матрицы P_1, P_2, P_3 для обобщения должны иметь одинаковую размерность. При разбиении множества (6.3) на $k_1 \leq k, k_2 \leq k, k_3 \leq k$ классов необходимо матрицы P_1, P_2, P_3 привести к размерности $k = \max\{k_1, k_2, k_3\}$. Матрицы приводятся к требуемой размерности k дополнением столбцами с нулевыми относительными частотами попадания объекта в дополнительные классы.

Шаг 5. Определение классов ЦФ.

В итоге кластеризации множества (6.3) тремя разными методами построена обобщенная матрица $P = \|p_{ij}\|, i=1, \dots, n, j=1, \dots, k$, анализ которой позволяет определить принадлежность ЦФ определенному классу. ЦФ CF_i принадлежит классу K_{j_0} тогда, когда относительная частота принадлежности ЦФ классу K_{j_0} в построенной обобщенной матрице P максимальна.

Пример построения классов целевых функций. Пусть приняты тестовые ЦФ $CF_i, i=1, \dots, n$ ($n=5$): CF_1 – Easom, CF_2 – Six-hump, CF_3 – Branin RCos, CF_4 – Modified Branin RCos, CF_5 – Goldstein&Price.

Алгоритм оптимизации – генетический алгоритм (ГА).

Параметры ГА: X_1 – вероятность кроссовера; X_2 – вероятность мутации; X_3 – доля родителей от размера популяции; X_4 – математическое ожидание для распределения количества точек разрыва. Количество параметров АО $m=4$.

Отклики ГА: Y_1 – лучшее решение поколения; Y_2 – худшее решение поколения; Y_3 – среднее значение приспособленности поколения;

Y_4 – дисперсия приспособленности поколения; Y_5 – лучшее решение; Y_6 – количество запусков функции качества. Таким образом, количество откликов АО с использованием ГА $r = 6$.

Значения параметров ГА заданы на трех уровнях (таблица 6.6), где <"-1"> означает "-1"-й уровень эксперимента, <"0"> – соответственно "0"-й, <"1"> – "1"-й уровень эксперимента.

Таблица 6.6 – Значения параметров АО на уровнях эксперимента

Уровень / параметр	X_1	X_2	X_3	X_4
<"-1">	0,01	0,01	0,5	2
<"0">	0,5	0,5	0,75	5
<"1">	0,99	0,99	1	8

Необходимо определить классы, к которым принадлежат целевые функции CF_1 – Easom, CF_2 – Six-hump, CF_3 – Branin RCos, CF_4 – Modified Branin RCos, CF_5 – Goldstein&Price [20].

Для решения поставленной задачи в ПТКИ BelSim выполняются шаги методики определения классов целевых функций.

Элементы S_i множества (6.3), полученные в результате предварительного анализа АО на шаге 1 методики, представлены $4 \cdot 6 = 24$ -мерными векторами действительных чисел. Для каждой из пяти ЦФ проведен $t = 31$ эксперимент по анализу чувствительности, общее количество элементов S_i множества (6.3) составляет $|\{S_i\}| = 5 \cdot 31 = 155$.

Для разбиения на классы множества векторов (6.3) (см. шаг 2 методики) приняты методы кластеризации K-means, Tree Clustering и FRC. Кластерный анализ методами K-means и Tree Clustering выполняется с использованием математического пакета Statistica. Метод кластеризации FRC реализован в составе ПТКИ BelSim в программном модуле BelSim Cluster. Гипотеза относительно числа классов $k = 3$.

Метод K-means делит кластеризуемые данные ровно на $k = 3$ класса. По запросу пакет Statistica предлагает построить график средних значений для каждого класса (Graph of means); вывести отдельно члены каждого класса (members of each cluster& distances) и др.

Полная классификация (Save classifications and distances), или соответствие номерам $i = 1, \dots, m$ элементов множества (6.3) номеров классов K_j , $j \in 1, \dots, k$, сохраняется для последующего анализа.

При использовании метода Tree Clustering в пакете Statistica построены дендрограммы (Horizontal hierarchical tree plot), матрицы расстояний D (Distance matrix), матрицы Amalgation Schedule. Полную классификацию получают в результате анализа данных матрицы расстояний D и матрицы Amalgation Schedule.

Для использования метода кластеризации FRC данные передаются в модуль кластеризации многомерных данных BelSim Cluster в составе ПТКИ BelSim. Задается значение параметра $\alpha \in [0;1]$ метода FRC. С учетом варьирования значений параметра $\alpha \in [0;1]$ метод FRC позволяет построить ожидаемое количество классов. Полная классификация, построенная методом FRC, сохраняется для последующего анализа.

Используемые методы кластеризации разбили множество (6.3) на три класса.

В соответствии с методикой определения классов ЦФ разбиение выполняется двумя методами: K-means и Tree Clustering. Интерпретированные результаты кластеризации элементов множества (6.3) (см. шаг 3 методики) – матрицы P_1, P_2 относительных частот принадлежности ЦФ CF_1, \dots, CF_5 классам $K_j, j = 1, \dots, 3$, в соответствии с методами K-means и Tree Clustering – приведены далее. Элементы матриц P_1, P_2 рассчитаны как отношение количества попаданий элементов множества (6.3) в определенный класс к общему количеству экспериментов.

При получении разных результатов разбиений, как в данном случае (см. элементы p_{22}, p_{23} матриц P_1, P_2), применяется метод FRC. Интерпретированный результат кластеризации методом FRC множества (6.3) (см. шаг 3 методики) – матрица P_3 – приведен далее.

$$P_1 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,03 & 0,90 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}; \quad P_2 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,94 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}; \quad P_3 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,94 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix};$$

$$P = \begin{pmatrix} 1,00 & 0,00 & 0,00 \\ 0,01 & 0,88 & 0,11 \\ 0,00 & 1,00 & 0,00 \\ 0,10 & 0,90 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}.$$

Матрица P – обобщенная матрица относительных частот принадлежности ЦФ определенным классам. Элементы матрицы P рассчитаны на основании (6.8) (см. шаг 4 методики). Например, элементы второй строки p_{21} , p_{22} , p_{23} матрицы $P = \|p_{ij}\|, i = 1, \dots, 5, j = 1, \dots, 3$, рассчитываются следующим образом:

$$p_{21} = (1 - 0,06) \cdot 0,06 \cdot 0,06 + 0,06 \cdot (1 - 0,06) \cdot 0,06 +$$

$$+ 0,06 \cdot 0,06 \cdot (1 - 0,06) + 0,06 \cdot 0,06 \cdot 0,06 = 0,01;$$

$$p_{22} = (1 - 0,03) \cdot 0,94 \cdot 0,94 + 0,03 \cdot (1 - 0,94) \cdot 0,94 +$$

$$+ 0,03 \cdot 0,94 \cdot (1 - 0,94) + 0,03 \cdot 0,94 \cdot 0,94 = 0,88;$$

$$p_{23} = 1 - 0,01 - 0,88 = 0,11.$$

В начале исследования ожидалось выделить три класса ЦФ. Методика определения классов ЦФ выявила наличие двух классов (см. шаг 5 методики):

– классу K_1 (с наименьшим значением чувствительности) со 100 % относительной частотой принадлежат функции Easom и Goldstein&Price;

– классу K_2 принадлежат функции Six-hump, Branin RCos, Modified Branin RCos с относительными частотами принадлежности классу 88 %, 100 % и 90 % соответственно.

Поведение ЦФ в каждом из выделенных классов отличается следующими особенностями: в класс K_1 попадают одномодальные ЦФ с пологими участками; в класс K_2 – многомодальные ЦФ с множеством локальных экстремумов.

7 Практическое применение интеллектуального анализа данных для оптимизации распределения ресурсов в производственном процессе

7.1 Моделирование реального производственного процесса крашения тканей набивным способом

Экспериментальное исследование ПП КТНС проведено на примере данных реального технологического процесса производства ткани с набивным печатным рисунком. Рассматриваемый артикул ткани предназначен для кожгалантерейной промышленности. ПП КТНС представлен в виде взвешенного ориентированного ациклического графа, вершинами которого являются технологические режимы (ТР) (рисунок 7.1).

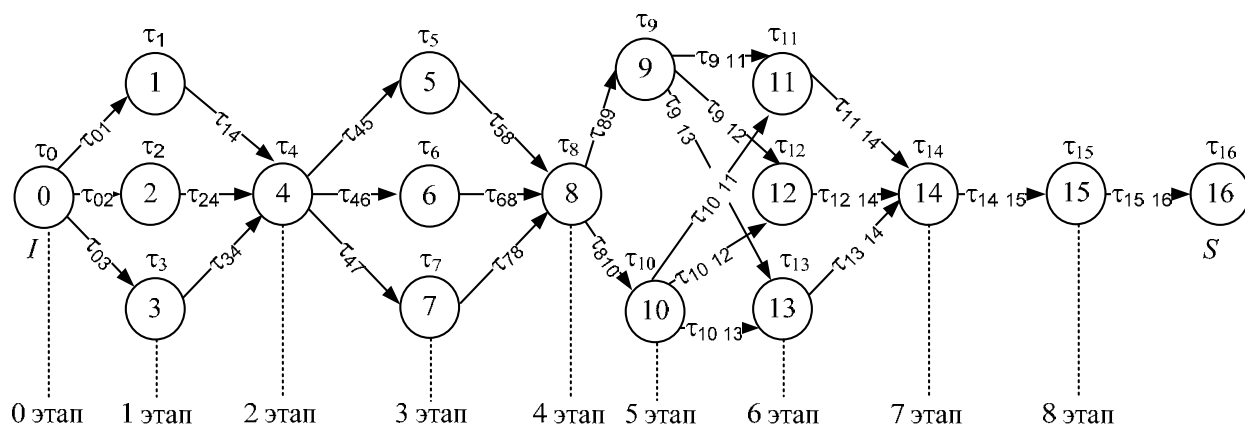


Рисунок 7.1 – Ориентированный граф с технологическими режимами

Процесс производства *TexPr* (2.1) рассматриваемого артикула ткани включает $N = 8$ технологических операций [18] Op_i , $i = 1, \dots, 8$: Op_1 – отварка; Op_2 – стабилизация; Op_3 – печать на ткани; Op_4 – зреление; Op_5 – промывка; Op_6 – вид отделки «стабилизация»; Op_7 – вид отделки «придание наполненного гифа»; Op_8 – заключительный этап: браковка ткани. При этом на каждом из этапов технологическая операция выбирается из допустимого количества ТР (этап 1: $M_1 = 3$; этап 2: $M_2 = 1$; этап 3: $M_3 = 3$; этап 4: $M_4 = 1$; этап 5: $M_5 = 2$; этап 6: $M_6 = 3$; этап 7: $M_7 = 1$; этап 8: $M_8 = 1$). Суммарные стоимости используемых в ТР ресурсов [13, 14]: химикатов, красителей, энергоресурсов (пара, газа, электроэнергии) составляют себестоимости ТР. Химикаты, применяемые в ТР отварки (см. рисунок 7.1, вершины 1 и 3) и промывки (см. рисунок 7.1, вершины 9 и 10), разделяются на основные и

аналоги (концентрация таких химикатов одинакова, различна стоимость). На рисунке 7.2 представлен расширенный граф ПП КТНС – граф ПП, построенный с учетом основных химикатов, химикатов-аналогов для ТР.

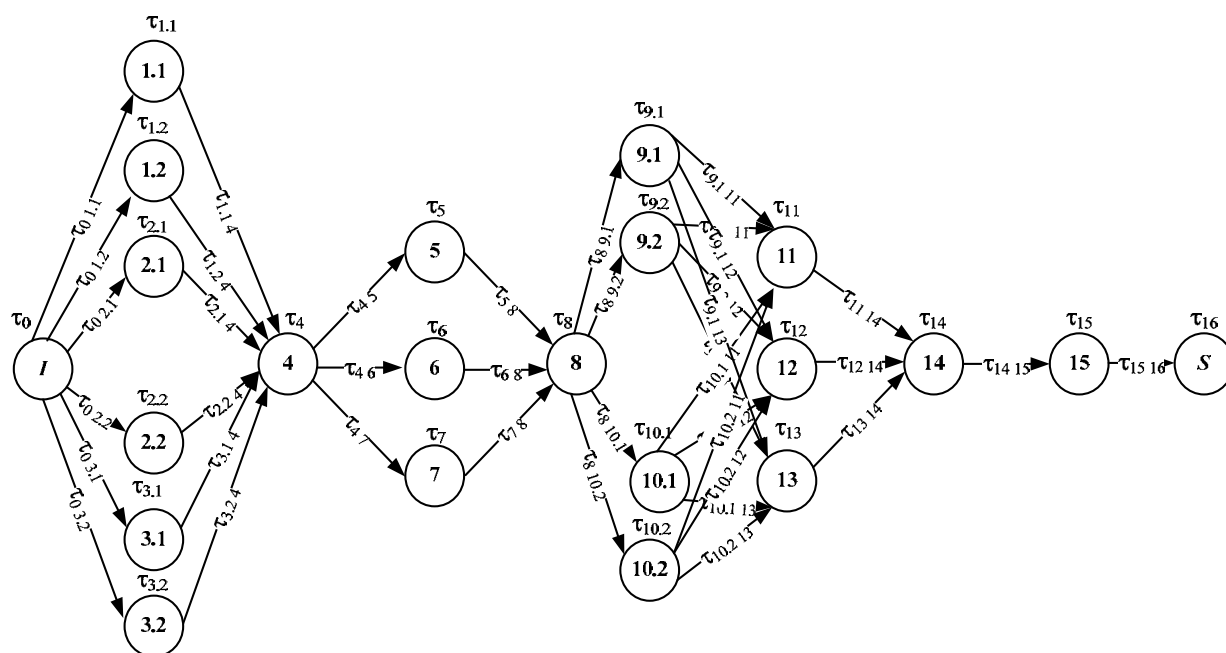


Рисунок 7.2 – Ориентированный граф с технологическими режимами, данные с учетом неосновных химикатов

Для уточнения нагрузок на дуги графа ПП КТНС (см. рисунки 7.1 и 7.2) и определения оптимальных ТР в среде PowerSim построена модель имитации (ИМ) временных норм расхода и норм расхода ресурсов, затрачиваемых для выпуска необходимого заданного объема готового текстильного полотна в разрезе отдельных технологических режимов. Стадии ПП КТНС представлены на отдельных листах ИМ. Выбор оптимальных технологических режимов ПП КТНС в имитационной модели осуществляется на вкладке итогов.

