

Методы прогнозирования и согласования временных рядов

Ю. И. Журавлев, академик РАН,
К. В. Рудаков, член-корреспондент РАН,
А. Д. Корчагин, кандидат экономических наук,
М. П. Кузнецов, А. П. Мотренко, М. М. Стенина,
В. В. Стрижов, доктор физико-математических наук

При решении задач планирования в системах железнодорожного транспорта возникают проблемы связанные с нестационарностью, неравномерностью и высокой зашумленностью данных о грузоперевозках. Для повышения эффективности управления необходимо создание интеллектуальных систем, опирающихся на математические модели, исторические данные и формализованный опыт экспертов. Данная статья посвящена описанию проекта по созданию системы прогнозирования, направленной на повышение качества управления грузовыми железнодорожными перевозками путем выявления взаимосвязи объемов погрузки и спроса на грузовые железнодорожные перевозки с учетом экзогенных факторов.

Для повышения эффективности транспортировки грузов выполняется прогноз потребностей заказчиков ОАО «РЖД» в вагонах в узлах погрузки/разгрузки с учетом временных интервалов доставки. Системы железнодорожного транспорта являются сложными системами, которые характеризуются такими свойствами как слабая структурированность, нестационарность, высокая зашумленность информации. Поэтому эффективное управление такими системами предполагает создание интеллектуальных систем, которые наряду с точными математическими моделями используют данные и знания, накопленные в процессе функционирования, а также формализованный опыт экспертов. В текущей экономической ситуации дальнейшее использование экстенсивных методов развития производства и сферы услуг не приводит к желаемой экономической отдаче. Использование интенсивного подхода, связанного с привлечением современных научных методов анализа данных и оптимизации, должно улучшить показатели эффективности работы индустриального партнера и увеличить величину добавленной стоимости услуг железнодорожных грузоперевозок.

В рамках федеральной целевой программы «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2014 – 2020 годы» был поддержан проект по разработке интеллектуальной подсистемы прогнозирования объемов спроса на грузовые железнодорожные перевозки. Проект направлен на решение задачи повышения эффективности бизнес-процессов индустриального партнера – ОАО «РЖД». Целью проекта является разработка и адаптация математических методов и алгоритмов прогнозирования с учетом специфики задач, решаемых в интересах ОАО «РЖД», и создание системы прогнозирования на их основе. Применение разработанных методов и алгоритмов, проектирование и внедрение подсистемы прогнозирования, позволят повысить обоснованность, объективность и эффективность принятия решений в бизнес-процессах, связанных с заключением контрактов и планированием грузовых железнодорожных перевозок.

В настоящее время задача прогнозирования решается в основном на экспертном уровне на основе долгосрочных соглашений с потребителями и накопленного специалистами

опыта. При принятии управленческих решений на основе прогнозов спроса учитываются также такие факторы, как стоимость перевозки различных грузов, контрактные сроки доставки до места назначения и т.п. Отдельную проблему при планировании грузоперевозок порождают высокоскоростное приоритетное движение и нештатные ситуации.

В рамках реформы железнодорожного транспорта грузоотправителям была предоставлена возможность выбора поставщиков и видов транспорта. Эти изменения привели к необходимости прогнозирования спроса на перевозки грузов при планировании перевозок, то есть замене оперативного планирования перевозок их прогнозированием для определения реальных потребностей грузоотправителей в перевозке грузов.

Системы планирования грузоперевозок

Модель прогнозирования должна формировать прогноз спроса на отправление/погрузку грузов в заданном периоде: на месяц подекадно; на квартал помесечно; на год поквартально; на период больше года; и прогноз спроса на отправление/погрузку грузов с разложением: по группам грузов. Прогнозы разрабатываются для номенклатуры грузов, включающей 41 наименование типов грузов. Основным источником данных являются учетные системы индустриального партнера. Данные содержат информацию об отправлениях грузов: дату погрузки, станцию отправления, станцию назначения, количество вагонов, которые прошли по маршруту от станции отправления до станции назначения, код груза, род вагонов, суммарный вес груза в тоннах и признак маршрутной отправки.

