

ПРОГНОЗИРОВАНИЕ СУТОЧНЫХ ОБЪЕМОВ ПАССАЖИРСКИХ ПЕРЕВОЗОК В МОСКОВСКОМ МЕТРОПОЛИТЕНЕ

В. В. Иванов^{а, б, 1}, Е. С. Осетров^{а, в, 2}

^а Объединенный институт ядерных исследований, Дубна

^б Национальный исследовательский ядерный университет «МИФИ», Москва

^в ФКУ «Ространсmodernизация», Москва

Нами развита методика среднесрочного прогнозирования суточных объемов пассажирских перевозок в Московском метрополитене. Она включает три варианта прогноза: 1) на основе искусственных нейронных сетей (ИНС); 2) с использованием сингулярно-спектрального анализа, реализованного в пакете «Гусеница»-SSA; 3) при совместном использовании ИНС и подхода «Гусеница»-SSA. Разработанные методы и алгоритмы позволяют проводить с приемлемой точностью среднесрочное прогнозирование пассажиропотока в Московском метрополитене.

We developed a methodology for the medium-term prediction of daily volumes of passenger traffic in the Moscow Metro. It includes three variants of the forecast: 1) on the basis of artificial neural networks (ANN); 2) using the singular-spectral analysis implemented in the “Caterpillar”-SSA package; 3) joint use of the ANN and the “Caterpillar”-SSA approach. The developed methods and algorithms allow one to conduct with an acceptable accuracy a medium-term forecasting of the passenger traffic in the Moscow Metro.

PACS: 02.50.-r; 05.45.Tr; 07.05.Mh; 89.60.-k

ВВЕДЕНИЕ

В 2015 г. отмечалось 80-летие Московского метрополитена (ММ), который является основой транспортной системы столицы. Он надежно связывает центр города с промышленными районами и жилыми массивами. В 2012 г. среднесуточный пассажиропоток составил 6,73 млн пассажиров, а пиковый — 9,28 млн (29 ноября 2012 г.). В настоящее время доля Московского метрополитена в перевозке пассажиров среди предприятий городского пассажирского транспорта столицы составляет около 56 % [1].

С учетом динамики развития ММ и строительства в Москве современной транспортной инфраструктуры роль метрополитена в пассажироперевозках будет только расти. Для обеспечения эффективного функционирования ММ необходимы постоянный мониторинг

¹E-mail: ivanov@jinr.ru

²E-mail: esosetrov@gmail.com

пассажирского трафика и прогнозирование объемов пассажиропотока на ближайшую и долгосрочную перспективу.

Получение оперативной информации о состоянии спроса на услуги пассажирского транспорта является актуальной задачей современного общества. Ее решение будет способствовать не только удовлетворению потребностей населения в передвижении при оптимальном использовании транспорта, но и повысит эффективность управления транспортной системой [2]. Ключевая роль в решении указанной проблемы отводится надежному прогнозированию объемов пассажирских перевозок на основе обработки соответствующих временных рядов [3–5].

В работе [6] была продемонстрирована принципиальная возможность краткосрочного прогноза суточных объемов пассажирских перевозок в Московском метрополитене с помощью искусственных нейронных сетей (ИНС). Из-за того что исходный временной ряд был сильно зашумлен, для исключения шума из анализируемого ряда была применена вейвлет-фильтрация [7]. И несмотря на то, что указанная процедура позволила более чем в четыре раза повысить точность прогноза, добиться приемлемого уровня прогнозирования нам не удалось.

В работах [6, 7] для обучения и прогнозирования с помощью ИНС на вход нейронной сети подавалась выборка, составленная из факторов, от которых зависит суточный трафик пассажироперевозок в метро. Один из ключевых факторов — это суточное потребление электрической энергии в Московском регионе.

В этой связи, для прогнозирования пассажирских перевозок в метрополитене, нужно предварительно получить прогноз суточного потребления электроэнергии в Московском регионе. Решению указанной задачи посвящены наши последующие работы [8, 9]. В работе [8] нами развит подход, позволивший успешно решить проблему среднесрочного прогнозирования суточного потребления электроэнергии в Московском регионе с помощью искусственной нейронной сети прямого типа — многослойного перцептрона (МСП). При этом на этапе прогнозирования обученным МСП использовались данные прогноза, построенного с помощью метода, реализованного в подходе «Гусеница»-SSA [10–12].

В настоящей работе нами исследована возможность применения развитого в [8] подхода к решению проблемы среднесрочного прогнозирования суточных объемов пассажирских перевозок в Московском метрополитене.

1. АНАЛИЗ ВРЕМЕННОГО РЯДА, ОПИСЫВАЮЩЕГО ТРАФИК ПАССАЖИРСКИХ ПЕРЕВОЗОК

На рис. 1 приведен временной ряд, отражающий динамику суточного объема пассажирских перевозок Московским метрополитеном за последние 5 лет. Виден очень большой разброс в реализованных объемах перевозок. Главная причина такого разброса состоит в том, что пассажирские перевозки существенно различаются по своим объемам в рабочие и выходные и праздничные дни.

Из приведенных на рис. 2 распределений суточного объема пассажирских перевозок в рабочие (гистограмма справа) и выходные и праздничные (гистограмма слева) дни видно, что они сильно разнесены по своему среднему значению и, несмотря на большие дисперсии, практически не перекрываются.

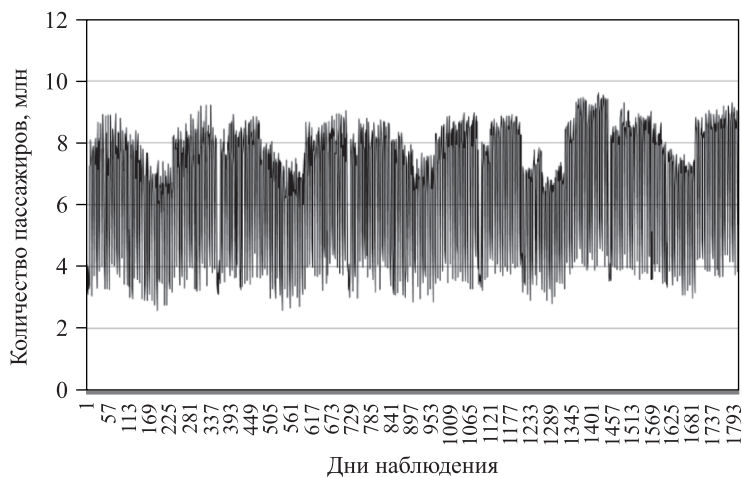


Рис. 1. Динамика суточного объема пассажирских перевозок в течение 5 лет



Рис. 2. Распределения суточного объема пассажирских перевозок в рабочие (справа) и выходные и праздничные (слева) дни

На рис. 3 и 4 представлены временные ряды, описывающие динамику суточного объема пассажирских перевозок в рабочие (рис. 3) и выходные и праздничные (рис. 4) дни соответственно.

Из сравнения указанных временных рядов можно сделать вывод о том, что динамика суточного объема пассажирских перевозок в рабочие дни носит более регулярный характер и не столь сильно зашумлена, как это имеет место в выходные и праздничные дни.

В связи с этим имеет смысл исследовать по отдельности временные ряды, описывающие динамику суточного объема пассажирских перевозок в рабочие и выходные и праздничные дни. В работах [6, 7] нами анализировался временной ряд, отвечающий суточным объемам пассажирских перевозок в рабочие дни. Исследованию указанного ряда посвящена настоящая работа.