Выходными параметрами ИМ ПП КТНС являются: T – время; стоимость ресурсов: S – общая стоимость ресурсов, S_{Xim} – стоимость химикатов, S_{EnRes} – стоимость энергоресурсов, S_{ElEn} – стоимость электроэнергии, S_{Par} – стоимость пара, S_{Gaz} – стоимость газа:

$$S = S_{Xim} + S_{EnRes} ; \quad (7.1)$$

$$S_{EnRes} = S_{ElEn} + S_{Par} + S_{Gaz} ; \quad (7.2)$$

$$S = S_{Xim} + S_{EnRes} = [S_{EnRes} = S_{ElEn} + S_{Par} + S_{Gaz}] = S_{Xim} + S_{ElEn} + S_{Par} + S_{Gaz} . \quad (7.3)$$

7.2 Оптимизация производственного процесса крашения тканей набивным способом

Нагрузки на дуги графа ТП (временные затраты, стоимости ресурсов) уточнены в ИМ (с использованием метода Монте-Карло и/или методов статистической обработки данных), известны и достоверны.

Для всех технологических режимов ПП КТНС определен расход времени T_i (в часах), $i = 1, \dots, 8$, затрачиваемого на отделку 1000 п. м текстильного полотна: этап 1 – $T_1 = (0,52; 0,52; 0,52)$; этап 2 – $T_2 = (0,6)$; этап 3 – $T_3 = (1,01; 1,40; 1,26)$; этап 4 – $T_4 = (0,87)$; этап 5 – $T_5 = (1,07; 1,01)$; этап 6 – $T_6 = (0,86; 0,86; 0,43)$; этап 7 – $T_7 = (1,90)$; этап 8 – $T_8 = (1,28)$.

Для всех технологических режимов ПП КТНС определены стоимости расхода ресурсов S_i (в рублях), $i = 1, \dots, 8$, затрачиваемых на отделку 1000 п. м текстильного полотна: этап 1 – $S_1 = (149913,95; 154348,75; 148572,68)$; этап 2 – $S_2 = (593499,34)$; этап 3 – $S_3 = (987782,50; 959879,82; 942975,85)$; этап 4 – $S_4 = (243148,13)$; этап 5 – $S_5 = (557763,48; 522017,79)$; этап 6 – $S_6 = (858455,08; 854431,31; 228772,21)$; этап 7 – $S_7 = (1653922,81)$; этап 8 – $S_8 = (1348,71)$.

Требуется: определить оптимальные ТР ТП (2.1) в соответствии с заданными критериями оптимальности.

При моделировании ПП КТНС рассмотрены приведенные далее критерии оптимизации.

1 Итоговые время и стоимость расхода ресурсов (T, S) определены для применяемого на производстве сочетания ТР ПП (фиксированные ТР ИМ).

2 Итоговые время и стоимость расхода ресурсов (T^*, S^*) определяются для оптимизации по одному из критериев: $T \rightarrow \min$; $T \rightarrow \max$; $S \rightarrow \min$; $S \rightarrow \max$.

3 Итоговые время и стоимость расхода ресурсов (T^*, S^*) определяются для оптимизации по двум лексикографически упорядоченным критериям (рисунки 7.3–7.5) [18]:

- критерий 1: $S \rightarrow \min$; критерий 2: $T \rightarrow \min$;
- критерий 1: $T \rightarrow \min$; критерий 2: $S \rightarrow \min$;
- критерий 1: $S \rightarrow \max$; критерий 2: $T \rightarrow \max$;
- критерий 1: $T \rightarrow \max$; критерий 2: $S \rightarrow \max$;
- критерий 1: $S \rightarrow \max$; критерий 2: $T \rightarrow \min$;
- критерий 1: $T \rightarrow \min$; критерий 2: $S \rightarrow \max$;
- критерий 1: $S \rightarrow \min$; критерий 2: $T \rightarrow \max$;
- критерий 1: $T \rightarrow \max$; критерий 2: $S \rightarrow \min$.

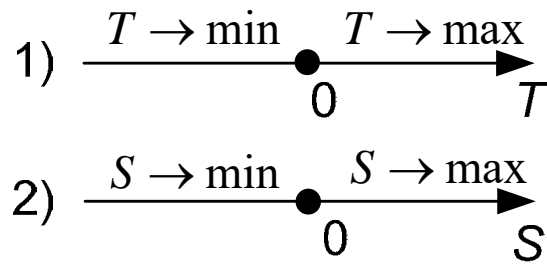
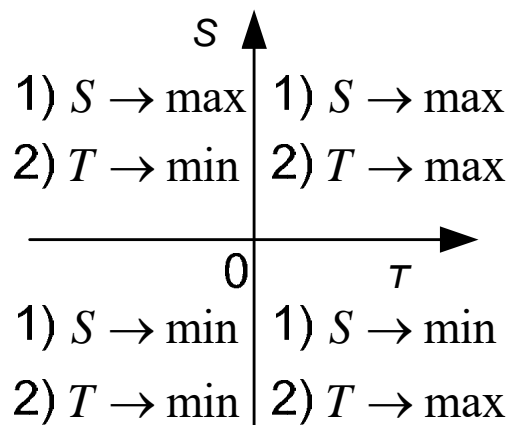
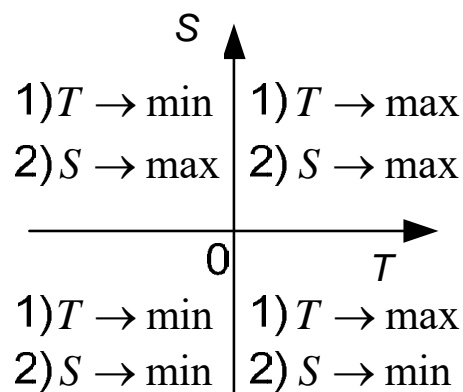


Рисунок 7.3 – Оптимизация по одному критерию



1 – стоимость; 2 – время

Рисунок 7.4 – Оптимизация по двум лексикографически упорядоченным критериям



1 – время; 2 – стоимость

Рисунок 7.5 – Оптимизация по двум лексикографически упорядоченным критериям

При этом S может рассматриваться как общая стоимость ресурсов, стоимость химикатов S_{Xim} , стоимость энергоресурсов S_{EnRes} , стоимость электроэнергии S_{ElEn} , стоимость пара S_{Par} , стоимость газа S_{Gaz} .

С учетом $T_i = (t_{i1}, ..., t_{iM_i})$, $S_i = (s_{i1}, ..., s_{iM_i})$, $i = 1, ..., 8$, нагрузки на дуги графа ПП КТНС для $\tau_{ij} \in R^2$ принимают значения: $\tau_{01} = \tau_{02} = \tau_{03} = (0,0)$; $\tau_{14} = (t_{11}, s_{11})$; $\tau_{24} = (t_{12}, s_{12})$; $\tau_{34} = (t_{13}, s_{13})$; $\tau_{45} = \tau_{46} = \tau_{47} = (t_{21}, s_{21})$; $\tau_{58} = (t_{31}, s_{31})$; $\tau_{68} = (t_{32}, s_{32})$; $\tau_{78} = (t_{33}, s_{33})$; $\tau_{89} = \tau_{810} = (t_{41}, s_{41})$; $\tau_{911} = \tau_{912} = \tau_{913} = (t_{51}, s_{51})$; $\tau_{1011} = \tau_{1012} = \tau_{1013} = (t_{52}, s_{52})$; $\tau_{1114} = (t_{61}, s_{61})$; $\tau_{1214} = (t_{62}, s_{62})$; $\tau_{1314} = (t_{63}, s_{63})$; $\tau_{1415} = (t_{71}, s_{71})$; $\tau_{1516} = (t_{81}, s_{81})$.

Определение оптимальных ТР ПП КТНС (2.1) производится согласно методу динамического программирования. Для этого по формуле (4.18) осуществляется построение последовательности двумерных состояний на этапах ПП КТНС и, соответственно, последовательности управлений ПП КТНС, определяемых предикатами вида (4.26).

Формулы определения экономического эффекта оптимизации $(\Delta T, \Delta S)$

$$\Delta T = T^* - T; \quad (7.4)$$

$$\Delta S = S^* - S, \quad (7.5)$$

где (T, S) , (T^*, S^*) – итоговые время и стоимость расхода ресурсов для применяемого на производстве (рисунок 7.6) и оптимального сочетаний ТР ТП соответственно.

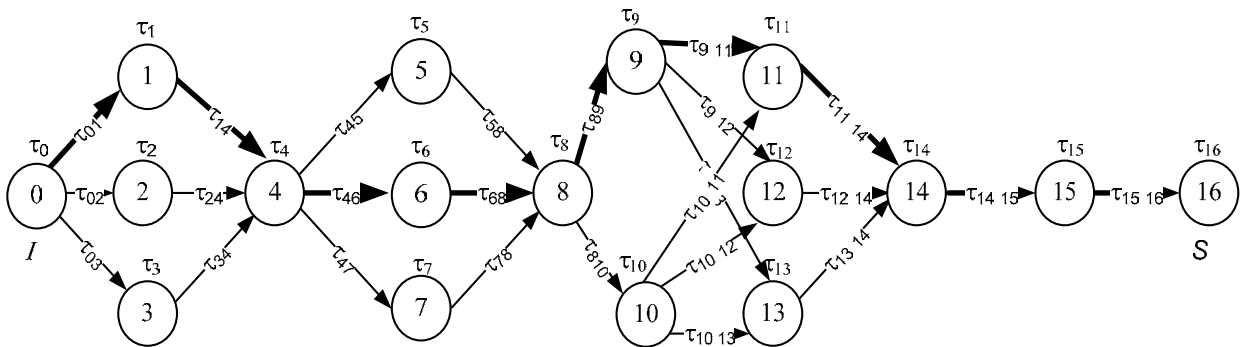


Рисунок 7.6 – Применяемое на производстве сочетание технологических режимов производственного процесса

В результате проведенного исследования получены следующие оптимальные технологические режимы ПП КТНС.

Случай 1. Для применяемого на производстве сочетания ТР исследуемого ПП КТНС (см. рисунок 7.6) маршрут по графу: $\langle 0-1-4-6-8-9-11-14-15 \rangle$; имеют место следующие показатели: $T = 7,50$ ч; $S = 5017931$ р.

Случай 2. Критерий 1: $S \rightarrow \min$; критерий 2: $T \rightarrow \min$ (приоритет: оптимизация по стоимости ресурсов, затем – по времени). Оптимальный маршрут: $\langle 0-3-4-7-8-10-13-14-15 \rangle$; итоги: $T^* = 6,87$ ч, $S^* = 4334258$ р.; экономический эффект оптимизации: $|\Delta T| = 0,63$ ч (8 %), $|\Delta S| = 683673$ р. (14 %).

Случай 3. Критерий 1: $T \rightarrow \min$; критерий 2: $S \rightarrow \min$ (приоритет: оптимизация по времени ресурсов, затем – по стоимости ресурсов). Оптимальный маршрут: $\langle 0-3-4-5-8-10-13-14-15 \rangle$; итоги: $T^* = 6,62$ ч, $S^* = 4379064$ р.; экономический эффект оптимизации: $|\Delta T| = 0,88$ ч (12 %), $|\Delta S| = 638867$ р. (13 %).

В результате исследования получен экономический эффект от оптимизации (при производстве 1000 п. м рассматриваемой группы тканей) за счет уменьшения стоимости расхода ресурсов на 14 % и времени на 8 %.