Для решения поставленной научной задачи было проведено изучение ведущего мирового и отечественного опыта в области прогнозирования и оптимизации железнодорожных грузоперевозок, по результатам которого были выявлены ключевые методы с оценкой сферы применения и эффективности.

В настоящее время на рынке информационных технологий присутствует достаточное число программных продуктов по оптимизации тех или иных процессов, в том числе программных пакетов в области оптимизации логистических процессов и процессов планирования.

В частности, система поддержки принятия решений КСУМ (корпоративная система управления маркетингом) [1] опирается на результаты анализа экономической ситуации, оценки эффективности работы ОАО «РЖД» на рынке транспортных услуг и моделирования состояний развития рынка и его элементов. При этом система объединяет ключевые блоки, связанные с различными комплексами задач, центральное место в решении каждого из которых отводится эксперту.

Автоматизированная система оформления перевозочных документов ЭТРАН (Электронная ТРАНспортная Накладная) [2] также включает модуль планирования, который позволяет формировать месячные планы перевозок грузов в вагонах и тоннах в разрезе по видам сообщений и железным дорогам отправления, по номенклатуре грузов. Возможно также формирование оперативного плана перевозок грузов на основе принятых заявок грузоотправителей на перевозки грузов по календарным датам планируемого периода погрузки.

Данные системы основаны на базе продуктов компании SAS [3]. Программные продукты, предлагаемые компанией SAS, используются для объединения разрозненных данных из различных оперативных баз – от создания информационного хранилища отрасли и основанных на нем информационно-аналитических систем, до единой отраслевой интегрированной аналитической платформы.

Существующие системы прогнозирования ориентированы на долгосрочные и среднесрочные прогнозы с грубой детализацией: по месяцам, кварталам, полугодиям и годам,

необходимым для стратегического планирования, поэтому в них особое внимание уделяется макроэкономическому анализу [4, 5, 6] товарного и транспортного рынков.

Компоненты системы

Задача прогнозирования спроса на грузовые перевозки была поставлена для оперативного планирования перевозок по историческим данным с временным интервалом от суток до года. Для решения задачи необходимо решить ряд подзадач, связанных со спецификой прикладной области.

Исследования по разработке модели прогнозирования

Прогнозируемые временные ряды отличаются высокой волатильностью, наличием сезонности или тренда. Пример временного ряда с высокой волатильностью изображен на рисунке 1(а). Для разработки математической модели прогнозирования объемов спроса на грузовые железнодорожные перевозки необходимо провести исследование алгоритмов прогнозирования высоко волатильных и нестационарных временных рядов. Был предложен алгоритм непараметрического прогнозирования, основанный на минимизации ожидаемых потерь, аппроксимируемых сверткой плотности распределения исторических значений временного ряда с функцией ошибки [7]. Схему алгоритма отражает рисунок 1(б). Алгоритм позволяет достичь оптимального качества прогнозов при прогнозировании стационарных временных рядов, в том числе в случае несимметричных функций потерь, учитывающих экспертные оценки потерь при недопрогнозе и перепрогнозе. Для прогнозирования нестационарных временных рядов используются параметрические методы: векторной авторегрессии [8], авторегрессионного скользящего среднего [9, 10], интегрированная модель авторегрессионного скользящего среднего с учетом сезонности [11] в случае симметричных функций потерь, и комбинация этих методов с предложенным алгоритмом непараметрического прогнозирования в случае несимметричных функций потерь.

Учет влияния экзогенных факторов

Предполагается, что объемы спроса на грузовые железнодорожные перевозки существенно зависят от внешних факторов, таких как биржевые цены на основные инструменты, факторы климатического и административного характера, поэтому при создании моделей, методов и алгоритмов прогнозирования было решено учитывать как предысторию самих грузоперевозок, так и предысторию экзогенных факторов [12].

Поставлена задача выявления и исследования экзогенных факторов, а также исследования их влияния на объемы спроса на грузовые железнодорожные перевозки. Предполагается наличие ряда факторов, влияющих на динамику перевозок.

