

ОТРАСЛЕВЫЕ И МЕЖОТРАСЛЕВЫЕ КОМПЛЕКСЫ

Для цитирования: Синицын Е. В., Толмачев А. В., Брусянин Д. А. Математическое моделирование региональных грузо- и пассажиропотоков // Экономика региона. — 2019. — Т. 15, вып. 4. — С. 1212-1225
<https://doi.org/10.17059/2019-4-19>
УДК 656.025

Е. В. Синицын^{а)}, А. В. Толмачев^{а, б)}, Д. А. Брусянин^{в)}

^{а)} Уральский федеральный университет им. первого Президента России Б. Н. Ельцина (Екатеринбург, Российская Федерация; e-mail: sinitsyn_ev@mail.ru)

^{б)} ООО Датател-Урал (Екатеринбург, Российская Федерация)

^{в)} Министерство транспорта и дорожного хозяйства Свердловской области (Екатеринбург, Российская Федерация)

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ РЕГИОНАЛЬНЫХ ГРУЗО- И ПАССАЖИРОПОТОКОВ¹

Разработка и реализация стратегий экономического и социального развития регионов России на период до 2035 года делает актуальным адекватное развитие транспортных услуг, затрагивающих все отрасли экономики и слои населения. В связи с этим в работе предложена модель, связывающая характеристики пассажирских и грузовых потоков с параметрами экономического и социального развития и демографической ситуацией в регионе. Такая модель позволяет конкретизировать нагрузку на транспортную систему, вытекающую из планов социально-экономического развития, а также планируемых решений в области хозяйственной деятельности. Для разработки модели были отобраны характеристики, описывающие экономическую ситуацию, рынок труда, демографические характеристики, уровень жизни и социальную ситуацию в анализируемом субъекте, имеющие наибольшие коэффициенты корреляции с анализируемыми параметрами транспортной инфраструктуры. Затем пошагово проводился регрессионный анализ, при этом на каждом шаге к уже имеющимся переменным модели добавлялась новая, дающая наибольший прирост коэффициента детерминированности R². Показано, что основным фактором, определяющим количество перевезенных автобусами общего пользования пассажиров, является среднегодовая численность занятых; на пассажирооборот наиболее сильно влияет численность населения. Количество перевезенных грузов определяется параметрами, характеризующими уровень развития производства (инвестициями в основной капитал и основными фондами в экономике, а также объемом отгруженных товаров собственного производства). Использование нелинейных моделей и сетей не привело к существенному уменьшению ошибок моделей. В работе также проведена кластеризация регионов РФ по показателям социально-экономического развития и характеристикам транспортной инфраструктуры, влияющим на транспортные потоки, оценена эффективность использования транспортной инфраструктуры в различных кластерах. Это позволит осуществлять целенаправленный бенчмаркинг — выбор субъектов для сравнения с анализируемым регионом.

Ключевые слова: пассажирские и грузовые потоки, пассажирооборот, грузооборот, социально-экономическое развитие, коэффициенты корреляции, многомерная регрессия, коэффициент детерминированности, интеллектуальный анализ данных, кластеризация, самоорганизующаяся карта Кохонена, метод *k*-средних, иерархическая структура кластеров

1. Введение

В настоящее время завершаются процессы разработки стратегических направлений раз-

вития различных отраслей экономики на период до 2035 г., в том числе транспорта. Это, в частности, порождает необходимость взаимной увязки стратегических целевых установок в различных отраслях, а также задачи прогнозирования пассажиро- и грузопотоков, целью которых является обеспечение со-

¹ © Синицын Е. В., Толмачев А. В., Брусянин Д. А. Текст. 2019.

ответствия транспортной инфраструктуры, маршрутов общественного транспорта, наличия подвижного состава поставленным задачам и заданным показателям экономического и социального развития. Необходимо принимать во внимание и социальную значимость транспортных услуг, затрагивающих практически все отрасли экономики и все население [1], ввиду чего проблемы с перевозками пассажиров и грузов могут стать источником социальной напряженности и макроэкономических проблем.

Необходим переход от простого отслеживания при планировании логической взаимосвязи между установочными параметрами социально-экономического развития региона и показателями деятельности транспорта к применению более надежных статистически установленных математических зависимостей. Не претендуя на полное решение всех возникающих на этом пути проблем, в данной работе мы рассмотрим достаточно простую модель, позволяющую связать показатели деятельности транспорта с отдельными показателями, характеризующими состояние экономики, демографию и другие характеристики региона. Это позволяет прогнозировать результаты деятельности транспорта, отталкиваясь от данных социально-экономических прогнозов и планов по развитию транспортной инфраструктуры. Подобные модели использовались нами, в частности, при разработке стратегии развития транспорта Свердловской области.

Математическое моделирование транспортных систем традиционно привлекает внимание исследователей (см., например, [2], где обобщены математические модели прогнозирования транспортных заторов, выбора оптимальной топологии транспортной сети, подходы к моделированию транспортных систем на основе теории графов). Подходы к решению задач краткосрочного прогнозирования транспортных потоков для конкретных дорожных сетей на основе регрессионного анализа рассматривались в [3–5], метода Бокса — Дженкинса [6] и его аналогов для временных рядов в [7–9]. Наконец, в ряде работ для целей краткосрочного прогнозирования потоков в транспортных системах использовались методы интеллектуального анализа данных: например, метод опорных векторов [10], нейронных сетей [11–14].

Отметим, что в перечисленных работах речь идет преимущественно о решении задач краткосрочного прогнозирования транспортных потоков в условиях конкретной транс-

портной инфраструктуры, например, улично-дорожной сети. Это, безусловно, интересно с позиций математического моделирования и, например, организации оперативного управления движением, однако для решения задачи долгосрочного планирования развития транспортных систем описанные выше подходы неудобны. Получение необходимых результатов, например, для пассажиро- и грузопотоков, на уровне области наталкивается на большие сложности как при сборе исходной информации для расчета, так и при выполнении самих вычислительных процедур.

В связи с этим, мы, как и в работе [15], для установления взаимосвязи характеристик пассажиро- и грузопотоков с экономическими и другими показателями региона, отражаемыми в программах социально-экономического развития, предлагаем использовать наглядную и достаточно простую модель на основе многомерной регрессии [16–19]. Данная модель будет описана в разделе 4.1. Будет также рассмотрено ее расширение для учета возможной нелинейной связи перечисленных параметров.

В самой общей постановке необходимо исследовать различные виды транспорта, в том числе общественный и личный, сухопутный, воздушный, речной и другие. Однако в этом случае для построения адекватных моделей необходимы сбор и обработка данных, которые, к сожалению, в полном виде отсутствуют, а организация их сбора, обеспечения целостности и полноты является самостоятельной и достаточно сложной задачей. Поэтому в данной работе мы ограничимся рассмотрением пассажирских и грузовых потоков, связанных с автомобильным транспортом.

Следует отметить еще одну задачу, возникающую в связи с долгосрочным планированием развития транспорта: бенчмаркинг — поиск ориентиров для сравнения, то есть регионов, близких в каком-то смысле к рассматриваемому, так что его сравнение с такими ориентирами может служить основой для управленческих решений. В частности, при анализе показателей развития транспортных услуг в Свердловской области сравнение часто производится с Тюменской и Челябинской областями, Пермским краем и республикой Татарстан. Разумеется, при выборе подобных ориентиров неизбежно возникает риск волюнтаризма, поэтому целесообразно проанализировать возможность подобной кластеризации и выделения пригодных для сопоставления регионов на основе методов интеллектуального

анализа данных [19–25]. Это будет сделано в разделе 4.2.

Таким образом, целью данной работы является построение математических моделей, увязывающих характеристики пассажирских и грузовых потоков, связанных с деятельностью автомобильного транспорта с параметрами, характеризующими регион. В качестве источника данных ниже будут использованы данные Росстата по субъектам РФ¹ на конец 2017 г.

