

Прогнозирование потоков груженых вагонов по базам заявок на перевозки грузов



А. Н. Баушев,
к.ф.-м.н., доцент кафедры
«Информационные
и вычислительные
системы» Петербургского
государственного
университета путей
сообщения (ПГУПС)



А. Т. Осьминин,
д.т.н., заместитель
председателя
Объединенного ученого
совета ОАО «РЖД»



Л. А. Осьминин,
к.т.н., научный сотрудник
кафедры «Управление
эксплуатационной
работой» ПГУПС

Задачи прогнозирования и планирования размеров движения груженых вагонов по сети являются фундаментальными в области системного управления грузовым железнодорожным транспортом. Особое место среди них занимают задачи месячного планирования, поскольку их решение определяет совокупность всех количественных показателей технического нормирования эксплуатационной работы на плановый период.

Значимые отклонения фактических размеров движения от плановых сопровождаются, как правило, увеличением напряженности работы железнодорожного транспорта, дополнительными ресурсными и финансовыми затратами. Для уменьшения этих затрат практикуются подекадная корректировка месячных планов и пересчет технических норм [1].

В связи с общей тенденцией к автоматизации процессов управления железнодорожным транспортом возникает задача создать полностью автоматизированную систему месячного планирования размеров грузового движения и расчета показателей технического нормирования. Возможно ли решение этой задачи при современном уровне информатизации и организации процессов перевозки грузов по железным дорогам? Ответ на этот вопрос кардинальным образом зависит от математических моделей и методов, используемых для прогнозирования.

Современная теория прогнозирования временных рядов включает в себя широкий спектр математических моделей и соответствующих методов построения прогнозных значений рассматриваемых временных рядов [2], которые успешно применяются в задачах прогнозирования значений рыночных показателей. Однако непосредственное использование этих методов для решения указанной задачи не даст желаемых результатов, так как величины рыночных показателей оказывают лишь косвенное воздействие на систему управления перевозочным процессом. Колебания значений рыночных показателей отражаются прежде всего на рас-

пределении номенклатур перевозимых грузов, а не на размерах движения.

Мы предлагаем альтернативный подход к решению задач прогнозирования и планирования, основанный на обнаруженной нами тесной связи между совокупностью заявок на перевозки грузов, находящихся в базе данных в момент планирования, и совокупностью исполненных заявок.

В настоящее время отправка каждого груженого вагона осуществляется в строгом соответствии с предварительной заявкой (формы ГУ-12), согласованной с перевозчиком. В заявке, в частности, указываются: станция отправления, станция назначения, дата отправления, количество вагонов, номенклатура груза, род подвижного состава и тип заявки (маршрутная, повагонная, групповая и др.). Предварительные заявки подаются не ранее чем за 45 суток до предполагаемого момента отправки груза и в течение 45 суток с момента подачи могут корректироваться грузоотправителем. Кроме того, с течением времени могут появляться и реализовываться новые заявки, которых в момент запроса не было в базе данных. Эти два процесса приводят к тому, что фактический поток груженых вагонов может в несколько раз отличаться от заявленного в момент запроса потока.

Однако наши исследования посutoчных данных о фактических и заявленных потоках выявили сильную корреляцию между ними. Коэффициент корреляции в различных ситуациях оказывается весьма близким к единице, причем с увеличением интенсивности потока заявок этот коэффициент возрастает.

Согласно результатам исследований, в качестве базовой модели для

построения прогноза можно использовать модель линейной регрессии между фактическим и заявленным потоками. Параметры модели оцениваются на основании предыстории рассматриваемых потоков. Используя такой подход, можно прогнозировать достаточно интенсивные потоки (300 и более вагонов в сутки) с относительной ошибкой прогноза до 5 % независимо от спецификации потока. Для потоков с малой интенсивностью точность прогноза линейной модели уменьшается, однако она может быть повышена в случае применения более сложных моделей регрессионной зависимости.

