

## **Глава 4. МЕТОДЫ ВОССТАНОВЛЕНИЯ МАТРИЦ КОРРЕСПОНДЕНЦИЙ ПО ДАННЫМ ИНТЕНСИВНОСТИ ДВИЖЕНИЯ**

### **4.1. Модели оценки матриц корреспонденций по данным интенсивности движения**

При выполнении проектов ОДД, реконструкции УДС одним из важнейших видов данных, на основе которых должны приниматься решения, является информация о существующем распределении транспортных потоков. Получить такую информацию можно, решая задачу оценки существующей матрицы корреспонденций, в которой значения корреспонденций представлены в виде потоков транспортных средств. В зарубежной теории и практике проектирования транспортных систем уже с 1970-х гг. уделялось большое внимание методам восстановления этого вида матриц, отнесенным специалистами “Мировой дорожной ассоциации” (PIARC) к числу важнейших инструментов анализа УДС. В Руководстве HCM 2000 [118] также уделено внимание восстановлению существующих матриц корреспонденций (гл. 29 “Corridor analysis”), изложен метод такой оценки применительно к магистральным дорогам и приведен его алгоритм.

В нашей стране в первую очередь получила развитие теория расчетов пассажиропотоков, в том числе были предложены методы оценки существующих матриц корреспонденций на маршрутном пассажирском транспорте [3,39,65]. Собственно методам восстановления матриц корреспонденций в виде потоков транспортных средств было посвящено лишь несколько исследований [22,37], что объяснялось относительно меньшим интересом к проблемам проектирования ОДД и реконструкции УДС. Продолжающийся рост автомобильного парка и объемов движения в российских городах требует совершенствования инструментария проектирования и

оценки УДС. Поэтому представляется интересным рассмотреть возможности решения указанной задачи применительно к российским условиям с учетом прежде всего используемых методов обследования УДС.

Приступая к краткому обзору моделей оценки существующих матриц корреспонденций, следует привести очень важную для понимания сути рассматриваемой задачи цитату из Руководства НСМ 2000 [118]: “Количество элементов матрицы корреспонденций всегда превосходит количество сегментов сети. Цель оценивания состоит не в определении точной матрицы корреспонденций, а в нахождении такой, которая достаточно близка к ней и соответствует данным интенсивности движения” (гл. 29 “Corridor analysis” с. 29-30). Следует пояснить, что под сегментами в оригинальном тексте понимаются дуги графа, которым представлено описание сети.

Поскольку количество зарубежных публикаций по рассматриваемому вопросу очень велико, отметим крупнейшие из них, а также работы, в которых приведены обзоры библиографии [75,77,81-84,86,90,92-94,98,106,108, 111,113,115,158,164,169,191,213-215,217-219,224-227]. По этой тематике наиболее часто публиковались и выступали на конференциях следующие авторы: M. G. H. Bell, D. E. Boyce, E. Cascetta, G. Davis, S. Erlander, C. Fisk, M. Florian, R. Hamerslag, C. Hendrickson, D. O. Jornsten, Y. Iida, J. T. Lundgren, M. J. Maher, S. McNeil, S. Nguyen, N. L. Nihan, T. Sasaki, H. Spiess, J. Van der Zijpp, H. J. Van Zuylen, L. G. Willumsenl, H. Yang.

Один из наиболее полных обзоров моделей восстановления матриц корреспонденций, в котором рассмотрены работы по данной тематике почти за 30 лет, представлен в отчете IIASA [75], в котором предлагается следующая классификация методов оценки существующих матриц корреспонденций:

методы, основанные на моделировании распределения транспортных потоков (traffic modeling based approaches);

статистические методы (statistical interference approaches);

градиентный метод (gradient based solution techniques).

Автор отчета [75] Т. Abrahamson полагает, что в наиболее общем виде модель восстановления существующей матрицы корреспонденций формулируется как

$$\min F(g, v) = \gamma_1 F_1(g, \mathfrak{E}) + \gamma_2 F_2(v, \mathfrak{E}), \quad (4.1)$$

$$v, g \geq 0; v = \text{assign}(g),$$

где  $g$  – восстанавливаемая матрица корреспонденций;  $g$  – старая матрица корреспонденций;  $v$  – вектор значений наблюдаемых потоков;  $F_1, F_2$  – некоторые меры расстояния, например, евклидова норма, или минимум информации (максимум энтропии);  $\text{assign}(g)$  – распределение потоков  $v$  по сети, разделяющее матрицу корреспонденций  $g_{ij}$  на потоки, следующие разными маршрутами.

В случаях, когда отсутствует “старая” матрица, используют методы, основанные на моделировании распределения потоков. В этих методах наиболее сложным является установление маршрутов, по которым следуют корреспонденции  $g_{ij}$  из зоны  $i$  в зону  $j$ . Такая информация представляется в виде матрицы  $P$ , элементы  $p_{ij}^a$  которой являются долей корреспонденций между зоной  $i$  и зоной  $j$ , использующих дугу  $a$ :

$$0 \leq p_{ij}^a \leq 1. \quad (4.2)$$

Соответственно интенсивность движения  $v_a$  на данной дуге  $a$

$$v_a = \sum_{ij} p_{ij}^a g_{ij}, \quad a \in A. \quad (4.3)$$

В задачах восстановления матриц корреспонденций выделяют два типа распределения потоков:

1. *Пропорциональное распределение.* Потоки  $v_a$  и доли  $p_{ij}^a$  считаются не зависимыми друг от друга. Значения  $p_{ij}^a$  могут задаваться до начала моделирования распределения. Допускается использование принципа “все или ничего” (all-or-nothing), при этом пред-

полагается, что корреспонденции используют пути минимальной стоимости.

2. *Равновесное распределение.* Моделирование выполняется в соответствии с принципами Водропа, т.е. значения  $p_{ij}^a$  зависят от величин потоков на всех дугах  $p_{ij}^a = p_{ij}^a(v)$  и не могут определяться заранее.

С. Fisk [111] предложил использовать сочетание процедуры моделирования равновесного распределения, аналогичной той, которая применяется в известной программе SATURN, и максимизацию энтропии:

$$\min \sum_{ij} g_{ij} (\ln(g_{ij}) - 1) \quad (4.4)$$

при ограничениях

$$\begin{aligned} C(h) \cdot (f - h) &\geq 0, & \forall f; \\ \sum_k h_{ijk} &= g_{ij}, & \forall i, j; \\ \epsilon_a &= \sum_k p_{ijk}^a h_{ijk}, & \forall a \in A; \\ g_{ij}, h_{ijk} &\geq 0. \end{aligned}$$

Здесь  $C(h)$  – цена передвижения по данному пути при интенсивности движения  $h$ ;  $f$  – вектор потоков, полученных в результате моделирования;  $p_{ijk}^a$  – доля корреспонденций из  $i$  в  $j$ , использующих путь  $k$ . Если ребро  $a$  не принадлежит пути  $k$ , то  $p_{ijk}^a = 0$ .

