

# Оценка пассажиропотока для новых линий железной дороги в Московском регионе

Д.Е. Намиот, О.Н. Покусаев, А.Е. Чекмарев

**Аннотация**— В статье рассматриваются вопросы, связанные с оценкой (прогнозированием) пассажиропотока для городских железных дорог в Московском регионе. Речь идет о так называемых Московских центральных диаметрах (МЦД) - проекте реконструкции существующих сквозных железнодорожных линий в Москве и Московской области и организации на них диаметральных маршрутов пригородных электропоездов. В проекте предполагается движение поездов повышенной комфортности с интервалами 5-6 минут. Очевидно, что такая схема движения вызовет увеличение количества пассажиров железной дороги на данных линиях. Вместе с тем, количественные прогнозные оценки этого увеличения остаются неизвестными. Схожие проекты (в России) отсутствуют, опросных значений также нет. Соответственно, модель пассажиропотока представляет собой комбинацию граничных значений (минимальный и максимально возможный пассажиропоток) и эвристик, предполагающих различные сценарии увеличения пассажиропотока. В работе рассмотрены методы измерения и оценки пассажиропотока, а также эвристики, использовавшиеся для оценки будущего спроса. В качестве граничных значений выступают текущие (реальные) данные о валидации проездных документов на существующих линиях и данные мобильных операторов о миграционных потоках.

**Ключевые слова**—цифровая урбанистика, транспортное планирование, моделирование, прогнозирование.

## I. ВВЕДЕНИЕ

В работе рассматривается задача прогнозирования пассажиропотока для новых линий городских железных дорог. Это так-называемые Московские центральные диаметры (МЦД) - проект реконструкции существующих сквозных железнодорожных линий в Москве и Московской области и организации на них диаметральных маршрутов пригородных электропоездов (рис. 1). В средствах массовой информации проект также называют «наземным метро», хотя фактически метрополитеном он не является [1]. Предполагается движение поездов повышенной комфортности с интервалами 5-6 минут.

Новые поезда запускаются по уже существующим

линиям железной дороги. Изменяется только связность системы. Появляется сквозное движение, которое отсутствовало ранее. Это сквозное движения соединяет две существовавшие до этого ветки железной дороги. До ввода такой единой схемы движения разные ветки железной дороги соединялись посредством городского транспорта уже в Москве.



Рис. 1. Центральные диаметры [2]

Помимо увеличения связности сообщения, на новых линиях будет увеличена частота движения. Также здесь будут использоваться поезда повышенной комфортности, к которым будет добавлена удобная форма оплаты (единая транспортная карта с городским пассажирским транспортом). По факту, это выглядит как новое метро (основной вид общественного транспорта в Москве). Естественно, что можно ожидать роста пассажиропотока для линий железных дорог, которые легли в основу новых сквозных маршрутов. Оценка этого прироста и есть задача, которая рассматривается в данной работе. Необходимость такой оценки вызвана как экономическими соображениями (нужно понимать экономику нового проекта), так и техническими соображениями – необходимо ответить на вопрос, достаточно ли будет существующей инфраструктуры (станций, переходов для систем городского общественного транспорта и т.д.) для обслуживания нового потока пассажиров? В первую очередь нас интересует прирост пассажиров при движении из

Статья получена 21 сентября 2018.

Д.Е. Намиот - МГУ имени М.В. Ломоносова; РУТ (МИИТ) (e-mail: dnamiot@gmail.com).

О.Н. Покусаев - Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ) (email: o.pokusaev@rut.digital)

А.Е. Чекмарев - Центр цифровых высокоскоростных транспортных систем РУТ (МИИТ) (email: a.chekmarev@rut.digital)

области в Москву.

Оставшаяся часть работы структурирована следующим образом. В разделе 2 рассматриваются имеющиеся средства измерения пассажиропотоков. В разделе 3 мы обсуждаем нашу модель, которая положена в оценку увеличения пассажиропотока после запуска сквозного движения.

## II. КАК ИЗМЕРЯЕТСЯ ПАССАЖИРОПОТОК

В настоящее время для железных дорог в Москве существующий пассажиропоток можно измерить довольно точно. По существующей схеме, на большинстве станций пассажиры должны предъявлять проездные документы дважды – при входе и при выходе. В терминах социальных сетей это соответствует (является аналогом) отметок check-in (при входе) и check-out (при выходе) [3]. Это повышает ценность информации об использовании железнодорожных станций в ряду других транспортных данных. Из информации об использовании сразу можно восстановить маршрут пассажира. Традиционно, проездные документы в Москве (в московском регионе) отмечаются только при входе (check-in). Соответственно, для восстановления маршрута необходимо пользоваться эвристическими алгоритмами для восстановления маршрутов. Обзор таких алгоритмов есть, например, в нашей работе [4].