Сегодня заявки подаются и формируются в электронной форме через автоматизированную систему ЭТРАН, причем данные о посutoчных поступлениях заявок и фактических отправках хранятся в течение шести месяцев. При необходимости (для прогнозирования потоков малой интенсивности и учета сезонных колебаний потоков) период хранения может быть увеличен до одного года.

Использование посutoчных данных для построения месячных прогнозов может показаться парадоксальным. Однако с помощью такого подхода можно прогнозировать не только объемы месячных перевозок, но и изменения интенсивности потоков груженых вагонов в течение планового периода, что важно для оперативного перераспределения ресурсов перевозчика и локального переопределения среднесуточных норм эксплуатационной работы.

Кроме того, данные о посutoчных планируемых отправках в скрытом виде содержат информацию о состоянии рынка железнодорожных грузоперевозок и ожидаемых критических событиях (например, о закрытии Северного морского пути для судоходства). Во время экономических кризисов суточные данные о заявках и отправках за несколько предшествующих месяцев довольно точно отражают общие тенденции, тогда как многолетние данные о ежемесячных отправках на протяжении периодов экономической стабильности могут приводить к составлению ошибочных планов.

Отмеченная взаимосвязь между фактическими и заявленными на начало планового периода потоками составляет основу предлагаемых методов решения актуальных задач прогнози-

рования и планирования на железных дорогах.

Из показателей, характеризующих размеры грузового движения, мы акцентируем внимание только на количестве перевозимых вагонов с грузом, поэтому в дальнейшем мы будем говорить о потоках груженых вагонов или просто о потоках.

Задачей прогнозирования мы называем задачу построения прогнозных значений для потоков без учета ограничений инфраструктуры и ресурсов перевозчика. Задачей планирования мы называем задачу прогнозирования с учетом ограничений. Предлагаемый подход к решению задач прогнозирования опирается на методы регрессионного анализа и в сущности не зависит от спецификации рассматриваемых потоков. Напротив, задачи планирования существенно различаются по своему содержанию и методам выделения потоков, удовлетворяющих ограничениям задачи.

Отметим, что фактические потоки всегда удовлетворяют инфраструктурным и ресурсным ограничениям, так же как и совокупности заявок, согласованных с перевозчиком. Процедура согласования заявки между участниками перевозочного процесса (на которую отводится 5 суток с момента подачи заявки) фактически заключает в себе решение локальной задачи планирования. Если бы все заявки планового периода подавались (и не изменялись) до его начала, то центр проблемы разработки автоматизированной системы планирования переместился бы с задачи прогнозирования на задачу полной автоматизации процесса согласования заявок.

Математические модели прогнозирования потоков груженых вагонов. Линейная модель построения прогнозов

В этом разделе мы будем считать, что спецификация прогнозируемого потока задана выделением совокупности информационных признаков, которые имеются в заявках. Речь может идти о потоках от станции назначения до станции отправления, о потоках между группами станций отправления и назначения, о междорожных потоках и т. д. Для удобства можно считать, что имеются две виртуальные станции: погрузки (отправления) (*A*) и выгрузки (назначения) (*B*).

Предположим, что сутки рассматриваемой истории процессов погрузок и заявок пронумерованы числами $1, \dots, T$. Заданы моменты построения прогнозов $1 = \tau_1 < \tau_2 < \dots < \tau_{k-1} < T = \tau_k$. Для $t \in [\tau_p, \tau_p^+)$ обозначим через x_t общее количество вагонов, которое должно быть отправлено в сутки с номером t согласно информации в базе заявок на момент τ_p . Пусть y_t — фактическое количество отправленных вагонов в сутки с номером t .

На рис. 1 показаны траектории процессов x_t (пунктирная линия) и y_t (сплошная линия), в которых *A* — это станция Ерунаково Западно-Сибирской железной дороги, *B* — станция Лужская (эксп.) Октябрьской железной дороги. В спецификации потока также задана номенклатура (каменный уголь). Параметр t пробегает значения 1, 2, ..., 120; нумеруются при этом сутки периода февраль — май 2013 г., а моменты построения прогнозов соответствуют первым числам месяцев указанного периода.

