Вниманию комиссии представлен проект системы идентификации временных рядов и передаточных функций.

**Практически** в каждой области знания человека встречаются процессы, которые необходимо изучать в развитии во времени. Разнообразные данные в экономике, торговле, промышленности и других сферах человеческой деятельности поступают в форме временных рядов, в которых наблюдения зависимы и зачастую интересен характер этих зависимостей. Методы анализа таких рядов зависимых наблюдений называются анализом временных рядов. И анализ временных рядов является важнейшим инструментом изучения различных процессов (*дает возможность прогнозирования будущего развития исследуемых процессов*) Он позволяет получить наглядное представление развития процесса и получить краткое описание его характерных особенностей.

**Важной прикладной областью** анализа временных рядов (*моделей временных рядов и динамических систем*) является прогнозирование будущих значений временного ряда по его текущим и прошлым значениям.

Предметом исследования данного проекта являются:

\* временные ряды различной природы,

\* методика идентификации моделей,

\* построение прогнозов (будущих значений временного ряда) будущего развития процессов.

**Разработанная** (*автоматизированная*) система предназначена для решения задачи:

- идентификации временных рядов и передаточных функций, т. е. получения математических моделей развивающихся во времени процессов,

- и задачи прогнозирования будущего развития процесса (*ряда*) по полученным моделям (*временных рядов*).

**Основной целью создания** данной системы является автоматизация процессов:

+идентификации (*параметров моделей*) временных рядов, и передаточных функций

+и процесса прогнозирования по этим моделям.

Главной идеей создания данной системы являлось создание универсального, открытого для расширения и совершенствование, инструмента анализа временных рядов (*для идентификации моделей различных процессов, развивающихся со временем, и прогнозирования по этим моделям развития процессов в будущем*), который может быть полезен исследователям (*специалистам*) , которые в своей работе встречаются с анализом и прогнозированием (*, изменяющихся со времени*) процессов.

**Схема функциональной структуры** автоматизированной системы идентификации временных рядов и передаточных функций приведена на плакате 1.

Элементами автоматизированной системы являются подсистемы, реализующие соответствующие автоматизированные функции. В составе данной системы выделены следующие основные подсистемы:

* подсистема ввода данных;
* подсистема идентификации временных рядов;
* подсистема идентификации передаточных функций;
* подсистема прогнозирования;
* подсистема представления (*отображения*) результатов.

Каждая из подсистем также включает большое число специфичных функций, таких как:

- функции сведения к стационарности,

- функции определения сезонности,

- функции расчета параметров различных моделей и другие.

**По функциональному** назначению система относится к классу информационно-советующих систем. Система функционирует в автоматизированном режиме, т.е. в режиме, предусматривающем совместные действия пользователя и разрабатываемой системы для достижения результата работы.

**Общий алгоритм функционирования** представлен на плакате № 3. Алгоритм соответствует основным стадиям анализа временных рядов. (*Одна из ветвей соответствует прогнозированию с использованием модели передаточной функции.)*

Стадии (этапы) анализа временных рядов включают:

- графическое представление и описание поведения временного ряда;

- выделение и удаление закономерных составляющих временного ряда, зависящих от времени: тренда, сезонных и циклических составляющих;

- исследование случайной составляющей временного ряда, оставшейся после удаления перечисленных выше составляющих;

- построение (подбор) математической модели для описания слу­чайной составляющей;

- прогнозирование будущего развития процесса, представленного временным рядом;

- исследование зависимостей между временными рядами и прогнозирование при помощи упреждающих индикаторов (*с использованием модели передаточной функции*).

На этапе проектирования системы был рассмотрен ряд методов анализа временных рядов, из которых были выбраны для реализации наиболее эффективные модели, обладающие достаточной простой и минимальным числом параметров и при этом адекватно описывающие наблюдения.