2. Методы анализа

Рассматриваемые модели базируются на использовании методов статистического анализа:

- корреляционный анализ и многомерная регрессия [16–19],
- методы интеллектуального анализа данных [20–25], в том числе:
- кластеризация методом k -средних [19, 23];
- кластеризация с помощью самоорганизующихся карт Кохонена [25];
- технология нейронных сетей [20–22].

3. Результаты

3.1. Общая информация и выбор показателей для анализа

В качестве характеристик пассажирских и грузовых потоков мы рассматривали следующие показатели:

- перевозки грузов, млн т;
 - грузооборот, млн т/км;
 - перевозки пассажиров автобусами общего пользования, млн чел.;
 - пассажирооборот автобусов общего пользования, млн пасс/км
- (1)

На первом шаге из показателей регионов, представленных на сайте Федеральной службы государственной статистики, были отброшены показатели, явно не имеющие логической связи с вышеперечисленными характеристиками грузовых и пассажирских потоков (1). Кроме того, были рассчитаны коэффициенты корреляции остальных показателей, характеризующих регион, и характеристик пассажирских и грузовых потоков (1) по всем субъектам РФ. Показатели, имеющие нулевые коэффициенты корреляции с параметрами (1) на уровне

значимости 5 % [26], также не принимались во внимание. В итоге для анализа был отобран набор показателей по каждому субъекту, представленный в таблице 1.

Как видно, показатели можно разделить на пять групп:

1. Характеристики транспортной инфраструктуры (показатели 1–4).
2. Общие характеристики субъекта (5–7).
3. Макроэкономические характеристики (8–15).
4. Показатели, характеризующие уровень жизни населения (16–18).
5. Отдельные социальные показатели, способные оказывать влияние на характеристики пассажирских потоков (19–22).

Плотность железнодорожных путей включена в число анализируемых показателей, поскольку доставка грузов до конечного потребителя зачастую осуществляется с помощью автомобильного транспорта и, следовательно, потенциально развитие железнодорожной инфраструктуры может одновременно содействовать увеличению грузооборота автомобильного транспорта. Коэффициент корреляции между этими показателями, действительно положителен, хотя и невелик (0,15).

Отметим, что показатель 3 из таблицы 1, как ни странно, не имеет значимых коэффициентов корреляции с показателями пассажиропотока, что заставляет задуматься об эффективности использования имеющегося автобусного парка.

3.2. Эффективность использования транспортной инфраструктуры

Содержание и развитие различных элементов транспортной инфраструктуры требуют существенных текущих и капитальных затрат, в том числе финансируемых из средств бюджетов различных уровней. В связи с этим актуальным является вопрос об эффективности использования этих элементов. Например, для оценки эффективности использования дорожной сети в различных субъектах РФ можно использовать характеристику, представленную ниже:

$$I = \frac{Q}{L}. \quad (2)$$

Здесь I — интенсивность использования (нагрузка) дорожной сети общего пользования: количество перевезенных грузов (Q), приходящее в среднем на 1 км дорог; L — длина дорог. Следует отметить, что характеристика (2) и сопоставление ее величин в различных субъек-

¹ Регионы России. Социально-экономические показатели. 2018: Р32 Стат. сб. М., Росстат. 2018 г. 1162 с. [Электронный ресурс]. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156 (дата обращения: 01.08.2019).

Таблица 1

Показатели субъектов РФ, используемые для построения математических моделей

№ пп	Показатель	Единицы измерения
1	Плотность железнодорожных путей на конец года	км путей на 10000 км ² территории
2	Плотность автомобильных дорог общего пользования с твердым покрытием, на конец года	км дорог на 10000 км ² территории
3	Число автобусов общего пользования на 100 000 чел. населения, на конец года	штук
4	Число дорожно-транспортных происшествий на 100000 чел. населения	штук
5	Площадь территории	тыс. км ²
6	Численность населения на 01.01.2018 г.	тыс. чел.
7	Количество муниципальных образований в субъекте — всего	штук
8	ВРП по субъектам Российской Федерации в текущих основных ценах*	млн руб.
9	Инвестиции в основной капитал	млн руб.
10	Основные фонды в экономике (по полной учетной стоимости; на конец года)	млн руб.
11	Объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами	млн руб.
12	Продукция сельского хозяйства — всего	млн руб.
13	Ввод в действие жилых домов,	тыс. м ² общей площади жилых помещений
14	Оборот розничной торговли	млн руб.
15	Сальдированный финансовый результат (прибыль минус убыток) деятельности организаций)	млн руб.
16	Среднегодовая численность занятых	тыс. чел.
17	Среднедушевые денежные доходы (в месяц)	руб.
18	Потребительские расходы в среднем на душу населения (в месяц)	руб.
19	Число собственных легковых автомобилей	штук на 1000 чел. населения
20	Студенты организаций — всего (бакалавриат, специалитет, магистратуры)	тыс. чел.
21	Численность пенсионеров	пенсионеры на 1000 чел. населения
22	Численность врачей всех специальностей	врачи на 10 000 чел. населения

* Данные на конец 2016 г.

тах могут оказаться полезными при определении объемов средств, выделяемых субъектам на поддержание дорожной сети в нормативном состоянии по федеральным программам.

При анализе перевозок пассажиров целесообразно начать с оценки транспортной мобильности населения в различных субъектах. Для этого будет использоваться параметр, показывающий количество пассажиро-километров, приходящееся на 1 человека:

$$M = \frac{PT}{N_p}, \quad (3)$$

где PT — пассажирооборот (пасс/км), а N_p (чел.) — численность населения соответствующего субъекта. Распределение данной величины по субъектам РФ представлено на рисунке 1.

Следует обратить внимание на то, что распределение на рисунке 1 достаточно симметрично, это позволяет предположить, что

транспортная мобильность в значительной степени складывается под воздействием случайных причин.

Наконец, эффективность использования автобусного парка можно охарактеризовать количеством (q) пассажиров, перевозимых в среднем за год одним автобусом:

$$q = \frac{Q_p}{N_B}, \quad (4)$$

где Q_p — общее количество пассажиров, перевезенных автобусами в субъекте РФ, а N_B — количество автобусов. Кроме того, можно оценить количество километров, пройденных за год одним автобусом:

$$l_B = \frac{PT}{q}. \quad (5)$$

Здесь, как и в (3), PT — пассажирооборот, а q определено (4).

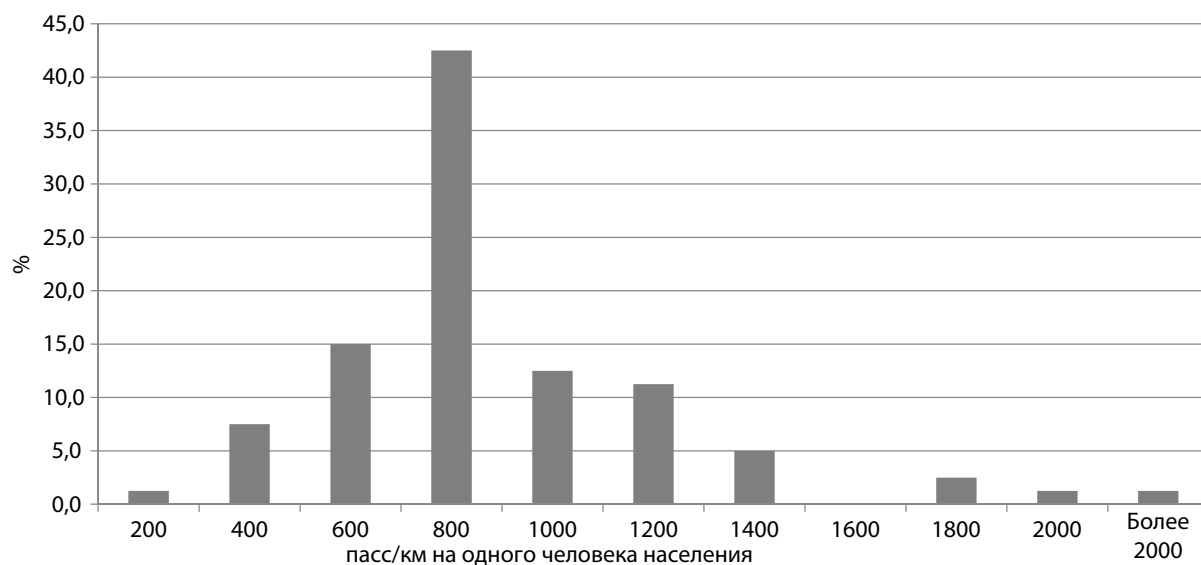


Рис. 1. Распределение субъектов РФ по мобильности населения (2)

Распределение величины l_B по субъектам РФ представлено на рисунке 2.