На рис. 1 не видно зависимости между процессами x_t и y_t . Выборочный

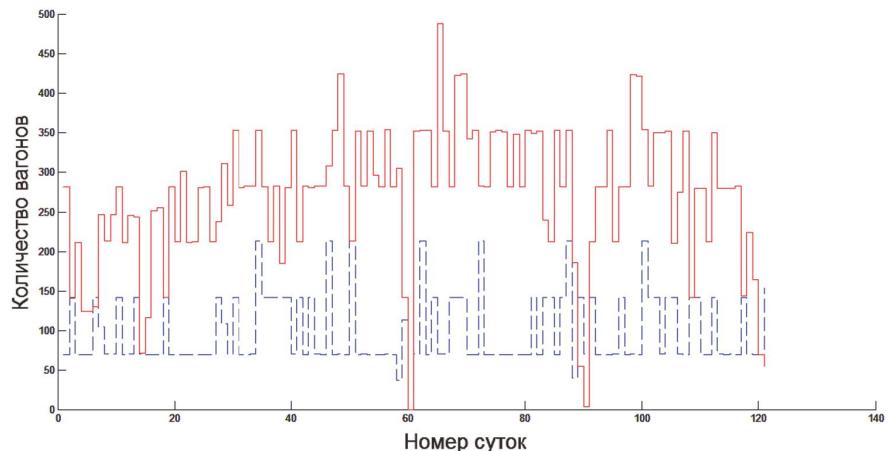


Рис. 1. Суточное количество заявленных на начало месяца (пунктирная линия) и фактически погруженных (сплошная линия) вагонов с номенклатурой «каменный уголь» в назначении Ерунаково — Лужская (эксп.) за февраль — май 2013 г. (120 сут.)

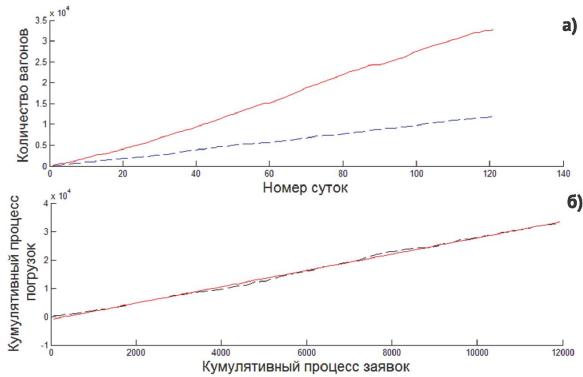


Рис. 2. Кумулятивные процессы заявок и погрузок (см. рис. 1):

а) зависимость процессов от времени; б) зависимость кумулятивного процесса погрузок от кумулятивного процесса заявок; прямая линия получена аппроксимацией зависимости методом наименьших квадратов.

сами x_t и y_t . Выборочный коэффициент корреляции [3] между заявленными и фактическими суточными погрузками равен 0,182.

Введем в рассмотрение кумулятивные процессы:

$$X_t = \sum_{s=1}^t x_s, \quad (1)$$

$$Y_t = \sum_{s=1}^t y_s, \quad (2)$$

называемые, соответственно, кумулятивным процессом заявок и кумулятивным процессом погрузок.

Траектории процессов X_t (пунктирная линия) и Y_t (сплошная линия) для рассмотренного примера показаны на рис. 2а).

Из графиков рис. 2а можно сделать вывод, что имеются приближенные равенства

$$X_t \approx \alpha_1 t + \alpha_0 \quad (3)$$

и

$$Y_t \approx \gamma_1 t + \gamma_0. \quad (4)$$

Коэффициенты α_1 и γ_1 представляют собой среднесуточное количество соответственно заявленных (на момент планирования) и фактических погрузок.

Из соотношений (3) и (4) следует, что имеется приближенное равенство

$$Y_t \approx \beta_1 X_t + \beta_0. \quad (5)$$

Отметим на плоскости в декартовой системе координат точки с координатами (X_t, Y_t) , откладывая по оси абсцисс значения кумулятивного процесса X_t , а по оси ординат — значения кумулятивного процесса Y_t . Эти точки будут лежать на кривой — графике зависимости процесса Y_t от процесса X_t .