- Экономические факторы, связанные с объектами перевозок: цены на нефть и нефтепродукты, черные металлы, цветные металлы, удобрения, уголь.
- Курс рубля к доллару и/или к евро связан с объемом перевозок грузов, отгружаемых на экспорт (нефть и нефтепродукты, металлы, уголь).
- Сезонность производства природного-климатического характера, связанная с объемом перевозок зерна, овощей, бахчевых культур.
- Сезонность спроса на продукцию. Группы грузов, на которые оказывается влияние: строительные грузы (щебень, кирпич), цемент, промсырье. Помимо перечисленных факторов, на динамику перевозок влияют факторы нормативно-правового характера, связанные с ограничениями экспорта тех или иных родов груза решениями государственных органов, экспортными (импортными) пошлинами и конвенционными запретами и ограничениями ОАО «РЖД».

Для учета экзогенных факторов в модель прогнозирования включаются временные ряды, испытывающие влияние выделенных факторов. На рисунке 2 более светлой линией показано суммарное количество вагонов на фиксированной станции, более темной – цена за

баррель нефти. Временные ряды нормированы на отрезок $[-1,1]$. Цена за баррель является фактором, используемым для прогнозирования объема поставок. Таким образом, модель должна содержать возможность учета включения дополнительных временных рядов.

Прогноз с учетом топологии

При прогнозировании объемов спроса на грузовые железнодорожные перевозки требуется построить, во-первых, прогноз количества вагонов, проходящих через заданную станцию, и, во-вторых, прогноз количества вагонов, проходящих с одной станции на другую. Предполагается, что второй вариант прогноза, называемый экспертами ``шахматкой'', более точен и более востребован, т.к. уточняет расписание сложившихся грузоперевозок, в связи с чем необходимо разработать методы прогнозирования по парам пунктов отправления и назначения с учетом топологии железнодорожной сети.

Согласование прогнозов

Технические требования к системе прогнозирования предполагают возможность формирования прогнозов в различных разрезах: по типам грузов, пунктам отправления и назначения, и различной детализации по времени. При этом должно выполняться условие согласованности прогнозов. Например, сумма прогнозов объемов спроса на отдельные типы грузов должна совпадать с прогнозом суммарного объема спроса на грузоперевозки. Для независимых прогнозов это условие в общем случае не выполняется, в связи с чем поставлена задача согласования прогнозов временных рядов, обладающих иерархической структурой [13].

Под иерархией понимается наличие уравнения, задающего связь между рядами [13], в связи с чем ряды нельзя прогнозировать независимо. В частности, каждый прогнозируемый временной ряд может быть либо разбит на несколько временных рядов для получения прогнозов с большей детализацией (нижний уровень иерархии), либо агрегирован по времени или суммирован с другими временными рядами для получения прогноза с меньшей детализацией (верхний уровень иерархии). Таким образом сумма прогнозов временных рядов нижнего уровня иерархии должна совпадать с прогнозом их суммы – временного ряда высшего уровня иерархии. Схематически иерархическая структура прогнозируемых временных рядов отражена рисунком 3: вершинам неплоского графа, изображенного на рисунке, соответствуют временные ряды. Дочерним вершинам соответствуют временные ряды более низкого уровня иерархии, для которых фиксированы типы грузов (районы отправления/назначения), родительским – более высокого, описывающие суммарные перевозки всех типов грузов (по всем районам). Задача согласования прогнозов поставлена как задача минимизации отклонения скорректированных прогнозов с линейными ограничениями, соответствующими условиям согласованности и неувеличения ошибки прогнозирования. Предложено решение задачи согласования независимых прогнозов с учетом иерархии, гарантированно неухудшающее качество прогноза.