7.3 Кластеризация состояний объекта управления по итогам решения задачи многокритериального управления производственным процессом

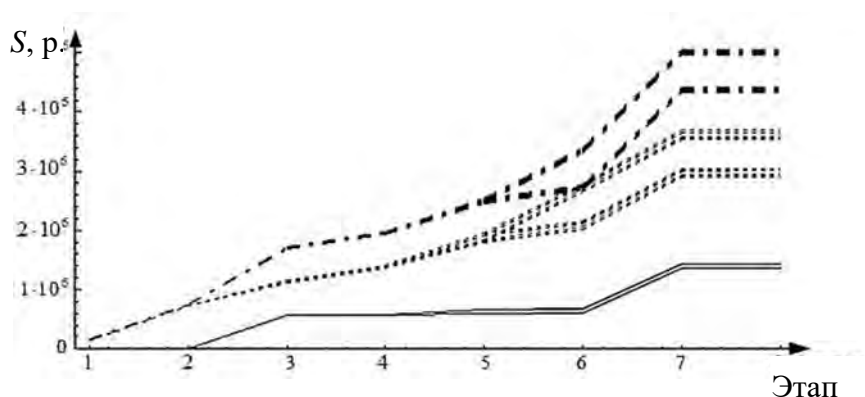
Апробация методики кластеризации многомерных данных несколькими методами кластерного анализа проведена на реальных массивах данных с известной структурой; искусственных массивах данных с известной структурой, генерируемых на ПК и обрабатываемых в режиме статистического моделирования; реальных массивах данных с неизвестной структурой. В частности, апробация программного модуля кластеризации многомерных данных методом Fuzzy Relation Clustering в составе программно-технологического комплекса имитации (ПТКИ) сложных систем BelSim [108] и методики разделения многомерных наблюдений на кластеры с использованием нескольких методов кластерного анализа проведена на реальных массивах данных при статистическом анализе и исследовании поведения пользователя с целью выявления зависимостей монетизации пользователя от его действий в программных продуктах. Результаты научных исследований позволяют предсказать уход пользователя из продукта, выявить группы пользователей со схожим поведением и формировать изменения в

продукте на основе расчетов по построенным зависимостям и результатов кластеризации пользователей [19, 21, 44].

Для решения задачи определения состояний, близких к оптимальным, в соответствии с заданной мерой близости применяется процедура анализа выборок данных для возможных состояний ТП. Общее количество всех возможных маршрутов по графу ТП (см. рисунок 7.1) $\eta = 54$.

Этап 1. Построение многомерных выборок данных для состояний ТП.

На рисунках 7.7–7.9 представлены построенные в математическом пакете Maple графики накопления стоимостей расхода соответствующих ресурсов на стадиях рассматриваемого ПП КТНС, граф которого изображен на рисунке 7.1. На рисунках 7.10–7.12 – графики накопления стоимостей расхода соответствующих ресурсов на стадиях модифицированного ПП (2.1), граф которого изображен на рисунке 7.2.



Общие расходы (· · ·), энергоресурсы (---), химикаты (—)

Рисунок 7.7 – Временные ряды состояний производственного процесса (см. рисунок 7.1) по «стоимостям расхода ресурсов»

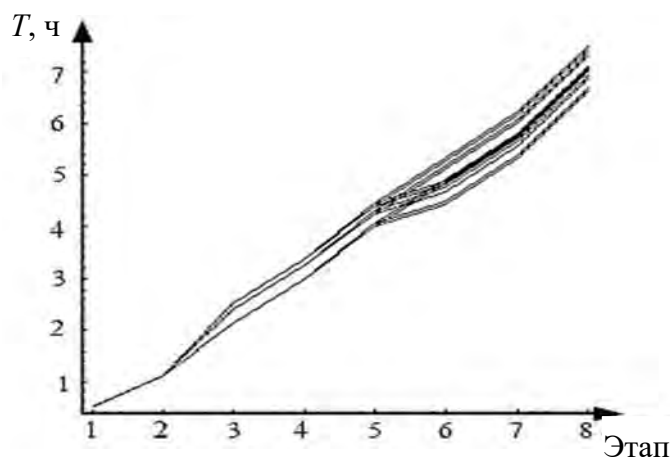
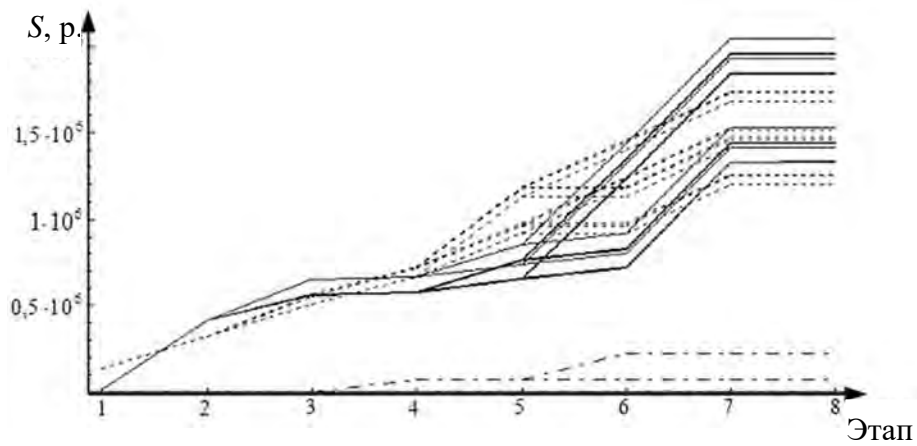
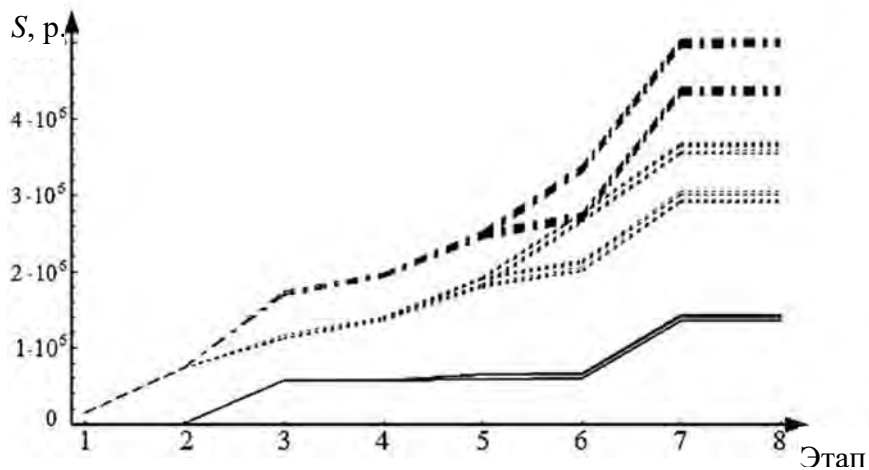


Рисунок 7.8 – Временные ряды состояний производственного процесса (см. рисунок 7.1) по «времени» на соответствующих стадиях



Электроэнергия (—), пар (----), газ (- · -)

Рисунок 7.9 – Временные ряды состояний производственного процесса (см. рисунок 7.1) по «стоимостям расхода ресурсов»



Общие расходы (- · -), энергоресурсы (---), химикаты (—)

Рисунок 7.10 – Выборка нагрузок графа производственного процесса (см. рисунок 7.2)

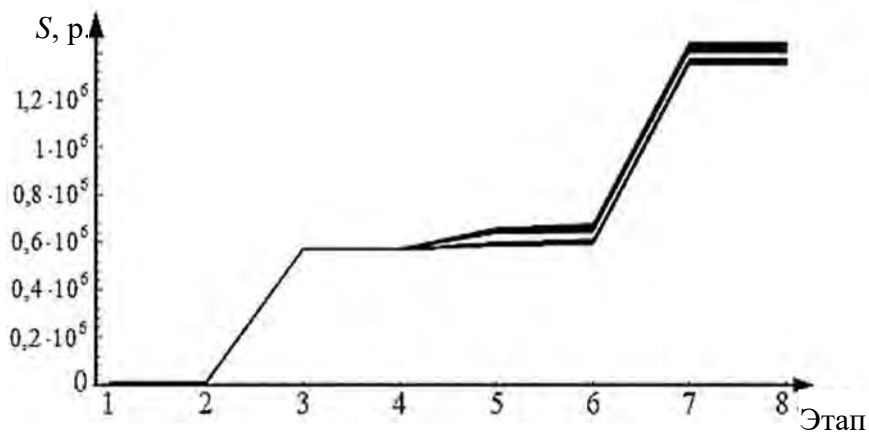


Рисунок 7.11 – Химикаты. Выборка нагрузок графа производственного процесса (см. рисунок 7.2)

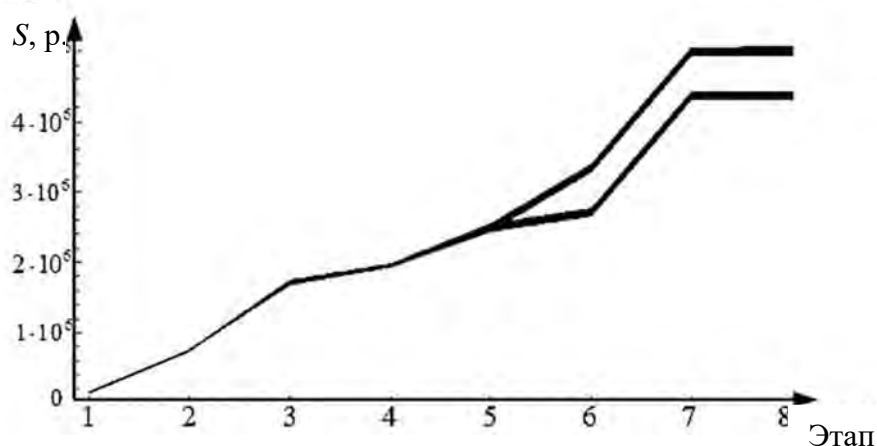


Рисунок 7.12 – Общий расход ресурсов. Выборка нагрузок графа производственного процесса (см. рисунок 7.2)

Построены следующие многомерные выборки всех возможных на этапе 8 состояний ТП (2.1) при обработке 1000 п. м ткани в соответствии с заданными сочетаниями технологических режимов ($l = 1, \dots, 54$):

1) двумерная: $(y_l^{(1)}, y_l^{(2)})$, где $y_l^{(1)}$ – итоги по времени; $y_l^{(2)}$ – итоговые стоимости расхода ресурсов ТП;

2) трехмерная: $(y_l^{(1)}, y_l^{(2)}, y_l^{(3)})$, где $y_l^{(1)}$ – итоги по времени; $y_l^{(2)}$ – итоговые стоимости расхода энергоресурсов ТП; $y_l^{(3)}$ – итоговые стоимости расхода химикатов ТП;

3) пятимерная: $(y_l^{(1)}, y_l^{(2)}, y_l^{(3)}, y_l^{(4)}, y_l^{(5)})$, где $y_l^{(1)}$ – итоги по времени; $y_l^{(2)}$ – итоговые стоимости расхода электроэнергии; $y_l^{(3)}$ – итоговые стоимости расхода пара; $y_l^{(4)}$ – итоговые стоимости расхода газа; $y_l^{(5)}$ – итоговые стоимости расхода химикатов ТП. Данные уточнены методом Монте-Карло, известны и достоверны.

Требуется разделить построенные многомерные множества на кластеры; определить состояния ТП, близкие к наиболее и наименее оптимальному в соответствии с критериями оптимизации состояниям ТП.

Этап 2. Кластерный анализ многомерных данных.

Результаты применения методов K-means и Tree Clustering для рассматриваемых множеств (двумерного (рисунок 7.13), трехмерного и пятимерного случаев состояний ТП) идентичны. В каждом из случаев построено по два кластера: K_1 , K_2 , где элементы K_1 – состояния ТП, наиболее близкие к оптимальному; элементы K_2 – наиболее удаленные от оптимального. Применения третьего метода для уточнения результатов кластеризации не требуется.

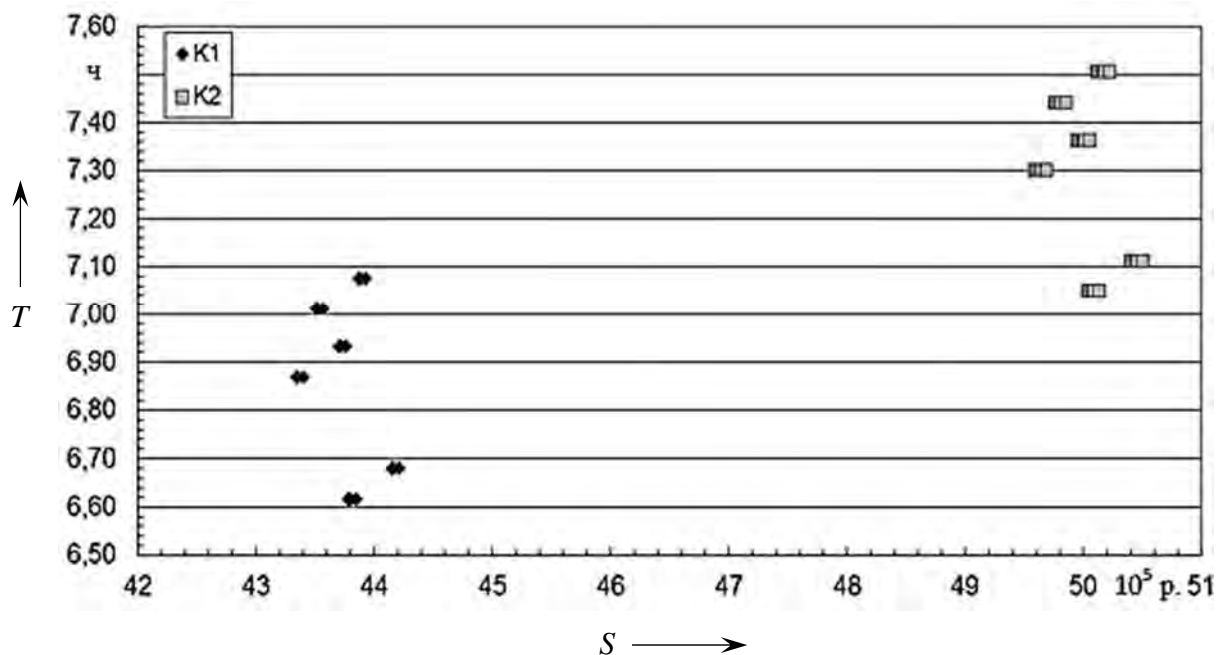


Рисунок 7.13 – Выборка всех возможных на n -м этапе двумерных состояний производственного процесса (см. рисунок 7.1)

Этап 3. Дополнительный анализ структур построенных кластеров.