Оценим параметры модели (5) методом наименьших квадратов (МНК) и проведем прямую, задаваемую соотношением (5). В результате мы получим прямую, представленную на рис. 2б.

Отметим, что модель (5) остается адекватной и в том случае, когда зависимости процессов X_t и Y_t от параметра t носят существенно нелинейный характер, т. е. соотношения (3) и (4) не выполняются.

На рис. 3а показаны траектории процессов X_t (пунктирная линия) и Y_t (сплошная линия), в которых А — это станция Трехсвятская Юго-Восточной железной дороги, В — станция Новороссийск (эксп.) Северо-Кавказской железной дороги.

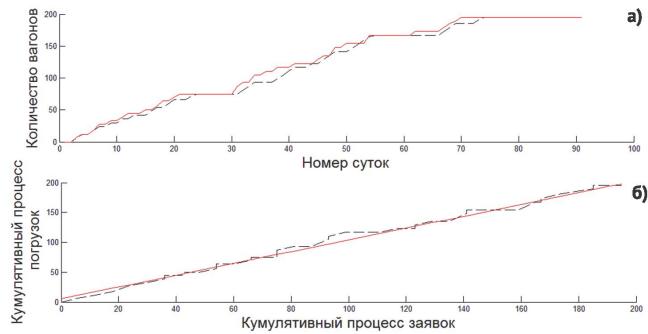


Рис. 3. Кумулятивные процессы заявок и погрузок по отправке зерна в назначении Трехсвятская — Новороссийск (эксп.) за сентябрь — ноябрь 2012 г.: а) зависимость процессов от времени; б) зависимость кумулятивного процесса погрузок от кумулятивного процесса заявок; прямая линия получена аппроксимацией зависимости методом наименьших квадратов

В спецификации потока также задана номенклатура (зерно). Параметр t пробегает значения 1, 2, ..., 91; нумеруют при этом сутки периода сентябрь — ноябрь 2012 г., а моменты построения прогнозов соответствуют первым числам месяцев этого периода.

Резюмируя эти наблюдения, можно прийти к заключению, что адекватной моделью для построения прогнозов потоков груженых вагонов по базе заявок может служить модель простой линейной регрессии между кумулятивными процессами фактических и заявленных отправок в момент построения прогноза. Эта модель имеет вид

$$Y_t = \beta_1 X_t + \beta_0 + \varepsilon_t, \quad (6)$$

где ε_t — ошибка модели, а параметры β_1 и β_0 оцениваются методом наименьших квадратов.

Обозначим соответствующие оценки $\hat{\beta}_1$ и $\hat{\beta}_0$.

Пусть $\tau \leq t_1 < t_2$. Тогда прогноз общего количества вагонов, которое будет отправлено на временном промежутке $[t_1, t_2]$, составит

$$\hat{Y}(t_1, t_2) = \hat{\beta}_1 X(t_1, t_2), \quad (7)$$

где $X(t_1, t_2)$ — общее количество заявленных на момент τ вагонов, подлежащих отправке на этом временном промежутке. Коэффициент масштаба β_1 в модели (6) называется коэффициентом искажения заявленного к моменту прогноза плана.

Естественная мера точности прогноза (7) — относительная ошибка

$$\delta(t_1, t_2) = \frac{|Y(t_1, t_2) - \hat{Y}(t_1, t_2)|}{Y(t_1, t_2)} \cdot 100\%. \quad (8)$$

Относительная ошибка прогноза по формуле (7) общего количества вагонов с углем в назначении Ерунаково — Лужская на май 2013 г. по предыдущим трем месяцам составила 8,2 %. Для прогноза общего количества вагонов с зерном в назначении Трехсвятская — Новороссийск на ноябрь 2012 г. по предыдущим двум месяцам относительная ошибка составила 33 %.