Анализ и сравнение моделей прогнозирования

Так как в различных условиях, определяемых заданными пользователем системы параметрами прогноза – детализацией, временным масштабом и горизонтом прогнозирования – оптимальными могут оказаться различные методы прогнозирования, одной из задач проекта являлось собрать воедино существующие и предложенные методы прогнозирования временных рядов и описать методику выбора оптимального алгоритма прогнозирования в заданном разрезе. При этом масштаб прогнозирования варьируется: рассматривается детализация прогнозов по дням, по неделям и по месяцам, а также по парам станций и по парам районов. Рисунок 4 обобщает результаты сравнения ряда алгоритмов – векторной авторегрессии, авторегрессионного скользящего среднего, а также прогнозов с

помощью выборочных медианы и среднего. Сравнение проводилось на основе среднего абсолютного отклонения MAE и среднего относительного абсолютного отклонения MAPE, вычисленных при ретроспективном прогнозе. Прогноз выполнялся по парам станций и парам районов, в различных временных масштабах. Согласно рисунку 4, хотя простейшие прогнозы с помощью выборочных медианы и среднего доставляют наибольшую ошибку в среднем, решение о преимуществе алгоритмов векторной авторегрессии и авторегрессионного скользящего среднего зависит от выбранного временного масштаба и разреза по станциям/районам. Каждому из перечисленных масштабов может соответствовать свой оптимальный алгоритм прогнозирования. Кроме того, результаты сравнения также зависят от выбранных критериев качества.

Оценки точности прогнозов должны отражать потери индустриального партнера при недостаточно точном прогнозе. В связи с этим для оценки качества прогнозов используются среднее абсолютное отклонение MAE и среднее относительное абсолютное отклонение MAPE, наиболее часто используемые при сравнении результатов прогнозирования [14]. Тем не менее, так ни один из способов оценки качества не является общепринятым, были предложены дополнительные критерии качества для выбора оптимальной модели, в том числе при прогнозировании временных рядов с переменным масштабом. Эти критерии включают статистическую оценку значимости различий между вычисленными значениями потерь различных методов прогнозирования и оценку максимального горизонта прогнозирования, при котором достигается заданный уровень достоверности прогноза. Кроме того, был предложен критерий согласованности, согласно которому при выборе из двух или нескольких методов прогнозирования стоит отдать предпочтение методу, для которого значения прогнозов меньше всего изменяются при согласовании.

Описан проект по созданию системы прогнозирования объемов спроса на грузовые железнодорожные перевозки, учитывающей влияние экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки, а также специфику бизнес-процессов и нормативов индустриального партнера – ОАО «РЖД». Прогнозирование потребностей в вагонах у заказчиков ОАО «РЖД» в узлах погрузки/разгрузки с учетом временных интервалов доставки, а также загруженности железнодорожных узлов является проблемой, которую необходимо решить для повышения эффективности транспортировки грузов.

Основными результатами исследования являются:

- 1) описание алгоритмов для решения задачи оценки влияния экзогенных факторов на объемы спроса на железнодорожные грузовые перевозки и методические материалы по использованию и внедрению алгоритмов прогнозирования объемов спроса,
- 2) макет модуля прогнозирования объемов спроса на грузовые железнодорожные перевозки,
- 3) программа и методика тестирования макета модуля прогнозирования объемов спроса на грузовые железнодорожные перевозки.

На основании разработанных алгоритмов в дальнейшем планируется создание промышленной подсистемы прогнозирования, позволяющей сократить издержки индустриального партнера бизнес-процессах планирования и оптимизации железнодорожных грузовых перевозок.

Работа выполнена финансовой поддержке Министерства образования и науки Российской Федерации (соглашение номер RFMEFI60414X0041).