**В системе были реализованы следующие методы**:

1- методы оценивания случайности данных (критерий поворотных точек, критерий длин фаз, критерий знаков разностей, критерий ранговой корреляции);

2- методы выделения неслучайной составляющей ряда (полиномиальная модель, скользящее среднее, сглаживание Фурье, взятие соседних и сезонных разностей);

3- методы вычислений характеристик ряда (автокорреляционная функции, частная автокорреляционная функция, взаимная корреляционная функция, гистограмма, спектр Фурье);

4- методы идентификации модели АРПСС ( линейно сходящийся процесс, квадратически сходящийся процесс, идентификация мультипликативной сезонной модели АРПСС);

5- методы прогнозирования (по модели АРПСС, методом экспоненциального сглаживания);

6- методы идентификации передаточных функций ( с предварительным выравниванием спектра входа, без выравнивания входа, идентификация многомерных передаточных функций) и прогнозирование по модели передаточной функции.

**Основные модели** представлены на плакате № 4

1 Модели тренда: полиномиальная модель, скользящее среднее, сглаживание Фурье

2 Модели стационарных временных рядов АР СС

3 Модели нестационарных временных рядов АРПСС

4 Сезонная мультипликативная модель

5 Модели передаточных функций

В некоторых случаях прогнозы ряда yt можно значительно улучшить, если использовать информацию, поступающую из какого-либо связанного с y ряда xt. Это особенно сильно проявляется в тех случаях, когда изменения x имеют тенденцию предварять изменения y, и в этом случае x называют «упреждающим индикатором» для y, а модель связи этих величин моделью передаточной функции.

Модели такого типа могут отражать не только ход технологических процессов, но и экономические, торговые и другие системы.

Для нахождения весов передаточной функции используются взаимная корреляционная функция входа и выхода и автокорреляционная функция входов.

Существуют алгоритмы идентификации передаточной функции с выравниванием и без выравнивания спектра входа. Эти алгоритмы отображены на соответствующих плакатах.

Модель, связанная с реальными данными, должна учитывать не только динамические соотношения входа и выхода, но и шум, существующий в системе. Такие совместные модели можно получить, комбинируя передаточную функцию со стохастической моделью шума.

*Временной ряд это последовательность наблюдений x(t1), x(t2), …, x(tN) анализируемой случайной величины, произведенных в последовательные моменты времени t1, t2,…, tN.*

1. Графическое представление и описание поведения временного ряда.

По графику временного ряда можно сделать довольно много выводов, которые в дальнейшем могут быть проверены и уточнены с помощью расчетов.

* наличие тренда и его характер;
* наличие сезонных и циклических компонент;
* степень плавности или прерывистости изменений последова­тельных значений ряда после устранения тренда (можно судить о характере и величине корреляции между соседними элементами).

Графическое представление ряда обычно задает направление его даль­нейшего анализа.

Для проверки временного ряда на случайность имеется ряд критериев, их перечисление приведено в работе [3]: *поворотные точки, длины фаз, критерий, основанный на знаках разностей, критерий, основанный на ранговой корреляции.* Выбор критерия зависит от того, какие гипотезы проверяются. Более подробно этот вопрос рассмотрен в [2], там же описаны требования, предъявляемые к подобным критериям, и приводятся основные методы сведения к стационарности.

Далее проводится статистический анализ данных , определение таких параметров как: минимальное и максимальное значение ряда, математическое ожидание, дисперсия, среднеквадратическое отклонение. Эти вопросы описываются в [1,4,5].

1. Выделение и удаление закономерных составляющих временного ряда, зависящих от времени: тренда, сезонных и циклических составляющих.

*Метод скользящего среднего – этот метод применяется для рядов с сезонными колебаниями и с неясной моделью тренда. Применяется для выявления общей тенденции изменения ряда, и основан на переходе от начальных значений ряда к средним на некотором интервале, который смещается вдоль ряда.* Скользящие средние подробно описаны в [2].

Метод вычисления *полиномиальной модели описан в [1,6] , в этом методе значения временного ряда x(t) рассматриваются как отклик, а t – как независимая переменная.* Параметры модели вычисляются методом наименьших квадратов, который подробно описан в [1, 5].

*Фурье сглаживание используется для выделения сезонной составляющей, метод описан в [7, 8].*

*Разностные операторы – метод заключается в переходе от исходного ряда к ряду разностей соседних значений ряда*, либо взятых через определенный интервал – сезон. [1, 4]

1. Исследование случайной составляющей временного ряда.