На основании результатов статистического анализа примеров была выявлена определенная закономерность: чем больше среднесуточная интенсивность заявленного в момент прогнозирования потока и чем меньше его дисперсия, тем меньше ожидаемая относительная ошибка прогноза.

Это можно объяснить следующим образом. Большая среднесуточная интенсивность и маленькая дисперсия означают запланированную высокую ритмичность суточных отправок вагонов, обеспеченную действующим графиком движения. Даже в случае отклонений фактических суточных отправок от заявленных ритмичность сохраняется. Таким образом, заявленный и фактический кумулятивные процессы отправок «почти линейно» зависят от времени. Поэтому числитель дроби (абсолютная ошибка) в (7) оказывается сравнительно малым, а знаменатель — большим.

Сегментированная кусочно-линейная модель прогнозирования

Статистический анализ примеров позволяет сделать заключение, что ошибки модели (6) имеют нулевое математическое ожидание, а их распределение может в первом приближении считаться нормальным. Однако ошибки являются попарно коррелированными и имеют дисперсию, зависящую от параметра t . Следовательно, возникает задача уточнения модели (6).

Замена линейной функции полиномом более высокой степени или каким-либо другим параметрическим семейством гладких функций для этой цели не подходит. Во-первых, использование таких параметрических семейств не имеет под собой физической основы. Во-вторых, хорошо известно, что чем точнее с помощью таких семейств функций решается задача интерполяции функции на заданном промежутке, тем большие ошибки могут возникать при использовании интерполирующей функции для решения задачи экстраполяции за пределы промежутка, т. е. для задачи прогнозирования.

В работе железнодорожного транспорта имеются естественные ритмы, определяющие жизненные циклы его функционирования и периоды планирования его работы. На графике кумулятивного процесса фактических отправок (рис. 2а) можно заметить, что первым числом месяцев плановых периодов соответствует «плато», за которым следует рост отправок. Детальный анализ месячных кумулятивных процессов фактических отправок в разных ситуациях позволяет и внутри месяца выделить фазы, которые условно можно назвать первой, средней и последней частями месяца.

Следует отметить, что не существует универсального деления на фазы всех потоков. Для некоторых потоков подходящее деление — разбиение месяца на декады, для других — на первую неделю месяца, среднюю часть, состоящую из двух следующих недель, и завершающую часть месяца. В общем, подходящие разбиения могут сильно зависеть от рассматриваемого потока и определяются особенностями соответствующего потоку действующего графика движения грузовых поездов.

Моменты перехода от некоторой фазы к следующей за ней фазе называются *моментами разладки* наблюдаемого процесса. В эти моменты изменяется характер процесса. Задачи определения таких моментов возникают в различных областях науки, техники и медицины [4–6]. В отечественной литературе совокупность этих задач получила название «задача о разладке», а в зарубежной — «change-point process problem».

Отметим, что универсального метода решения задачи о разладке не существует. Все известные методы ее решения опираются на априорные предположения как о наблюдаемом процессе, так и о характере разладки. Предлагаемый нами метод определения моментов разладки в процессах фактических отправок подробно проиллюстрирован в [7] на примере так называемой «задачи о подмене монеты».

Рассмотрим задачу уточнения модели (6). Пусть прогнозируемый период — месяц, моменты построения прогнозов τ_1, \dots, τ_k совпадают с первыми числами месяцев μ_1, \dots, μ_k рассматриваемой предыстории, T_1, \dots, T_k — длительности (сутки) соответствующих месяцев. Пусть τ — первое число прогнозного месяца μ , а T — его длительность. Для суток внутри месяцев мы будем использовать также внутреннюю (естественную) нумерацию.

Положим $N = \min\{T, T_1, \dots, T_k\}$ и рассмотрим в качестве возможных моментов разладки множество $\Sigma = \{\sigma = (\sigma_1, \sigma_2) | \sigma_2 \geq \sigma_1 + 2, 1 \leq \sigma_1, \sigma_2 \leq N\}$.