ЛИТЕРАТУРА

- [1] М.М. Ковшова. Конкурентный маркетинг: стратегический подход. *Железнодорожный транспорт*, (2):2--4, 2002.
- [2] ЭТРАН (Электронная ТРАнспортная Накладная).
Url: <http://www.intellex.ru/projects/etran/>.
- [3] Официальный сайт компании SAS. Url: <http://www.sas.com/>.
- [4] А. А. Шананин. Вычислимая модель железнодорожных грузоперевозок с учетом коммуникационных ограничений. In *Тезисы докладов 10-й Международной конференции Интеллектуализация обработки информации. Греция, о. Крит, 2014.*, page 192--192.
- [5] А.В. Гасников. Заметка об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях. *Математическое моделирование*, page arXiv:1410.3123, 2015.
- [6] М.П. Ващенко, А.В. Гасников, Е.Г. Молчанов, Л.Я. Пospelова, А.А. Шананин. Вычислимые модели и численные методы для анализа тарифной политики железнодорожных грузоперевозок. In *Сообщения по прикладной математике*, pages 1--51. ВЦ РАН Москва, 2014.
- [7] А.С. Вальков, Е.М. Кожанов, М.М. Медведникова, Ф.И. Хусаинов. Непараметрическое прогнозирование загруженности системы железнодорожных узлов по историческим данным. *Машинное обучение и анализ данных*, 1(1):448--465, 2012.
- [8] Robert F. Engle and Clive W. J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):251--276, 1987.
- [9] P. Whittle. *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press, 1983.
- [10] M. Roopaei, M. Zolghadri, and A. Emadi. Economical forecasting by exogenous variables. In *IEEE International Conference on Fuzzy Systems*, pages 1491--1495, 2008.
- [11] George Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice-Hall, (third ed.) edition, 1994.
- [12] А.П. Мотренко, В.В. Стрижов. Построение агрегированных прогнозов объемов железнодорожных грузоперевозок с использованием расстояния Кульбака-Лейблера. *Информатика и ее применения*, 8(2):86--97, 2014.
- [13] М.М. Стенина, В.В. Стрижов. Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования. *Системы и средства информатики*, 24(2):21--34, 2014.

[14] J.T. Yokum and J.S. Armstrong. Beyond accuracy: Comparison of criteria used to select forecasting methods. *International Journal of Forecasting*, 11(4):591--597, 1995.

Сведения об авторах:

Ю. И. Журавлев,
академик РАН,
ФИЦ ИУ РАН, ВЦ РАН
zhur@ccas.ru

К. В. Рудаков,
член-корреспондент РАН,
ФИЦ ИУ РАН, ВЦ РАН
rudakov@ccas.ru

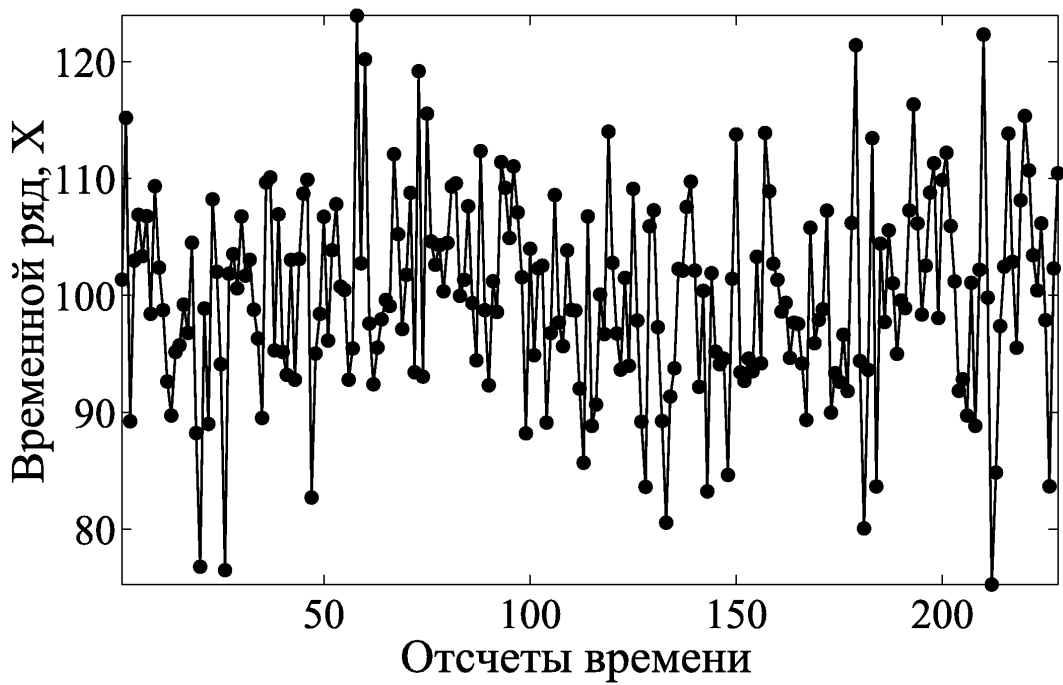
А. Д. Корчагин,
кандидат экономических наук,
Центр инновационного развития ОАО "РЖД",