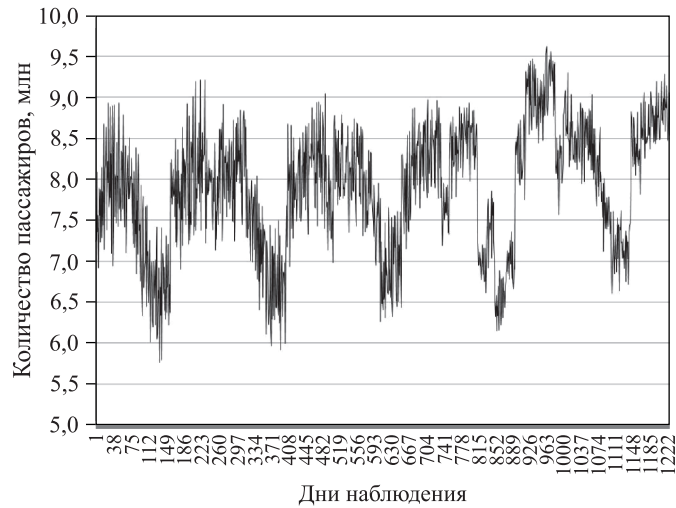


Рис. 3. Динамика суточного объема пассажирских перевозок в рабочие дни

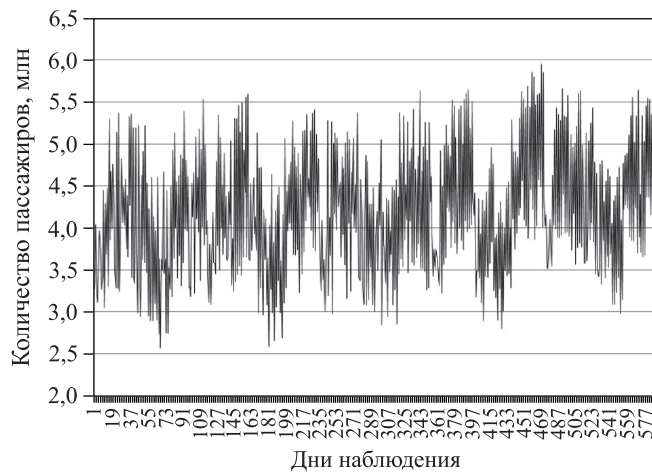


Рис. 4. Динамика суточного объема пассажирских перевозок в выходные и праздничные дни

В работе [6] приведен перечень из 15 переменных величин, от которых может зависеть динамика суточных объемов перевозок пассажиров в Московском метрополитене, а также подробно описана процедура на основе корреляции Пирсона [13], позволившая существенно уменьшить их число.

В результате проведенного анализа для дальнейшего рассмотрения были оставлены только такие переменные:

- для которых не наблюдается явления мультиколлинеарности [14, 15] — сильной корреляционной связи ($> 0,8$) между разными переменными;
- которые имеют значимый характер для ИНС, обусловленный сезонностью прогнозируемой величины;

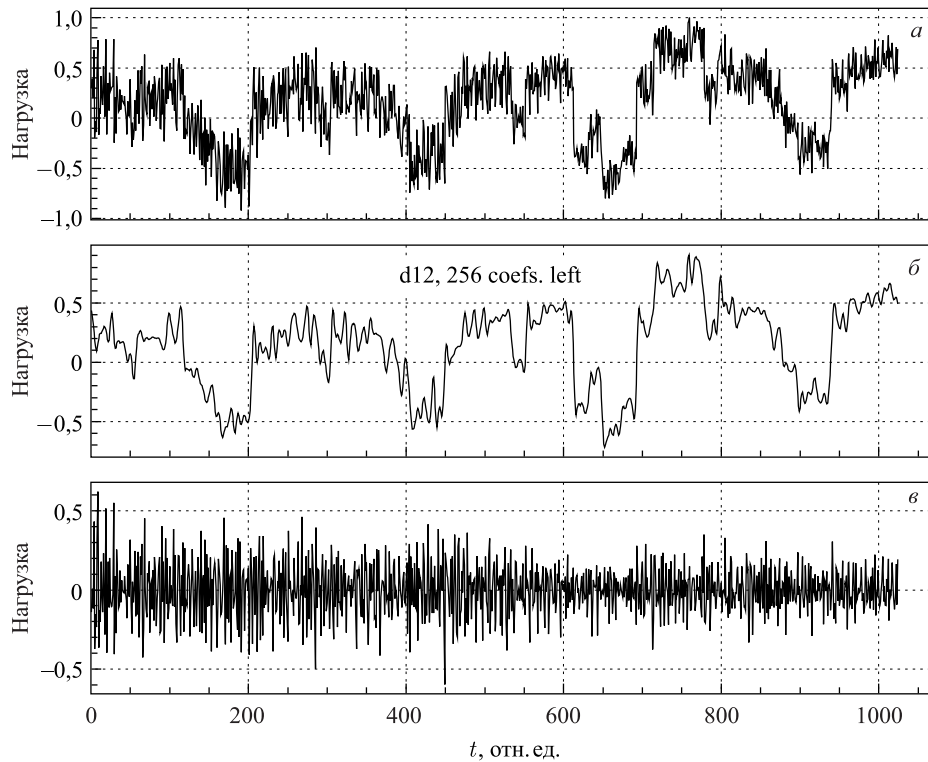


Рис. 5. *a*) Исходный временной ряд, содержащий данные о пассажиропотоке за последние 1024 рабочих дня (из полного ряда в 1229 рабочих дней); *б*) указанный ряд после применения к нему вейвлет-фильтрации (оставлено 256 коэффициентов из 1024); *в*) исключенная из исходного ряда шумовая компонента

— для которых на практике можно получить достоверные прогнозные значения с использованием альтернативных методов получения информации.

Из всего набора переменных для последующего использования были отобраны следующие: Var1 (год наблюдения), Var2 (месяц), Var3 (день недели), Var4 (тип дня), Var5 (отклонение дневной температуры от нормы) и Var6 (суточное потребление электрической энергии в Московском регионе).

Из указанных величин и переменной Var7 (суточный объем пассажирских перевозок в ММ) формировались выборки данных, которые использовались в работе [6] для обучения и прогнозирования с помощью ИНС. В отличие от [6] в работе [7] ряд, отвечающий переменной Var7, предварительно подвергался вейвлет-фильтрации [16]. Делалось это для того, чтобы исключить из исходных данных высокочастотную шумовую компоненту.

Опуская детали, связанные с вейвлет-фильтрацией исходных данных (ряд, отвечающий переменной Var7), вкратце обсудим полученные результаты.

На рис. 5 представлены: 1) исходный временной ряд, содержащий измерения пассажиропотока (последние 1024 рабочих дня из полного ряда в 1229 рабочих дня) (рис. *a*); 2) указанный ряд после применения к нему вейвлет-фильтрации (оставлено 256 коэффициентов из 1024) (рис. *б*); 3) исключенная из исходного ряда шумовая компонента (рис. *в*).

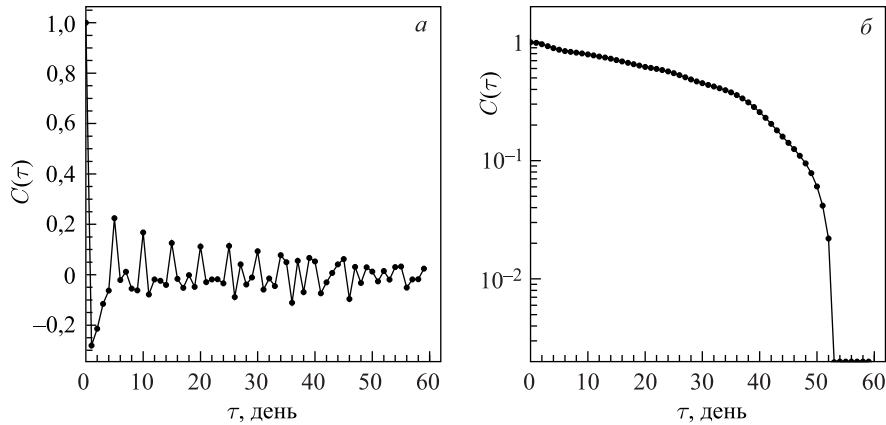


Рис. 6. Поведение автокорреляционной функции $C(\tau)$ в зависимости от интервала корреляции τ для шумовой (а) и регулярной (б) составляющих временного ряда

Проведенный в [7] анализ с использованием критерия согласия χ^2 (известного также как критерий согласия Пирсона) [17] и критерия симметрии на основе статистики ω_n^2 [18] показал, что исключенная из исходного ряда компонента строго следует гауссовскому закону.

Заметим, что прогнозирование временного ряда возможно только тогда, когда существует связь последующих значений ряда с предыдущими. Горизонт прогноза анализируемого нами ряда можно оценить с помощью линейной автокорреляционной функции [19]:

$$C(\tau) = \frac{\sum_{i=1}^N (x_{i+\tau} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1)$$

где N — это число измерений в анализируемом временном ряде $\{x_i\}$.

В качестве оценки интервала корреляции τ — горизонта прогноза принимается величина первого пересечения автокорреляционной функцией временной оси в доверительном интервале, отвечающем белому шуму [20, 21].