Несколько иной подход был предложен К. Jornsten, S. Ngueng [75], а позднее L. LeBlanc, К. Farhangian [75]. Задача восстановления матрицы использует равновесное распределение с ограничениями, аналогичными ограничениям задачи (4.4), и критерий наименьших квадратов (т.е. евклидову норму):

$$\min \sum_{ij} (g_{ij} - \epsilon_{ij})^2. \quad (4.5)$$

Восстановление матрицы корреспонденций с использованием обобщенного метода наименьших квадратов (GLS) предполагает наличие матрицы, полученной в результате оценки  $\mathfrak{E}$  и наблюдений потоков  $v$ :

$$\mathfrak{E} = g + \eta; \quad (4.6)$$

$$v = v(g) + e, \quad (4.7)$$

где  $\eta$  – ошибка оценки матрицы  $\mathfrak{E}$ ;  $e$  – ошибка наблюдений значений потоков.

Задача сводится к нахождению матрицы, максимально близкой к “правильной” матрице  $g$ , с использованием имеющихся  $\mathfrak{E}$  и  $v$ . В самом общем виде GLS-оценка

$$\min \left[ \frac{1}{2} (\mathfrak{E} - g)' Z^{-1} (\mathfrak{E} - g) + \frac{1}{2} (v - v(g))' W^{-1} (v - v(g)) \right] \quad (4.8)$$

при ограничениях на переменные  $g_{ij} \geq 0$ ,

где  $Z$  – вариационно-ковариационная матрица, заданная или оцененная для матрицы  $g$ ;  $W$  – матрица дисперсий измеренных потоков  $v$ .

Для выполнения задачи (4.8) необходимо предварительно задать или оценить  $Z$  и  $W$ . Матрицу  $Z$  определяют разными способами. Например, E. Cascetta [92] для определения ковариаций использовал данные выборочных исследований корреспонденций. Матрицу  $W$  часто задают диагональной и даже принимают значения элементов диагонали, равными 1 [92].

Методы восстановления матриц корреспонденций стали объектом внимания специалистов не только в области транспорта, но и в области информационных сетей. Одним из самых новых исследований по данной тематике является опубликованная в июле 2003 г. работа Y. Zhang, M. Roughan, N. Duffield, A. Greenberg [227]. В ней рассматривается возможность оценки корреспонденций по значениям нагрузок элементов сети, передающей информацию. На первом шаге предлагаемого в работе алгоритма [227] осуществляется

предварительная оценка матрицы корреспонденций с использованием гравитационной модели. Применение методов томографии рассматривается авторами как определение корреспонденций по данным нагрузок на элементах сети. На следующем шаге полученная квадратная матрица корреспонденций преобразуется в вектор-столбец, при этом исключаются нулевые элементы. Таким образом, в результате использования гравитационной модели и преобразования получают вектор-столбец  $\mathbf{t}_g = (t_{g1}, t_{g2}, \dots, t_{gM})^T$ , где  $t_{gr}$  – поток корреспондирующей пары вершин  $r$ ,  $M$  – количество ненулевых элементов матрицы корреспонденций. Данные о потоках на ребрах графа представлены вектором-столбцом  $\mathbf{x} = (x_1, x_2, \dots, x_L)^T$ . Здесь  $x_i$  – замеры нагрузок на элементах сети; в других терминах  $x_i$  – потоки на ребрах графа сети (или дугах, если граф ориентированный). Принадлежность корреспонденций ребрам (дугам) описывается матрицей

$$A_{ir} = \begin{cases} F_{ir}, & \text{если корреспонденция } r \text{ проходит по ребру } i, \\ 0 & \text{в другом случае.} \end{cases}$$

здесь  $F_{ir}$  – доля корреспонденции  $r$ , проходящей по ребру  $i$ .

Определяемый вектор корреспонденций  $\mathbf{t}$  образует систему линейных уравнений

$$\mathbf{x} = \mathbf{A} \mathbf{t}, \quad (4.9)$$

которые составляют линейные ограничения задачи наименьших квадратов.

Оценка вектора корреспонденций  $\mathbf{t}$  сводится к хорошо известной задаче линейного регрессионного анализа при наличии линейных ограничений [227]. Поскольку в обычной форме регрессионный анализ не содержит ограничений на знак оцениваемых переменных, авторы предусмотрели замену отрицательных значений

значениями 0. Следует отметить, что прием замены отрицательных значений уже осуществлен в функции `nls` для решения задачи наименьших квадратов при условии только положительных значений, входящей в состав библиотеки Optimization Toolbox пакета MATLAB, который использован авторами.

Необходимо подчеркнуть, что в публикациях по информационным сетям [164,227] указывается, что в рассматриваемых задачах количество оцениваемых корреспонденций  $N$  (корреспонденции  $x_i$ ,  $i = 1, 2, \dots, N$ ) значительно превосходит количество дуг  $K$  (или ребер, если граф неориентированный), на которых известны нагрузки, т.е.  $N \gg K$ .

В нашей стране исследования в области восстановления матриц корреспонденций (А.П. Артынов, Г.А. Варепуло, М.Л. Дыдаюк, В.Ш. Крупник, В.С. Огай, В.В. Скалецкий) проводились в основном для маршрутов общественного транспорта [3]. Исключением являются исследования Н.В. Булычевой и С.Л. Сена [39], которые выполнялись в составе комплексной транспортной схемы Ленинграда (ЛенНИИПИгенплана 1987–1989 гг.) с целью установления матрицы грузовых корреспонденций. Математические модели оценки маршрутных матриц корреспонденций классифицируют как основанные на вероятностных методах, основанные на аналитических методах и энтропийные [3].

Кроме теоретических аспектов при выборе модели оценки матриц корреспонденций необходимо учитывать качество исходной информации и возможные способы ее получения.