Смысл состоит в том, что перерывы в использовании проездных документов как раз и являются разграничителями поездок. Например, если мы отметили использование проездной карты (в Москве это может быть карта Тройка) на станции метро А, а далее – после относительно длительного перерыва на станции В, то мы можем предположить, что пассажир проехал от А до В, далее, например, находился на работе и после ее завершения совершает новую поездку. Выделение маршрутов по данным телекоммуникационных операторов также базируется на эвристиках [5]. Например, место, откуда совершаются звонки утром и поздним вечером, считается “домом”, в дневное время – “работой” и так далее. Здесь возможно также наложение дополнительных условий. Например, задание минимального суммарного времени пребывания в течение месяца, чтобы признать данное место работой. В случае же информации об использовании железнодорожных станций маршрут известен [6].

Данные по валидациям (предъявлениям/проверкам проездных документов) представлены в виде отдельных текстовых файлов, каждый из которых описывает проходы по конкретной станции за один месяц. Одна запись (строка в файле) соответствует одному проходу (на вход или выход). Данные полностью анонимны и не содержат никакой персональной информации. Может присутствовать тип используемого проездного документа (например, это льготный тариф или нет, является проездной документ одноразовым или переиспользуемым), но никакой идентификации

документов в принципе нет. Размер каждого такого лога зависит, конечно, от использования конкретной станции в конкретном месяце и колеблется в интервале 20 – 70 Мб.

Поля, которые содержатся в записях:

- Дата и время
- Характеристика цены (полный или льготный билет)
- Тип льготы (федеральная, РЖД и т.д.)
- Тип билета (разовый в одну сторону, разовый туда и обратно, абонемент и т.д.)
- Начальная станция
- Конечная станция

Такого рода информация позволяет иметь как сводные данные по станциям (вход/выход), так и, при необходимости, анализировать распределение пассажиров по входам/выходам. Например, ниже следующие рисунки показывают часовые распределения входов и выходов на одной подмосковной станции.

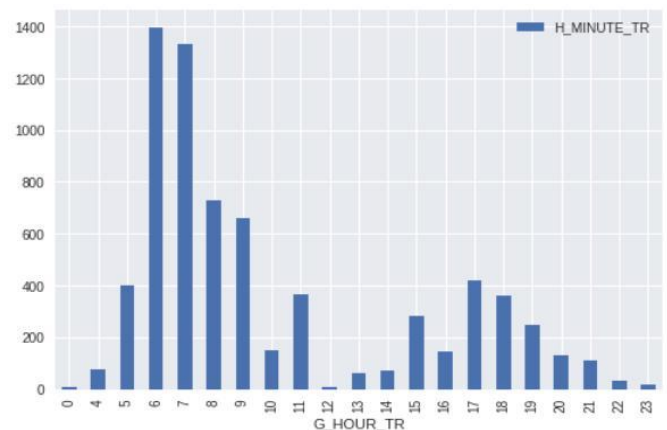


Рис. 2. Входы на станцию (Подмосковье)

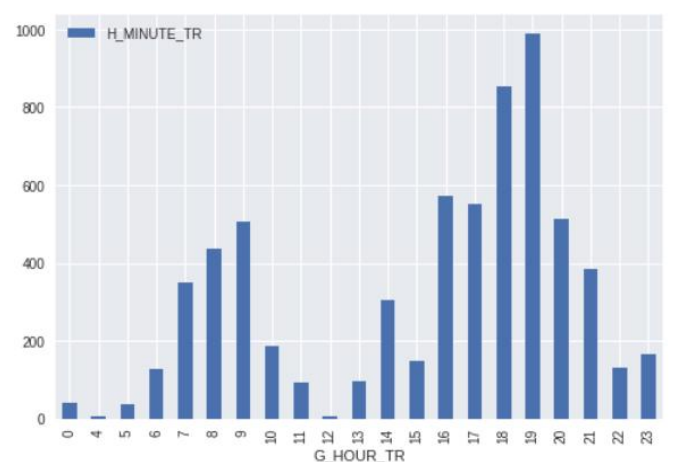


Рис.3. Выходы со станции (Подмосковье)

Это типичная картина с рабочим трафиком, где присутствует пик по входам утром и пик по выходам вечером (утром едут в Москву, вечером возвращаются с работы). Отметим, что такая картина характерна не для всех станций.