Среднее количество пассажиров, перевозимых за год одним автобусом, иллюстрируется рисунком 3.

Разумеется, интенсивное использование автобусного парка (большое количество перевозимых пассажиров, большое расстояние, проходимое в среднем одним автобусом за год) приводит к его повышенному износу, поэтому данные, представленные на рисунках 2, 3, целесообразно использовать при планировании выделения средств бюджетов различного уровня на обновление и модернизацию парка подвижного состава.

4. Математические модели пассажирских и транспортных потоков. Обсуждение результатов

Как видно из результатов предыдущего раздела, основные показатели, характеризующие

эффективность использования транспортной инфраструктуры, так или иначе, вычисляются на основании базовых характеристик транспортных потоков, представленных в (1). Поэтому в данном разделе мы рассмотрим возможные математические модели для их вычисления.

4.1. Многомерная регрессия

Для определения зависимости характеристик (1) будем использовать пошаговую многомерную линейную регрессию. При этом на первом шаге в качестве исходной переменной, используемой для вычисления искомой характеристики, применяется показатель из таблицы 1, имеющий наибольший коэффициент корреляции с рассматриваемой характеристикой. Затем, на втором шаге по очереди рассматриваются остальные оставшиеся показатели таблицы 1 и выбирается показатель, дающий наибольшее увеличение коэффици-

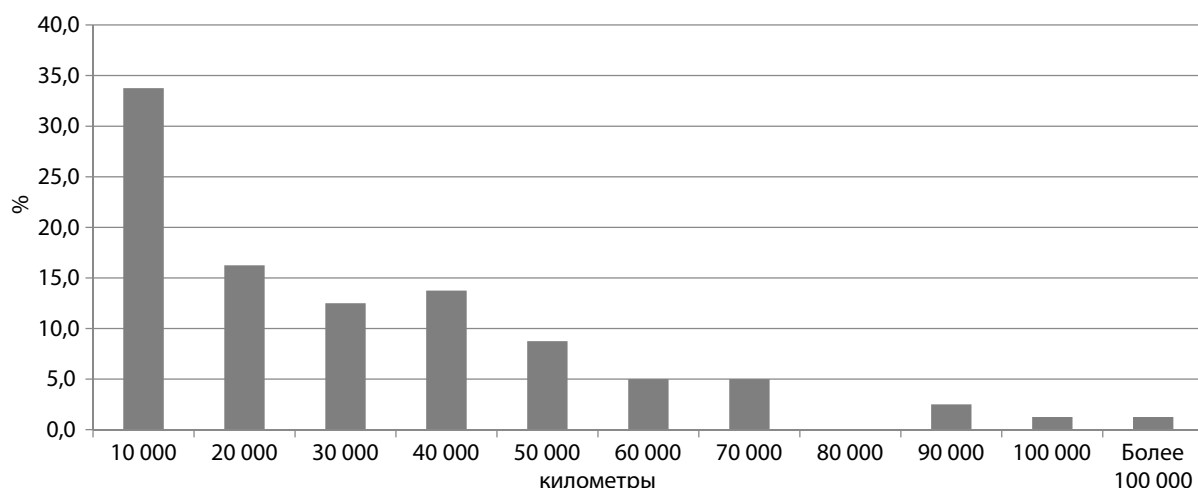


Рис. 2. Распределение расстояния, проходимого в среднем за год одним автобусом (5) в субъектах РФ

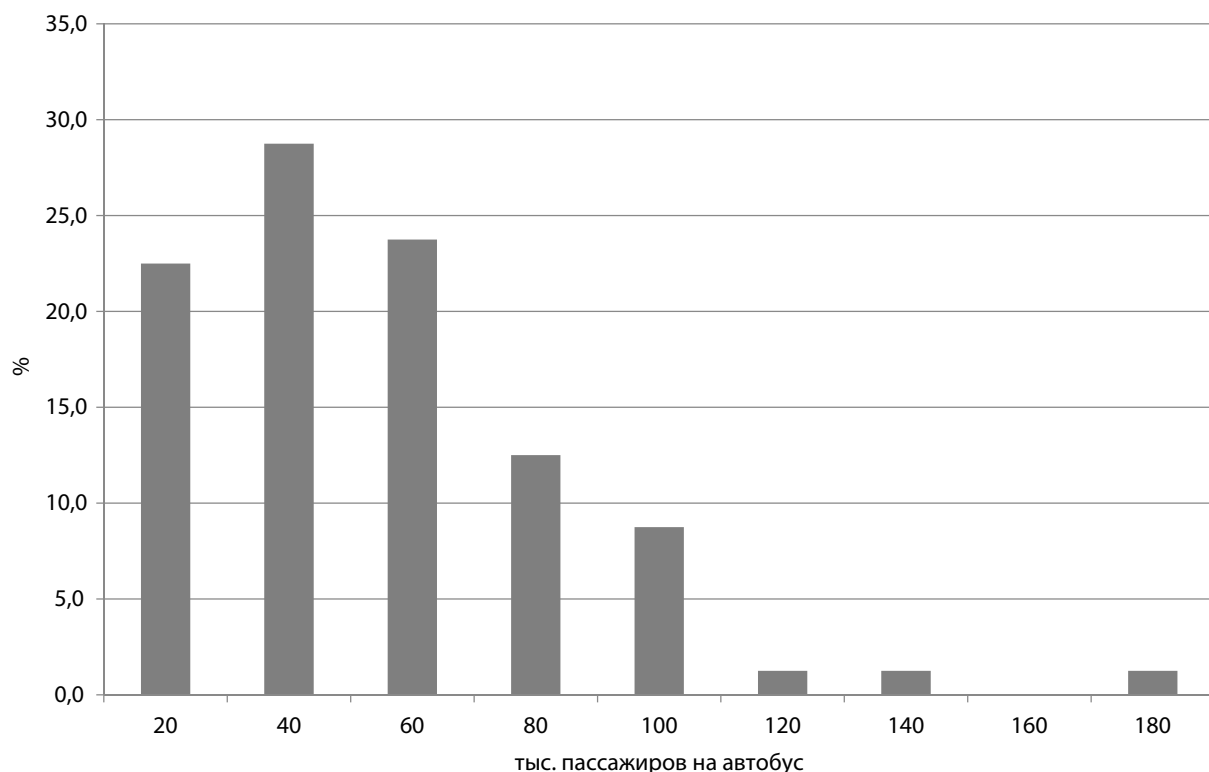


Рис. 3. Среднее количество пассажиров, перевозимых за год одним автобусом

ента детерминированности и уменьшение остаточной регрессии [16–19]. Затем эта процедура повторяется для оставшихся показателей таблицы 1 до тех пор, пока увеличение коэффициента детерминированности при добавлении нового показателя является значимым. Полученные результаты представлены в таблице 2.