#### **4.2. Методы обследований интенсивности движения при выполнении оценки матриц корреспонденций**

Выполненный выше анализ моделей восстановления матриц корреспонденций позволяет отметить их общую характеристику – наличие в той или иной форме информации о существующем рас-

пределении потоков или о “старой” матрице. В [113,115] подчеркивается, что любая предварительная информация о распределении транспортных потоков (ТП) существенно повышает точность оценки восстановления матрицы корреспонденций. В целом такую предварительную информацию можно классифицировать следующим образом:

данные частичного обследования маршрутов следования, которые распространяются на генеральную совокупность;

“старая” матрица корреспонденций, корректируемая на основе данных интенсивности движения [формулы (4.1),(4.6)–(4.8)];

данные моделирования распределения потоков по УДС, корректируемые с использованием существующих значений интенсивности движения [формулы (4.2)–(4.5),(4.9)].

В зарубежной практике большое количество исследований [83,213] базируется на использовании периферийного оборудования (детекторов), позволяющего собирать подробные данные об интенсивности движения, в том числе в режиме реального времени. Такие данные по необходимости дополняются обследованиями маршрутов следования, анкетированием водителей и т.д. В нашей стране преобладают ручные методы получения исходной информации (подсчет интенсивности движения наблюдателями, анкетирование). Поэтому остановимся лишь на методах обследований, которые применялись, применяются или рекомендуются для применения в нашей стране.

Одним из немногих исследований по восстановлению корреспонденций по значениям потоков транспортных средств (ТС) является работа Н.В. Булычевой и С.Л. Сена [39]. Авторы выполняли построение матрицы потоков грузовых транспортных средств между расчетными транспортными районами. В общих чертах это обследование проводилось следующим образом:

на основании схемы организации движения и данных интенсивности движения назначались пункты учета грузового движения;



в пунктах учета проводился опрос водителей грузовых автомобилей в течение 2–4 ч интенсивного грузового движения, а в ряде пунктов – в течение суток без перерыва;

при опросе фиксировались адреса пунктов отправления и прибытия, марка автомобиля, вид и масса груза;

дополнительно для ряда пунктов учета обследования проводились непрерывно в течение суток.

Результаты выборочного обследования (выборка не менее 5% интенсивности движения грузовых автомобилей) распространялись на всю генеральную совокупность. Далее на основе данных традиционных обследований транспортных потоков (картограммы транспортных потоков, распределения интенсивности по часам суток) и данных опроса получали следующую информацию:

величину суммарной интенсивности движения грузовых автомобилей и распределение по часам суток;

количество грузовых автомобилей, проходящих через определенный учетный пункт за сутки, и распределение их по часам суток в этом учетном пункте;

распределение по направлениям автомобилей, проходящих через определенный учетный пункт за сутки, и распределение их по часам суток в этом учетном пункте по каждому из направлений.

На основе всех этих данных строилась матрица суточных матриц потоков между расчетными районами и оценивались характеристики корреспонденций. Таким образом, предложенная методика предполагает два типа обследований:

1) детальное обследование интенсивности движения для построения картограммы интенсивности движения;

2) анкетирование водителей в определенных учетных пунктах, результаты которого (частичная выборка корреспонденций) распространяются на всю генеральную совокупность корреспонденций.

Несколько позднее, в 1990 г., в нашей стране для проектов организации движения была предложена методика определения матрицы корреспонденций для транспортных потоков [40], которая имеет много общего с методикой [37]:

территория города в пределах административных границ разбивается на учетно-транспортные районы (УТР);

посты учета движения располагают на границах УТР, при этом желательно в качестве границ принимать естественные и искусственные рубежи, затрудняющие движение (реки, железные дороги, незастроенные территории и т.д.);

распределение транспортных потоков по УДС наблюдается в течение времени, равного трехкратной продолжительности движения легкового автомобиля по большому диаметру городской территории, но не менее 2 ч; рекомендуемое время обследований – межпиковые периоды (ориентировочно 13.00—16.00) будних дней;

группа учетчиков регистрирует тип транспортного средства и четыре цифры номерного знака; через каждую минуту столбцы учетных данных подчеркиваются горизонтальной линией, 30-минутные интервалы отмечаются соответствующим временем (13.30, 14.00 и т.д.).

По результатам обследования должны быть сделаны следующие графические материалы [40]:

матрицы основных корреспонденций (отдельно для легковых и грузовых автомобилей);

схемы основных маршрутов транспортных потоков корреспонденций (отдельно для легковых и грузовых автомобилей);

картограмма интенсивности движения по УДС.

Рассмотренные выше методы сбора данных для общей матрицы корреспонденций трудновыполнимы в пиковые периоды для наиболее загруженных участков УДС по следующим причинам:

обследование должно охватывать все виды транспортных средств, исключая маршрутный пассажирский транспорт;

требуется получение данных о пунктах отправления и прибытия, а также детальных сведений о маршруте следования, что увеличивает продолжительность заполнения анкеты;

требуется не создавать помехи движению.

Избежать дополнительных помех движению и задержек транспортных средств, которые должны учитываться в стоимости обследований, можно, если выполнять анкетирование в более благоприятных условиях:

во внепиковый период с распространением результатов на пиковый период;

в пиковый период в учетных пунктах, размещающихся на участках УДС, с низкой интенсивностью движения;

в пиковый период в учетных пунктах на участках УДС с высокой интенсивностью движения, но низким уровнем загрузки.

Характерным примером такого обследования является анкетирование водителей, выполнявшееся в соответствии с распоряжением мэра от 18 октября 1999 г. №1168-РМ “О проведении комплексного обследования условий движения на улично-дорожной сети Москвы”. Опросы проводились с 10.00 до 14.00; при выборе расположения постов руководствовались следующими требованиями:

пост должен располагаться на некотором расстоянии от пересечений так, чтобы ТС еще не могли разогнаться после перекрестка либо уже начали тормозить перед перекрестком;

должно обеспечиваться место для остановленных 3–4 автомобилей (в том числе и грузовых), причем остановленные автомобили не должны затруднять движение ТП и пешеходам; движение учетчиков к опрашиваемому водителю автомобиля должно быть максимально безопасным;

мест опроса для конкретного поста должно быть не менее двух для обоих направлений; на ряде постов должно быть четыре

места опроса (по два для каждой из пересекающихся улиц перекрестка).

### **4.3. Особенности оценки матрицы корреспонденций с использованием только значений интенсивности движения**

При выполнении обследований в соответствии с рассмотренными в предыдущем параграфе требованиями велика вероятность того, что не будет получена репрезентативная информация о маршрутах следования в пиковые периоды на наиболее загруженных участках УДС. Такая информация представляет наибольший интерес для оценки пропускной способности УДС и разработки проектных решений. Поэтому авторы настоящей монографии задались целью разработать метод, пригодный для получения оценок корреспонденций для условий, указанных ниже:

рассматривается детализированная улично-дорожная сеть, включая отдельные потоки по разным направлениям на перекрестках;

исходными данными являются замеры интенсивности движения в отдельных пунктах улично-дорожной сети (т.е. на перекрестках, на перегонах), выполняемые в определенное время в будние дни (например, вечерний час пик 17.00—19.00);

данные об интенсивности движения содержат ошибки, вызванные проведением замеров в разные дни, и ошибки самих подсчетов интенсивности движения.