Эта информация, помимо суммарных данных, дает

еще возможность определить шаблоны использования станций. Насколько, например, представленная здесь картина использования станции утром и вечером стабильна в течение месяца? Технически, это сводится к проверке подобия временных рядов (например, часовые данные по входам за разные дни должны быть похожи, чтобы можно было бы говорить о шаблоне). Поскольку в нашем случае временные ряды имеют всегда равную длину, то могут использоваться самые разные методы сравнения подобия временных рядов. В простейшем случае можно просто считать Евклидово расстояние.

Анализ активности пассажиров по станциям показывает очень высокий уровень постоянства. Временные ряды по разным дням недели оказались очень похожи. Степень подобия рядов, измеряемая по разным метрикам – довольно большая. Это означает, что железнодорожный трафик довольно стабилен. Железная дорога имеет своих постоянных пассажиров, которые ей пользуются. И колебания тут (до запуска новых условий движения) довольно малы.

Для анализа данных использовался облачный инструмент от Google – Collaboratory [7]. Технически, это есть облачная реализация Jupyter notebook. Она не требует никакой установки программного обеспечения, все доступно через браузер. Файлы данных хранились на Google Drive.

В агрегированном виде, обработка данных валидаторов позволяет получить матрицу корреспонденций: для каждой пары станций указать количество пассажиров, перемещавшихся между этими двумя станциями. И такого рода матрицы могут быть построены в любом временном срезе: за час, день, месяц и т.д. Здесь для дальнейших рассуждений нас будет интересовать распределение пассажиров по станциям. Например, на маршруте (плече) Лобня – Тимирязевская, как распределяются пассажиры по промежуточным станциям? В свете указанного выше исследования шаблонов использования станций, это распределение остается довольно стабильным (хотя есть сезонные изменения).

Этот набор представляет собой объективные данные об имеющихся пассажирах железной дороги. Что может послужить основой для оценки потенциального спроса на такую услугу? В нашем случае мы использовали данные, которые собирают мобильные операторы. Телекоммуникационные операторы для того, чтобы выставлять пользователям счета за предоставленные услуги постоянно собирают информацию об активности данного пользователя (мобильного устройства). Это не связано ни с каким слежением за абонентами, это чисто экономическое мероприятие (нужно знать длительность совершенных звонков, чтобы подсчитывать их стоимость, номера, на которые совершались звонки, поскольку от этого также может зависеть их стоимость и т.д.). В этот же журнал попадает информация о том, какая сотовая вышка обслуживает в данный момент конкретное мобильное устройство. Соответственно,

отражается и факт изменения этой информации. А каждая сотовая вышка (базовая станция) имеет вполне конкретное местоположение (конкретные географические координаты). Соответственно, местоположение мобильного устройства (изменение местоположения) можно связать с географическими координатами. Это на самом деле, выглядит более сложно, чем описано здесь (необходимо моделировать распространение радиосигнала, знать расположение других базовых станций, диаграммы направленности и т.д.), но итог будет именно таким – можно оценивать географическое положение мобильного устройства.

Естественно, что эта оценка производится с какой-то точностью. Эта точность, очевидным образом, связана с количеством базовых станций, которые участвуют в расчетах. Поэтому можно сказать, что в городских условиях (большая плотность базовых станций) эта точность будет выше, чем там, где живет меньше народу и базовых станций меньше. В целом, в текущих условиях мы можем ожидать точности в несколько сотен метров (например, квадрат 500x500 метров). Отметим, что определение местоположения с какой-то закругленной точностью соответствует также необходимости соблюдения интересов приватности мобильных абонентов. Если перемещения выдаются для всех мобильных абонентов из некоторого географического квадрата, то нет возможности отслеживать таким образом перемещения любого отдельно взятого абонента.