Данные таблицы 2 могут использоваться для оценки характеристик транспортных потоков, например, в процессах стратегического планирования, когда потребности в перевозках пассажиров и грузов должны соответствовать поставленным задачам экономического и социального развития, демографическим прогнозам и т. д. Как показывают данные таблицы 2, коэффициенты детерминированности для всех моделей достаточно высоки. Отметим, что вероятности распределения остатков (разниц между фактическими значениями показателей и значениями, задаваемыми моделями таблицы 2) с уровнем значимости 5 % подчинены нормальному распределению. Для иллюстрации на рисунке 4 представлено распределение таких остатков для количества перевезенных пассажиров в различных регионах.

Таким образом, модели, представленные в таблице 2, можно использовать для оценки характеристик транспортных потоков, в том числе в процессах долгосрочного стратеги-

ческого планирования. В этом случае заданные стратегией социально-экономического развития региона показатели позволяют оценить ожидаемые значения показателей пассажиро- и грузопотоков и, соответственно, адекватно планировать развитие транспортной системы.

Вместе с тем, достаточно большая величина случайной компоненты в показателях пассажиро- и грузоперевозок (рис. 4) говорит о наличии не учитываемых моделью случайных процессов, влияющих на транспортные потоки. Проведенный дополнительный выборочный корреляционный анализ величины случайных остатков с рядом показателей для регионов России, не включенных в таблицу 1, не выявил существенных корреляций. Можно предположить, что причины процессов, приводящих к появлению случайных компонент в показателях пассажиро- и грузоперевозок, не лежат в сфере экономики и демографии. К сожалению, необходимая информация для проверки этого предположения пока отсутствует.

Можно предположить, что качество моделей улучшится при рассмотрении нелинейных зависимостей показателей транспортных потоков от представленных в таблице 2 параметров. Для оценки возможного наличия гетероскедастичности [18] была проверена логарифмическая модель (линейная зависимость логарифма

Таблица 2

Результат построения многомерных регрессионных моделей характеристик пассажирских и грузовых потоков (1)

Показатели	Коэффициенты при независимых переменных в уравнениях регрессии								Константа
	Плотность железнодорожных путей	Плотность автомобильного общего пользования с твердым покрытием	Ввод в действие жилых домов	Инвестиции в основной капитал	Основные фонды в экономике	Объем отгруженных товаров собственного производства	Среднегодовая численность занятых в народном хозяйстве	Число ППС на 100 000 населения	Численность населения
1. Перевозки пассажиров автобусами общего пользования					-1.08E-05	0.171	2.94-05		-8.98
Коэффициент детерминации R^2					0.933	0.930	0.925		
F-статистика					511.5	355.0	964.7		
Отношение критического значения F-статистики при 1 % уровне значимости к ее фактическому значению					0.011	0.010	0.007		
2. Пассажиро-оборот автобусов общего пользования						0.00020	-2.14	5.23	-796.7
Коэффициент детерминации R^2						0.76	0.74	0.75	0.65
F-статистика						60.3	110.5	77.3	150.5
Отношение критического значения F-статистики при 1 % уровне значимости к ее фактическому значению						0.059	0.044	0.052	0.046
3. Перевозки грузов	-0.026			0.00011	-1.04 E-05	3.26E-05			
Коэффициент детерминации R^2	0.764			0.531	0.651	0.705			
F-статистика	60.6			88.3	71.8	60.6			
Отношение критического значения F-статистики при 1 % уровне значимости к ее фактическому значению	0.059			0.079	0.068	0.067			
4. Грузооборот		-2.02	2.07			0.00044			
Коэффициент детерминации R^2		0.792	0.738			0.817			
F-статистика		146.2	219.6			112.9			
Отношение критического значения F-статистики при 1 % уровне значимости к ее фактическому значению		0.033	0.031			0.036			

Примечание: возрастание коэффициента R^2 соответствует последовательности добавления независимых переменных при пошаговой многомерной регрессии.

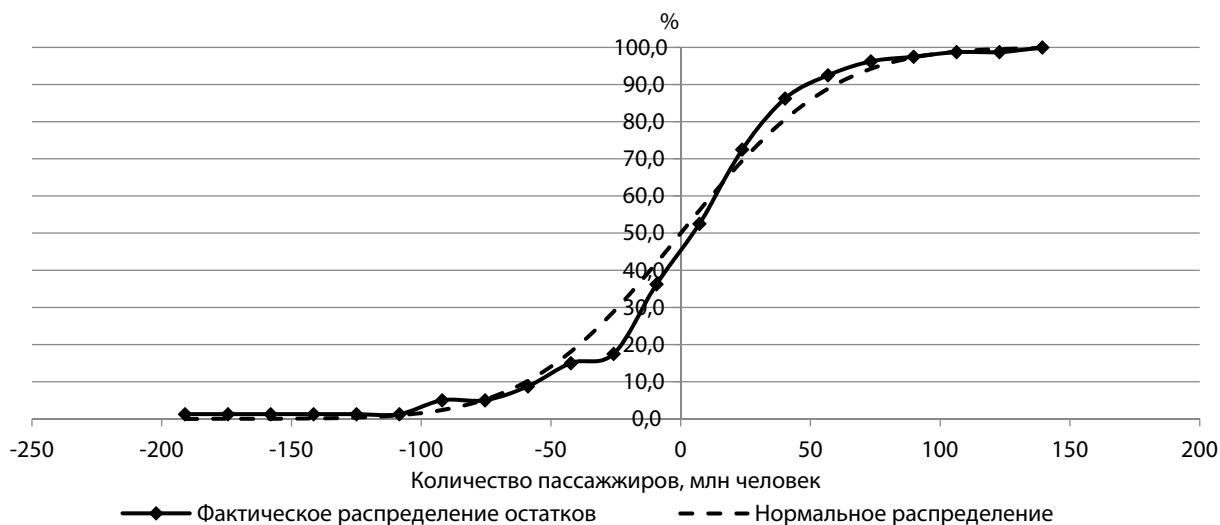


Рис. 4. Интегральная функция распределения остатков для модели перевозки пассажиров таблицы 2. Для сравнения представлена функция нормального распределения со средним значением, равным нулю, и соответствующей фактическому значению дисперсией

рифма исследуемого показателя от логарифмов независимых переменных). Коэффициент детерминации логарифмической модели, как правило, был ниже, чем у соответствующей модели таблицы 2, распределение же случайных остатков практически не изменялось.

Для нелинейного моделирования характеристик пассажирских и грузовых потоков могут использоваться и технологии нейронных сетей. В соответствии с известной теоремой Колмогорова — Арнольда [27, 28], при наличии функциональной зависимости между рассчитываемым показателем и переменными модели можно построить аппроксимацию моделируемого показателя с помощью многослойного персептрона [20–22]. Для построения таких персептронов использовалось ПО Deductor Studio (разработка Base Groop) академической версии 5.3 0.88.

Архитектура используемых многослойных персептронов не приводится ввиду громоздкости рисунка. В качестве входных переменных персептронов использовались независимые переменные, представленные в таблице 2. Коэффициенты детерминации для нейросетевых моделей и функции распределения случайных остатков по сравнению с моделями таблицы 2 также существенно не улучшились.

Таким образом, представленные результаты свидетельствуют о возможности оценки характеристик пассажирских и грузовых перевозок, осуществляемых автомобильным транспортом, на основании ограниченного набора параметров состояния транспортной инфраструктуры, экономического и демографического состояния региона, с использованием

простых линейных моделей, представленных в таблице 2.