На этом этапе анализируется ряд остатков, полученный из исходного временного ряда после ис­ключения детерминированной компоненты. Для проверки стационарности ряда остатков и оценки его дисперсии используются выборочная автокорреляционная [9] и частная автокорреляционная функция. По ним делаются выводы о порядках моделей скользящего среднего и авторегрессии. Методы получения оценок этих характеристик временных рядов приведены в [2, 9,10].

1. Построение (подбор) математической модели для описания слу­чайной составляющей и проверка ее адекватности.

В качестве модели стационарных временных рядов чаще всего используются процессы авторегрессии, скользящего среднего и их комбинации. Эти модели и методы их идентификации описаны в источниках [1, 8, 11].

*Модель авторегрессии (АР) является исключительно полезной стохастической моделью для описания некоторых встречающихся на практике рядов.[11] В этой модели текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и им­пульса at. Порядок модели АР оценивается с помощью критерия Кенуя по частной автокорреляционной функции. Оценивание порядков моделей описано в [1, 4, 11]. Параметры модели получают из уравнений Юла-Уокера, через оценки автокорреляций.*

Другой тип моделей, имеющий большое значение в описании наблюдаемых временных рядов, — это *процесс скользящего среднего. Порядок модели оценивают по автокорреляционной функции критерием Бартлета.*

Для достижения большей гибкости в подгонке моделей к на­блюдаемым временным рядам иногда целесообразно объеди­нить в одной модели и авторегрессию, и скользящее среднее. Это приводит к комбинированной модели авторегрессии — сколь­зящего среднего. Метод идентификации такой модели подробно описан в [11], также описание методов идентификации этой модели имеется в [1, 8, 12]. Решение о выборе порядков (p,q) модели также принимают руководствуясь частной автокорреляционной функцией и автокорреляционной функцией [4].

Многие ряды обнаруживают нестационарный характер и, в част­ности, не колеблются относительно фиксированного среднего, но их свойства могут быть в некотором смысле однородными – поведение рядов с учетом различий в уровне оказывается во многом сходным. В книге [11] показано, что такой ряд может быть представлен обобщенным оператором авторегрессии. *Нестационарный процесс может быть описан моделью, которая требует, чтобы d-я разность про­цесса была стационарной. Такая модель эффективна для описания стационарных и неста­ционарных временных рядов и называется моделью авторе­грессии – проинтегрированного скользящего среднего (АРПСС)* порядка (р, d, q).[11]. Идентификация модели заключается в первую очередь в подборе порядка разности к модели. После выбора порядка d идентификация сводится к идентификации модели АРСС(p,q).

1. Прогнозирование будущего развития процесса, представленного временным рядом.

Наиболее эффективный подход к решению задач кратко и среднесрочного прогноза – это прогнозирование, основанное на использовании идентифицированных моделей типа АРПСС(p,q,k), включая в качестве частных случаев модели АР, СС, АРСС. Методы экспоненциального сглаживания подробно описаны в книге [12-14].

1. Исследование зависимостей между временными рядами и прогнозирование при помощи упреждающих индикаторов.

*В некоторых случаях прогнозы ряда Y можно значительно улучшить, если использовать информацию поступающую из какого-либо связанного с Y ряда X. Это особенно сильно проявляется в тех случаях, когда изменения X имеют тенденцию предварять изменения Y, и в этом случае X называют «упреждающим индикатором» для Y. В книгах [11,14] вводятся модели передаточных функций, связывающие выходы системы с одним или несколькими ее входами, и рассматриваются методы их идентификации.*

В [15] подробно рассматривается методики идентификации передаточной функции как без предварительного выравнивания спектра входа так и с предварительным выравниванием спектра входа, и построение модели шума. В этих алгоритмах используются взаимные ковариационная и корреляционная функции входа и выхода. Вычисление оценок этих характеристик подробно описывается [15-17].

Обычно при анализе временных рядов имеют дело с отсчетами временной области, если в работе нас интересует частотная информация о сигнале, получить ее позволяет дискретное преобразование Фурье [20, 21]. Преобразование Фурье является математической основой, которая связывает временной или пространственный сигнал с его представлением в частотной области [8].