Пусть $\sigma \in \Sigma$, μ_i — произвольный месяц из предыстории ($i \in (1, \dots, k)$). Положим $\eta_i^1 = (1, \dots, \sigma_1)$, $\eta_i^2 = (\sigma_1 + 1, \dots, \sigma_2)$, $\eta_i^3 = (\sigma_2 + 1, \dots, N)$. Последовательности $\eta_i^1, \eta_i^2, \eta_i^3$ представляют собой, соответственно, первую, вторую и третью (возможно, без одного-двух последних дней) фазы месяца μ_i при условии, что моменты разладки совпадают с моментами σ_1 и σ_2 .

Введем в рассмотрение соответствующие конкатенации фаз: $\eta^1 = [\eta_1^1, \dots, \eta_k^1]$, $\eta^2 = [\eta_1^2, \dots, \eta_k^2]$, $\eta^3 = [\eta_1^3, \dots, \eta_k^3]$.

Пусть $X^1(t)$ и $Y^1(t)$, ($t = 1, \dots, k\sigma_1$) — кумулятивные процессы заявок и фактических погрузок, отвечающие конкатенации первых фаз η^1 ; $X^2(t)$ и $Y^2(t)$, ($t = 1, \dots, k(\sigma_2 - \sigma_1)$) — кумулятивные процессы заявок и фактических погрузок, отвечающие конкатенации вторых фаз η^2 ; $X^3(t)$ и $Y^3(t)$, ($t = 1, \dots, k(N - \sigma_2)$) — кумулятивные процессы заявок и фактических погрузок, отвечающие конкатенации третьих фаз η^3 .

Пусть $\hat{\beta}_1^i, \hat{\beta}_0^i$ — МНК-оценки параметров линейной регрессионной зависимости (6) процесса $Y^i(t)$ от процесса $X^i(t)$ и $R_i(\sigma) = \sum_t (Y^i(t) - \hat{\beta}_1^i X^i(t) - \hat{\beta}_0^i)^2$, $i = 1, 2, 3$.

Положим $R(\sigma) = R_1(\sigma) + R_2(\sigma) + R_3(\sigma)$ и рассмотрим оптимизационную задачу

$$R(\sigma) \rightarrow \min, \sigma \in \Sigma. \quad (9)$$

Пусть $\sigma^* = (\sigma_1^*, \sigma_2^*)$ — решение задачи (9) и $\beta_1^{i*}, \beta_0^{i*}$ — соответствующие оценки параметров β_1^i, β_0^i , $i = 1, 2, 3$.

Используя полученные оценки, построим прогноз кумулятивного процесса отправок на месяц μ :

$$\hat{Y}(t) = \begin{cases} \beta_1^{1*} X(t) + \beta_0^{1*}, & \text{если } 1 \leq t \leq \sigma_1^*; \\ \hat{Y}(\sigma_1^*) + \beta_1^{2*} X(t) + \beta_0^{2*}, & \text{если } \sigma_1^* + 1 \leq t \leq \sigma_2^*; \\ \hat{Y}(\sigma_2^*) + \beta_1^{3*} X(t) + \beta_0^{3*}, & \text{если } \sigma_2^* + 1 \leq t \leq T. \end{cases} \quad (10)$$

Прогноз количества отправок вагонов на промежутке $t_1 < t \leq t_2$ определяется по формуле

$$\hat{Y}(t_1, t_2) = \hat{Y}(t_2) - \hat{Y}(t_1). \quad (11)$$

В рассмотренных нами примерах прогноз, рассчитанный по формулам (10), (11), давал более точные оценки, чем прогноз, рассчитанный по линейной модели, причем зачастую это улучшение было весьма существенным. Так, при определении количества вагонов с углем, отправленных со станции Ерунаково на станцию Лужская (эксп.), относительная ошибка прогноза была снижена с 8,2 до 2,0 %. В случае оценки



ФОТО: СЕРГЕЙ ТЮРИН

В случае оценки количества вагонов с зерном, отправленных со станции Тресвятская на станцию Новороссийск (эксп.), относительную ошибку прогноза удалось снизить с 33 до 14 %.