4.2. Кластеризация и бенчмаркинг

Как уже отмечалось во введении, распространенной практикой при определении стратегических ориентиров региона является изучение опыта его аналогов с целью выполнения отраслевых сравнений и принятия необходимых управленческих решений. По сути дела, такой выбор является решением задачи классификации — поиска близких, в каком-то смысле, к рассматриваемому иных субъектов РФ. Ввиду того, что задачи развития транспорта вытекают из поставленных социально-экономических целей, а соответствующие показатели, как следует из результатов предыдущего раздела, определяются экономическими и демографическими показателями, попробуем провести классификацию субъектов РФ, используя в качестве переменных, описывающих субъект, именно параметры, представленные в таблице 2. Разумеется, такая классификация, выделяет субъекты «близкие» в смысле схожести транспортных систем и не носит абсолютного характера, поэтому ее распространение на весь спектр социально-экономических характеристик субъектов РФ необоснованно.

В качестве методов решения задач классификации, будем использовать алгоритм k -средних [19, 23, 24].

На рисунке 5 представлена иерархическая структура кластеров, объединяющих субъекты РФ в группы такие, что элементы каждой группы ближе к элементам, принадлежащим к этому же кластеру, чем к элементам других групп. При этом мерой близости является рас-

стояние, рассчитываемое в пространстве всех независимых переменных, характеризующих субъект (табл. 2). Каждый кластер может быть описан с помощью так называемого центроида — «центра тяжести» кластера, задаваемого многомерным вектором $\bar{X}_c = \{X_{1c}, X_{2c}, \dots, X_{nc}\}$, где

$$X_{ic} = \sum_{j=1}^M X_j(i). \quad (6)$$

Здесь X_i — независимые переменные таблицы 2, индекс j — нумерует субъекты (M — количество субъектов), индекс i нумерует переменные (n — количество переменных).

Чем ближе субъект расположен к центру кластера, тем с большей уверенностью можно говорить о правильности его отнесения к данному кластеру — надежности классификации. На рисунке 5 субъекты РФ перечисляются в порядке возрастания расстояния от центроида.

Кластеры в алгоритме k -средних формируются итерационно. На первом шаге центроиды выбираются случайно, затем идет процесс последовательного добавления точек — векторов $\bar{X}(j)$ в пространство признаков. После добавления каждой точки положение центроидов

пересчитывается. Процесс прекращается после стабилизации положений центроидов (перемещение на очередном шаге меньше заданной величины). Количество кластеров в нашем случае определялось автоматически, посредством алгоритма g -средних (последовательной проверки того, что данные внутри каждого выделенного кластера подчиняются определенному гауссовому распределению). Если такая проверка дает отрицательный результат, то кластер разбивается на два новых кластера. Процесс повторяется на каждом шаге процедуры до стабилизации количества кластеров и положения их центроидов.

Сами иерархические структуры на рисунке 5 построены по правилу минимальности расстояния [19, 29]. На первом шаге наиболее близкие кластеры объединяются в группы, затем к ним добавляются более удаленные кластеры до полного формирования иерархии. Расстояние между кластерами (группами кластеров) соответствует высоте вертикальной линии до ближайшего узла иерархии. Близкие результаты получаются и при кластеризации субъектов с помощью самоорганизующейся карты Кохонена [25]

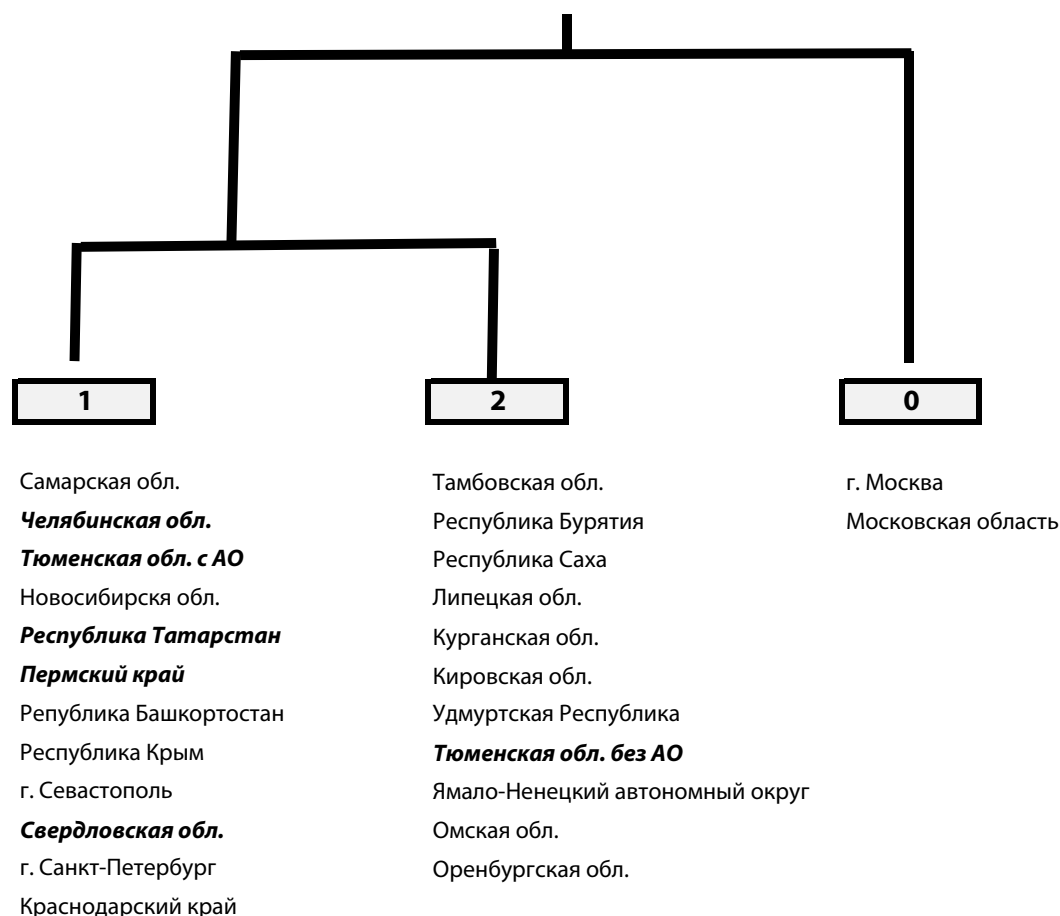


Рис. 5. Результаты кластеризации субъектов РФ в пространстве независимых переменных (табл. 2) методом k -средних. Выделены субъекты, традиционно сравниваемые со Свердловской областью

Следует обратить внимание на то, что области, традиционно сравниваемые со Свердловской областью, действительно, при любом алгоритме кластеризации располага-

ются либо в одном и том же, либо в близких кластерах. Обращает на себя внимание выделение Москвы и Московской области в отдельный кластер.