Каждому элементу из рассматриваемых многомерных выборок соответствует информация о последовательности состояний ТП. Объекты исследования – одиночные независимые ТР: вершины графа (см. рисунок 7.1) с номерами u , $u = 1, \dots, 15$, 0-я и 16-я вершины – не учитываются.

Получены следующие результаты *дополнительного анализа структур построенных кластеров*. С вероятностью 100 % ТР, соответствующий вершине 13 графа ТП, принадлежит кластеру K_1 (состояния ТП, наиболее близкие к оптимальному). С вероятностью 100 % ТР, соответствующие вершинам 11, 12 графа ТП, а также с вероятностью 67 % ТР, соответствующие вершинам 1–10, 14, 15 графа ТП, принадлежат кластеру K_2 (состояния, наименее близкие к оптимальному) (рисунок 7.14).

Все сочетания ТР исследуемого ТП, включающие ТР, соответствующий вершине графа 13, обеспечивают в конечном итоге обработку 1000 п. м ткани за оптимальное время и с оптимальной стоимостью расхода ресурсов по сравнению с другими сочетаниями ТР ТП.

Согласно рассчитанным оценкам характеристик кластеров K_1 для пятимерного случая $(T, S_{ElEn}, S_{Par}, S_{Gaz}, S_{Xim})$ состояний ТП (рисунок 7.15) средняя стоимость газа в K_1 (состояния ТП, наиболее близкие к оптимальному) превышает среднюю стоимость газа в K_2 (за счет объема потребления), но при

этом для K_1 имеет место меньшее потребление более дорогостоящих ресурсов: пара и электроэнергии.

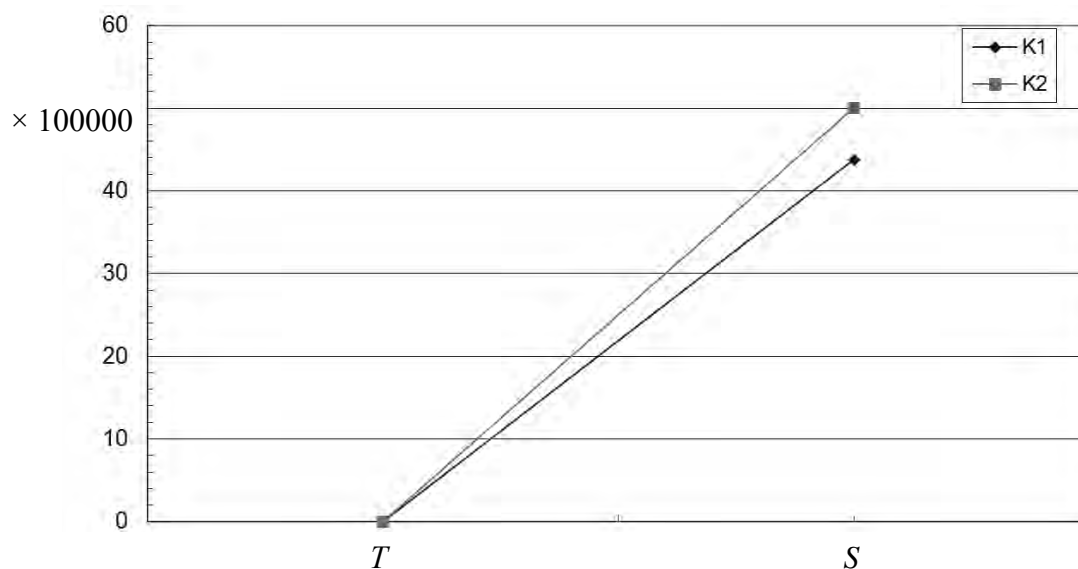


Рисунок 7.14 – График средних значений координат (T, S) для кластеров

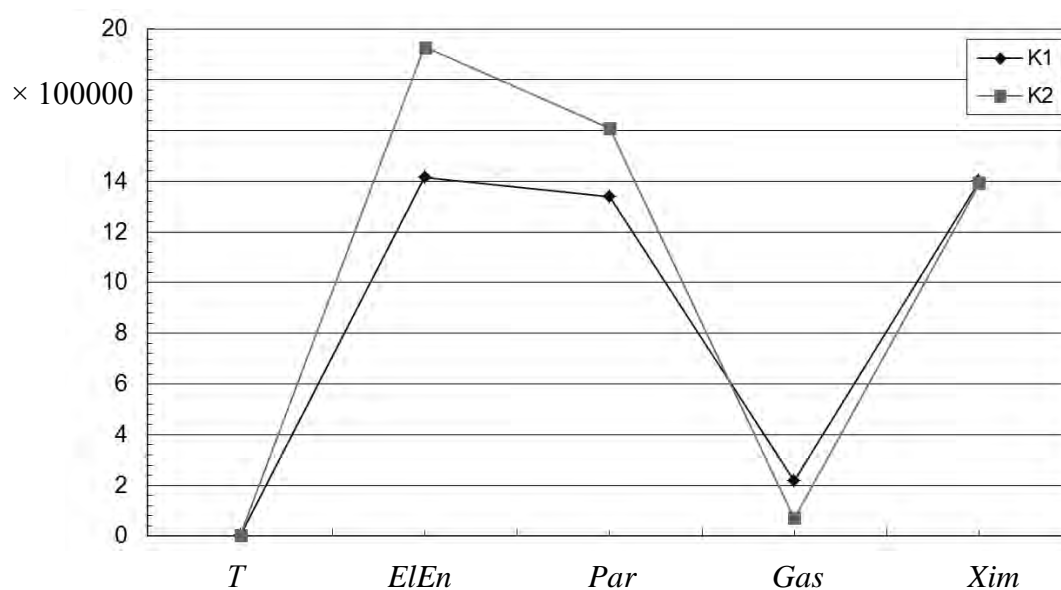


Рисунок 7.15 – График средних координат $(T, S_{ElEn}, S_{Par}, S_{Gaz}, S_{Xim})$ для кластеров K_1, K_2

Результаты кластеризации состояний: в кластере K_1 математическое ожидание экономического эффекта за счет уменьшения стоимости расхода ресурсов – на 13 % и времени – на 8 % со среднеквадратическим отклонением соответственно 0,53 % ($p = 0,95$) и 2,28 % ($p = 0,95$); в кластере K_2 математическое ожидание экономического эффекта за счет уменьшения стоимости

расхода ресурсов – на 0 % и времени – на 3 % со среднеквадратическим отклонением соответственно 0,53 % ($p = 0,95$) и 2,25 % ($p = 0,95$).

7.4 Определение чувствительности общей стоимости обработки ткани к замене основных химикатов на химикаты-аналоги и к изменению цен на ресурсы

Моделирование ПП проведено по данным одного из двух рассматриваемых расчетных временных периодов: 1 и 2.

Введем обозначения:

$S^{(1)}, S^{(2)}$ – общие стоимости ресурсов за периоды 1 и 2 соответственно (стоимости ТР – с учетом только основных химикатов);

$S_A^{(1)}, S_A^{(2)}$ – общие стоимости ресурсов за периоды 1 и 2 соответственно (стоимости ТР – с учетом только химикатов-аналогов);

$S_{Xim}^{(1)}, S_{Xim}^{(2)}$ – стоимости химикатов за периоды 1 и 2 соответственно (стоимости с учетом только основных химикатов);

$S_{XimA}^{(1)}, S_{XimA}^{(2)}$ – стоимости химикатов за периоды 1 и 2 соответственно (стоимости с учетом только химикатов-аналогов);

$S_{EnRes}^{(1)}, S_{EnRes}^{(2)}$ – стоимости энергоресурсов за периоды 1 и 2 соответственно;

$S_{ElEn}^{(1)}, S_{ElEn}^{(2)}$ – стоимости электроэнергии за периоды 1 и 2 соответственно;

$S_{Par}^{(1)}, S_{Par}^{(2)}$ – стоимости пара за периоды 1 и 2 соответственно;

$S_{Gaz}^{(1)}, S_{Gaz}^{(1)}$ – стоимости газа за периоды 1 и 2 соответственно;

$S_1^{(2)}$ – суммарная стоимость химикатов, электроэнергии, газа за период 1, соответственно, с учетом стоимости пара за период 2;

$S_2^{(2)}$ – суммарная стоимость химикатов, пара, газа за период 1, соответственно, с учетом стоимости электроэнергии за период 2;

$S_3^{(2)}$ – суммарная стоимость химикатов, электроэнергии, пара за период 1, соответственно, с учетом стоимости газа за период 2;

$S_4^{(2)}$ – суммарная стоимость пара, электроэнергии, газа за период 1, соответственно, с учетом стоимости химикатов за период 2;

$S_5^{(2)}$ – суммарная стоимость химикатов за период 1 и энергоресурсов за период 2.

$$S^{(1)} = S_{Xim}^{(1)} + S_{ElEn}^{(1)} + S_{Par}^{(1)} + S_{Gaz}^{(1)} ; \quad (7.6)$$

$$S^{(1)} = S_{Xim}^{(1)} + S_{EnRes}^{(1)} ; \quad (7.7)$$

$$S_1^{(2)} = S_{Xim}^{(1)} + S_{ElEn}^{(1)} + S_{Par}^{(2)} + S_{Gaz}^{(1)} ; \quad (7.8)$$

$$S_2^{(2)} = S_{Xim}^{(1)} + S_{ElEn}^{(2)} + S_{Par}^{(1)} + S_{Gaz}^{(1)} ; \quad (7.9)$$

$$S_3^{(2)} = S_{Xim}^{(1)} + S_{ElEn}^{(1)} + S_{Par}^{(1)} + S_{Gaz}^{(2)} ; \quad (7.10)$$

$$S_4^{(2)} = S_{Xim}^{(2)} + S_{ElEn}^{(1)} + S_{Par}^{(1)} + S_{Gaz}^{(1)} ; \quad (7.11)$$

$$S_5^{(2)} = S_{Xim}^{(1)} + S_{EnRes}^{(2)} . \quad (7.12)$$

Относительные погрешности изменения стоимостей ресурсов:

$\Delta S^{(A1)}$, $\Delta S^{(A2)}$ – относительные погрешности изменения общих стоимостей обработки ткани при замене основных химикатов на химикаты-аналоги за периоды 1 и 2 соответственно;

$\Delta S_{Xim}^{(A1)}$, $\Delta S_{Xim}^{(A2)}$ – относительные погрешности изменения стоимостей химикатов при замене основных химикатов на химикаты-аналоги за периоды 1 и 2 соответственно.

$$\Delta S^{(A1)} = (S^{(1)} - S_A^{(1)}) \cdot 2 / (S^{(1)} + S_A^{(1)}) \cdot 100 ; \quad (7.13)$$

$$\Delta S^{(A2)} = (S^{(2)} - S_A^{(2)}) \cdot 2 / (S^{(2)} + S_A^{(2)}) \cdot 100 ; \quad (7.14)$$

$$\Delta S_{Xim}^{(A1)} = (S_{Xim}^{(1)} - S_{XimA}^{(1)}) \cdot 2 / (S_{Xim}^{(1)} + S_{XimA}^{(1)}) \cdot 100 ; \quad (7.15)$$

$$\Delta S_{Xim}^{(A2)} = (S_{Xim}^{(2)} - S_{XimA}^{(2)}) \cdot 2 / (S_{Xim}^{(2)} + S_{XimA}^{(2)}) \cdot 100 . \quad (7.16)$$

Определены **чувствительности** [43] $\Delta S^{(A1)} / \Delta S_{Xim}^{(A1)}$, $\Delta S^{(A2)} / \Delta S_{Xim}^{(A2)}$ **общей стоимости обработки ткани к замене основных химикатов на химикаты-аналоги** за периоды 1 и 2 соответственно. На рисунках 7.16 и 7.18 представлены графики соотношений $(\Delta S_{Xim}^{(A1)}, \Delta S^{(A1)})$, где $\Delta S_{Xim}^{(A1)}$ – погрешности

изменения стоимостей химикатов; $\Delta S^{(A1)}$ – погрешности общих стоимостей ресурсов ТП при замене основных химикатов на аналоги, данные за периоды 1 и 2 соответственно. На рисунках 7.17 и 7.19 – временные ряды чувствительности общей стоимости ресурсов ТП к замене основных химикатов на химикаты-аналоги, данные за периоды 1 и 2 соответственно.