В итоге, данные мобильных операторов дают возможность построить матрицу корреспонденции (перемещений) для географических квадратов. Для каждой пары квадратов (допустим, 500x500 метров) можно получить количество человек, перемещающихся из одного квадрата в другой. Опять-таки – во временном срезе: час, день, месяц и т.д. Общий размер такой географической сетки может варьироваться, от него, естественно, зависит размер данных. В нашем случае он охватывал достаточную зону в пригородах Москвы. Достаточность в данном случае предполагает либо непосредственное наличие железной дороги, проходящей через Москву, либо возможность добраться транспортом до станции этой дороги в разумное время.

Информация по географическим квадратам (в любом временном срезе) может быть агрегирована, так что мы можем получить информацию по перемещениям для административных районов. Ценность такой агрегации состоит в том, что по административным районам есть, как правило, статистика по населению, и это позволяет верифицировать собранные данные.

Итогом является получение объективной картины перемещений мобильных устройств (абонентов) из районов, примыкающих к железнодорожным станциям [8]. Из рассмотрения можно сразу исключить перемещения на близкие расстояния (предполагаем, что

это пешеходы). То, что останется – это пользователи транспорта. Собственно говоря, здесь и начинается транспортное моделирование. Предположим, что мы рассматриваем конкретный район в пригороде Москвы. Все эти перемещающиеся мобильные абоненты должны быть как-то распределены по видам транспорта. Точно известно, сколько их воспользовалось железной дорогой. Если есть информация по использованию автобусов в пригородах (например, валидации транспортных карт), то разница между общим числом абонентов и суммой пассажиров железной дороги и автобусов даст количество тех, кто будет добираться до Москвы автотранспортом. Это текущая картина. После ввода городской железной дороги она будет меняться. Потоки должны перераспределиться между видами транспорта (собственно, основная идея как раз состоит в том, что какая-то часть автомобилистов переседет на железнодорожный транспорт), возможно, появятся новые пассажиры, которые не ездили до этого из-за того, что добираться в город было неудобно.

### III. О МОДЕЛЯХ ОЦЕНКИ ПАССАЖИРОПОТОКА

Во-первых, из анализа операторских данных по перемещениям, например, за месяц можно определить характер перемещений по типу “дом-работа”. Это стандартный подход, который использовался еще в самых первых работах по анализу такого рода данных [9]. Идея состоит в соотношении времени суток и пользовательской активности. Если мобильный абонент постоянно проявляет активность (звонит, пишет SMS, etc.) из определенного региона в вечернее время, то этот регион может быть признан его “домом”. Аналогично, постоянная активность из конкретного региона в дневное время ведет к классификации этого региона как “рабочего”. Термин “постоянно” определяет здесь время, необходимое для классификации. Например, 7 дней в месяце. Задается это значение эмпирически.

После такой классификации становятся доступными два момента. Во-первых, мы можем описывать перемещения в терминах работа-дом. Это может быть более устойчивым шаблоном перемещения. Во-вторых, появляется возможность оценить реальную численность населения по географическим объектам (административным единицам, если агрегировать информацию из квадратов).

Оценка численности в таком случае является первым моментом оценки пассажиропотока. Это, очевидно, верхняя оценка возможного трафика железной дороги. Это есть теоретически возможное количество пассажиров. Больше их просто нет.

При этом мы предполагаем, что распределение “новых” пассажиров по станциям будет точно таким же, как и текущее распределение железнодорожных пассажиров по станциям. Иными словами, каждая станция получит свой пропорциональный прирост пассажиров.

С практической точки зрения, по большинству направлений (станций), полученные цифры превышали физические возможности (емкости) соответствующих

станций.

Здесь есть одно важное замечание по поводу классификации дом-работа. Вовсе не обязательно, что будут классифицированы оба элемента для каждого конкретного абонента. Например, абоненты с разъездной работой. Они не задерживаются подолгу на одном месте, чтобы классифицировать их активности там. Но они, тем не менее, создают нагрузку на транспортную систему. С другой стороны, могут быть абоненты, которые реально никуда не перемещаются. Например, пожилые люди или, наоборот, дети. При этом домашний район для них будет определен – будет реально зарегистрированная телефонная активность по месту жительства.

Отметим, что другой информации о поведении пассажиров нет. Исторические данные для такого рода проектов – отсутствуют по очевидным причинам. Опросные оценки также отсутствуют. Соответственно, для оценки будущего пассажиропотока применялись несколько эвристик, основанных на характеристиках текущего и максимально возможного пассажиропотоков. Примеры этих эвристик приведены ниже.