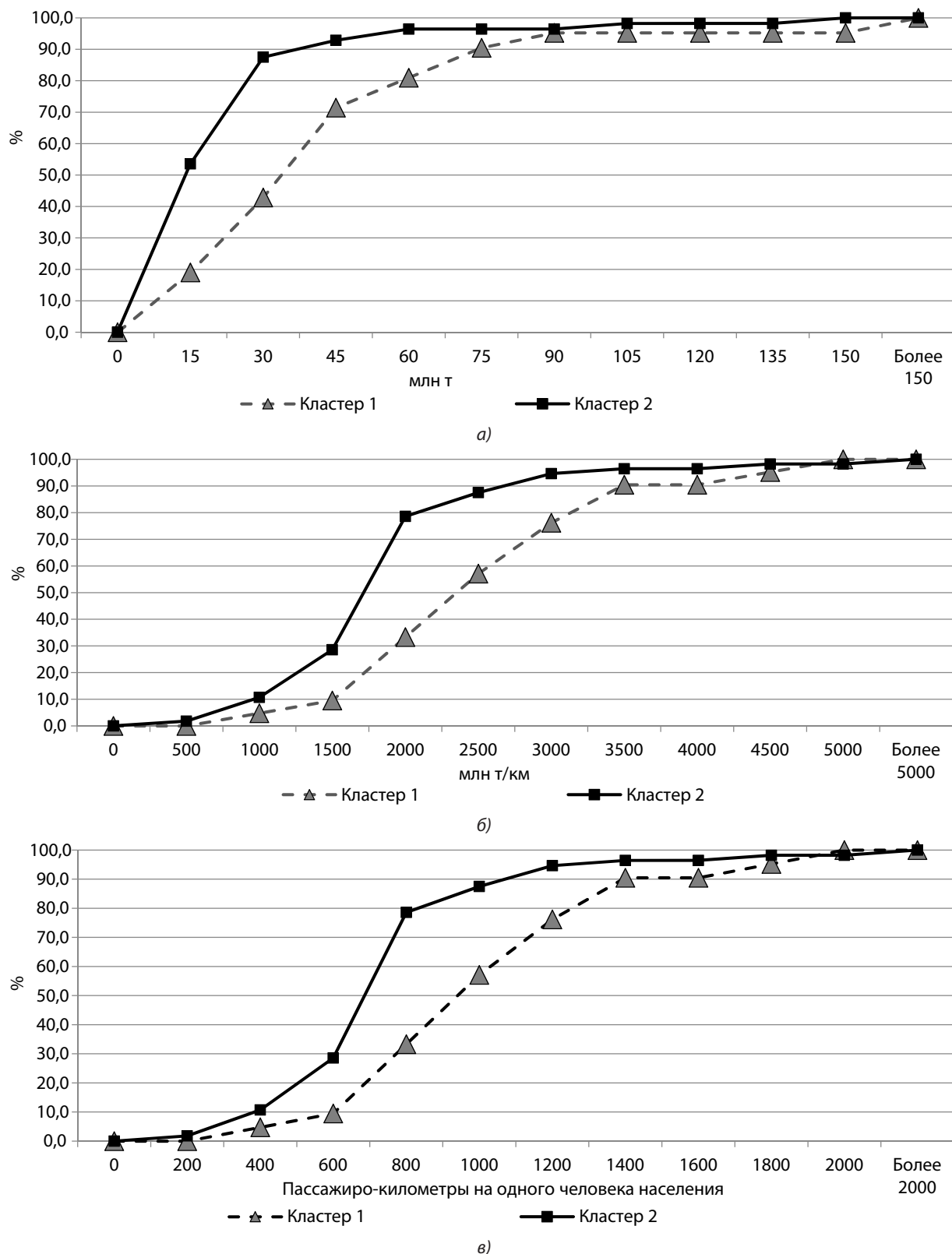


Рис. 6. Распределения показателей использования транспортной инфраструктуры в кластерах 1 и 2 (а) перевозки грузов; б) грузооборот; в) транспортная мобильность населения (для перевозок автобусами общего пользования)

Рассмотрим различия в деятельности транспорта в субъектах, принадлежащих разным кластерам. При этом, поскольку Москва и Московская область формируют отдельный кластер, существенно отличающийся по основным показателям от кластеров 1 и 2, сравним только эти кластеры.

Функции распределения параметров, характеризующих транспортную инфраструктуру в кластерах 1 и 2, достаточно близки.

Тем не менее, результаты использования транспортной инфраструктуры в кластерах 1 и 2 различаются значительно. Так, например, на рисунке 6 представлены данные по перевозке грузов, грузообороту и транспортной мобильности населения (для автобусного транспорта общего пользования). Хорошо видно, что распределения этих показателей в кластере 1 значительно смещены в сторону больших значений по сравнению с распределениями в кластере 2.

Таким образом, различия между группами субъектов РФ, выделенными в процессе использования алгоритма k -средних, обусловлены показателями использования близких по своим параметрам транспортных инфраструктур. Это свидетельствует о логичности применения данного алгоритма кластеризации при бенчмаркинге — выборе субъектов для сравнения с анализируемым.

5. Заключение

В данной работе предложена модель косвенной оценки характеристик пассажирских и грузовых потоков, связанных с автомобильным транспортом, позволяющая оценивать объемы перевозок, пассажирооборот и грузооборот не напрямую, посредством фиксации перевозимых грузов и пассажиров, что достаточно сложно, дорого и не всегда возможно, а по косвенным параметрам, характеризующим социальное и экономическое развитие территории, на которой осуществляются перевозки. Подобные модели имеют широкую область применения — от задач проводимого в настоящее время стратегического планирования развития территорий на период до 2035 г., до решения вопросов управления транспортом на

тактическом и оперативном уровне. Их придется решать, например, для ответа на вопрос о том, какие задачи перед транспортной инфраструктурой возникнут вследствие выполнения целевых установок в области экономического и социального развития (строительства новых предприятий, модернизации оборудования, развития жилищного строительства и т. д.).

Проведенный анализ показал, что в первую очередь на характеристики пассажирских и грузовых перевозок автомобильным транспортом влияют показатели, характеризующие уровень именно экономического развития территории (см. табл. 2):

- инвестиции в основной капитал;
- основные фонды в экономике;
- объем отгруженных товаров собственного производства;
- ввод в действие жилых домов.

И лишь во вторую очередь на них влияют параметры, характеризующие собственно транспортную инфраструктуру и демографические процессы. Влияние социальных показателей незначительно.

Вместе с тем, подобный макроанализ не дает возможности ответить на ряд принципиальных для организации перевозок вопросов, например, таких как формирование оптимальных маршрутов и остановочной сети, расписания рейсов автобусов, маршрутов грузового транспорта, минимизирующих нагрузку на дорожную сеть и т. д. Для решения подобных задач необходимо устранить целый комплекс проблем по обеспечению сбора и хранения полной и адекватной информации о текущем состоянии транспортных потоков, причем, как на внутри-, так и на межрегиональном уровне. Обсуждение связанных с этим проблем может составить предмет самостоятельной работы.

Следует также отметить перспективы использования методов интеллектуального анализа данных в анализе пассажирских грузовых потоков как для целей их прогнозирования, так и для выявления наиболее типичных сценариев связи экономических, транспортных и социальных систем с целью обеспечения их сбалансированного развития.

Список источников

1. Leontief W. W. Input-Output Economics. 2nd ed. — New York: Oxford University Press, 1986. — 448 p.
2. Гасникова А. В. Введение в математическое моделирование транспортных потоков под редакцией. — М.: МЦНМО, 2013. — 427 с.
3. Use of Local Linear Regression Model for Short-Term Traffic Forecasting / Sun H., Liu H., Xiao H., He R., Ran B. // Journal of Transportation Research Board. — 2003. — Vol. 1836. — P. 143–150. — DOI: 10.3141/1836–18.
4. Oswald R., Scherer T., Smith B. L. Traffic flow forecasting using approximate nearest neighbor nonparametric regression. — USA: Center for Transportation Studies, University of Virginia, 2001. — 115 p.