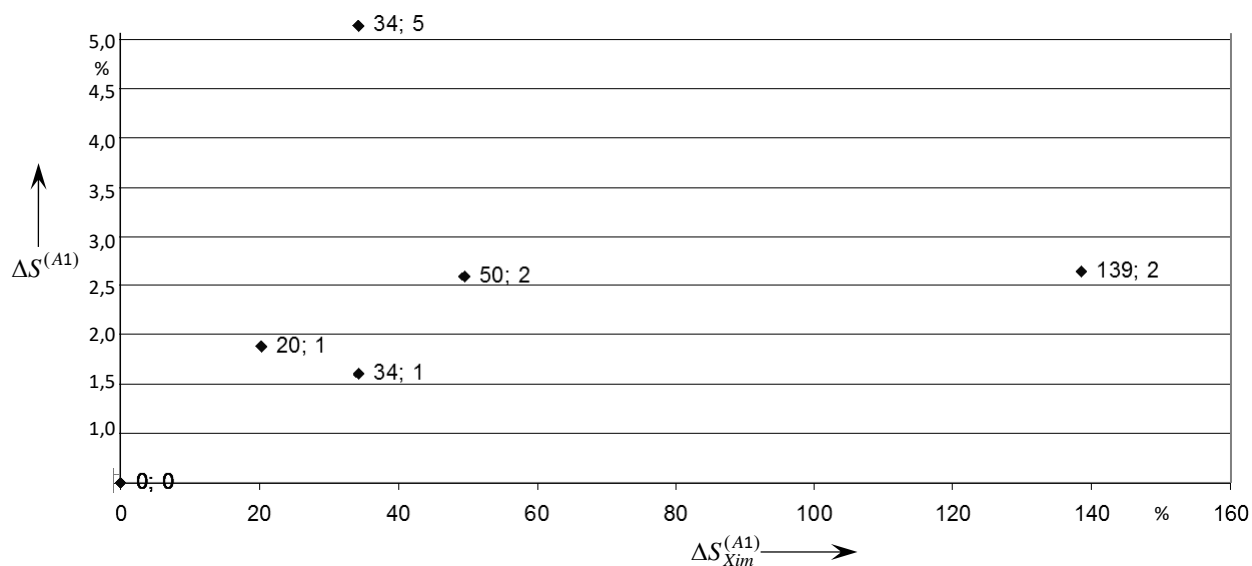


Рисунок 7.16 – Соотношение $(\Delta S_{Xim}^{(A1)}, \Delta S^{(A1)})$, данные за период 1

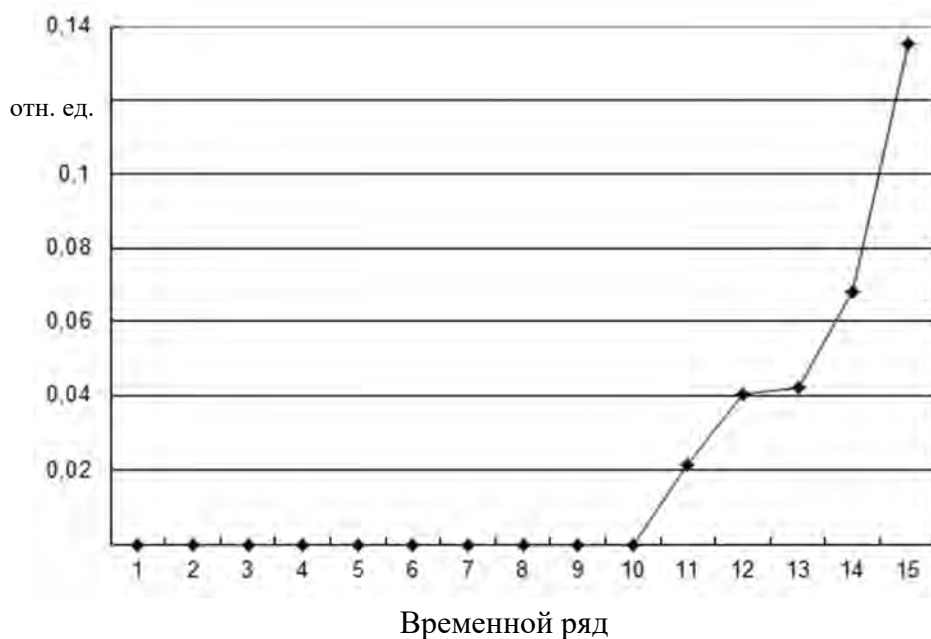


Рисунок 7.17 – Чувствительность общей стоимости ресурсов к замене основных химикатов на химикаты-аналоги

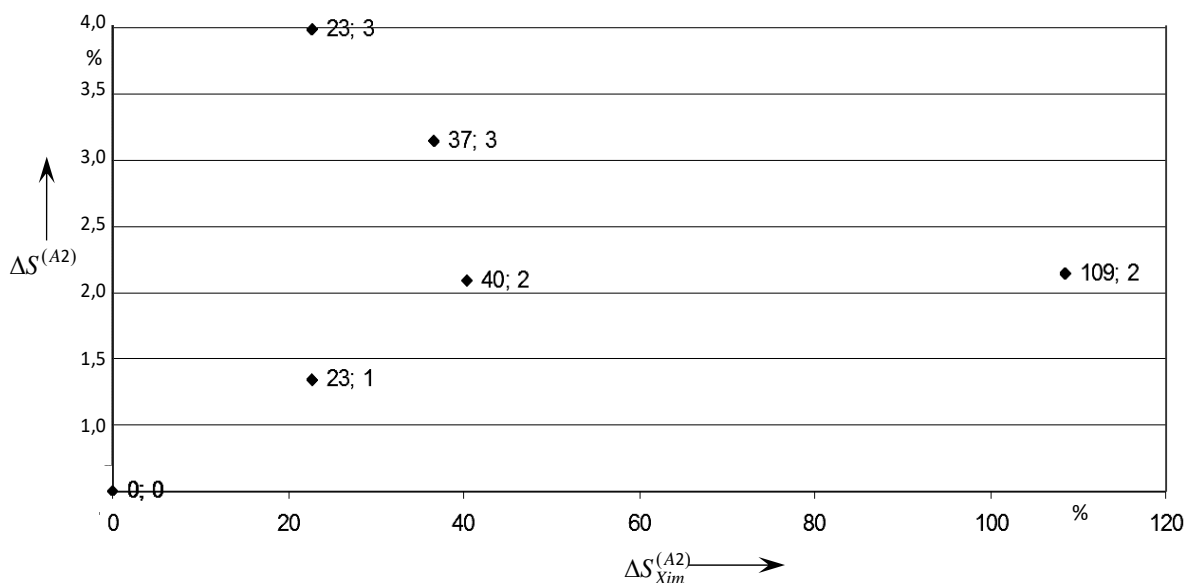


Рисунок 7.18 – Соотношение $(\Delta S_{Xim}^{(A2)}, \Delta S^{(A2)})$, данные за период 2

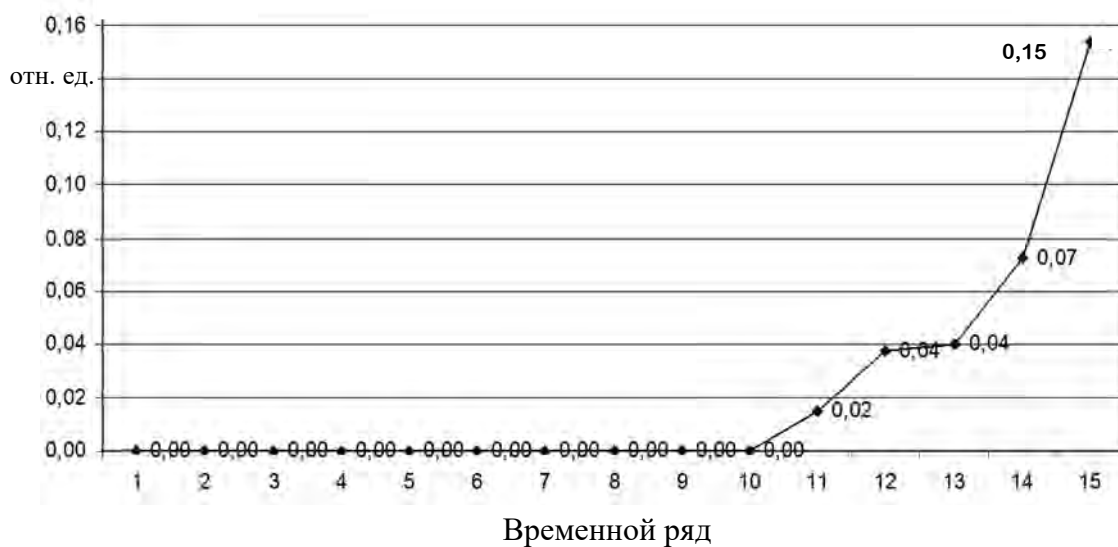


Рисунок 7.19 – Чувствительность общей стоимости ресурсов к замене основных химикатов на химикаты-аналоги

ΔS_{Par} – относительная погрешность изменения стоимости пара за период 1 по отношению к стоимости пара за период 2.

ΔS_{ElEn} – относительная погрешность изменения стоимости электроэнергии за период 1 по отношению к стоимости электроэнергии за период 2.

ΔS_{Gaz} – относительная погрешность изменения стоимости газа за период 1 по отношению к стоимости газа за период 2.

ΔS_{Xim} – относительная погрешность изменения стоимости химикатов за период 1 по отношению к стоимости химикатов за период 2 (стоимости с учетом только основных химикатов).

ΔS_{EnRes} – относительная погрешность изменения стоимости энергоресурсов за период 1 по отношению к стоимости энергоресурсов за период 2.

$$\Delta S_{Par} = (S_{Par}^{(2)} - S_{Par}^{(1)}) \cdot 2 / (S_{Par}^{(2)} + S_{Par}^{(1)}) \cdot 100; \quad (7.17)$$

$$\Delta S_{ElEn} = (S_{ElEn}^{(2)} - S_{ElEn}^{(1)}) \cdot 2 / (S_{ElEn}^{(2)} + S_{ElEn}^{(1)}) \cdot 100; \quad (7.18)$$

$$\Delta S_{Gaz} = (S_{Gaz}^{(2)} - S_{Gaz}^{(1)}) \cdot 2 / (S_{Gaz}^{(2)} + S_{Gaz}^{(1)}) \cdot 100; \quad (7.19)$$

$$\Delta S_{Xim} = (S_{Xim}^{(2)} - S_{Xim}^{(1)}) \cdot 2 / (S_{Xim}^{(2)} + S_{Xim}^{(1)}) \cdot 100; \quad (7.20)$$

$$\Delta S_{EnRes} = (S_{EnRes}^{(2)} - S_{EnRes}^{(1)}) \cdot 2 / (S_{EnRes}^{(2)} + S_{EnRes}^{(1)}) \cdot 100. \quad (7.21)$$

ΔS_{11} , ΔS_{12} , ΔS_{13} , ΔS_{14} , ΔS_{15} – относительные погрешности изменения общей стоимости ресурсов при изменении стоимости пара, электроэнергии, газа, химикатов, энергоресурсов соответственно.

$$\Delta S_{11} = (S_1^{(2)} - S_1^{(1)}) \cdot 2 / (S_1^{(2)} + S_1^{(1)}) \cdot 100; \quad (7.22)$$

$$\Delta S_{12} = (S_2^{(2)} - S_2^{(1)}) \cdot 2 / (S_2^{(2)} + S_2^{(1)}) \cdot 100; \quad (7.23)$$

$$\Delta S_{13} = (S_3^{(2)} - S_3^{(1)}) \cdot 2 / (S_3^{(2)} + S_3^{(1)}) \cdot 100; \quad (7.24)$$

$$\Delta S_{14} = (S_4^{(2)} - S_4^{(1)}) \cdot 2 / (S_4^{(2)} + S_4^{(1)}) \cdot 100; \quad (7.25)$$

$$\Delta S_{15} = (S_5^{(2)} - S_5^{(1)}) \cdot 2 / (S_5^{(2)} + S_5^{(1)}) \cdot 100. \quad (7.26)$$

$\Delta S_{11} / \Delta S_{Par}$, $\Delta S_{12} / \Delta S_{ElEn}$, $\Delta S_{13} / \Delta S_{Gaz}$, $\Delta S_{14} / \Delta S_{Xim}$, $\Delta S_{15} / \Delta S_{EnRes}$ – чувствительности [43] стоимости обработки ткани к изменению цен на пар, электроэнергию, газ, химикаты, все энергоресурсы одновременно соответственно.

На рисунке 7.20 представлены графики чувствительности общей стоимости ресурсов (в процентах) к изменению цен на ресурсы (в процентах): соотношения $(\Delta S_{Par}, \Delta S_{11})$, $(\Delta S_{ElEn}, \Delta S_{12})$, $(\Delta S_{Gaz}, \Delta S_{13})$, $(\Delta S_{Xim}, \Delta S_{14})$. На рисунке 7.21 – временные ряды чувствительности общей стоимости ресурсов к изменению цен на химикаты и энергоресурсы. На рисунке 7.22 – графики чувствительности общей стоимости ресурсов (в процентах) к изменению цен на ресурсы (в процентах): соотношения $(\Delta S_{Xim}, \Delta S_{14})$, $(\Delta S_{EnRes}, \Delta S_{15})$.