Можно указать ряд отличительных особенностей сформулированной задачи восстановления матрицы корреспонденций, которые связаны с видом и качеством исходных данных:

детальное рассмотрение улично-дорожной сети в виде ориентированного графа, для каждой дуги которого обследованиями устанавливается значение интенсивности движения;

используются данные обследований только интенсивности движения, которые традиционно применяются в практике ОДД для построения картограмм интенсивности движения (подсчет количества транспортных средств учетных видов с последующим определением приведенной интенсивности движения);

матрицы корреспонденций между выделенными пунктами улично-дорожной сети определяются методами, устойчивыми к выбросам, т.е. с помощью так называемого робастного регрессионного анализа.

Отсутствие предварительной информации о маршрутах движения требует представить исходные данные для оценки матрицы корреспонденций так, чтобы каждая пара взаимно корреспондирующих пунктов связывалась одним маршрутом. Это можно осуществить разделением матрицы корреспонденций участка УДС на группу матриц корреспонденций, каждая из которых будет представлять “отдельный маршрут”. Аналогами являются модель оценки межостановочных корреспонденций отдельного маршрута, хорошо изученная в нашей стране [3], и методика определения существующей матрицы для случая отдельной магистральной дороги, которая приведена в НСМ 2000 [118] и имеет много общего с моделью изолированного маршрута пассажирского транспорта.

В модели изолированного маршрута [3] рассматривается одно из двух направлений движения и используются данные о количестве входящих и выходящих пассажиров на остановочных пунктах (ОП); межостановочная матрица корреспонденций  $X$  представляется треугольной матрицей с нулевой диагональю

$$\begin{matrix}
 0 & x_{12} & x_{13} & \dots & x_{1,n} \\
 & 0 & x_{23} & \dots & x_{2,n} \\
 & & & \dots & \dots \\
 & & & & 0 & x_{n-1,n} \\
 & & & & & 0
 \end{matrix} , \tag{4.10}$$

где  $x_{ij}$  – количество пассажиров, проехавших между  $i$  и  $j$  ОП,  $x_{ij} \geq 0$ ,  $i \leq j$ .

Сумма элементов  $i$  строки матрицы (4.10) соответствует количеству пассажиров, вошедших на  $i$  ОП, а сумма элементов  $j$  столбца матрицы – количеству пассажиров, вышедших на  $j$  ОП:

$$a_i = \sum_{j=i}^n x_{ij}; \quad b_j = \sum_{i=1}^j x_{ij}; \quad i, j = 1, \dots, n. \quad (4.11)$$

При этом  $a_i$  и  $b_j$  удовлетворяют условию

$$\sum_{i=1}^n a_i = \sum_{j=1}^n b_j. \quad (4.12)$$

Для матрицы  $X$ , максимизирующей взвешенную энтропию и отвечающей условию

$$H(x) = \sum_{ij} x_{ij} \ln \frac{a_i}{x_{ij}} \rightarrow \max,$$

было получено решение в следующем виде [3]:

$$x_{12} = b_2;$$

$$x_{ij} = \begin{cases} \frac{\left( a_i - \sum_{m=i+1}^{j-1} x_{im} \right) b_j}{C_{j-1}}, & \text{если } C_{i-1} > 0; \\ 0, & \text{если } C_{i-1} = 0; \end{cases} \quad (4.13)$$

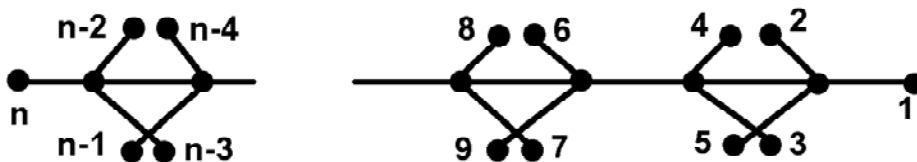
$$x_{ij} = \begin{cases} \frac{a_{j-1} b_j}{C_{j-1}}, & \text{если } C_{i-1} > 0; \\ 0, & \text{если } C_{i-1} = 0; \end{cases}$$

где  $C_{j-1} = \sum_{k=1}^{j-2} \left( a_k - \sum_{r=k+1}^{j-1} x_{kr} \right) + a_{j-1}$ .

Определенные положения рассмотренной модели оценки межостановочной матрицы корреспонденций изолированного маршрута (4.10)–(4.13) могут быть использованы в модели восстановления матрицы корреспонденций транспортных потоков.

Как уже указывалось выше, УДС разделяется на отдельные “маршруты”, или (как в НСМ 2000 [118]) “направления движения”. Отдельно рассматриваемое направление движения (маршрут) представляется в виде графа (рис. 4.1, 4.3). Соответственно начальными и конечными пунктами корреспонденций являются поворотные потоки на перекрестках, т.е. входящие потоки на улицу или выходящие с нее (см. рис. 4.1). Однако при этом оценка корреспонденций (4.13) уже неприменима.

В данной задаче восстановления матрицы корреспонденций вместо центроидов расчетных транспортных районов используются “нагрузочные” вершины сети (рис. 4.2), которые образуются на границах рассматриваемого участка УДС и в которых начинаются и оканчиваются корреспонденции.



**Рис. 4.1.** Представление одного из направлений движения на улице или дороге в виде графа для определения матрицы корреспонденций:  $1, 2, \dots, n$  - вершины графа, в которых возникают и заканчиваются корреспонденции

Таким образом, рассматриваемое формализованное описание сети сводится к разделению изучаемого участка УДС на графы (см. рис 4.1, 4.3), в которых каждая корреспондирующая пара вершин будет иметь только один соединяющий маршрут, тогда для определения матриц каждого из графов можно использовать распределение поездок по принципу “все или ничего” (all or nothing).