5. Min W., Wynter L. Real-time road traffic prediction with spatiotemporal correlations // *Transportation Research Part C: Emerging Technologies*. — 2011. — Vol. 19, iss. 4. — P. 606–616. — DOI: 10.1016/j.trc.2010.10.002.
6. Box G. E., Jenkins G. M., Reinsel G. C. *Time Series Analysis: Forecasting and Control*. — 4th ed. — USA: Wiley, 2008. — 784 p.
7. Mai T., Ghosh B., Wilson S. Short-term traffic-flow forecasting with auto-regressive moving average models // *Proceedings of the Institution of Civil Engineering-Transport*. — 2014. — Vol. 167, iss. 4. — P. 232–239. — DOI: 10.1680/tran.12.00012.
8. Stathopoulos A., Karlaftis M. G. A multivariate state space approach for urban traffic flow modeling and prediction // *Transportation Research Part C: Emerging Technologies*. — 2003. — Vol. 11, iss. 2. — P. 121–135. — DOI: 10.1016/S0968-090X(03)00004-4.
9. The application of space-time ARIMA model on traffic flow forecasting / Lin S.-H., Huang H.-Q., Zhu D.-Q., Wang T.-Z // *Machine Learning and Cybernetics, 2009 International Conference on*. — 2009. — Vol. 6. — P. 3408–3412. — DOI: 10.1109/ICMLC.2009.5212785.
10. Jin X., Zhang Y., Yao D. Simultaneously Prediction of Network Traffic Flow Based on PCA-SVR // *Lecture Notes in Computer Science*. — 2007. — Vol. 4492. — P. 1022–1031. — DOI: 10.1007/978-3-540-72393-6_121.
11. Vlahogianni E. I., Karlaftis M. G., Golias J. C. Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*. — 2014. — Vol. 43, Part 1. — P. 3–19. — DOI: 10.1016/j.trc.2014.01.005.
12. Guorong G., Yanping L. Traffic Flow Forecasting based on PCA and Wavelet Neural Network // *Information Science and Management Engineering (ISME)*. — 2010. — Vol. 1. — P. 158–161. — DOI: 10.1109/ISME.2010.10.
13. Zheng W., Lee D.-H., Shi Q. Short-term freeway traffic flow prediction: bayesian combined neural network approach // *Journal of Transportation Engineering*. — 2006. — Vol. 132, № 2. — P. 114–121. — DOI: 10.1061/(ASCE)0733-947X(2006)132:2(114).
14. Zhang X., He G. Forecasting Approach for Short-term Traffic Flow based on Principal Component Analysis and Combined Neural Network // *Systems Engineering: Theory & Practice*. — 2007. — Vol. 27(8). — P. 167–171. — DOI: 10.1016/S1874-8651(08)60052-6.
15. Sinitsyn E., Vikharev S., Brusyanin D. Economic and Mathematical Model for Forecasting Passenger Traffic on a Long Term Basis Case of Study Russia // *Journal of Engineering and Applied Sciences*. — 2019. — 14(3). — P. 773–779. — DOI: 10.3923/jeasci.2019.773.779.
16. Дрейнер Н., Смит Г. Прикладной регрессионный анализ, изд. 2-е, перераб. и доп. в 2-х кн. / Пер. с англ. Ю. П. Адлера и В. Н. Горского. — М.: Финансы и статистика, 1986. — 366 с.
17. Дугерти К. Введение в эконометрику. Пер. с англ. — М.: Инфра-М, 1999. — 402 с.
18. Ханк Д. Э., Уичерн Д. У., Райтс А. Дж. Бизнес прогнозирование, 7-е изд. : пер. с англ. — М.: Издательский дом «Вильямс», 2003. — 656 с.
19. Технологии анализа данных. Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А. А., Куприянов М. С., Степаненко В. В., и др. — СПб.: БХВ-Петербург, 2007. — 384 с.
20. Осовский С. Нейронные сети для обработки информации / Пер. с пол. И. Д. Рудинского. — М.: Финансы и статистика, 2002. — 344 с.
21. Николенко С. И., Кадури А. А., Архангельская Е. О. Глубокое обучение. Погружение в мир нейронных сетей. — СПб.: Питер, 2018. — 480 с.
22. Хайкин С. Нейронные сети. Полный курс : пер. с англ., 2-е изд. испр. — М.: ООО «ИД Вильямс», 2006. — 1104 с.
23. Leskovec J., Rajaraman A., Ullman J. D. Mining of massive datasets. — Stanford Univ., Millway Labs, 2014. — 516 p.
24. Замятин А. В. Интеллектуальный анализ данных: учеб. пособие. — Томск: Издательский дом Томского государственного университета, 2016. — 120 с.
25. Дебок Г., Кохонен Т. Анализ финансовых данных с помощью самоорганизующихся карт : пер. с англ. — М.: Альпина, 2001. — 317 с.
26. Митропольский А. К. Техника статистических вычислений: 2-е изд., перераб. и доп. — М.: Наука, 1971. — 576 с.
27. Колмогоров А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного и сложения // *Известия АН СССР*. — 1957. — № 114. — С. 953–956.
28. Арнольд В. И. О функции трех переменных // *Доклады АН СССР*. — 1957. — № 114. — С. 679–681.
29. Фор А. Восприятие и распознавание образов. — М.: Машиностроение, 1989. — 272 с.

Информация об авторах

Синицын Евгений Валентинович — доктор физико-математических наук, профессор, кафедра анализа систем и принятия решений, Высшая школа экономики и менеджмента, Уральский федеральный университет имени первого Президента России Б. Н. Ельцина (Российская Федерация, 620002, г. Екатеринбург, ул. Мира, 19; e-mail: sinitsyn_ev@mail.ru).

Толмачев Александр Владимирович — генеральный директор ООО «Датател-Урал»; старший преподаватель, кафедра анализа систем и принятия решений, Высшая школа экономики и менеджмента, Уральский феде-

ральный университет (Российская Федерация, 620100, г. Екатеринбург, Сибирский тракт, 12Б, офис 311; 620002, г. Екатеринбург, ул. Мира, 19; e-mail: at@idtu.ru).

Бруснянин Дмитрий Алексеевич — заместитель Министра транспорта и дорожного хозяйства Свердловской области (Российская Федерация, 620031, г. Екатеринбург, пл. Октябрьская, д.1; dbrusyanin@mail.ru).

For citation: Sinitsyn, E. V., Tolmachev, A. V. & Brusyanin, D. A. (2019). Mathematical Modelling of Regional Cargo and Passenger Flows. *Ekonomika regiona* [Economy of region], 15(4), 1212-1225

E. V. Sinitsyn^{a)}, A. V. Tolmachev^{a, b)}, D. A. Brusyanin^{c)}

^{a)} Ural Federal University (Ekaterinburg, Russian Federation; e-mail: sinitsyn_ev@mail.ru)

^{b)} Datatel-Ural LLC (Ekaterinburg, Russian Federation)

^{c)} Ministry of Transport of the Sverdlovsk Region (Ekaterinburg, Russian Federation)

Mathematical Modelling of Regional Cargo and Passenger Flows

The creation and implementation of the strategies for economic and social development in the Russian regions for the period up to 2035 implies an adequate development of transport services affecting all economic sectors and segments of the population. In this regard, we propose a model connecting the characteristics of passenger and cargo flows with the parameters of economic and social development, as well as with the region's demography. This model allows specifying the congestion of the transport system resulting from the implementation of plans for social and economic development and planned decisions in the sphere of economic activity. For developing the model, we selected parameters describing the economic situation, labour market, demography, living standards and social situation in the analysed subject. These parameters have the highest correlation coefficients with the analysed characteristics of the transport infrastructure. Further, we conducted a step-by-step regression analysis, adding to the already existing variables new ones that gave the greatest increase in the determinacy coefficient R². The model shows that the main factor determining the amount of passengers transported by public buses is the annual average number of employed persons. The passenger turnover is mostly affected by the population size. The volume of goods transported by trucks is determined by parameters characterising the level of the production development (investments in fixed assets, fixed capital in the economy, and the volume of shipped goods of domestic production). The use of nonlinear models and networks did not significantly reduce the model's errors. Additionally, we clustered the Russian regions by indicators of socio-economic development and the characteristics of transport infrastructure affecting traffic flows. Then we assessed the efficiency of transport infrastructure's exploitation in various clusters. This allows the targeted benchmarking, namely the selection of regions mostly appropriate for comparison with the analysed one.