На рис. 6 приведены зависимости автокорреляционной функции от величины интервала корреляции τ для шумовой (а) и регулярной (б) составляющих временного ряда. На рис. 6, а видно, что для шумовой компоненты $C(\tau)$ при $\tau = 1$ обращается в нуль. Из этого следует, что соседние члены рассматриваемого временного ряда некоррелированы. Абсолютные значения $C(\tau)$ для $\tau > 1$ малы и с ростом τ быстро спадают по абсолютной величине. Учитывая то, что реализации рассматриваемого ряда распределены согласно гауссовскому закону и симметричны относительно нуля, можно сделать вывод о том, что данный ряд представляет собой гауссовский шум.

Исходя из поведения зависимости автокорреляционной функции от величины интервала корреляции τ для регулярной компоненты (см. рис. 6, б) анализируемого ряда, можно сделать вывод о том, что указанный ряд предоставляет неплохие возможности для среднесрочного прогноза.

2. ПРОГНОЗИРОВАНИЕ ПАССАЖИРСКИХ ПЕРЕВОЗОК С ПОМОЩЬЮ ИНС

Наиболее подходящей архитектурой ИНС для нашей задачи является многослойный перцептрон (МСП) (см., например, [22, 23]). Такая сеть содержит несколько слоев нейронов: слой входных нейронов, один или несколько скрытых слоев и слой выходных нейронов. Связи в МСП организованы между соседними слоями, а передача информации происходит в одном направлении: от входного слоя к выходному.

Анализируемые данные подаются на входные нейроны, число которых определяется объемом анализируемой выборки. В нашем случае она включала семь переменных: Var1–Var7. Теоретически одного скрытого слоя в МСП достаточно для аппроксимации любой непрерывной функции [24]. Однако на практике для описания таких сложных временных рядов, как в нашем случае, лучший результат дает МСП с двумя скрытыми слоями (см., например, [25] и ссылки в ней). Числа нейронов в скрытых слоях подбирались исходя из результатов обучения сети на анализируемых данных.

Перед тем как приступить к работе с МСП, переменные Var1–Var7 (обозначим их через переменную z) приводились к диапазону $[-1; +1]$ с помощью следующего преобразования:

$$y_i = \frac{(z_i - A_i) \cdot 2}{B_i - A_i} - 1, \quad i = 0, 1, \dots, 6, \quad (2)$$

где z_i — исходное значение переменной z ; A_i и B_i — минимальное и максимальное значения переменной z_i , а y_i — значение нормированной переменной.

В настоящей работе использовался многослойный перцептрон из пакета TMVA 4.2.0 [26] в среде ROOT [27]. Структура сети была следующей: 7 нейронов на входе, первый скрытый слой содержал 16 нейронов, второй — 8 нейронов, 1 выходной нейрон. Для обучения сети использовалась выборка, включающая в зависимости от горизонта прогнозирования разное количество дней наблюдения: 994 или 974 при прогнозе на 30 или 50 дней соответственно. В результате проведенного нами анализа для обучения был выбран метод Fletcher–Reeves (см., например, [28]), обеспечивший наилучшие результаты как на этапе обучения, так и при тестировании МСП. Процедура обучения включала 1500 эпох.

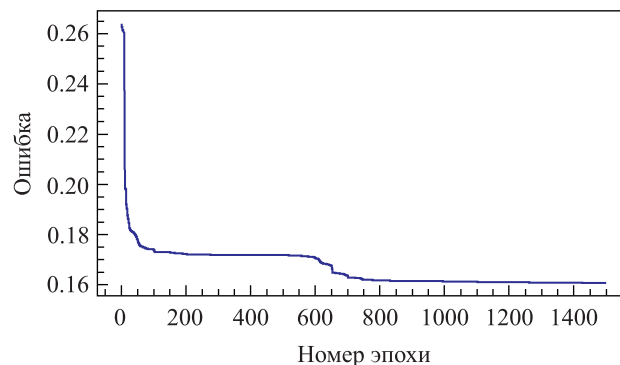


Рис. 7. График изменения ошибки при обучении ИНС в зависимости от номера эпохи

На рис. 7 приведен типичный график изменения ошибки при обучении ИНС в зависимости от номера эпохи.

После завершения процедуры обучения МСП проводилось тестирование обученной сети — оценка качества прогнозирования. На этом этапе в качестве переменной Var7 на вход МСП подавалось то значение, которое выдавалось сетью в качестве прогноза на предыдущем шаге вычислений.

На рис. 8 представлены результаты прогнозирования с помощью МСП на 30 (а) и 50 (б) дней наблюдения в сравнении с реальными данными. На рис. 9 показано поведение относительной ошибки величины R_y в зависимости от порядкового номера дня наблюдения при прогнозировании на 30 (а) и 50 (б) дней.

Ошибка R_y вычислялась по формуле

$$R_y = \frac{y_{\text{real}} - y_{\text{prog}}}{y_{\text{real}}}, \quad (3)$$

где y_{real} — наблюдаемое (реальное) значение объема пассажироперевозок, а y_{prog} — спрогнозированное значение (полученное в данном случае с помощью МСП).

На рис. 10 приведены гистограммы распределения относительной ошибки R_y при прогнозировании на 30 (а) и 50 (б) наблюдений. Из приведенных на рис. 9 и 10 резуль-

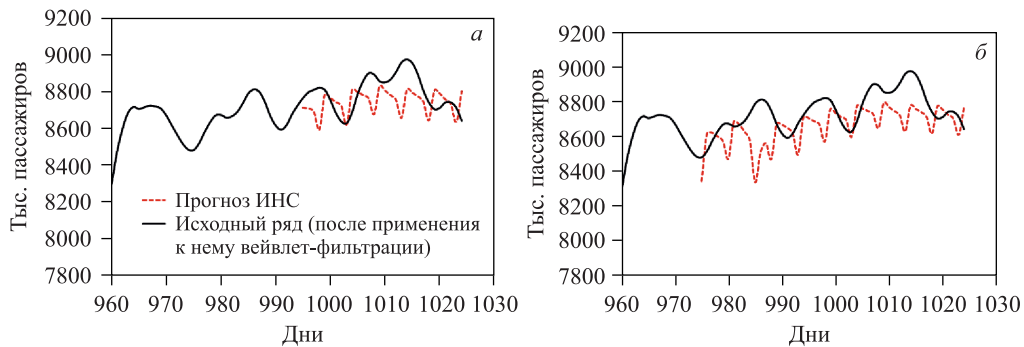


Рис. 8. Результаты прогнозирования с помощью МСП на 30 (а) и 50 (б) дней наблюдения в сравнении с реальными данными

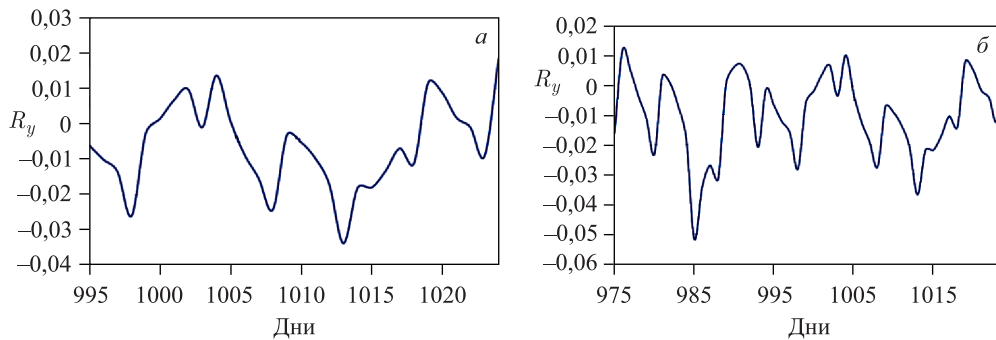


Рис. 9. Динамика величины R_y в зависимости от порядкового номера дня наблюдения при прогнозировании с помощью МСП на 30 (а) и 50 (б) дней