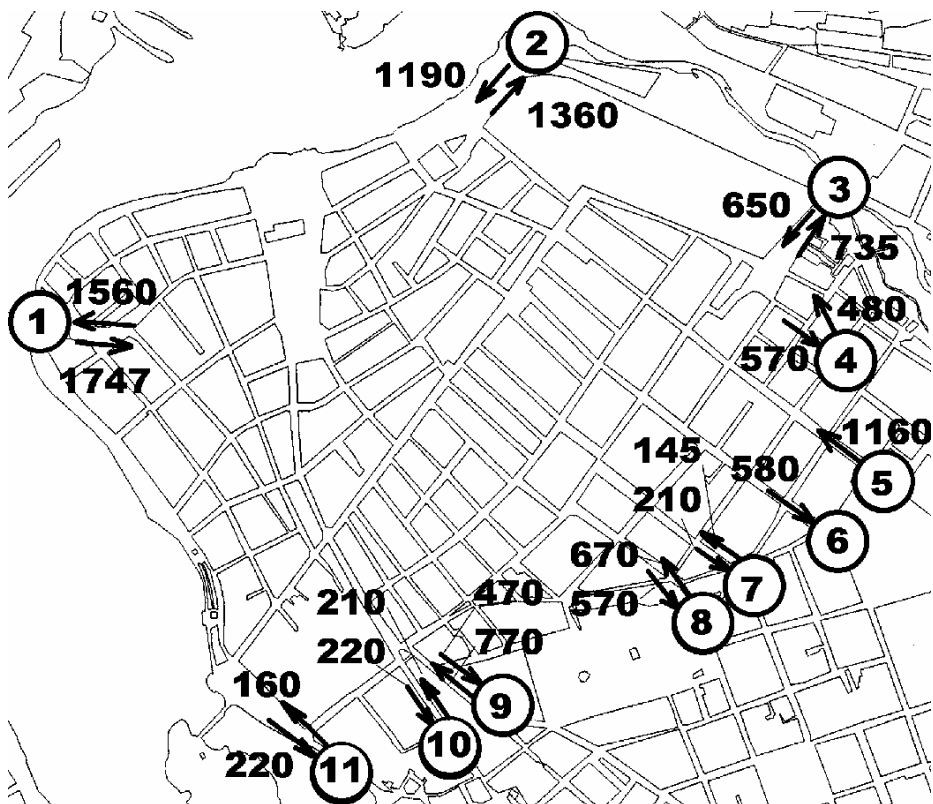
Прием определения матрицы корреспонденций, который можно адаптировать для случая УДС (рис. 4.1 – 4.4), приведен в Руководстве НСМ 2000 [118], в соответствии с которым направле-

ния движения двухсторонней дороги рассматриваются отдельно; соответственно оцениваются две отдельные треугольные матрицы корреспонденций. Исходными данными являются значения интенсивности движения на въездах на магистраль и выездах с нее. Кроме того, для протяженных участков производятся замеры интенсивности движения всех поворотных потоков и на перегонах, эти дополнительные данные используются для аппроксимации.

Значения корреспонденций  $T_{ij}$  из точки  $i$  в точку  $j$  (см. рис. 4.1) определяются как

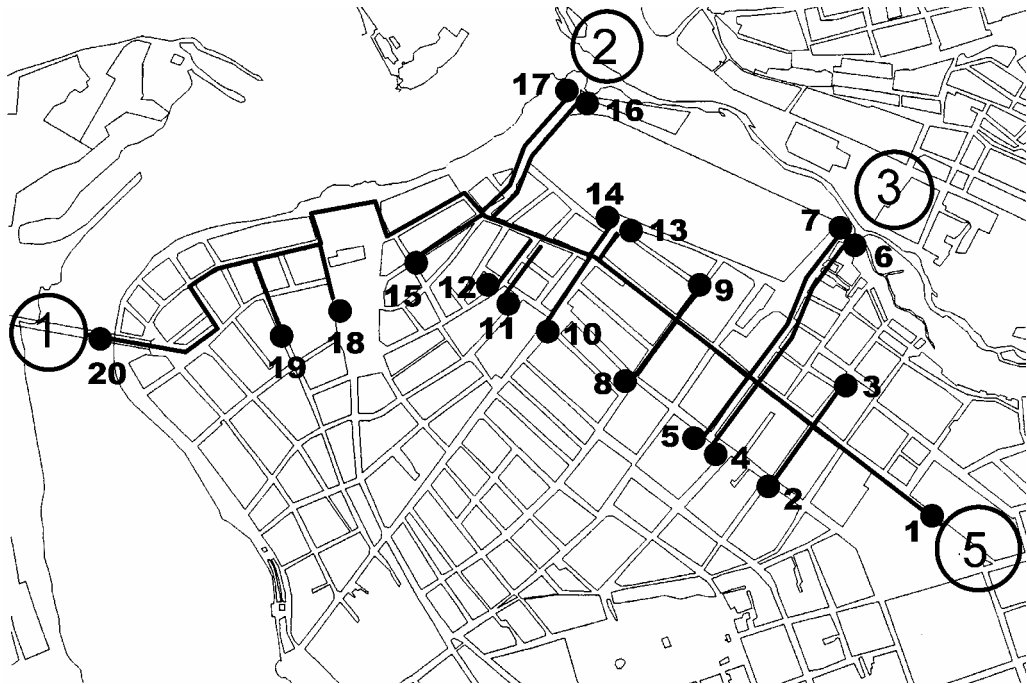
$$T_{ij} = \left[ \frac{T_i \cdot T_j}{\sum_j T_j} \right], \quad (4.14)$$

где  $T_i$  – сумма корреспонденций, выходящих из точки  $i$ ;  $T_j$  – сумма корреспонденций, входящих в точку  $j$ .