Keywords: passenger and cargo flows, passenger turnover, cargo turnover, socio-economic development, correlation coefficients, multidimensional regression, determinacy coefficients, data mining, clustering, Kohonen self-organizing map, k-means method, hierarchical structure of clusters

References

1. Leontief, W. W. (1986). *Input-Output Economics*. 2nd ed. New York: Oxford University Press, 448.
2. Gasnikova, A. V. (2013). *Vvedenie v matematicheskoe modelirovanie transportnykh potokov* [Introduction to mathematical modeling of traffic flows]. Moscow, 427. (In Russ.)
3. Sun, H., Liu, H., Xiao, H., He, R. & Ran, B. (2003). Use of Local Linear Regression Model for Short-Term Traffic Forecasting. *Journal of the Transportation Research Board*, 1836, 143–150. DOI: 10.3141/1836–18.
4. Oswald, R., Scherer, T. & Smith, B. L. (2001). *Traffic flow forecasting using approximate nearest neighbor nonparametric regression*. USA: Center for Transportation Studies, University of Virginia, F115.
5. Min, W. & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, 19(4), 606–616. DOI: 10.1016/j.trc.2010.10.002
6. Box, G. E., Jenkins, G. M. & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*. 4th edition. Wiley, 784.
7. Mai, T., Ghosh, B., & Wilson, S. (2014). Short-term traffic-flow forecasting with auto-regressive moving average models. *Proceedings of the Institution of Civil Engineers — Transport*, 167(4), 232–239. DOI: 10.1680/tran.12.00012
8. Stathopoulos, A. & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, 11(2), 121–135. DOI: 10.1016/S0968–090X(03)00004–4
9. Lin, S.-H., Huang, H.-Q., Zhu, D.-Q. & Wang, T.-Z. (2009). The application of space-time ARIMA model on traffic flow forecasting. *2009 International Conference on Machine Learning and Cybernetics*, 6, 3408–3412. DOI: 10.1109/ICMLC.2009.5212785
10. Jin, X., Zhang, Y. & Yao, D. (2007). Simultaneously Prediction of Network Traffic Flow Based on PCA-SVR. *Lecture Notes in Computer Science*, 1022–1031. DOI: 10.1007/978–3–540–72393–6_121
11. Vlahogianni, E. I., Karlaftis, M. G. & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, 43, 3–19. DOI: 10.1016/j.trc.2014.01.005.
12. Guorong, G. & Yanping, L. (2010). Traffic Flow Forecasting based on PCA and Wavelet Neural Network. *Information Science and Management Engineering (ISME)*, 1, 158–161. DOI: 10.1109/ISME.2010.10.

13. Zheng, W., Lee, D.-H. & Shi, Q. (2006). Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach. *Journal of Transportation Engineering*, 132(2), 114–121. DOI: 10.1061/(ASCE)0733-947X(2006)132:2(114)
14. Zhang, X. & He, G. (2007). Forecasting Approach for Short-term Traffic Flow based on Principal Component Analysis and Combined Neural Network. *Systems Engineering: Theory & Practice*, 27(8), 167–171. DOI: 10.1016/S1874-8651(08)60052-6
15. Sinitsyn, E., Vikharev, S. & Brusyanin, D. (2019). Economic and Mathematical Model for Forecasting Passenger Traffic on a Long Term Basis Case of Study Russia. *Journal of Engineering and Applied Sciences*, 14(3), 773–779. DOI: 10.3923/jeasci.2019.773.779.
16. Draper, N. & Smith, G. (1986). *Prikladnoy regressionnyy analiz [Applied Regression Analysis]*. Trans. from English. Moscow: Finance and statistics, 366. (In Russ.)
17. Dougherty, K. (1999). *Vvedenie v ekonometriku [Introduction to Econometrics]*. Trans. from English. Moscow: Infra-M, 402. (In Russ.)
18. Hanke, J. E., Reitsch, A. G. & Wichern, D. W. (2003). *Biznes prognozirovanie [Business Forecasting]*. Trans. from English. Moscow: Williams, 656. (In Russ.)
19. Barsegyan, A. A., Kupriyanov, M. S., Stepanenko, V. V. & Kholod, I. I. (2007). *Tekhnologii analiza dannykh. Data Mining, Visual Mining, Text Mining, OLAP [Data Analysis Technologies. Data Mining, Visual Mining, Text Mining, OLAP]*. St. Petersburg: BHV, 384. (In Russ.)
20. Osovskiy, S. (2002). *Neyronnye seti dlya obrabotki informatsii [Neural networks for information processing]*. Moscow, 344. (In Russ.)
21. Nikolenko, S. I., Kadurin, A. A. & Arkhangelskaya, E. O. (2018). *Glubokoe obuchenie. Pogruzhenie v mir neyronnykh setey [Deep learning. Immersion into the world of neural networks]*. St. Petersburg: Piter, 480. (In Russ.)
22. Haykin, S. S. (2006). *Neyronnye seti. Polnyy kurs [Neural Networks: A Comprehensive Foundation]*. Trans. from English. Moscow: Williams, 1104. (In Russ.)
23. Leskovec, J., Rajaraman, A. & Ullman, J. D. (2014). *Mining of massive datasets*. Stanford University: Millway Labs, 516.
24. Zamyatin, A. V. (2016). *Intellektualnyy analiz dannykh [Intellectual data analysis]*. Tomsk: Tomsk State University Publishing House, 120. (In Russ.)
25. Deboeck, G. & Kohonen, T. (2001). *Analiz finansovykh dannykh s pomoshchyu samoorganizuyushchikhsya kart [Visual explorations in finance with self-organizing maps]*. Trans. from English. Moscow: Alpina, 317. (In Russ.)
26. Mitropolskiy, A. K. (1971). *Tekhnika statisticheskikh vychisleniy [Statistical Computation Technique]*. Moscow: Science, 576. (In Russ.)
27. Kolmogorov, A. N. (1957). O predstavlenii nepreryvnykh funktsiy neskol'kikh peremennykh superpozitsiyami nepreryvnykh funktsiy odnogo peremennogo i slozheniya [On the representation of continuous functions of several variables as a superposition of continuous functions of one variable and addition]. *Doklady AN SSSR [Reports of the Academy of Sciences of the USSR]*, 114, 953–956. (In Russ.)
28. Arnold, V. I. (1957). O funktsii trekh peremennykh [About the function of three variables]. *Doklady AN SSSR [Reports of the Academy of Sciences of the USSR]*, 114, 679–681. (In Russ.)
29. Faure, A. (1989). *Vospriyatie i raspoznavanie obrazov [Perception et reconnaissance des formes]*. Trans. From French. Moscow: Mashinostroenie, 272. (In Russ.)

Authors

Evgeny Valentinovitch Sinitsyn — Doctor of Physics and Mathematics, Professor, Academic Department of Systems Analysis and Decision Making, Graduate School of Economics and Management, Ural Federal University; Scopus Author ID: 7003263555 (19, Mira St., Ekaterinburg, 620002, Russian Federation; e-mail: sinitsyn_ev@mail.ru).

Alexander Vladimirovich Tolmachev — CEO, Datatel-Ural LLC; Senior Lecturer, Academic Department of Systems Analysis and Decision Making, Graduate School of Economics and Management, Ural Federal University; Scopus Author ID: 57204904436 (12B, Sibirskiy tract, Ekaterinburg, 620100; 19, Mira St., Ekaterinburg, 620002, Russian Federation; e-mail: at@idtu.ru).

Dmitrii Alekseevitch Brusyanin — Deputy Minister of Transport of the Sverdlovsk Region; Scopus Author ID: 56369168600 (1, Oktyabrskaya Sq., Ekaterinburg, 620031, Russian Federation; e-mail: dbrusyanin@mail.ru).