Автоматизированная система прогнозирования потоков груженых вагонов

Прогнозирование — сложная задача в любой сфере жизни. Кто может сказать, что произойдет завтра? Знаменитый французский ученый Пьер-Симон Лаплас в работе [8] рассматривал задачу оценки вероятности, что завтра взойдет солнце. Разумеется, Лапласа, которого по праву считают «отцом небесной механики», интересовал не факт восхода солнца, а необходимость применения теории вероятностей и математической статистики при решении научных проблем. Лаплас был убежденным детерминистом, само понятие вероятности он рассматривал как следствие неполноты наших знаний о детерминированных законах, управляющих мировыми процессами. Если законы небесной механики известны, то вопрос о восходе солнца решается без привлечения теории вероятности, в противном случае можно только оценить вероятность восхода солнца по статистическим данным прошлых лет.

Похожая ситуация имеет место и в рассматриваемой нами задаче прогнозирования потоков вагонов по сети железных дорог. Если бы грузоотправители подавали все заявки на плановый месяц до его начала, то прогноз был бы абсолютно точным. Однако сегодня заявки поступают непрерывно, что приносит стохастичность и в работу грузового железнодорожного транспорта, и в задачу построения прогнозов.

Безусловно, закономерности в работе железнодорожного транспорта, формирующие ритмичность и устойчивость его работы, доминируют над стохастичностью. В модели построения прогнозов следовало бы учитывать действующие графики движения грузовых поездов. Однако разнообразие этих графиков привело бы к значительному усложнению моделей построения прогноза. Более того, число моделей прогноза выросло бы во много раз. На наш взгляд, при построении прогноза предпочтительнее иметь дело с простой универсальной моделью, чем с большим семейством сложных разнородных моделей.

Когда речь идет о создании автоматизированной системы прогнозирования и планирования, основной вопрос — точность прогнозов. Достаточна ли точность прогнозов, получаемых по описанной нами методике, чтобы взять ее за основу при разработке расчетного модуля автоматизированной системы?

Мы думаем, что ответ на этот вопрос положителен. Большое количество рассмотренных нами примеров позволяет заключить, что для потоков большой интенсивности относительная ошибка не превышает 5 %. Для потоков малой интенсивности ошибка возрастает. Кроме того, существуют спецификации потоков, при которых в момент построения прогноза заявок в базе данных нет, а фактические потоки в течение планового периода появляются. Для прогнозирования таких потоков нужны другие методы.

Тем не менее, предложенную нами методику можно положить в основу автоматизированной системы прогнозирования. Основное назначение такой системы — служить вспомогательным

инструментом для решения задач технического нормирования и распределения ресурсов перевозчика. Отметим, что показатели технического нормирования в основном определяются потоками большой интенсивности. Кроме того, возможно дальнейшее развитие автоматизированной системы, основанной на предложенной методике.

Литература

- Сметанин А. И. Техническое нормирование эксплуатационной работы железных дорог. М.: Транспорт, 1984. 295 с.
- Цыплаков А. Введение в прогнозирование в классических моделях временных рядов // Квантарь. 2006. № 1. С. 3–19.
- Лагутин М. Б. Наглядная математическая статистика. М.: БИНОМ. Лаборатория знаний, 2007. 472 с.
- Ширяев А. Н. Вероятностно-статистические методы в теории принятия решений. М.: МНЦМО, 2011.
- Polunchenko A. S., Sokolov G., Du W. Quickest Change-Point Detection: A Bird's Eye View // Proc. Joint Statist. Meet. Montreal, Quebec (Canada), 2013.
- Levielle M. Optimal segmentation of random processes // IEEE: Transactions on signal Proc. 1998. Vol. 46, № 5. P. 1365–1373.
- Баушев А. Н., Осьминин А. Т., Осьминин Л. А. и др. О моментах разладки в регрессионных моделях // Материалы XIX междунар. конф. по мягким вычислениям и измерениям (SCM-2016). В 2 т. Т. 1. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2016. С. 16–18.
- Лаплас П.-С. Опыт философии теории вероятностей. М.: ЛИБРОКОМ, 2011. 208 с.