*А.* Оцениваем долю железнодорожных пассажиров для каждой из станций нового маршрута. Иными словами: какой процент из всех перемещающихся из данного района (по статистике мобильного оператора) реально перевозится железной дорогой (по данным валидации станции железной дороги). Определяем максимальный процент для всех станций. И далее предполагаем, что такой процент будет на каждой станции и, соответственно, пересчитываем их потоки. Возможное объяснение: после запуска новой схемы движения 100% населения не переседет на железную дорогу, максимальный существующий процент охвата железной дорогой и есть реально достижимая доля. Таким образом мы определяем новое количество пассажиров (входов) для станций, а далее эти пассажиры распределяются по другим станциям в той же пропорции, какая существует для реальных данных валидации железной дороги.

*В.* Мы можем оценить долю новых пассажиров для станций железной дороги по данным сезонных трендов. Летом количество жителей, которые ездят в город из пригородов увеличивается за счет тех горожан, которые летом живут на дачах, но ездят в город. Соответственно, мы получим какое-то увеличение количества населения в пригороде, которое перемещается (по данным оператора) из области в город и обратно. Из этих поездок какая-то доля придется на железнодорожный транспорт, что можно объективно определить по данным валидаций. Долю новых “жителей”, которые выбрали железнодорожный транспорт можно экстраполировать на все население района и таким образом подсчитать возможную долю железной дороги

в перевозках жителей в данном районе. Возможное объяснение: новые “жители” не имеют устойчивых предпочтений по сравнению с постоянно живущими, и их выбор будет отражать объективное соотношение долей транспорта

*C.* Большая часть жителей пригородов, которые совершают поездки, по крайней мере, один раз пользовалась железной дорогой. Такое предположение выглядит разумным, поскольку железная дорога, в любом случае, представляет собой самый простой способ добраться из пригорода в город. Появление новых пассажиров у железной дороги возможно только за счет того, что те, кто ездил редко, станут ездить чаще. Соответственно, помимо подсчета количества поездок, необходимо подсчитать их распределение по частоте. Технически – это еще одна группировка данных о перемещениях между районами (территориями, ячейками) по идентификатору пользователя. Данные у операторов (которые по-факту есть основа для систем биллинга) всегда содержат какую-то идентификацию абонента (например, IMEI или хэш-код, который заменяет этот идентификатор для сохранения приватности). Если подсчитать количество перемещений пользователей за месяц из района в район, то можно получить полезную частотную информацию. Например, для каждой конкретной пары районов за месяц:

*N* перемещений совершили столько-то абонентов  
*N-1* перемещений совершили столько-то абонентов  
*N-2* перемещений совершили столько-то абонентов  
 и т.д.

Из соображений приватности здесь можно и кластеризовать данные по числу перемещений. Например, с шагом 5:

От 5 до 10 перемещений совершили столько-то абонентов  
 От 10 до 15 перемещений совершили столько-то абонентов  
 От 15 до 20 перемещений совершили столько-то абонентов  
 и т.д.

Это даст распределение жителей (которые совершали поездки/перемещения) по частоте поездок. Именно из такого распределения и будут видны резервы прироста в поездках. В этом распределении нас будет интересовать низкочастотный “хвост”. Фактически, нас интересуют агрегированные данные (количество поездок) но только для части жителей (для низкочастотников). Мы предполагаем просуммировать эти низкочастотные поездки и далее провести что-то вроде стресс-теста:

- Если количество поездок увеличится на 25%, то мы получим такой-то прирост в пассажирах железной дороги (по валидациям билетов железной дороги мы знаем оценку доли железной дороги в числе перемещений – это из

другого отчета)

- Если количество поездок увеличится на 50%, то мы получим такой-то прирост в пассажирах железной дороги
- Если количество поездок увеличится на 100% (это в два раза больше), то мы получим такой-то прирост в пассажирах железной дороги

И для каждой цифры прироста можно будет оценить, хватает ли поездов (емкости платформ, пропускной способности турникетов) и каков будет прирост пассажиров на станциях метро.