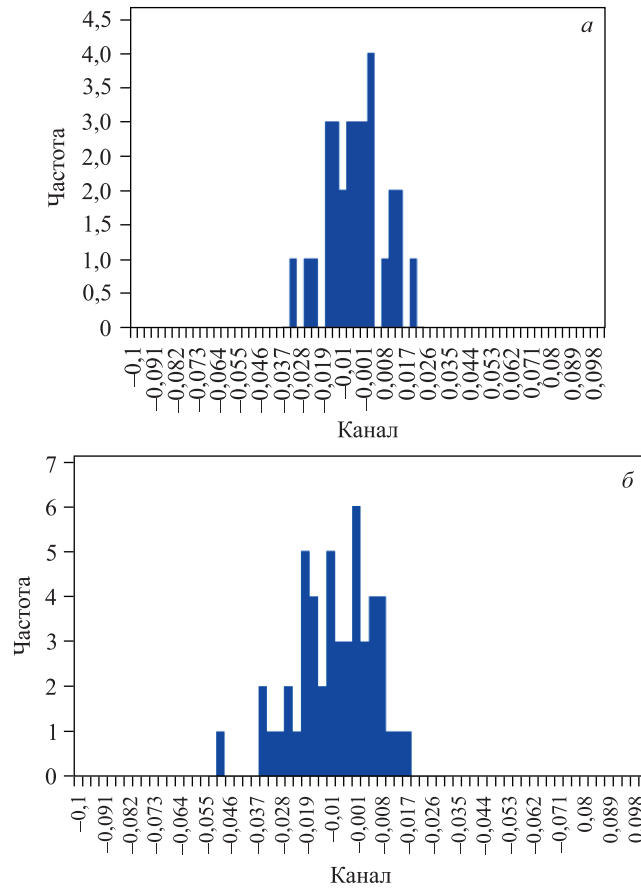


Рис. 10. Распределения переменной R_y при прогнозировании на 30 (а) и 50 (б) дней

татов видно, что ошибка R_y распределена симметрично относительно нуля и на всем интервале прогнозирования она чуть более 3% при прогнозе на 30 дней и не превышает 5% при прогнозе на 50 дней.

Таким образом, используя МСП, можно с хорошей точностью проводить среднесрочный прогноз объемов пассажирских перевозок в Московском метрополитене.

3. ПРОГНОЗИРОВАНИЕ ПАССАЖИРСКИХ ПЕРЕВОЗОК НА ОСНОВЕ СИНГУЛЯРНО-СПЕКТРАЛЬНОГО АНАЛИЗА

Как отмечалось в работе [8], нами было показано, что проблема среднесрочного прогнозирования суточного потребления электроэнергии в Московском регионе может быть успешно решена с помощью искусственной нейронной сети прямого типа — многослойного перцептрона. При этом на этапе прогнозирования обученного МСП использовались данные прогноза, построенного с помощью метода, реализованного в подходе «Гусеница»-SSA [10–12].

Метод «Гусеница»-SSA, развитый в работах [10, 11], может быть применен для анализа вещественнозначного временного ряда F , отвечающего произвольной функции $f(t)$, $t > 0$, определенной на равномерной сетке (в равноудаленных) точках:

$$f_i = f[t_i] = f(i\Delta t), \quad i = 0, 1, \dots, N - 1, \quad (4)$$

где Δt — временной интервал, в нашем случае равный 1.

На начальном этапе выполняется преобразование временного ряда (4) в траекторную матрицу [11, 12]. Далее проводится сингулярное разложение этой матрицы, анализ результатов ее разложения с помощью метода главных компонент и отбор тех компонент, которые могут представлять интерес для исследователя. При этом (несмотря на проведенную ранее вейвлет-фильтрацию исходного ряда) дополнительно отбрасываются компоненты, ответственные за высокочастотную составляющую анализируемого временного ряда. На заключительной стадии проводится реконструкция одномерного временного ряда на основе отобранных компонент.

Базовый алгоритм метода «Гусеница»-SSA состоит из двух дополняющих друг друга этапов: 1) разложения и 2) восстановления [11, 12].

3.1. Первый этап: разложение

Шаг 1: Вложение. Процедура вложения переводит исходный временной ряд (4) в последовательность многомерных векторов. Она представляет собой стандартную процедуру анализа временных рядов (см., например, [29]).

Пусть L — некоторое целое число (длина окна), $1 < L < N$. Тогда L -траекторная матрица ряда F — это матрица следующего вида:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} f_0 & f_1 & f_2 & \dots & f_{K-1} \\ f_1 & f_2 & f_3 & \dots & f_K \\ f_2 & f_3 & f_4 & \dots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \dots & f_{N-1} \end{pmatrix}. \quad (5)$$

Таким образом, процедура вложения образует $K = N - L + 1$ векторов вложения

$$\mathbf{X} = [X_1 : \dots : X_K]$$

из столбцов траекторной матрицы (5)

$$X_i = (f_{i-1}, \dots, f_{i+L-2})^T, \quad 1 \leq i \leq K,$$

имеющих размерность L .

Заметим, что в нашем случае величина окна L будет выбираться достаточно большой, но не более половины длины анализируемого временного ряда. В этом случае каждый из векторов L -вложения будет содержать в себе существенную часть поведения исходного ряда $F = (f_0, \dots, f_{N-1})$. Результатом такого шага является анализ именно матрицы \mathbf{X} , а не ее столбцов (векторов вложения), так как упорядоченность элементов внутри каждого вектора, как и упорядоченность самих векторов, играет ключевую роль в методе «Гусеница»-SSA [12].

Шаг 2: Сингулярное разложение. Сингулярное разложение — ключевая основа математической части метода. Результатом этого шага является сингулярное разложение (SVD = Singular Value Decomposition) траекторной матрицы ряда F .

Сингулярное разложение произвольной $L \times K$ -матрицы $\mathbf{X} = [X_1 : \dots : X_K]$ — это разложение \mathbf{X} в виде

$$\mathbf{X} = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T, \quad (6)$$

где $\lambda_i (i = 1, \dots, L)$ — собственные числа матрицы $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, упорядоченные по убыванию,

$$d = \max\{i, \text{таких, что } \lambda_i > 0\} = \text{rank } \mathbf{X},$$

где $\{U_1, \dots, U_d\}$ — соответствующие системы ортонормированных собственных векторов матрицы \mathbf{S} и векторов $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$.

В стандартной терминологии $\sqrt{\lambda_i}$ называются *сингулярными числами*, а U_i и V_i — *левыми и правыми сингулярными векторами* матрицы \mathbf{S} соответственно. Набор $(\sqrt{\lambda_i}, U_i, V_i)$ называется *i -й собственной тройкой* матрицы \mathbf{X} . Так как строки и столбцы траекторной матрицы являются отрезками исходного временного ряда, то левый и правый сингулярные векторы несут в себе привязку ко времени.

Если определить $\mathbf{X}_i = \sqrt{\lambda_i} U_i V_i^T$, то разложение (6) может быть переписано в виде

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad (7)$$

т. е. как представление \mathbf{X} в виде суммы элементарных матриц. При этом собственное число λ_i является характеристикой вклада матрицы \mathbf{X}_i в разложение (7).

3.2. Второй этап: восстановление

Шаг 3: Группировка. Целью этапа группировки является разделение аддитивных компонент анализируемого ряда. На основе разложения (7) процедура группировки делит все множество индексов $\{1, \dots, d\}$ на m непересекающихся подмножеств I_1, \dots, I_m .

Пусть $I = \{i_1, \dots, i_p\}$. Тогда результирующая матрица \mathbf{X}_I , соответствующая группе I , определяется как

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}.$$

Такие матрицы вычисляются для $I = I_1, \dots, I_m$, тем самым разложение (7) может быть записано в сгруппированном виде

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (8)$$

Процедура выбора множеств I_1, \dots, I_m называется группировкой собственных троек.

Для простоты примем $m = 2$, $I_1 = I = \{i_1, \dots, i_r\}$ и $I_2 = \{1, \dots, d\} \setminus I$, где $1 \leq i_1 < \dots < i_r \leq d$. Таким образом, мы полагаем, что временной ряд F равен сумме двух рядов, $F^{(1)}$ и $F^{(2)}$, т. е. $f_i = f_i^{(1)} + f_i^{(2)}$ для $i = 0, \dots, N - 1$.

Зафиксируем длину окна L и обозначим через \mathbf{X} , $\mathbf{X}^{(1)}$ и $\mathbf{X}^{(2)}$ соответствующие L -траекторные матрицы рядов F , $F^{(1)}$ и $F^{(2)}$. Рассмотрим сингулярное разложение (7) траекторной матрицы \mathbf{X} . Мы будем говорить, что ряды $F^{(1)}$ и $F^{(2)}$ (слабо) разделимы разложением (7), если существует такой набор индексов $I \subset \{1, \dots, d\}$, что $\mathbf{X}^{(1)} = \sum_{i \in I} \mathbf{X}_i$

и, следовательно, $\mathbf{X}^{(2)} = \sum_{i \in I} \mathbf{X}_i$ [12]. В случае разделимости вклад $\mathbf{X}^{(1)}$, первой компоненты в разложении $\mathbf{X} = \mathbf{X}^{(1)} + \mathbf{X}^{(2)}$, естественно выражается как доля соответствующих собственных чисел: $\sum_{i \in I} \lambda_i / \sum_{i=1}^L \lambda_i$.