Имеется временной ряд, и нас интересует вопрос, насколько сильно представлена синусоида частотой ω в этом зву­ке. ДПФ дает возможность ответить на этот вопрос. Основная идея ДПФ про­ста: для измерения амплитуды одной частоты нужно умножить имеющийся сигнал на синусоиду исследуемой частоты и сложить по­лученные отсчеты. По­вторяя ДПФ для разных частот, можно получить достаточно полное представление о спектре исследуемого ряда. Чтобы осуществлять точное обратное преобразование Фурье, помимо ампли­туды и частоты необходимо измерять фазу каждой частоты. Простое комп­лексное число - это двумерное значение, поэтому оно одновремен­но представляет и амплитуду, и фазу. В 60-х годах ХХ века Cooley и Tuckey открыли метод вычис­ления ДПФ, больше подходящий для использования на практике. Их алгоритм быстрого преобразования Фурье (БПФ) эффективно обсчитывает весь спектр сра­зу.

Подсистема *ввода-вывода* данных – данная подсистема обеспечивает ввод исходных данных из файлов различных форматов и сохранение результатов работы в удобном для пользователя формате.

Подсистема *идентификации временных* рядов – данная подсистема обеспечивает построение моделей временных рядов, проверку адекватности полученной модели.

Подсистема *идентификации передаточных* функций – данная подсистема обеспечивает построение моделей передаточных функции и проверку адекватности полученной модели.

Подсистема *прогнозирования* – данная подсистема обеспечивает прогнозирование будущего развития процесса, на основе полученной модели временного ряда, и прогнозирование будущего развития процесса (его выходного отклика) в зависимости от заданных входных воздействий, на основе полученной передаточной функции.

Подсистема *отображения результатов* – данная подсистема обеспечивает представление результатов работы системы в виде, понятном и удобном для пользователя (в виде графиков, таблиц).

**Структура ПО**, реализующего описанные методы, представлена на плакате № х.

Файлы исходных текстов, в соответствии со схемой функциональной структуры, могут быть разбиты на несколько частей:

Интерфейсную часть (ввод вывод, настройка параметров),

Реализации методов анализа временных рядов.

Реализации методов прогнозирования по моделям.

**В результате** работы над дипломным проектом была создана «система идентификации временных рядов и передаточных функций».

Разработанная система автоматизирует процессы идентификации моделей и вычисления прогнозов, что может существенно сократить время и усилия исследователя.

**Система была использована** для анализа данных о поступлении больных в наркологический диспансер с диагнозами зависимость и дилирей.

Решаемая задача в этом случае это идентификация модели временного ряда числа заболевших и прогнозирование по полученной модели числа обращений (тенденции распространения заболевания) на будущие периоды.

Были построены различные тренды по данным за все месяцы и линейные тренды для каждого в отдельности, построены гистограммы поступлений по дням недели, и построены прогнозы по модели АРПСС и сезонной модели.

Входными данными для автоматизированной подсистемы являются данные о предметной области, которые должны представлять собой временные ряды исследуемых параметров. Это должны быть ряды вещественных или целые чисел. Также данные, могут быть введены пользователем с клавиатуры или из буфера обмена при работе с системой.

Результатами работы пользователя с подсистемой являются вычисленные модели временных рядов и передаточных функций и прогнозы, построенные на основе этих моделей. Результаты предоставляются в виде:

* графиков прогнозов, которые могут быть сохранены файлах графического формата bmp;
* таблиц рассчитанных значений ( прогнозов, ошибок прогнозов, параметров моделей), которые могут быть сохранены столбцами значений в файлах форматов html, doc, xls, csv;
* текстовых отчетов (промежуточные результаты, и статистические характеристики рядов), которые могут быть сохранены в текстовых файлах.**Результат**

Разрабатываемая система должна обеспечивать дружественный интерфейс для взаимодействия пользователя с системой, обеспечивать возможность ввода и корректировки информации, а также возможность представлять результаты в различных формах, удобных для анализа и восприятия.

Полезным инструментом для специалистов, которые в своей работе встречаются с анализом и прогнозированием данных, изменяющихся со временем, может служить система идентификации моделей временных рядов.

В результате работы над дипломным проектом была создана «система идентификации временных рядов и передаточных функций».

По функциональному назначению система относится к классу информационно-советующих систем

В разработанной системе на основании идентифицированных моделей имеется возможность прогнозирования будущих значений временного ряда по его текущим и прошлым значениям, что может являться основой для планирования в экономике, производстве, коммерции.