М. П. Кузнецов,
аспирант, Московский физико-технический институт,
mikhail.kuznecov@phystech.edu

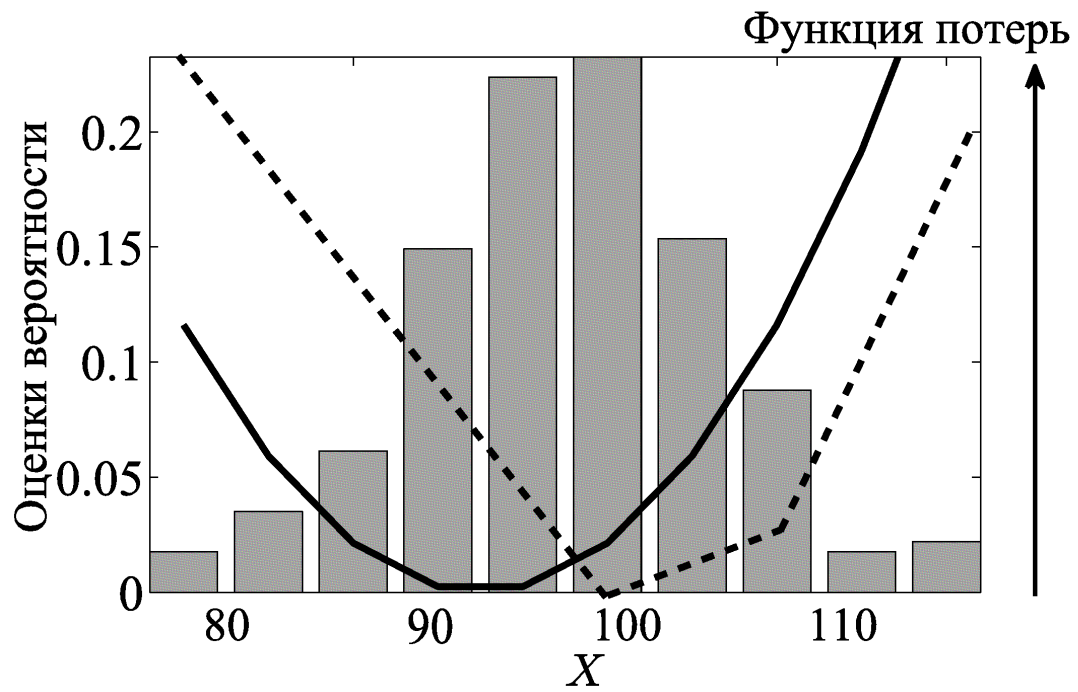
А. П. Мотренко,
аспирант, Московский физико-технический институт,
anastasiya.motrenko@phystech.edu

М. М. Стенина,
студент, Московский физико-технический институт,
mmedvednikova@gmail.com

В. В. Стрижов
доктор физико-математических наук,
ФИЦ ИУ РАН, ВЦ РАН
strijov@ccas.ru



(a)



(b)

Рис.1 – Демонстрация алгоритма гистограммного прогнозирования.