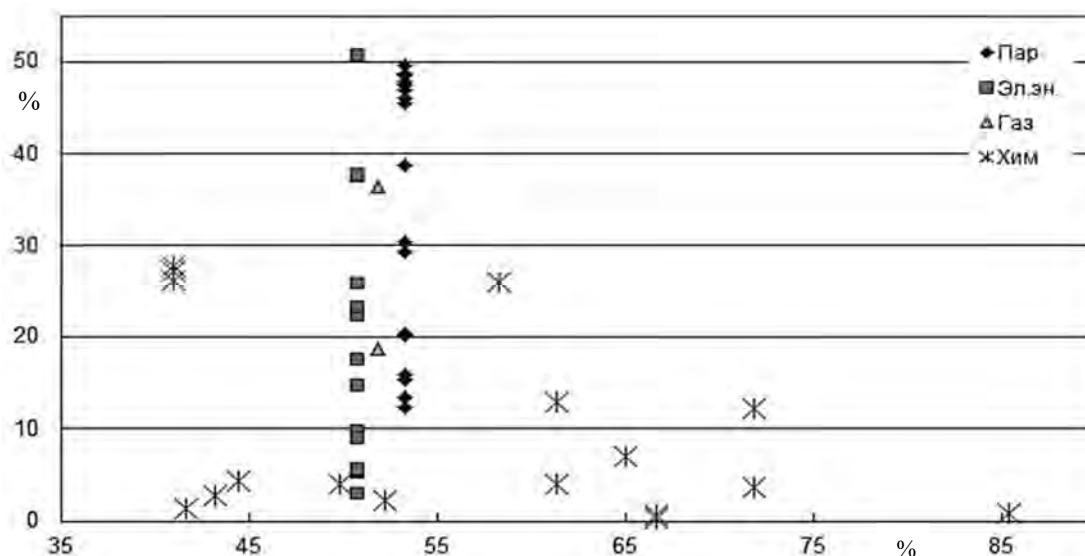


Рисунок 7.20 – Соотношения $(\Delta S_{Par}, \Delta S_{11})$, $(\Delta S_{ElEn}, \Delta S_{12})$, $(\Delta S_{Gaz}, \Delta S_{13})$, $(\Delta S_{Xim}, \Delta S_{14})$

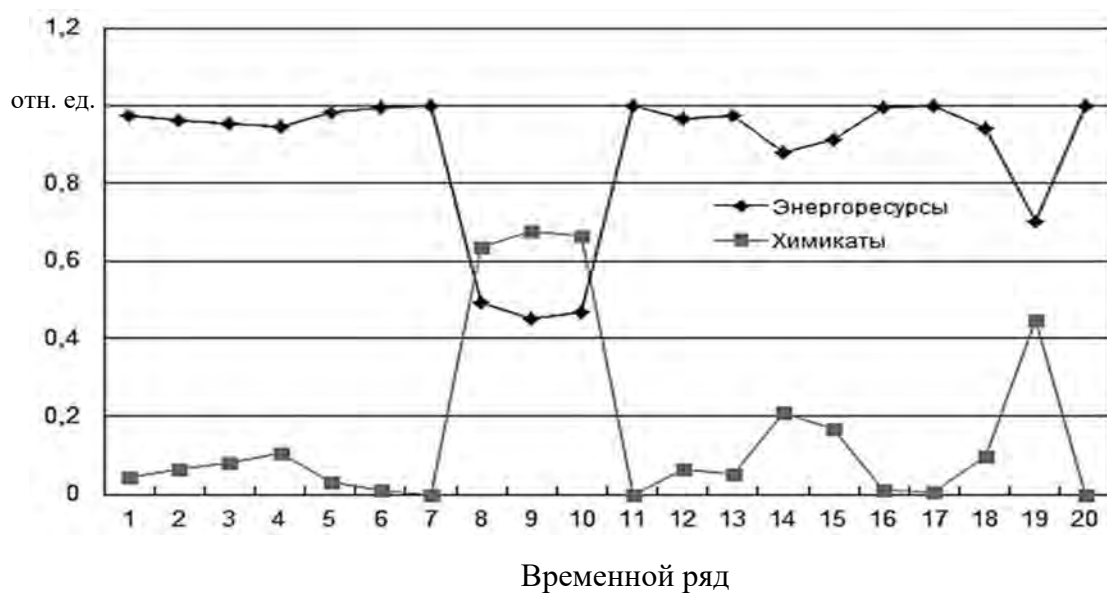


Рисунок 7.21 – Временные ряды чувствительности общей стоимости ресурсов к изменению цен на ресурсы

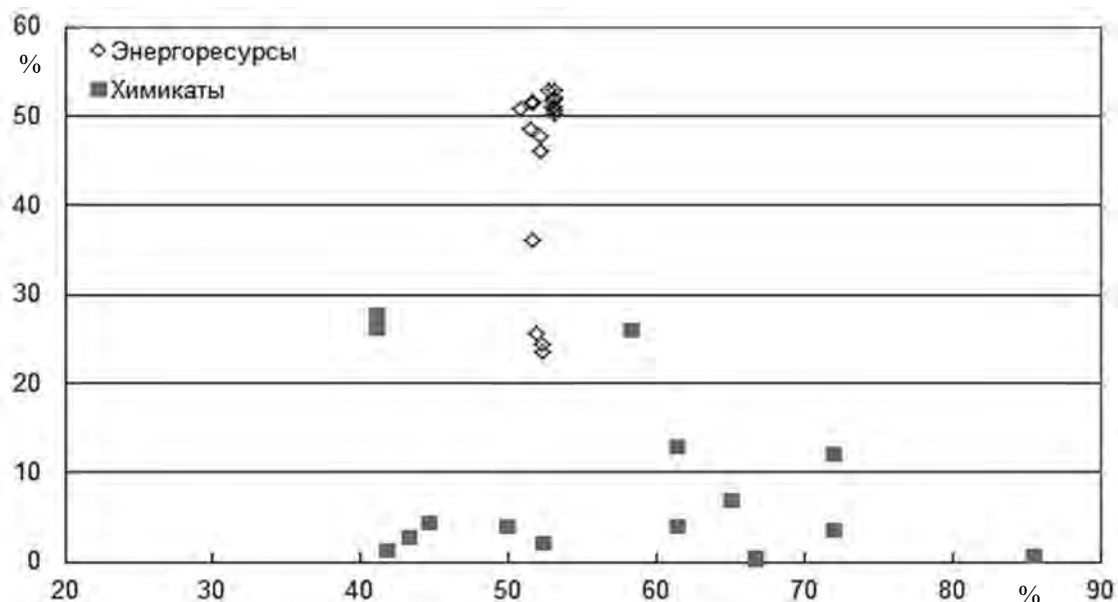


Рисунок 7.22 – Соотношения $(\Delta S_{EnRes}, \Delta S_{15})$, $(\Delta S_{Xim}, \Delta S_{14})$

Определены чувствительность общей стоимости обработки ткани к замене основных химикатов на химикаты-аналоги и чувствительность общей стоимости обработки ткани к изменению цен на ресурсы. При изменении цен на химматериалы в случае замены основных химикатов ТР на химикаты-аналоги до 139 % за период 1 исследования и до 109 % за период 2 общая стоимость ресурсов ПП КТНС изменяется до 5 % (период 1), до 3 % (период 2) соответственно. Изменение цен на химматериалы ТР ПП КТНС от периода 1 к периоду 2 оказывает меньшее влияние на изменение общей стоимости ресурсов ПП КТНС по сравнению с влиянием на общую стоимость ресурсов изменения цен на энергоресурсы.

Заключение

Предложен итерационный процесс M извлечения полезной информации из накопленных в базах данных комплексных информационных систем, представленный композицией операторов

$$P = T_6 \circ T_5 \circ T_4 \circ T_3 \circ T_2 \circ T_1,$$

где T_1 – отбор данных, накопленных в комплексных информационных системах предприятий, определение (с привлечением экспертов в исследуемой предметной области) переменных для решения поставленной задачи;

T_2 – предварительная обработка данных, устранение выбросов и пропущенных данных;

T_3 – выбор данных для исследования в соответствии с решаемой задачей, *отличается* формированием XML-файла для автоматизированной обработки;

T_4 – выбор метода поиска закономерностей, основанных главным образом на применении статистических исследований, *отличается* использованием метода сингулярного спектрального анализа и обобщенного закона распределения для оценки структуры распределения случайных величин;

T_5 – построение модели исходных данных для применения в имитационной модели на этапе ее эксплуатации;

T_6 – формирование XML-файла с моделью последовательности данных для последующего применения.

Для разработки технологии извлечения знаний из накопленных баз данных, представленных вещественными числами, исследован сингулярный спектральный метод (SSA-метод), включающий этапы: вложение; сингулярное разложение; группировка; диагональное усреднение.

Предложена методика исследования сингулярного спектрального анализа временных рядов с комплексным применением информационных технологий, основанная на использовании табличного процессора MS Excel, математического пакета Mathcad, пакета статистической обработки данных Statistica и *отличающаяся* применением сдвиговых процедур при формировании ганкелевой матрицы на этапе вложения и при восстановлении составляющих исходного временного ряда.

Показано, что шумовая составляющая восстановленного ряда для равномерного, экспоненциального и нормального распределений смещается в сторону нуля на величину математического ожидания шумовой составляющей исходного ряда с погрешностью до 6 % и возрастает до 30 % при росте дисперсии шума исходного ряда до 25 раз. Стандартное отклонение шумовой

составляющей восстановленного ряда отличается от исходного на 7 %...9 % во всех случаях.

Получено аналитическое выражение для нормированного временного ряда, включающее математическое ожидание элементов временного ряда, наибольшее по модулю значение элемента ряда, параметр масштаба, определяемый верхней границей нормированного временного ряда, параметром сдвига.

Построена математическая модель производственного процесса крашения тканей набивным способом в виде взвешенного ориентированного ациклического графа, вершинами которого являются технологические режимы (ТР), а нагрузки (стоимости ресурсов, временные затраты) на дуги графа в общем случае имеют стохастический характер.

В связи с существованием на этапах ПП КТНС множества допустимых технологических режимов ПП требуется группировка технологических режимов ПП, обеспечивающих схожие состояния ПП КТНС. Для выявления особенностей поведения ПП КТНС требуется анализ многомерных данных, в частности, решение задачи разделения множеств многомерных данных на непересекающиеся подмножества (кластеры). Для решения поставленной задачи приведены классификация и описание основных методов кластеризации.

Предложен метод проверки и уточнения результатов разделения многомерных наблюдений на кластеры с применением нескольких методов кластеризации данных. Метод состоит в последовательном применении нескольких методов кластерного анализа, а затем – в использовании специальной методики согласования полученных результатов кластеризации.

Предложена методика построения статистических моделей распределений для многомодальных выборок данных, сводящаяся к разделению исходной выборки методами кластерного анализа на несколько однородных выборок-кластеров с последующим построением на каждой из них своей функции плотности обобщенного распределения. Для проверки соответствия построенных моделей распределений данным применяются статистические критерии согласия Пирсона, Колмогорова–Смирнова, Мизеса.

Предложена методика определения последовательностей состояний/управлений ПП, обеспечивающих значения итоговых временных и/или стоимостных затрат, необходимых для обработки заданного количества продукции, близкие к оптимальному, в соответствии с критерием оптимальности, значению. Методика предполагает построение выборок итоговых стоимостей расхода ресурсов и времени для всех возможных вариантов сочетаний технологических режимов ПП с последующей их

обработкой методами статистического (построение статистических моделей распределения выборок) и кластерного (кластеризация многомерных данных с использованием нескольких методов кластерного анализа) анализа.

Проведено экспериментальное исследование метода, средств и методик оптимизации распределения ресурсов в производственном процессе крашения тканей набивным способом на примере данных действующего предприятия.

Список литературы

- 1 **Александров, Ф. И.** Разработка программного комплекса автоматического выделения и прогноза аддитивных компонент временных рядов в рамках подхода «Гусеница»–SSA: автореф. дис. ... канд. физ.-мат. наук: 05.13.18 / Ф. И. Александров; С.-Петерб. гос. ун-т. – Санкт-Петербург, 2006. – 15 с.
- 2 **Александров, Ф. И.** Разработка программного комплекса автоматического выделения и прогноза аддитивных компонент временных рядов в рамках подхода «Гусеница»–SSA: дис. ... канд. физ.-мат. наук: 05.13.18 / Ф. И. Александров. – Санкт-Петербург, 2006. – 152 л.
- 3 **Альховик, С. А.** Генетический алгоритм в задаче оптимизации плана грузоперевозок / С. А. Альховик, А. В. Сазоненко, А. А. Ковалевич // Изв. Гомел. гос. ун-та им. Ф. Скорины. – 2006. – № 4 (37). – С. 110–112.
- 4 **Андерсон, Т.** Статистический анализ временных рядов: пер. с англ. / Т. Андерсон. – Москва: Мир, 1976. – 755 с.: ил.
- 5 **Астер, Х.** Пигментная печать и экология. Мягкая химия: мечта и реальность / Х. Астер // Текстильная химия. – 1996. – № 1 (8). – С. 13–19.
- 6 Краткий курс химической технологии волокнистых материалов: учебник для технол. спец. вузов / Т. Д. Балашова [и др.]. – Москва: Легкая и пищевая пром-сть, 1984. – 199 с.
- 7 Методы и модели анализа данных: OLAP и Data Mining / А. А. Барсегян [и др.]. – Санкт-Петербург: БХВ-Петербург, 2004 – 336 с.: ил.
- 8 **Бельцов, В. М.** Оборудование текстильных отделочных предприятий: учебник / В. М. Бельцов. – 2-е изд., перераб. и доп. – Санкт-Петербург: СПГУТД, 2000. – 658 с.
- 9 **Битюкова, В. В.** Применение универсальных семейств распределений Пирсона для моделирования загруженности кабинетов лечебно-профилактических учреждений / В. В. Битюкова, А. А. Хвостов, Д. И. Ребриков // Вестн. ТГТУ. – 2008. – Т. 14, № 1. – С. 202–208.
- 10 **Большев, Л. Н.** Таблицы математической статистики / Л. Н. Большев, Н. В. Смирнов. – Москва: Наука, 1983. – 416 с.
- 11 **Борисов, В. В.** Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – 2-е изд., стер. – Москва : Горячая линия – Телеком, 2012. – 284 с. : ил.
- 12 **Борчик, Е. М.** Процедура построения кривых плотностей распределения Пирсона для многомодальных выборок / Е. М. Борчик, В. В. Башаримов, А. И. Якимов // Информационные технологии и системы 2011

(ИТС 2011): материалы Междунар. науч. конф., Минск, 26 окт. 2011 г. – Минск: БГУИР, 2011. – С. 181–182.