*D.* Для районов, где возможен подсчет автомобилистов (как разницы между всеми перемещениями и валидациями проездных документов на железной дороге и в общественном транспорте) можно воспользоваться историческими данными по запуску городских железных дорог за рубежом. Идея пересадить автомобилистов на железную дорогу, на самом деле, была всегда одной из основных. В литературе отмечалось, что железная дорога в города – это единственный на сегодня транспорт, который может передвигаться без пробок и, в то же время, обеспечивать довольно высокий уровень комфорта. Отмечается, что городская железная дорога – это единственный на сегодня транспорт, на который пересаживаются автомобилисты. Поскольку, например, использование ими автобуса часто означает те же самые пробки, что и на персональном автомобиле, но только с меньшим уровнем комфорта (опять-таки, по сравнению с персональным автомобилем). Но при этом очевидно, что 100% замещения все равно не происходит. Цифра, встречающаяся в литературе – это до 20% - 25% пользователей автотранспорта, переключившихся на железную дорогу. Это определялось, естественно, уже после запуска проектов в эксплуатацию. Вот именно эту цифру – 20% от числа автомобилистов и можно использовать для подсчета увеличения количества пассажиров железной дороги.

#### IV. ЗАКЛЮЧЕНИЕ

В работе рассмотрены модели оценки пассажиропотока для новых линий городской железной дороги, проектируемых в Москве. В качестве исходных данных для построения прогноза была взята информация о текущей загрузке железнодорожного транспорта, данные об использовании общественного транспорта (на некоторых участках вне Москвы) и измерения, базирующиеся на данных мобильных операторов и описывающие перемещения мобильных абонентов в Московском регионе. Модель предлагает ряд эвристических подходов к оценке трафика, которые комбинируются с реальными данными существующего железнодорожного трафика.

#### БИБЛИОГРАФИЯ

- [1] Московские центральные диаметры. Каким будет новое наземное метро

- <https://www.m24.ru/articles/stroitelstvo/16112017/151562> Retrieved: Oct, 2018
- [2] Преимущества МЦД: от Лобни до Одинцово можно будет доехать за 60 минут <https://www.metronews.ru/novosti/moscow/reviews/preimuschestva-mcd-ot-lobni-do-odincovo-mozhno-budet-doehat-za-60-minut-1342251/> Retrieved: Oct, 2018
- [3] Namiot, Dmitry, and Manfred Sneps-Sneppe. "Customized check-in procedures." *Smart Spaces and Next Generation Wired/Wireless Networking*. Springer, Berlin, Heidelberg, 2011. 160-164.
- [4] Namiot, Dmitry, and Manfred Sneps-Sneppe. "A Survey of Smart Cards Data Mining." In *Supplementary Proceedings of the Sixth International Conference on Analysis of Images, Social Networks and Texts (AIST 2017) Moscow, Russia, July 27 - 29, 2017*
- [5] Steenbruggen J. G. M. et al. Real-time data from mobile phone networks for urban incidence and traffic management-a review of application and opportunities. – 2010.
- [6] Calabrese F. et al. Real-time urban monitoring using cell phones: A case study in Rome //IEEE Transactions on Intelligent Transportation Systems. – 2011. – Т. 12. – №. 1. – С. 141-151.
- [7] Google Collaboratory <https://research.google.com/colaboratory/unregistered.html> Retrieved: Sep, 2018
- [8] Намиот Д. Е. и др. Об оценке социально-экономических эффектов городской железной дороги //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 1.
- [9] Намиот Д. Е., Покусаев О. Н., Лазуткина В. С. О моделях пассажирского потока для городских железных дорог //International Journal of Open Information Technologies. – 2018. – Т. 6. – №. 3.

# On the assessment of passenger traffic for new rail lines in the Moscow region

Dmitry Namiot, Oleg Pokusaev, Alexander Chekmarev

*Abstract*— The article deals with issues related to the assessment (forecasting) of passenger traffic for urban railways in the Moscow region. We are talking about the so-called Moscow central diameters (MCD) - the project of reconstruction of the existing through railway lines in Moscow and the Moscow region and the organization of the diameter routes of suburban electric trains on them. The project assumes the movement of luxury trains at intervals of 5-6 minutes. It is obvious that such a scheme will cause an increase in the number of passengers on the railways on these lines. At the same time, quantitative forecast estimates of this increase remain unknown. Similar projects (in Russia) are not available, there are also no survey values. Accordingly, the model of passenger traffic is a combination of boundary values (minimum and maximum possible passenger traffic) and heuristics, involving various scenarios for increasing passenger traffic. The paper discusses the methods of measuring and estimating passenger traffic, as well as heuristics used to estimate future demand. The limit values are the current (real) data on the validation of travel documents on existing lines and the data of mobile operators on migration flows.

*Keywords*— digital urbanism, transport planning, modeling, forecasting.