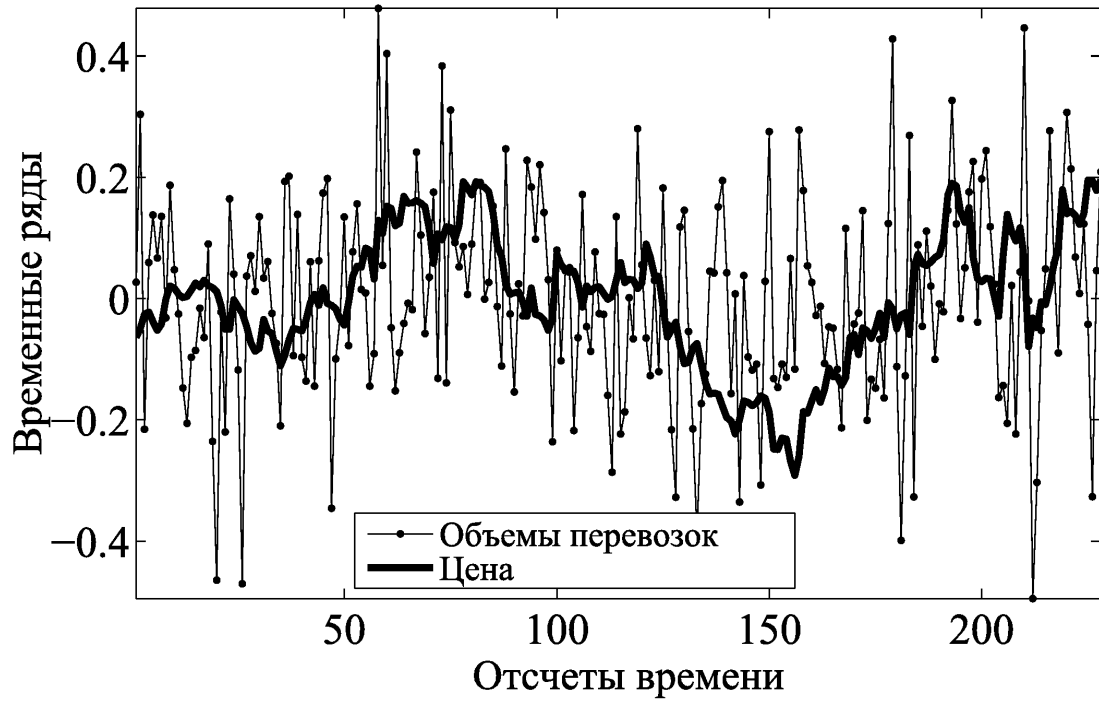


Рис.2 – Пример прогнозируемого временного ряда и внешнего временного ряда, испытывающего влияние экзогенного фактора. Временные ряды нормированы на отрезок $[-1,1]$.