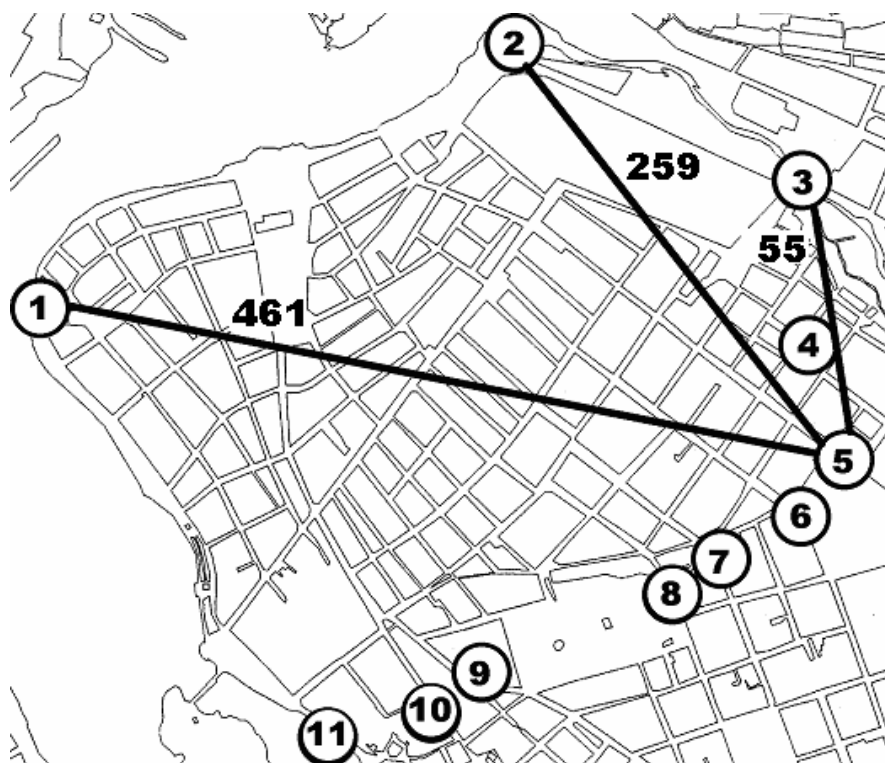


**Рис. 4.2.** “Загрузочные” вершины матрицы корреспонденций транзитных потоков центра Иркутска и интенсивности движения в вечерний час пик, авт./ч





*Рис. 4.3.* Граф для восстановления матрицы корреспонденций из вершины 1 в вершины 2,3, ..., 20, используемой для определения транзитных потоков из вершины 5 в вершины 1,2,3 (см. рис. 4.2)



*Рис. 4.4.* Транзитные потоки из вершины 5 (см. рис. 4.2) в вершины 1,2,3, полученные при оценке матрицы корреспонденций для графа УДС, представленного на рис. 4.3

Поскольку при определении матрицы корреспонденций  $T$  ее некоторая часть становится известной из замеров интенсивности, значения остальных элементов матрицы рассчитываются по несколько измененной формуле:

$$T_{ij} = \frac{(T_i - k_i)(T_j - k_j)}{\sum_j (T_j - k_j)}, \quad (4.15)$$

где  $T_i$  – сумма корреспонденций, выходящих из точки  $i$ ;  $T_j$  – сумма корреспонденций, входящих в точку  $j$ ;  $k_i$  – сумма известных корреспонденций, выходящих из точки  $i$ ;  $k_j$  – сумма известных корреспонденций, входящих в точку  $j$ .

Для участков большей протяженности предложен алгоритм оценки матрицы с применением регрессии. На первом шаге с использованием гравитационной модели (4.14) или (4.15) оценивается матрица корреспонденций  $T$ . По результатам оценки определяются значения потоков на участках дороги (на дугах графа), на которых производились замеры интенсивности движения.

Рассчитывается значение целевой функции – суммы квадратов градиентов (отклонений) – как

$$z = \sum_a g(a)^2, \quad (4.16)$$

где  $v'(a)$  – значение интенсивности движения по результатам оценки матрицы;  $v(a)$  – значение интенсивности движения по результатам обследований;  $g(a)$  – ошибка,  $g(a) = v'(a) - v(a)$ .

На основе значения целевой функции (4.16) определяется точность оценки матрицы  $T$ . Например, если задана среднеквадратическая ошибка оценки 100 авт./ч, то для дороги, представленной 10 участками, значение менее 100 000 является признаком достаточной точности оценки матрицы  $T$ . Если целевая функция  $z$  – сумма квадратов остатков регрессии (4.16) – не достигла порогового значения, ее минимизация рассматривается как задача квадратичного

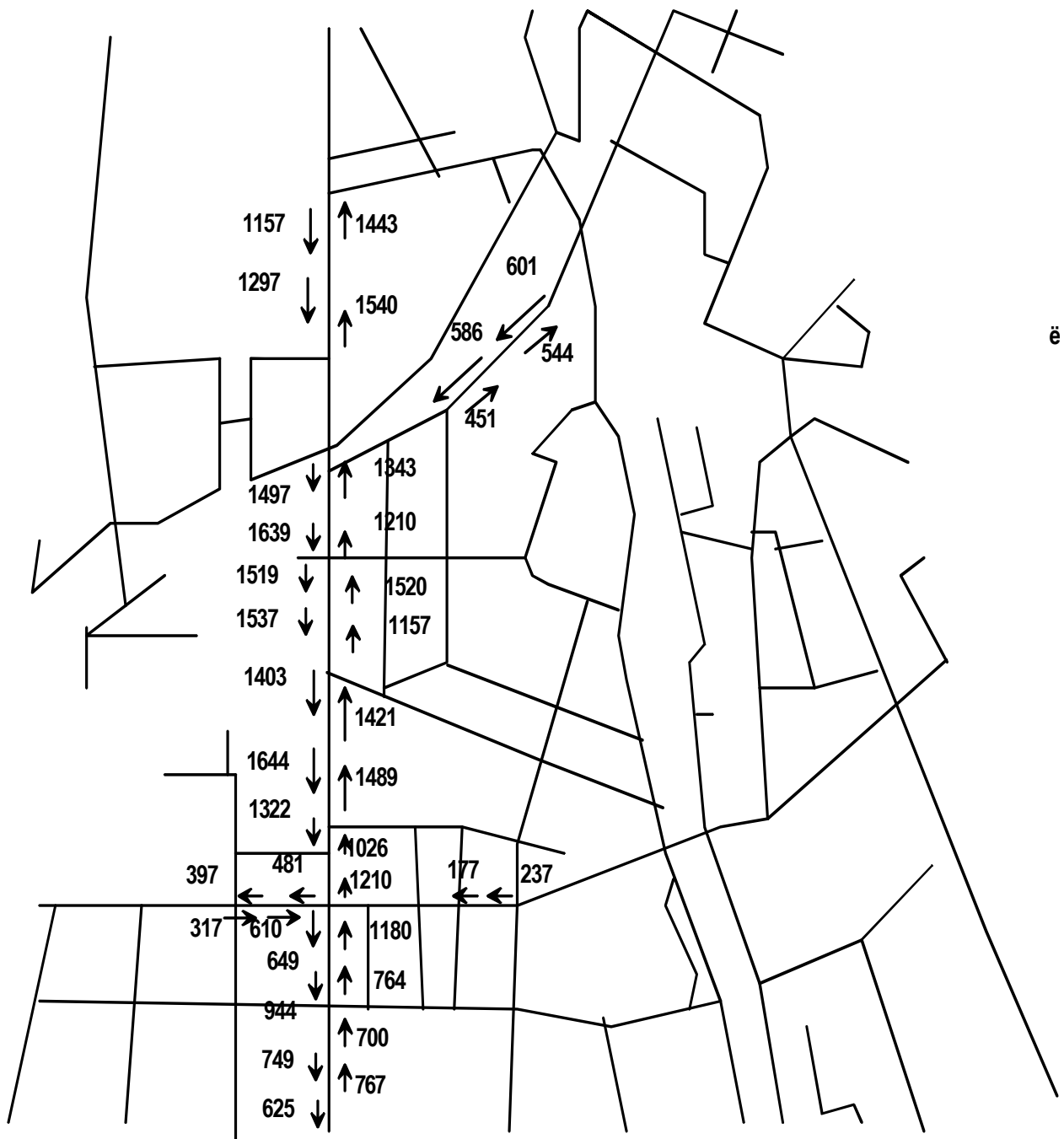
программирования, решаемая градиентным методом, в результате чего получается новая матрица  $T'$ .