13 **Борчик, Е. М.** Автоматизация расчета затрат ресурсов в процессе производства тканей / Е. М. Борчик // Автоматизир. технологии и пр-ва. – 2016. – № 4. – С. 9–17.

14 **Борчик, Е. М.** Автоматизация расчета затрат химикатов в производстве тканей ОАО «Моготекс» [Электронный ресурс] / Е. М. Борчик // Автоматизация и управление в технических системах. – 2016. – № 2. – Режим доступа: auts.esrae.ru/19-381. – Дата доступа: 26.01.2019.

15 **Борчик, Е. М.** Выбор методов кластерного анализа при решении задач оптимизации в имитационном моделировании / Е. М. Борчик, В. В. Башаримов, А. И. Якимов // Информационные технологии, энергетика и экономика (информационные технологии, математическое моделирование технологических процессов, электроника): сб. тр. 7 Межрегион. (Междунар.) науч.-техн. конф. студентов и аспирантов, Смоленск, 8–9 апр. 2010 г.: в 3 т. – Смоленск, 2010. – Т. 2. – С. 21–26.

16 **Борчик, Е. М.** Обработка многомерных данных несколькими методами кластерного анализа / Е. М. Борчик, В. В. Башаримов // Информационные технологии и системы 2011 (ИТС 2011): материалы Междунар. науч. конф., Минск, 26 окт. 2011 г. – Минск: БГУИР, 2011. – С. 183–184.

17 **Борчик, Е. М.** Алгоритм кластерного анализа на основе нечетких отношений / Е. М. Борчик, В. В. Башаримов // Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях: материалы X Респ. науч. конф. студентов и аспирантов, Гомель, 12–14 марта 2007 г. – Гомель: ГГУ им. Ф. Скорины, 2007. – С. 307–308.

18 **Борчик, Е. М.** Многокритериальное управление технологическим процессом с использованием принципа оптимальности Беллмана / Е. М. Борчик, В. И. Аверченков, А. И. Якимов // Изв. Волгоград. гос. техн. ун-та. – 2014. – № 25 (152). – С. 95–101.

19 **Борчик, Е. М.** О совместном использовании методов кластерного анализа многомерных данных / Е. М. Борчик, А. И. Якимов, В. В. Башаримов // Докл. БГУИР. – 2011. – № 5 (59). – С. 95–102.

20 **Борчик, Е. М.** Определение классов целевых функций методами кластерного анализа / Е. М. Борчик, А. И. Якимов, В. В. Башаримов // Весн. Гродзен. дзярж. ун-та імя Янкі Купалы. – 2012. – № 1 (126). – С. 80–89.

21 Построение плотностей распределения Пирсона для многомодальных выборок с применением кластеризации / Е. М. Борчик [и др.] // Информационные технологии. – 2012. – № 12. – С. 30–35.

22 Применение распределений Джонсона к задаче классификации аэрокосмических изображений / Ю. Б. Буркатовская [и др.] // Изв. Том. политехн. ун-та. – 2007. – Т. 311, № 5. – С. 76–80.

23 **Вадзинский, Р. Н.** Справочник по вероятностным распределениям / Р. Н. Вадзинский. – Санкт-Петербург: Наука, 2001. – 295 с.

24 **Воронин, С. М.** Метод производящих функций / С. М. Воронин, А. Г. Кулагин // Квант. – 1984. – № 5. – С. 11–15.

25 **Гантмахер, Ф. Р.** Теория матриц / Ф. Р. Гантмахер. – 2-е изд., доп. – Москва: Наука, 1966. – 576 с.

26 **Гламаздин, Е. С.** Управление корпоративными программами: информационные системы и математические модели / Е. С. Гламаздин, Д. А. Новиков, А. В. Цветков. – Москва: ИПУ РАН, 2003. – 159 с.

27 **Гмурман, В. Е.** Теория вероятностей и математическая статистика: учебное пособие для вузов / В. Е. Гмурман. – Москва: Высшая школа, 2003. – 478 с.: ил.

28 **Дорошенко, Ю. А.** Использование методологии динамического программирования в процессе разработки инновационно-технологической стратегии предприятия / Ю. А. Дорошенко, И. В. Сомина // Вестн. БГТУ им. В. Г. Шухова. – 2014. – № 6. – С. 147–149.

29 **Дюк, В. А.** Поиск сложных непериодических шаблонов в последовательностях чисел и символов [Электронный ресурс] / В. А. Дюк // Business intelligence – effective data mining & analysis. – 2006. – Режим доступа: <http://www.datadiver.nw.ru/Articles/TSA.htm>. – Дата доступа: 12.03.2018.

30 **Журавлёва, Н. В.** Колорирование текстильных материалов: учебное пособие для вузов / Н. В. Журавлёва, М. В. Коновалова, М. А. Куликова. – Москва: МГТУ им А. Н. Косыгина, 2007. – 368 с.

31 **Зайцев, И. В.** Проектирование оптимальных химико-технологических систем на основе двухэтапной задачи оптимизации: автореф. дис. ... канд. техн. наук: 05.13.01 / И. В. Зайцев; Казан. нац. исслед. технол. ун-т. – Казань, 2013. – 19 с.

32 **Зайцева, Е. А.** Проверка сложных гипотез о согласии с распределениями Джонсона на основе использования непараметрических критериев / Е. А. Зайцева, Б. Ю. Лемешко, С. Н. Постовалов // Вестн. СибГАУ. – 2002. – Вып 3. – С. 70–77.

33 **Ивановский, Р. И.** Компьютерные технологии в науке и образовании. Практика применения систем MathCAD Pro: учебное пособие / Р. И. Ивановский. – Москва: Высшая школа, 2003. – 431 с.: ил.

34 **Йенсен, П.** Потокное программирование: пер. с англ. / П. Йенсен, Д. Барнес. – Москва: Радио и связь, 1984. – 392 с.: ил.

35 **Кашкин, В. Б.** Применение сингулярного спектрального анализа для выделения слабо выраженных трендов / В. Б. Кашкин, Т. В. Рублева // Изв. Том. политехн. ун-та. – 2007. – Т. 311, № 5. – С. 116–119.

36 **Кендалл, М.** Многомерный статистический анализ и временные ряды / М. Кендалл, А. Стюарт. – Москва: Наука, 1976. – 736 с.

37 **Кендалл, М.** Теория распределений: пер. с англ. / М. Кендалл, А. Стюарт. – Москва: Наука, 1966. – 588 с.

38 **Козелков, О. А.** Модели многокритериального оценивания и выбора вариантов технологических процессов / О. А. Козелков // Вестн. БГТУ им. В. Г. Шухова. – 2015. – № 2. – С. 177–180.

39 **Кричевский, Г. Е.** Химическая технология текстильных материалов: учебник в 3 т. / Г. Е. Кричевский. – Москва: ВЗИТЛП, 2000. – Т. 2. – 540 с.

40 **Крошилин, А. В.** Методологические основы построения интеллектуальной поддержки принятия решений в медицинских системах: автореф. дис. ... д-ра техн. наук: 05.13.10, 05.13.01 / А. В. Крошилин ; Рязан. гос. радиотехн. ун-т. – Рязань, 2014. – 39 с.

41 **Круглов, В. В.** Искусственные нейронные сети. Теория и практика / В. В. Круглов, В. В. Борисов. – Москва: Горячая линия – Телеком, 2001. – 382 с.: ил.

42 **Ланкастер, П.** Теория матриц: пер. с англ. / П. Ланкастер. – Москва: Наука, 1982. – 272 с.

43 Задачи и модели исследования операций / И. В. Максимей [и др.]. – Гомель: БелГУТ, 1999. – 150 с.: ил.

44 **Мандель, И. Д.** Кластерный анализ / И. Д. Мандель. – Москва: Финансы и статистика, 1988. – 176 с.

45 **Мартко, Е. О.** Сингулярный спектральный анализ как метод моделирования электрической нагрузки / Е. О. Мартко, И. В. Белицын // Ползунов. вестн. – 2009. – № 4. – С. 76–85.

46 Математическая энциклопедия / Под ред. И. М. Виноградова. – Москва: Совет. энцикл., 1985. – 623 с.

47 **Мельников, Б. Н.** Применение красителей: учебное пособие / Б. Н. Мельников, Т. Л. Щеглова, Г. И. Виноградов. – Москва: БИНОМ; Лаборатория знаний, 2010. – 331 с.

48 **Муха, В. С.** Статистические методы обработки данных: учебное пособие / В. С. Муха. – Минск: БГУ, 2009. – 183 с.

49 **Нешитой, В. В.** Элементы теории обобщенных распределений : монография / В. В. Нешитой. – Минск: РИВШ, 2009. – 204 с.

50 **Новак, В.** Математические принципы нечеткой логики: пер. с англ. / В. Новак, И. Перфильева, И. Мочкрож. – Москва: Физматлит, 2006. – 352 с.

51 **Орлов, А. И.** Часто ли распределение результатов наблюдений является нормальным? / А. И. Орлов // Заводская лаборатория. – 1991. – Т. 57, № 7. – С. 64–66.

52 Отделка хлопчатобумажных тканей: справочник в 2 ч. / Н. В. Егоров [и др.]. – Москва: Легпромбытиздат, 1991. – Ч. 1. – 432 с.

53 Отделка хлопчатобумажных тканей: справочник / Под ред. Б. Н. Мельникова. – Иваново: Талка, 2003. – 484 с.

54 **Паклин, Н.** Алгоритмы кластеризации на службе Data Mining [Электронный ресурс] / Н. Паклин // Технологии анализа данных. – 2011. – Режим доступа: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>. – Дата доступа: 12.03.2011.

55 **Панкратьев, П. С.** Поддержка принятия решений при многокритериальном двухуровневом выборе пунктов размещения электростанций: автореф. дис. ... канд. техн. наук: 05.13.01 / П. С. Панкратьев; Брат. гос. ун-т. – Иркутск, 2015. – 22 с.

56 **Попов, А. В.** Разработка метода построения негауссовских статистических моделей экспериментальных данных / А. В. Попов, И. Н. Колесник // Радиоэлектронные и компьютерные системы. – 2009. – № 3 (37). – С. 33–39.

57 Программный модуль расчета кривой плотности распределения случайной составляющей в последовательности данных BelSim2#.random: свидетельство о регистрации компьютерной программы № 306: заявка ВУ С20110024 / Е. А. Якимов, А. А. Ковалевич, Е. М. Борчик, В. В. Башаримов. – Оpubл. 04.04.2011.

58 **Рунион, Р.** Справочник по непараметрической статистике. Современный подход: пер. с англ. / Р. Рунион. – Москва: Финансы и статистика, 1982. – 198 с.

59 **Соболева, М. Ю.** Прогнозирование как средство повышения эффективности деятельности организаций / М. Ю. Соболева // Программные продукты и системы. – 2007. – № 1. – С. 24–27.

60 **Таран, Т. А.** Искусственный интеллект. Теория и приложения: учебное пособие / Т. А. Таран, Д. А. Зубов. – Луганск: ВНУ им. В. Даля, 2006. – 240 с.: ил.

61 **Тишаева, И. Р.** Алгоритмическое обеспечение систем поддержки принятия решения по выбору наилучшей доступной технологии в химическом производстве: автореф. дис. ... канд. техн. наук: 05.13.01 / И. Р. Тишаева; Моск. гос. ун-т тонких хим. технологий им. М. В. Ломоносова (МИТХТ). – Москва, 2014. – 25 с.

62 **Тюрин, Ю. Н.** Непараметрические методы статистики / Ю. Н. Тюрин. – Москва: Знание, 1978. – 64 с.

63 **Черноруцкий, И. Г.** Методы принятия решений: учебник / И. Г. Черноруцкий. – Санкт-Петербург: БХВ-Петербург, 2005. – 416 с.

64 **Щепкин, А. В.** Внутрифирменное управление (модели и методы) / А. В. Щепкин. – Москва: ИПУ РАН, 2001. – 80 с.

65 **Ягур, В. Е.** Новый подход к статистическому анализу биомедицинских данных / В. Е. Ягур, В. В. Нешиной, И. И. Саливон // Здоровоохранение. – 2009. – № 8. – С. 8–13.

66 **Якимов, А. И.** Автоматизация эксперимента на примере исследования генетического алгоритма / А. И. Якимов, В. В. Башаримов, С. А. Альховик // Автоматизация и современные технологии. – 2006. – № 5. – С. 3–9.

67 **Якимов, А. И.** Имитационное моделирование в ERP-системах управления / А. И. Якимов, С. А. Альховик. – Минск: Беларус. навука, 2005. – 197 с.: ил.

68 **Якимов, А. И.** Кластеризация состояний объекта при решении задачи выбора оптимальных технологических режимов / А. И. Якимов, Е. М. Максимов, Е. М. Борчик // Информационные технологии и вычислительные системы. – 2016. – № 4. – С. 33–43.