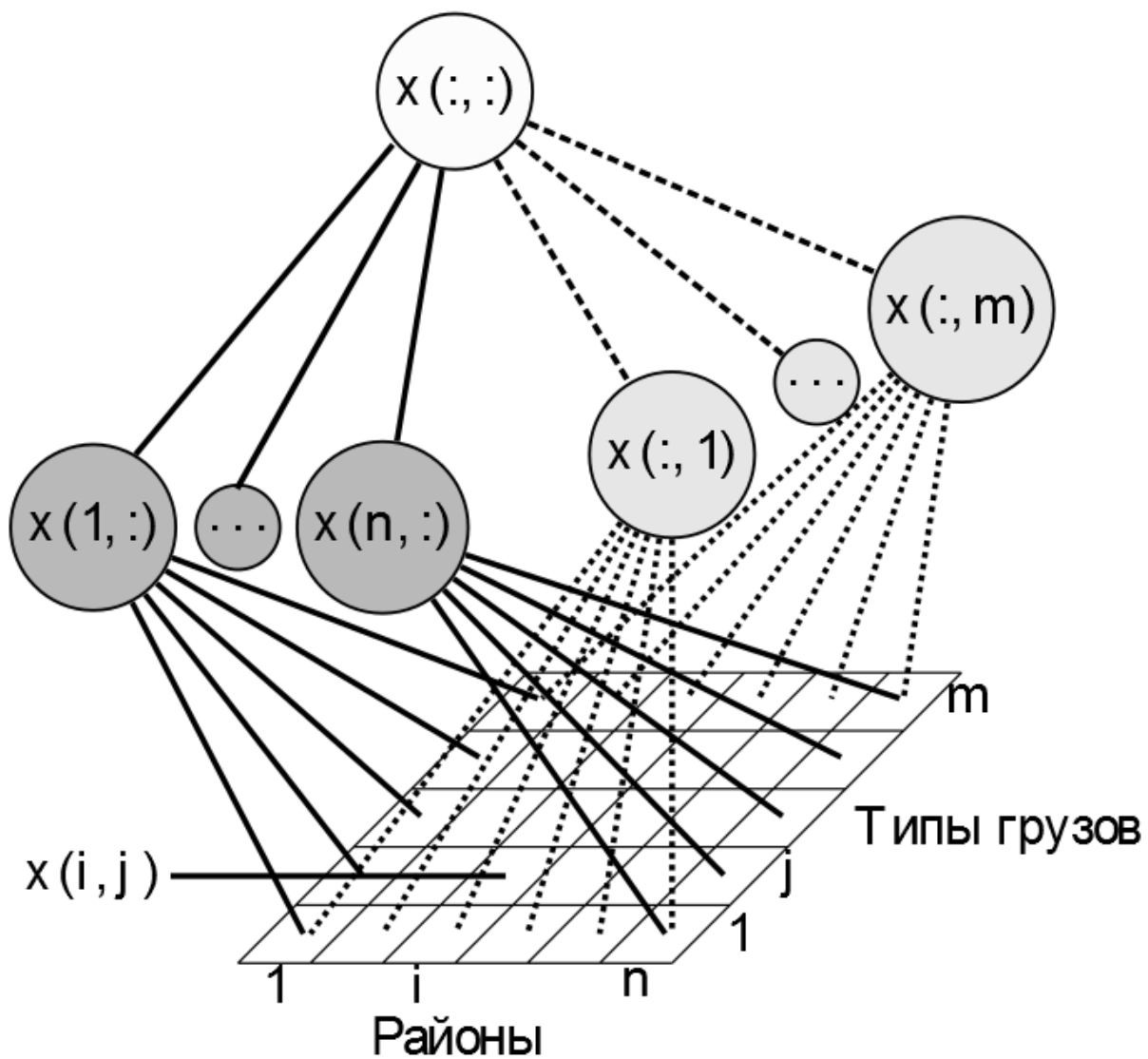


Рис.3 – Вид иерархической структуры прогнозируемых временных рядов.

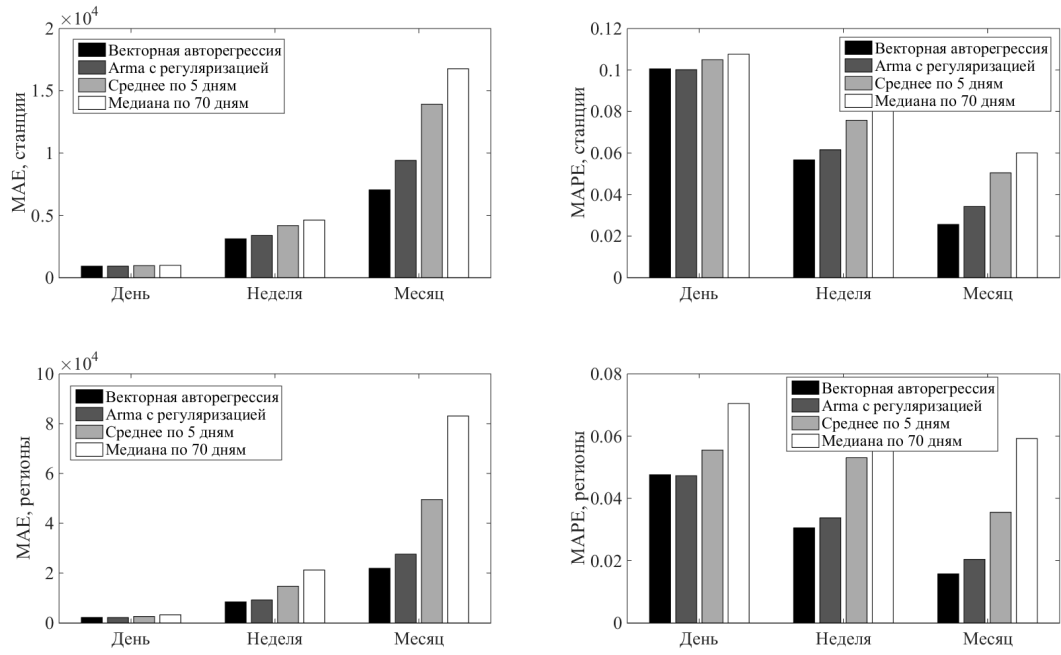


Рис. 4 – Сравнение алгоритмов прогнозирования в различных временных масштабах на основе ошибок MAE и MAPE, усредненных по парам станций/районов.