Заметим, что при описании динамики суточных объемов пассажирских перевозок в Московском метрополитене суммой двух рядов ($m = 2$) более естественно говорить не о разделении двух компонент, а о выделении регулярной составляющей ряда F и отбрасывании высокочастотного шума.

Шаг 4: Диагональное усреднение. На последнем шаге базового алгоритма каждая матрица сгруппированного разложения (8) переводится в новый ряд длины N .

Пусть \mathbf{Y} — матрица $L \times K$ с элементами y_{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 $y_{ij}^* = y_{ij}$, если $L < K$, и $y_{ij}^* = y_{ji}$ иначе.

Диагональное усреднение переводит матрицу \mathbf{Y} в ряд g_0, \dots, g_{N-1} по формуле

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m, k-m+2}^* & \text{для } 0 \leq k < L^* - 1, \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, k-m+2}^* & \text{для } L^* - 1 \leq k < K^*, \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m, k-m+2}^* & \text{для } K^* \leq k < N. \end{cases} \quad (9)$$

Выражение (9) соответствует усреднению элементов матрицы вдоль «диагоналей» $i + j = k + 2$: выбор $k = 0$ дает $g_0 = y_{11}$, для $k = 1$ $g_1 = (y_{12} + y_{21})/2$ и т. п.

Применив диагональное усреднение (9) к результирующим матрицам \mathbf{X}_{I_k} , мы получаем ряды $\tilde{F}^{(k)} = (\tilde{f}_0^{(k)}, \dots, \tilde{f}_{N-1}^{(k)})$, и, следовательно, исходный ряд f_0, \dots, f_{N-1} раскладывается в сумму m рядов:

$$f_n = \sum_{k=1}^m \tilde{f}_n^{(k)}. \quad (10)$$

Заметим, что процедура диагонального усреднения в общем случае должна приводить к улучшению разделимости разных компонент ряда (8) [12].

4. ПРОГНОЗИРОВАНИЕ В ПОДХОДЕ «ГУСЕНИЦА»-SSA

Для прогнозирования суточных объемов пассажирских перевозок в Московском метрополитене использовалась последняя версия программы CaterpillarSSA (version 3.40, Professional M Edition, см. детали на сайте [30]). Указанная версия программы предназначена для анализа и прогнозирования одномерных и многомерных временных рядов.

Остановимся на отдельных и важных моментах для рассматриваемого нами примера:

1) в качестве входной информации для программы CaterpillarSSA использовался отфильтрованный с помощью вейвлет-фильтрации временной ряд, т. е. ряд, из которого предварительно была удалена высокочастотная шумовая компонента (см. детали в разд. 1);