Рассмотренный выше метод отличают простота и возможность выполнения оценки только данных интенсивности движения. Вместе с тем в методике, предлагаемой в НСМ 2000 [118], вообще не рассматривается качество исходных данных, поскольку она предназначена для оценки магистральных дорог (corridor analysis), когда одновременные данные об интенсивности движения на въездах на магистраль и выездах с нее поступают с периферийного оборудования.

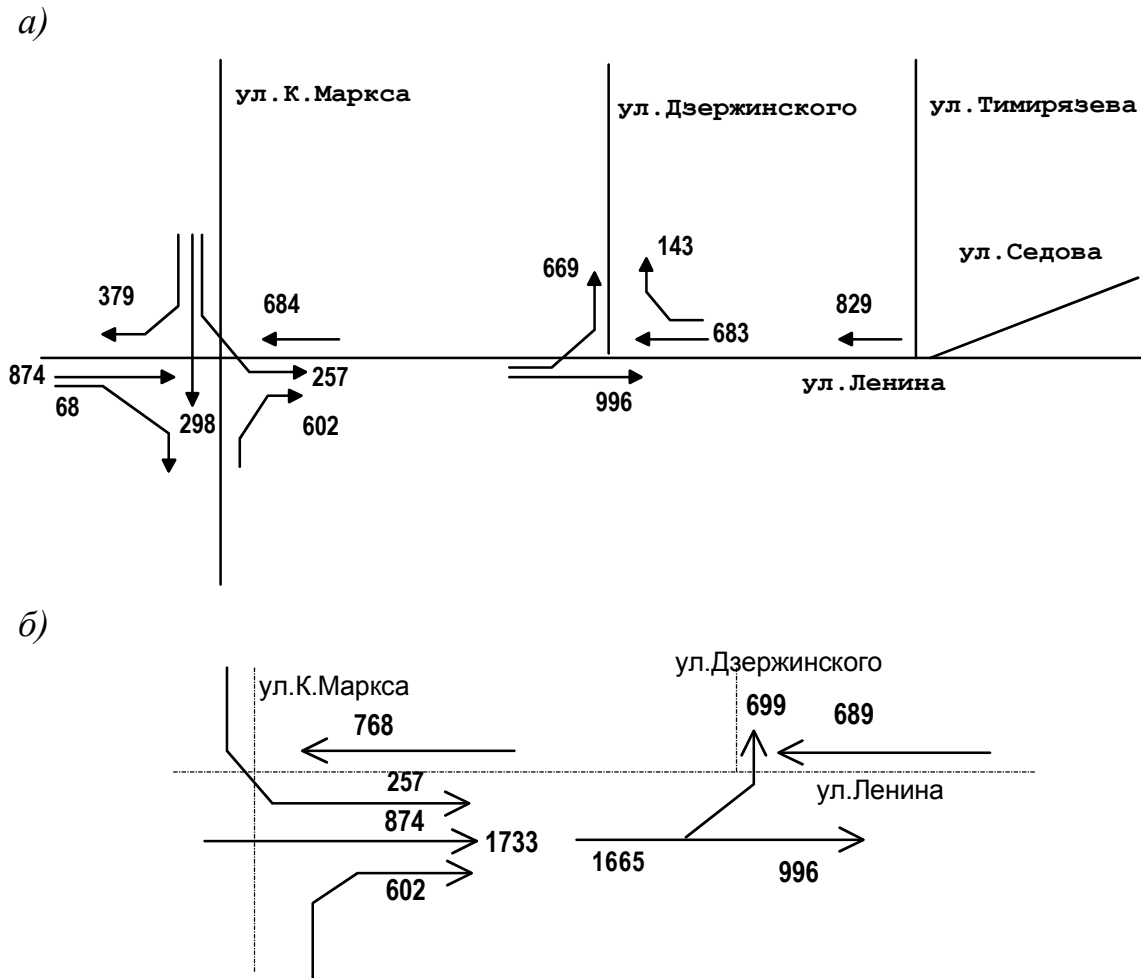
Рассматриваемый в данной работе случай принципиально иной, так как данные интенсивности движения получают в результате обследований УДС. Поскольку исходными данными являются замеры интенсивности движения на отдельных перекрестках УДС, выполняемые в определенное время в будние дни (например, вечерний час пик), то при сведении этих данных в одну выборку возникают расхождения данных на перегонах УДС (рис. 4.5, 4.6).

Расчитанные по данным замеров на смежных перекрестках величины входящего и выходящего потоков имеют разные значения (см. рис. 4.6). Эти ошибки вызваны проведением замеров в разные дни и ошибками самих подсчетов интенсивности движения. Оценка матриц корреспонденций, использующая указанные данные, должна иметь такое важное свойство, как робастность.

Следует отметить, что рассматриваемый вид ошибок (см. рис. 4.6) неизменно возникает в процессе построения картограмм интенсивности движения. В результате сведения в картограмму замеров интенсивности движения, выполненных на смежных пересечениях, получают последовательности перегонов, каждый из которых имеет пару значений интенсивности движения – “входящий” и “выходящий потоки” (см. рис. 4.5). При сведении результатов измерений интенсивности в картограмму возникает необходимость устранения невязки таких данных.



**Рис. 4.5.** Входящие и выходящие потоки на перегонах Московского проспекта в Санкт-Петербурге по результатам сведения данных обследований за 1986–1987 гг.