Разработанная система автоматизирует процессы идентификации моделей и вычисления прогнозов, что может существенно сократить время и усилия исследователя.

**Пример**

Пример применения данной системы не связанный с производством может быть ее использование в медицинском учреждении.

Решаемая задача в этом случае это идентификация модели временного ряда числа заболевших и прогнозирование по полученной модели числа обращений (тенденции распространения заболевания) на будущие периоды. Или может быть поставлена другая задача: задача идентификация модели зависимости между количеством обращений больных с различными заболеваниями от метеорологических условий окружающей среды и прогнозирование числа обращений в зависимости от метеорологических условий.

Еще одним примером информационного объекта может выступать технологический объект, для которого требуется идентифицировать модель передаточной функции.

Недостатки аналогичных систем обработки информации.

Существует несколько систем, решающих те же задачи: STATGRAPHICS, SPSS, STATISTICA. Это иностранные универсальные статистические пакеты обработки данных, куда методы анализа временных рядов входят в качестве одной из составных частей. Они достаточно сложны для широкого применения, требуют соответствующих курсов подготовки. И лицензионные версии данных систем являются весьма дорогостоящими.

Специализированные пакеты сократили свое самостоятельное присутствие и часто входят интегрированными модулями с состав более крупных систем.

Современные пакеты обработки данных достаточно сложны имеют встроенные командные языки и часто требуют специализированных курсов обучения и сопутствующей литературы. В случае решения небольшого числа задач, функции таких универсальных коммерческих систем можно считать избыточными, а необходимость их использования при их немалой стоимости – недостатком.

Задача идентификации передаточных функций в современных пакетах статистической обработки данных если ставится, то решается методом множественной линейной регрессии, при том что многими авторами такая задача выделяется в отдельный класс [1], имеющий собственные методы решения [5]. Соответственно не оцениваются модели ошибок, на необходимость чего указывается в литературе.

В ранее разработанных на нашей кафедре интегрированных системах управления предприятием [30] задачи идентификация моделей временных рядов и передаточных функций решались двумя различными подсистемами, при том что задачи имеют пересекающиеся методы решения.

Metod

На этапе проектирования системы был рассмотрен ряд методов анализа временных рядов, из которых были выбраны для реализации наиболее эффективные модели, обладающие достаточной простой и минимальным числом параметров.

Элементами автоматизированной системы являются подсистемы, реализующие соответствующие автоматизированные функции

При практическом изучении временных рядов чаще всего ставятся следующие цели:

краткое (сжатое) описание, характерных особенностей ряда;

подбор статистической модели (моделей), описывающей вре­менной ряд (модели временных рядов должны обладать максимальной простой, минимальным числом параметров и адекватно описывать наблюдения);

предсказание будущих значений на основе прошлых наблюде­ний.

Модели временных рядов должны обладать максимальной простой, минимальным числом параметров и адекватно описывать наблюдения. Получение таких моделей важно так как: они могут помочь понять природу системы, генерирующей временные ряды; они могут быть использованы для прогнозирования будущих значений ряда; если исследуется два или несколько связанных временных рядов, можно расширить модели так, чтобы они описывали динамические взаимосвязи между рядами (модель передаточной функции).

Модель, связанная с реальными данными, должна учитывать не только динамические соотношения входа и выхода, но и шум, существующий в системе. Такие совместные модели можно получить, комбинируя передаточную функцию со стохастической моделью шума.

Модели такого типа могут отражать не только ход технологических процессов, но и экономические, торговые и другие системы.

Различные источники [3,4] с разной степенью детальности приводят схему анализа временных рядов:

1. графическое представление и описание поведения временного ряда;
2. выделение и удаление закономерных составляющих временного ряда, зависящих от времени: тренда, сезонных и циклических составляющих;
3. исследование случайной составляющей временного ряда, оставшейся после удаления перечисленных выше составляющих;
4. построение (подбор) математической модели для описания слу­чайной составляющей;
5. прогнозирование будущего развития процесса, представленного временным рядом;
6. исследование зависимостей между временными рядами и прогнозирование при помощи упреждающих индикаторов.