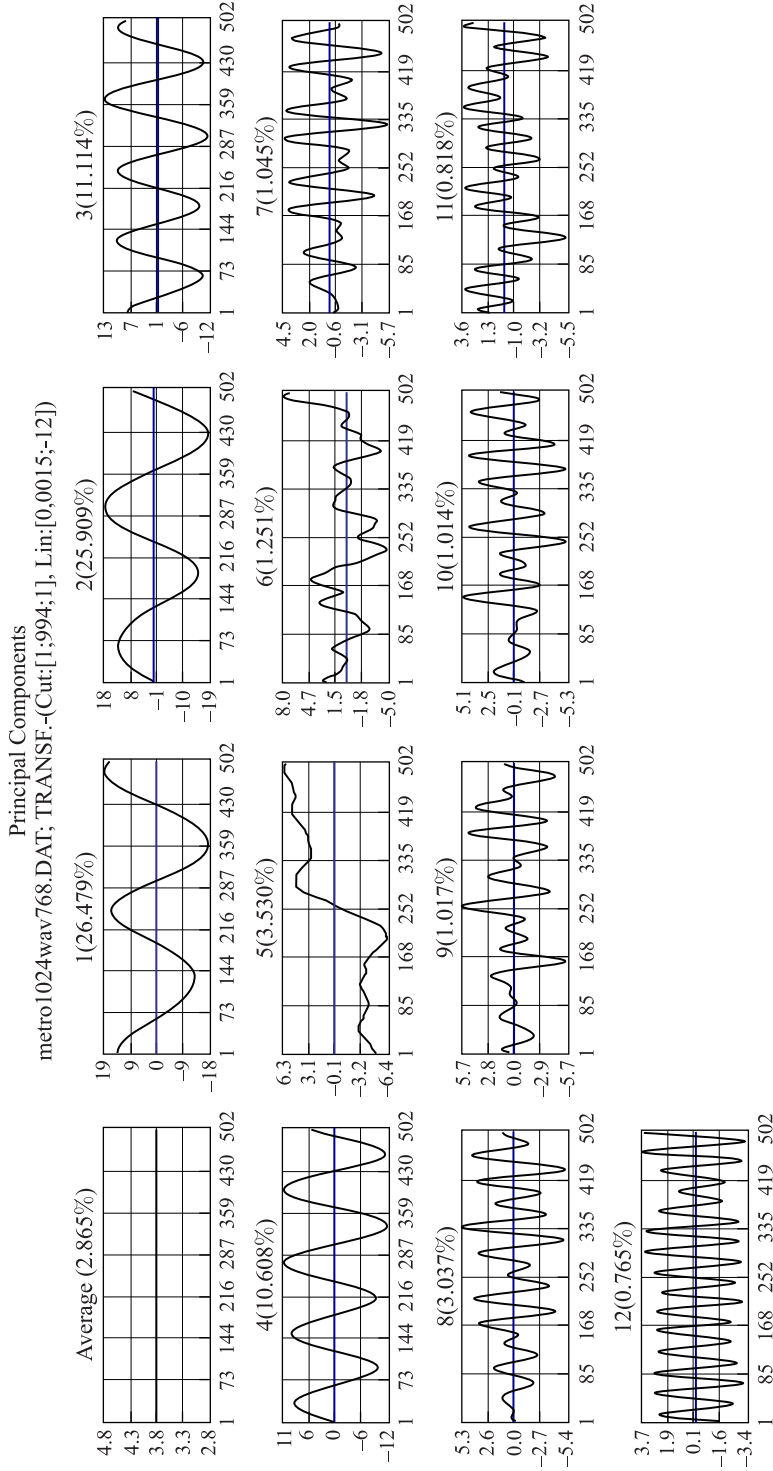


Рис. 11. Графики 13 первых компонент разложения при прогнозировании на 30 дней

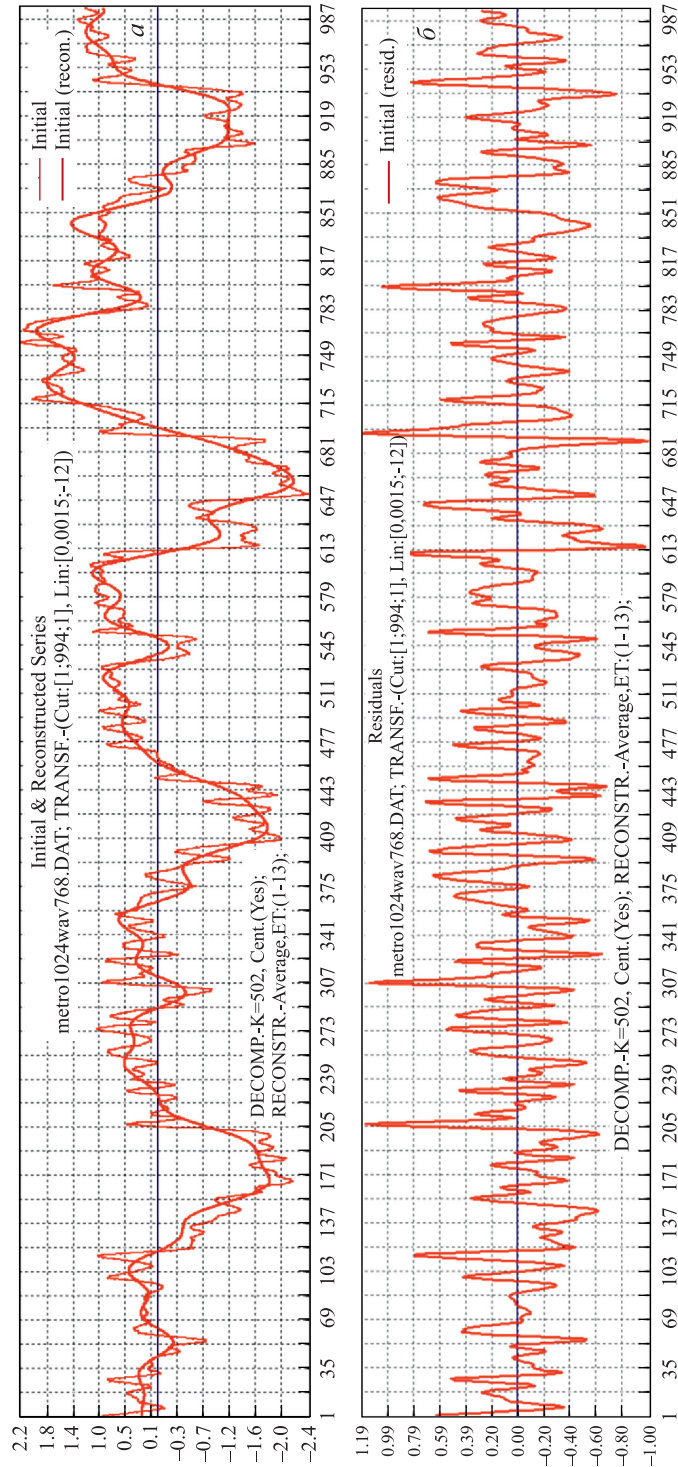


Рис. 12. а) Исходный ряд и его аппроксимация; б) ряд, отвечающий отбрасываемым компонентам

2) перед тем как проводить преобразование к многомерному виду (5), анализируемый временной ряд был стандартизован средствами CaterpillarSSA [30];

3) при переходе к многомерному виду (5) длина гусеницы бралась равной 492, что составляет примерно половину анализируемого ряда, позволяя выделить все характерные его особенности;

4) при реконструкции одномерного ряда использовалось 13 главных компонент; их общий вклад составил 99,91 %;

5) на этапе прогнозирования доверительный интервал был задан равным 0,25.

На рис. 11 приведены графики 13 первых компонент разложения анализируемого ряда при прогнозировании на 30 дней наблюдения. Из графиков видно, что рассматриваемые

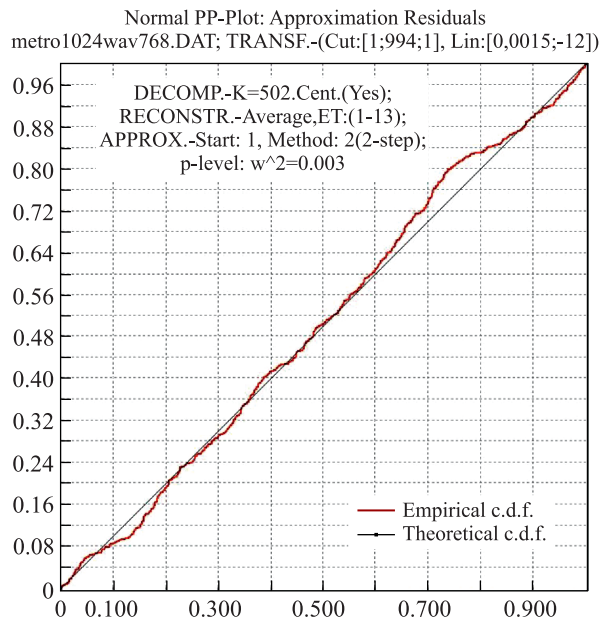


Рис. 13. Normal-PP-plot распределение, отвечающее отброшенным компонентам

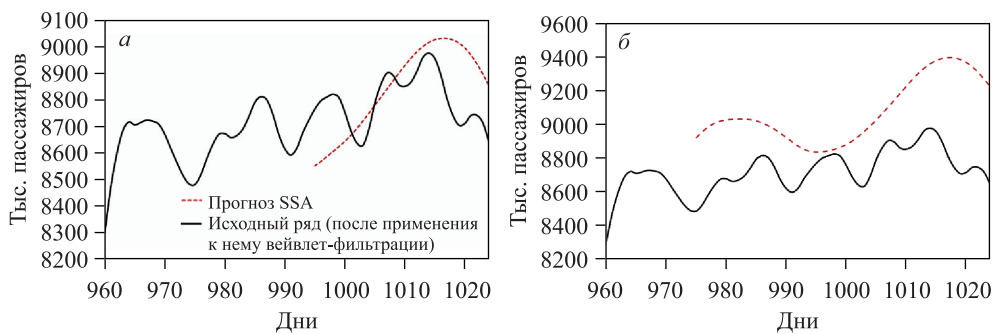


Рис. 14. Результаты прогнозирования в подходе «Гусеница»-SSA на 30 (а) и 50 (б) дней наблюдения в сравнении с реальными данными

компоненты отвечают за трендовую и периодические (сезонные) составляющие анализируемого ряда.

На рис. 12 приведены: исходный ряд с его аппроксимацией, реконструированной по первым 13 компонентам (рис. а) и ряд, восстановленный на основе отбрасываемых компонент (рис. б).

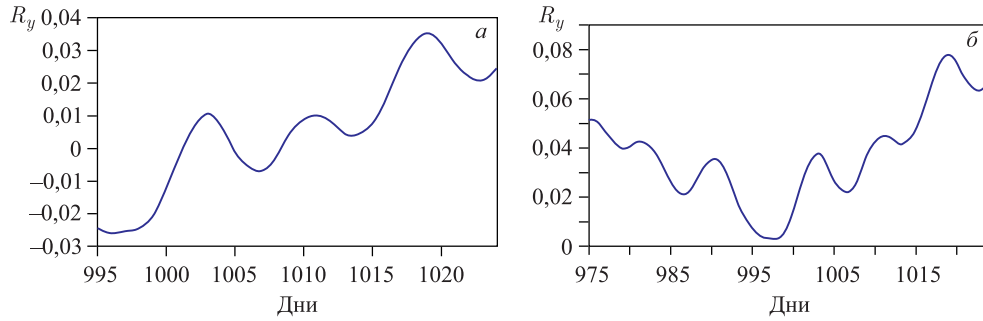


Рис. 15. Динамика величины R_y в зависимости от порядкового номера дня наблюдения при прогнозировании на основе метода, реализованного в подходе «Гусеница»-SSA на 30 (а) и 50 (б) наблюдений

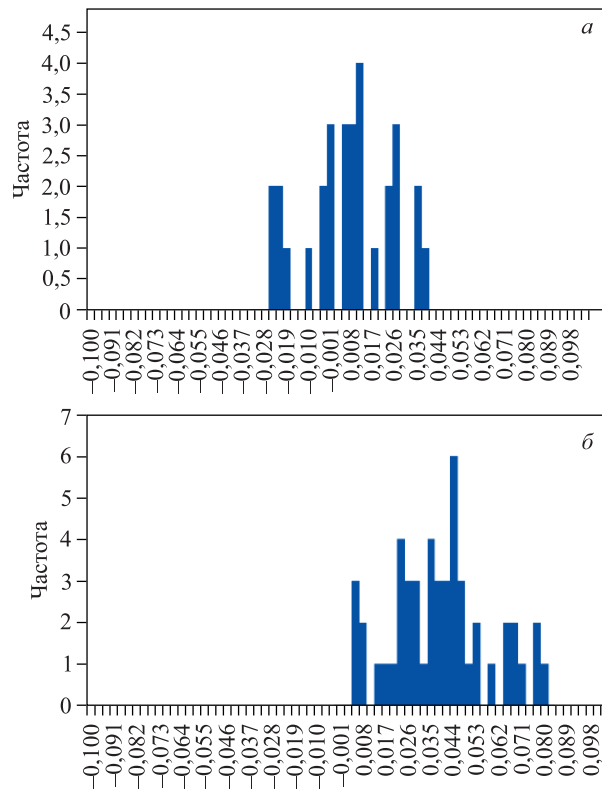


Рис. 16. Распределения переменной R_y при прогнозировании в подходе «Гусеница»-SSA на 30 (а) и 50 (б) наблюдений

На рис. 13 представлен PP-plot [31] функций накопленной вероятности для нормального распределения (изображенного отрезком прямой) и для данных, отвечающих отброшенным компонентам.

Видно, что отбрасываемые данные неплохо согласуются с нормальным распределением, из чего можно предположить, что отвечающий им процесс по своему поведению близок к гауссовскому шуму.