69 **Якимов, А. И.** Технология имитационного моделирования систем управления промышленных предприятий: монография / А. И. Якимов. – Могилев: Белорус.-Рос. ун-т, 2010. – 304 с.: ил.

70 **Якимов, Е. А.** Верификация организации распределенных вычислений в ПТКИ BelSim 2 / Е. А. Якимов, А. А. Ковалевич // Математика программных систем: межвуз. сб. науч. ст. – Пермь: Перм. гос. ун-т, 2008. – С. 103–109.

71 **Якимов, Е. А.** Исследование SSA-метода на основе комплексного применения информационных технологий / Е. А. Якимов // Докл. БГУИР. – 2010. – № 2 (48). – С. 77–83.

72 **Якимов, Е. А.** Исследование временных рядов с равномерным, нормальным и экспоненциальным шумом с помощью SSA-метода / Е. А. Якимов, В. Г. Замураев, А. И. Якимов // Вестн. Брест. гос. техн. ун-та. Сер. Физика, математика, информатика. – 2010. – № 5 (65). – С. 100–104.

73 **Якимов, Е. А.** Исследование детерминированных временных последовательностей данных на основе SSA-метода / Е. А. Якимов, И. В. Максимей // Информационные технологии, энергетика и экономика (информационные технологии, математическое моделирование технологических процессов, электроника): сб. тр. 7 Межрегион. (Международ.) науч.-техн. конф. студентов и аспирантов, Смоленск, 8–9 апр. 2010 г.: в 3 т. – Смоленск: Фил. ГОУ ВПО МЭИ (ТУ), 2010. – Т. 2. – С. 103–107.

74 **Якимов, Е. А.** К вопросу об исследовании SSA-метода для анализа временных последовательностей данных / Е. А. Якимов // Новые математические методы и компьютерные технологии в проектировании, производстве и научных исследованиях: материалы XIII Респ. науч. конф. студентов и аспирантов, Гомель, 15–17 марта 2010 г.: в 2 ч. – Гомель: ГГУ им. Ф. Скорины, 2010. – Ч. 1. – С. 61–62.

75 **Якимов, Е. А.** О преобразовании методом сингулярного спектрального анализа последовательностей данных с равномерным шумом / Е. А. Якимов, В. Г. Замураев, А. И. Якимов // Информатика. – 2011. – № 1. – С. 52–61.

76 О приведении последовательностей данных к нормированному виду для преобразования методом сингулярного спектрального анализа / Е. А. Якимов [и др.] // Изв. Гомел. гос. ун-та им. Ф. Скорины. – 2011. – № 6 (69). – С. 204–214.

77 **Ясницкий, Л. Н.** Введение в искусственный интеллект / Л. Н. Ясницкий. – Москва: Академия, 2005. – 176 с.: ил.

78 **Batarseh, O. G.** Reliable Simulation with Input Uncertainties Using an Interval-Based Approach [Electronic resource] / O. G. Batarseh, Y. Wang // Proceedings of 2008 Winter Simulation Conference. – 2008. – P. 344–352. – Mode of access: <http://www.informs-sim.org/wsc08papers/040.pdf>. – Data of access: 10.04.2019.

79 **Bause, F.** Comparison of Markovian Arrival and ARMA/ARTA Processes for the Modeling of Correlated Input Processes [Electronic resource] / F. Bause, P. Buchholz, J. Kriege // Proceedings of 2009 Winter Simulation Conference. – 2009. – P. 634–645. – Mode of access: <http://www.informs-sim.org/wsc09papers/061.pdf>. – Data of access: 11.04.2019.

80 Input Data Management Methodology for Discrete Event Simulation [Electronic resource] / N. Bengtsson [et al.] // Proceedings of 2009 Winter Simulation Conference. – 2009. – P. 1335–1344. – Mode of access: <http://www.informs-sim.org/wsc09papers/125.pdf>. – Data of access: 11.04.2019.

81 **Biller, B.** Capturing Parameter Uncertainty in Simulations with Correlated Inputs [Electronic resource] / B. Biller, C. Gunes // Proceedings of 2010 Winter Simulation Conference. – 2010. – P. 1167–1177. – Mode of access: <http://www.informs-sim.org/wsc10papers/107.pdf>. – Data of access: 11.04.2019.

82 **Biller, B.** Introduction to Simulation Input Modeling [Electronic resource] / B. Biller, C. Gunes // Proceedings of 2010 Winter Simulation Conference. – 2010. – P. 49–58. – Mode of access: <http://www.informs-sim.org/wsc10papers/006.pdf>. – Data of access: 10.04.2019.

83 **Elderton, W. P.** Frequency-curves and correlation / W. P. Elderton. – London: Charles and Edwin Layton, 1906. – 172 p.

84 **Elkan, C.** Using the triangle inequality to accelerate k-means / C. Elkan // Proc. Twentieth Int. Conf. on Machine Learning (ICML'03). – 2003. – P. 147–153.

85 **Frawley, W.** Knowledge Discovery in Databases: An Overview / W. Frawley, G. Piatetsky-Shapiro, C. Matheus // AI Magazine. – 1992. – P. 213–228.

86 **Golyandina, N.** Analysis of Time Series Structure: SSA and Related Techniques / N. Golyandina, V. Nekrutkin, A. Zhigljavsky. – Boca Raton: Chapman & Hall/CRC, 2001. – 310 p.

87 **Golyandina, N.** First-order SSA-errors for long time series: model examples of simple noisy signals / N. Golyandina, E. Vlassieva // Proceedings of the 6th St. Petersburg Workshop on Simulation, June 28 – July 4, 2009. – St. Petersburg: VVM com.Ltd., 2009. – Vol. 1. – P. 314–319.

88 **Hassani, H.** Singular Spectrum Analysis: Methodology and Comparison / H. Hassani // Journal of Data Science. – 2007. – № 5. – C. 239–257.

89 **Hastie, T.** The Elements of Statistical Learning: Data Mining, Inference, and Prediction / T. Hastie, R. Tibshirani, J. Friedman. – 2-nd ed. – Springer-Verlag, 2009. – 746 p.

90 **Hyndman, R. J.** The problem with Sturges' rule for constructing histograms [Electronic resource] / R. J. Hyndman. – 1995. – Mode of access: <http://www.robjhyndman.com/papers/sturges.pdf>. – Data of access: 12.04.2010.

91 **Kelton, D.** Representing and Generating Uncertainty Effectively [Electronic resource] / D. Kelton // Proceedings of 2009 Winter Simulation Conference. – 2009. – P. 40–44. – Mode of access: <http://www.informs-sim.org/wsc09papers/005.pdf>. – Data of access: 11.04.2019.

92 Introduction to Modeling and Generating Probabilistic Input Processes for Simulation [Electronic resource] / M. E. Kuhl [et al.] // Proceedings of 2009 Winter Simulation Conference. – 2009. – P. 184–202. – Mode of access: <http://www.informs-sim.org/wsc09papers/016.pdf>. – Data of access: 11.04.2019.

93 **Law, A. M.** ExpertFit Version 7 [Electronic resource] / A. M. Law // User's Guide.pdf/. – 2006. – 144 p. – Mode of access: <http://www.averill-law.com/>. – Data of access: 23.03.2020.

94 **Law, A. M.** Statistical Analysis of Simulation Output Data: The Practical State of the Art [Electronic resource] / A. M. Law // Proceedings of 2010 Winter Simulation Conference. – 2010. – P. 65–74. – Mode of access: <http://www.informs-sim.org/wsc10papers/008.pdf>. – Data of access: 10.04.2019.

95 **Nakayama, M.** Statistical Analysis of Simulation Output [Electronic resource] / M. Nakayama // Proceedings of 2008 Winter Simulation Conference. – 2008. – P. 62–72. – Mode of access: <http://www.informs-sim.org/wsc08papers/009.pdf>. – Data of access: 10.04.2019.

96 **Nelson, B. L.** The More Plot. Displaying Measures of Risk & Error From Simulation Output [Electronic resource] / B. L. Nelson // Proceedings of 2008 Winter Simulation Conference. – 2008. – P. 413–416. – Mode of access: <http://www.informs-sim.org/wsc08papers/048.pdf>. – Data of access: 10.04.2019.

97 **Perera, T.** Methodology for rapid identification of input data in the simulation of manufacturing systems / T. Perera, K. Liyanage // Simulation Practice and Theory. – 2000. – P. 645–656.

98 **Povinelli, R. J.** Time series data mining: Identifying temporal patterns for characterization and prediction of time series events / R. J. Povinelli // A Dissertation submitted of the Graduated School, Marquette University, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy. – Milwaukee, Wisconsin, 1999.

99 **Skoogh, A. A.** Methodology for Input Data Management in Discrete Event Simulation Projects [Electronic resource] / A. Skoogh, B. Johansson // Proceedings of 2008 Winter Simulation Conference. – 2008. – P. 1727–1735. – Mode of access: <http://www.informs-sim.org/wsc08papers/212.pdf>. – Data of access: 10.04.2019.

100 **Swain, J. J.** Least-squares estimation of distribution functions in Johnson's translation system / J. J. Swain, S. Venkatraman, J. R. Wilson // Journal of Statistical Computation and Simulation. – 1988. – № 29. – P. 271–297.

101 **Wagner, M. A. F.** Using univariate Bezier distributions to model simulation input processes / M. A. F. Wagner, J. R. Wilson // IIE Transactions. – 1996. – № 28 (9). – P. 699–711.

102 **Zouaoui, F.** Accounting for parameter uncertainty in simulation input modeling / F. Zouaoui, J. R. Wilson // IIE Transactions. – 2003. – № 35 (3). – P. 781–792.

Оглавление

Перечень сокращений и условных обозначений	3
Введение.....	4
1 Краткий обзор методов интеллектуального анализа данных в имитационном моделировании	6
1.1 Анализ эксплуатации программно-технологического комплекса имитации сложных систем BelSim	6
1.2 Исследование входных последовательностей данных в имитационном моделировании	10
1.3 Поиск сложных непериодических шаблонов в последовательностях данных	12
1.4 Data Mining в исследовании временных последовательностей данных.....	14
1.5 Особенности подготовки данных, накопленных в информационных системах	15
2 Моделирование производственного процесса	18
2.1 Содержательное описание и концептуальное моделирование производственного процесса крашения тканей набивным способом.....	18
2.2 Формализованное описание производственного процесса крашения тканей набивным способом	30
2.3 Обзор законов распределения статистических данных в производственном процессе крашения тканей набивным способом.....	33
2.4 Обзор методов кластерного анализа данных производственного процесса крашения тканей набивным способом	37
3 Метод извлечения полезной информации из последовательности числовых данных	39
3.1 Идея метода извлечения полезной информации из последовательности числовых данных для применения в имитационном моделировании	39
3.2 Содержательное описание SSA-метода	40
3.3 Методика исследования SSA-метода на основе информационных технологий.....	42
3.4 Технология исследования SSA-метода и пример его реализации	46
3.5 Предварительные результаты и их обсуждение	50

3.6 Исследование временных рядов с равномерным, нормальным и экспоненциальным шумом с помощью SSA-метода	53
4 Интеллектуальный анализ данных при оптимизации распределения ресурсов в процессе производства готовой продукции.....	64
4.1 Решение задачи многокритериального управления производственным процессом крашения тканей набивным способом	64
4.2 Построение кластеров состояний/управлений производственным процессом крашения тканей набивным способом по итогам решения задачи многокритериального управления производственным процессом.....	76
4.3 Кластерный анализ многомерных данных на этапах производственного процесса крашения тканей набивным способом.....	84
4.4 Статистические модели распределения данных многомодальных выборок при моделировании нагрузок на дуги графа производственного процесса крашения тканей набивным способом.....	97
5 Исследование методом сингулярного спектрального анализа числовых последовательностей данных.....	101
5.1 Исследование методом сингулярного спектрального анализа последовательностей числовых данных с равномерным шумом.....	101
5.2 Исследование преобразования методом сингулярного спектрального анализа последовательностей данных с нормальным шумом.....	110
5.3 Обоснование применения нормированного временного ряда.....	117
6 Реализация методов интеллектуального анализа данных в программном комплексе BelSim.....	130
6.1 Программный модуль для оценки кривой плотности распределения шумовой составляющей.....	130
6.2 Методика использования программного модуля.....	138
6.3 Кластеризация целевых функций	142
7 Практическое применение интеллектуального анализа данных для оптимизации распределения ресурсов в производственном процессе.....	151
7.1 Моделирование реального производственного процесса крашения тканей набивным способом	151

7.2 Оптимизация производственного процесса крашения тканей набивным способом	153
7.3 Кластеризация состояний объекта управления по итогам решения задачи многокритериального управления производственным процессом	156
7.4 Определение чувствительности общей стоимости обработки ткани к замене основных химикатов на химикаты-аналоги и к изменению цен на ресурсы	162
Заключение	169
Список литературы.....	172

Научное издание

Якимов Анатолий Иванович
Якимов Евгений Анатольевич
Борчик Екатерина Михайловна

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ ДЛЯ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ПРОИЗВОДСТВЕННЫХ СИСТЕМ

Редакторы *И. В. Голубцова, А. А. Подошевка*
Художественное оформление обложки *Е. М. Борчик, В. П. Бабичева*
Компьютерный дизайн *Н. П. Полевничая*

Подписано в печать 17.05.2021. Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. 10,70. Уч.-изд. л. 11,50. Тираж 100 экз. Заказ № 360.

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».

Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.

Пр-т Мира, 43, 212022, г. Могилев.