Заметим, что графики, представленные на рис. 11, 12 и 13, отвечают прогнозу на 30 дней наблюдения. Аналогичные графики для прогноза на 50 дней наблюдения имеют схожий характер.

На рис. 14 представлены результаты прогнозирования в подходе «Гусеница»-SSA на 30 (а) и 50 (б) дней наблюдения в сравнении с реальными данными.

На рис. 15 показана динамика величины R_y в зависимости от порядкового номера дня наблюдения при прогнозировании на основе метода, реализованного в подходе «Гусеница»-SSA на 30 (а) и 50 (б) дней наблюдения. На рис. 16 приведены распределения относительной ошибки R_y при прогнозировании в подходе «Гусеница»-SSA на 30 (а) и 50 (б) наблюдений.

Из представленных на рис. 15 и 16 результатов видно, что среднее значение ошибки R_y смещено относительно нуля в область положительных значений. При этом на всем интервале прогноза она не превышает 3% при прогнозировании на 30 дней наблюдения и приближается к 8% при прогнозировании на 50 дней наблюдения.

Приведенные результаты показывают, что хотя нам не удалось достичь точности прогнозирования, полученного с помощью МСП, тем не менее метод, реализованный в подходе «Гусеница»-SSA, позволяет проводить среднесрочный прогноз объемов пассажирских перевозок в Московском метрополитене с приемлемой точностью.

5. АНАЛИЗ РЕЗУЛЬТАТОВ И ВЫВОДЫ

В разд. 1 приведен набор переменных (Var1–Var6), которые использовались в качестве входных данных для МСП при прогнозировании суточных объемов пассажирских перевозок в Московском метрополитене. Очевидно, что для проведения такого прогноза нужно было предварительно решить задачи, связанные с прогнозированием двух переменных: Var5 (отклонение дневной температуры от нормы) и Var6 (суточное потребление электрической энергии в Московском регионе).

Задача прогнозирования переменной Var6 была успешно решена в работах [8, 9], однако вопрос с прогнозом переменной Var5 оставался открытым. В этой связи мы провели анализ того, как переменная Var5 влияет на результаты прогнозирования с помощью МСП.

На рис. 17 и 18 представлены сравнительные результаты прогнозирования с использованием развитых в настоящей работе подходов на 30 и 50 дней наблюдения, в том числе, с учетом и без учета вклада переменной Var5.

На этих рисунках приведены следующие графики: R — прогнозируемый ряд в исходных величинах; RF — прогнозируемый ряд, из которого с помощью вейвлет-фильтрации был удален высокочастотный шум; F1 — прогноз помощью МСП (без использования SSA-прогноза в качестве «подсказки»); F2 — прогноз с помощью SSA; F3 — прогноз с помощью МСП с использованием SSA-прогноза в качестве «подсказки» и без учета вклада переменной Var5; F4 — прогноз помощью МСП с включением SSA-прогноза в

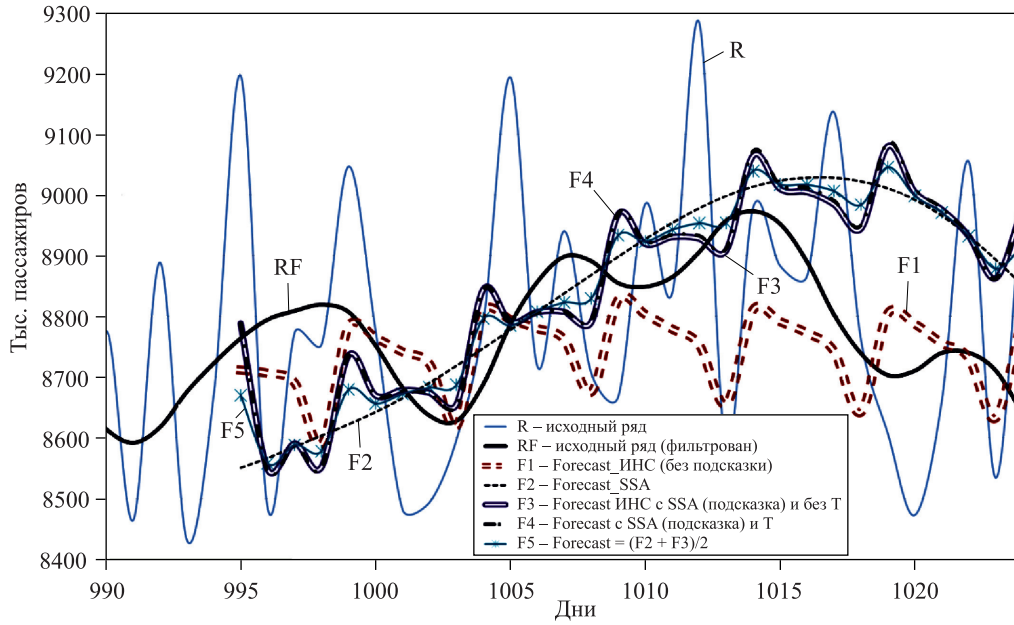


Рис. 17. Сравнительные результаты прогнозирования на 30 дней наблюдения с использованием развитых в настоящей работе подходов

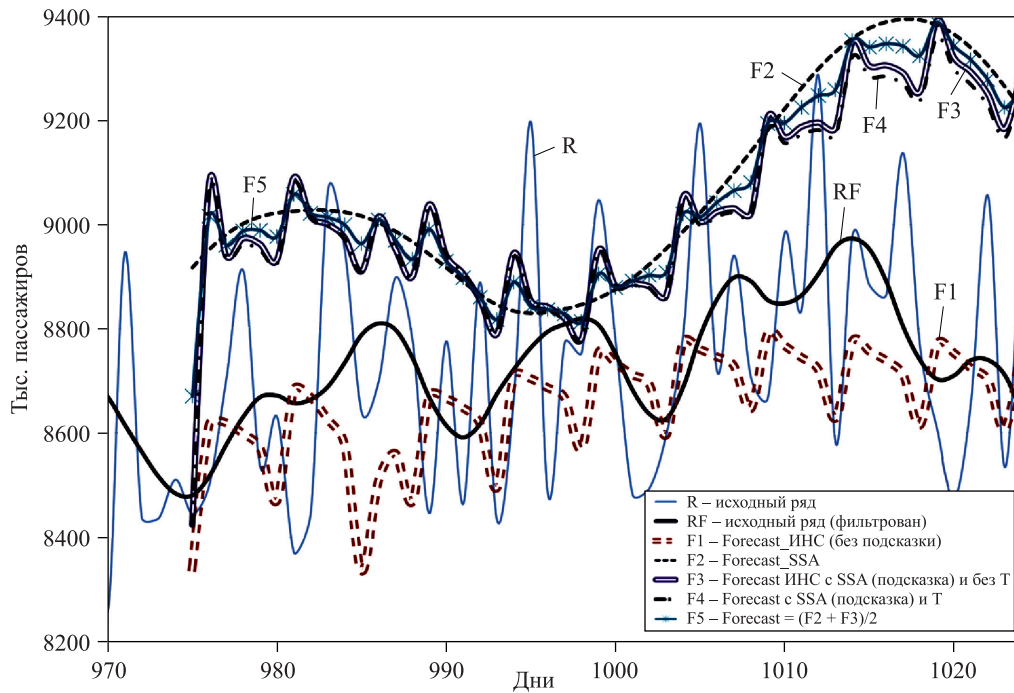


Рис. 18. Сравнительные результаты прогнозирования на 50 дней наблюдения с использованием развитых в настоящей работе подходов

качестве «подсказки», но без учета вклада переменной $Var5$; $F5$ — усредненный прогноз $(F2 + F3)/2$.

Анализ приведенных графиков позволяет сделать ряд выводов:

— наилучший вариант прогноза достигнут с помощью МСП (без использования SSA-«подсказки»);

— исключение переменной $Var5$ из набора входных данных для МСП не оказало заметного влияния на результат прогноза; небольшое отличие на графиках $F3$ и $F4$ можно увидеть только при прогнозе на 50 дней наблюдения (см. рис. 18);

— прогноз на основе подхода «Гусеница»-SSA оказался смещенным относительно прогнозируемых данных в область больших значений пассажиропотока; при использовании SSA-прогноза в качестве «подсказки» для МСП это же привело к смещению прогнозных кривых $F3$ и $F4$ в область больших значений;

— наряду с двумя рассмотренными вариантами прогноза (МСП и SSA) для надежности целесообразно использовать усредненный прогноз: $(F2 + F3)/2$.

ЗАКЛЮЧЕНИЕ

В настоящей работе разработана методика для среднесрочного прогнозирования суточных объемов пассажирских перевозок в Московском метрополитене, включающая три разных варианта прогноза: 1) на основе искусственных нейронных сетей; 2) с применением сингулярно-спектрального анализа, реализованного в подходе «Гусеница»-SSA; 3) при совместном использовании ИНС и подхода «Гусеница»-SSA. Нами показано, что развитые методы и алгоритмы позволяют с приемлемой точностью проводить среднесрочное прогнозирование пассажиропотока на ММ.

Наилучшего варианта прогноза удалось добиться с помощью ИНС прямого типа — многослойного перцептрона, на вход которого подавался набор факторов, влияющих на суточные объемы пассажирских перевозок в ММ. Проведенный нами анализ показал, что из указанного набора можно исключить один из факторов, а именно, отклонение дневной температуры от среднестатистической нормы, так как данный фактор не оказывает заметного влияния на результат прогноза.

Прогноз, построенный на основе сингулярно-спектрального анализа, реализованного в пакете «Гусеница»-SSA, при котором использовались только данные суточных объемов перевозок пассажиров в ММ, оказался смещенным относительно прогнозируемых данных в область больших значений пассажиропотока. Это же явилось причиной смещения прогнозных значений в область больших величин пассажиропотока в варианте совместного использования МСП и подхода «Гусеница»-SSA.

Развитая нами методика среднесрочного прогнозирования с достигнутой в настоящей работе точностью будет способствовать повышению эффективности и скорости принятия управленческих решений в зависимости от ситуации и на метрополитене, и в окружающей его внешней среде, включая климатические условия и динамику энергопотребления как индикатора промышленной, деловой и социальной активности населения региона, который использует метрополитен в качестве одного из основных видов транспорта. Это создаст необходимые условия для выработки предприятиями городского пассажирского транспорта эффективных стратегий и оперативных мер исходя из прогнозируемых объемов пассажиропотока в ММ, в том числе: 1) определения количества необходимого подвижного состава для перевозки пассажиров; 2) закупки электрической энергии для тяги

поездов; 3) подготовки и проведения специальных акций по привлечению пассажиров в определенные временные интервалы; 4) принятия решения о необходимости увеличения на определенных участках транспортной сети обслуживающего персонала и сотрудников метрополитена; 5) краткосрочного планирования различного рода мероприятий и др.

СПИСОК ЛИТЕРАТУРЫ

1. Московский метрополитен. Официальный сайт. <http://mosmetro.ru/> (дата обращения: 01.09.2015).
2. *Артемкина Е. Е.* Методы определения спроса на пассажирские автоперевозки и их значение для эффективной организации системы пассажирского автотранспорта // Актуальные вопросы экономических наук: Материалы III междунар. науч. конф., Уфа, Россия, июнь 2014 г. Уфа, 2014. С. 169–171.
3. *Правдин Н. Н., Негрей В. Я.* Прогнозирование пассажирских перевозок. М.: Транспорт, 1980.
4. *Леванова Д. С.* Прогнозирование пассажиропотока метрополитена на основе математических моделей. Автореф. дис. . . канд. техн. наук. СПб., 2005.
5. *Кударов Р. С.* Математические модели формирования входного потока пассажиропотока станций метрополитена. Автореф. дис. . . канд. техн. наук. СПб., 2009.
6. *Иванов В. В., Осетров Е. С.* Прогнозирование объемов пассажирских перевозок в Московском метрополитене с помощью искусственных нейронных сетей // Вестн. Нац. исслед. ядерного ун-та «МИФИ». 2016. Т. 5, № 1. С. 65–74.
7. *Иванов В. В., Осетров Е. С.* Прогнозирование пассажиропотока в Московском метрополитене на основе нейронных сетей с предварительной фильтрацией анализируемых данных // Там же. № 2. С. 162–169.
8. *Иванов В. В., Крянев А. В., Осетров Е. С.* Прогнозирование суточного потребления электроэнергии в Московском регионе с использованием искусственных нейронных сетей // Письма в ЭЧАЯ. 2017. Т. 14, № 4(209). С. 418–432.
9. *Иванов В. В., Крянев А. В., Осетров Е. С.* Прогнозирование суточного потребления электроэнергии в Московском регионе на основе сингулярно-спектрального анализа // Вестн. Нац. исслед. ядерного ун-та «МИФИ» (в печати).
10. Главные компоненты временных рядов: метод «Гусеница» / Ред.: Д. Л. Данилов, А. А. Жиглявский. СПб.: Изд-во СПбГУ, 1997.
11. *Golyandina N., Nekrutkin V., Zhigljavsky A.* Analysis of Time Series Structure: SSA and Related Techniques. Chapman & Hall/CRC, 2001.
12. *Галяндина Н. Э., Некруткин В. В., Браулов К. А.* Метод «Гусеница»-SSA: анализ временных рядов. Gistat Group, <http://www.gistatgroup.com/gus>. 2002.
13. Общая теория статистики: Учеб. / Под ред. Р. А. Шмойловой. 3-е изд., перераб. М.: Финансы и статистика, 2002. 560 с.
14. *Светульников И. С., Светульников С. Г.* Методы и модели социально-экономического прогнозирования: Учеб. и практикум для акад. бакалавриата. М.: Юрайт, 2014. 351 с.
15. <http://www.theignatpost.ru/magazine/index.php?mlid=2905>
16. *Antoniou I., Ivanov V. V., Ivanov V. V., Zrellov P. V.* Wavelet Filtering of Network Traffic Measurements. JINR Commun. E11-2002-223. Dubna, 2002. 21 p.; Physica A. 2003. V. 324. P. 733–753.
17. *Eadie W. T., Dryard D., James F. E., Roos M., Sadoulet B.* Statistical Methods in Experimental Physics. Amsterdam; London: North-Holland Publ. Comp., 1971.

18. *Мартынов Г. В.* Критерии омега-квадрат. М.: Наука, 1978.
19. *Broomhead D. S., King G. P.* Time-Series Analysis // Proc. Roy. Soc. London. 1989. V. 423. P. 103–110.
20. *Broomhead D. S., King G. P.* Extracting Qualitative Dynamics from Experimental Data // Physica D. 1986. V. 20. P. 217.
21. *Albano A. M., Muench J., Schwartz C., Mees A. I., Rapp P. E.* Singular Value Decomposition and the Grassberger Procaccia Algorithm // Phys. Rev. A. 1988. V. 38. P. 3017.
22. *Denby B.* Tutorial on Neural Networks Applications in High Energy Physics: 1982 Perspective // Proc. of the Second Intern. Workshop on “Software Engineering, Artificial Intelligence and Expert System in High Energy Physics”, Jan. 13–18, 1992. L’Agelaude France-Telecom La Londe-les-Maures, France. New Computing Techniques in Physics Research II / Ed. by D. Perret-Gallix. World Sci., 1992. P. 287.
23. *Fogelman Soulie F.* Neural Networks for Patterns Recognition: Introduction and Comparison to Other Techniques // Ibid. P. 277.
24. *Cybenko G.* Approximation by Superposition of a Sigmoidal Function // Math. Control Signals Syst. 1989. V. 2. P. 303.
25. *Peterson C., Rognvaldsson Th., Lönnblad L.* JETNET 3.0 — A Versatile Artificial Neural Network Package // Comp. Phys. Commun. 1994. V. 81. P. 185.
26. *Hoecker A., Speckmayer P., Stelzer J., Therhaag J., E. von Toerne, Voss H.* TMVA 4.2.0 — Toolkit for Multivariate Data Analysis with ROOT. arXiv:physics/0703039 [Data Analysis, Statistics and Probability]. CERN-OPEN-2007-007. TMVA version 4.2.0. Oct. 4, 2013; <http://tmva.sourceforge.net>.
27. *Brun R., Rademakers F.* ROOT — An Object Oriented Data Analysis Framework // Nucl. Instr. Meth. A. 1997. V. 389. P. 81.
28. *Lahmiri S.* A Comparative Study of Backpropagation Algorithms in Financial Prediction // Intern. J. Comp. Sci., Engin. Appl. 2011. V. 1, No. 4. P. 15–21.
29. *Elsner J. B., Tsonis A. A.* Singular Spectrum Analysis. A New Tool in Time Series Analysis. New York; London: Plenum Press, 1996.
30. <http://www.gistatgroup.com/cat/>
31. https://en.wikipedia.org/wiki/P-P_plot