**Рис. 4.6.** Ошибки, возникающие при сведении данных обследований интенсивности движения на смежных перекрестках: *а* – результаты замеров интенсивности движения (физ.ед./ч) на смежных пересечениях; *б* – расхождения замеров интенсивности на смежных перекрестках, ограничивающих перегон, и получаемые при этом пары значений интенсивности движения – “входящий” и “выходящий” потоки (соответственно на перегоне ул. Ленина между ул. К.Маркса и ул. Дзержинского 1733 – 1665 и 689 – 768)

Хотя построение картограмм является традиционным видом графического представления данных обследований интенсивности движения, процедура анализа возникающих при этом ошибок не рассматривалась в специальной литературе. Поскольку для построения картограмм не предлагались формализованное описание и необходимые для этого статистические оценки, такая работа вы-

полняется проектировщиками субъективно в соответствии с их профессиональными вкусами и опытом.

Таким образом, для оценки матрицы корреспонденций при использовании данных, содержащих ошибки, необходимо иметь:

метод и критерии выявления недопустимых значений ошибок исходных данных;

статистическую процедуру оценки корреспонденций на основе засоренных данных.

На первый взгляд можно утверждать, что рассматриваемой задаче восстановления матрицы корреспонденций наиболее отвечает регрессионное оценивание – обобщенный метод наименьших квадратов (Generalized Least Squares). Вопрос в том, насколько в действительности метод наименьших квадратов (МНК) соответствует такой задаче – оцениванию в случае грубых ошибок исходных данных.

#### **4.4. Модели робастного оценивания матриц корреспонденций**

Регрессионный анализ, использующий МНК, основан на предположении, что независимые и зависимая переменные, остатки регрессии  $e_i = y_i - \hat{y}_i$  (отклонений наблюдаемых значений  $y_i$  от аппроксимирующей регрессии  $\hat{y}_i$ ) подчинятся нормальному распределению. Как подчеркивается в [18,54], принятие такой гипотезы предполагает, что основная масса отклонений от регрессии сосредоточена на конечном интервале  $-3\sigma, 3\sigma$ , где  $\sigma$  – стандартное отклонение нормального распределения. Соответственно вероятность больших отклонений считается очень малой. В результате оценка МНК очень чувствительна к так называемым выбросам (т.е. грубым ошибкам исходных данных) и реагирует на них.

Сформулированная задача регрессионного оценивания корреспонденций требует применения статистических процедур, устойчи-

вых к выбросам. Робастные (от английского *robust*), или устойчивые, регрессионные методы достаточно подробно рассмотрены в работах российских авторов [18,61,64]. Главное общее свойство таких оценок заключается в том, что они малочувствительны к выбросам, но менее эффективны в случаях, когда вероятность выбросов мала [16].

Несколько видов робастных оценок было изучено П. Хьюбером [16]. Он предложил для оценки параметров регрессии вместо суммы квадратов отклонений регрессии  $e_i$

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \theta)^2 \quad (4.17)$$

использовать минимизацию суммы

$$\sum_{i=1}^n \rho(y_i - \theta), \quad (4.18)$$

где  $\rho$  – некоторая выпуклая функция.

Значения  $\theta$ , минимизирующие функцию (4.18), в специальной литературе называют М-оценкой. Е.З. Демиденко [18] отмечает, что М-оценку можно рассматривать как оценку метода максимального правдоподобия. Частным случаем оценок Хьюбера является функция  $\rho(x) = |x|^v$ ,  $1 \leq v < 2$ , что предполагает минимизацию

$$\sum_{i=1}^n |e_i|^v = \sum_{i=1}^n |y_i - \theta|^v. \quad (4.19)$$

Оценки (4.19) предложено называть  $L_v$ -оценками [18]. Устойчивость такой аппроксимации по сравнению с МНК (4.17) вызвана тем, что большие отклонения  $(y_i - \theta)$  меньше влияют на целевую функцию. Поэтому показатель степени  $v$  интерпретируется некоторыми авторами как фильтр выбросов [18]. При  $v=1$  оценка (4.19) сводится к минимизации суммы абсолютных значений отклонений (модулей), то есть

$$\sum_{i=1}^n |e_i| = \sum_{i=1}^n |y_i - \theta|. \quad (4.20)$$





что соответствует взвешенному МНК со значениями весовых коэффициентов  $w_i$ , которые могут определяться на основе предыдущих итераций.

С учетом вводимых обозначений задачи оценки корреспонденций нормальные уравнения ВВП (4.21) получают в матричной форме следующий вид:

$$A'WAx = A'WY, \quad (4.23)$$

где  $x$  – оцениваемый вектор значений корреспонденций размерности  $m \times 1$ ;  $A$  – матрица независимых переменных, которыми являются значения долей потоков на ребрах графа улично-дорожной сети; значения  $a_{ij}$  изменяются от 0 до 1; размерность матрицы  $n \times m$ ;  $y$  – вектор значений интенсивности движения на дугах улично-дорожной сети размерности  $n \times 1$ ;  $W$  – вектор весов, преобразованный в диагональную матрицу размерности  $n \times n$ .

С учетом вектора остатков  $e$  размерности  $n \times 1$  имеем следующие векторы и матрицы:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix},$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{bmatrix}, \quad W = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_n \end{bmatrix}.$$

Оценки корреспонденций методом ВВП в матричной форме выразятся как

$$x = (A'WA)^{-1}A'WY, \quad (4.24)$$

где  $W = \text{diag}[w_1, w_2, \dots, w_n]$  заменяет матрицу  $V^{-1}$  обобщенного метода наименьших квадратов, имеющего в зарубежной литературе название оценки Эйткена

$$x = (A' V^{-1} A)^{-1} A' V^{-1} Y. \quad (4.25)$$

Использование ВВП [формулы (4.22) – (4.24)] для оценки корреспонденций  $x$  соответствует минимизации суммы

$$\sum_{i=1}^n w_i e_i^2 = \sum_{i=1}^n w_i \left( y_i - \sum_{j=1}^m x_j a_{ij} \right)^2, \quad (4.26)$$

где  $a_{ij}$  – значения доли потока  $j$  на дуге  $i$  графа УДС;  $y_i$  – значение интенсивности движения по дуге графа УДС ;  $e_i$  – остатки регрессии.

Как уже указывалось выше, частным случаем  $L_\nu$ -оценки ( $\nu = 1$ ) является метод наименьших модулей (МНМ), сводящийся к минимизации суммы абсолютных модулей отклонений [см. формулу (4.20)]:

$$\sum_{i=1}^n |e_i| = \sum_{i=1}^n \left| y_i - \sum_{j=1}^m x_j a_{ij} \right|, \quad (4.27)$$

которая при использовании весовых коэффициентов имеет вид

$$\sum_{i=1}^n w_i |e_i| = \sum_{i=1}^n w_i \left| y_i - \sum_{j=1}^m x_j a_{ij} \right|. \quad (4.28)$$

Минимизация выражений (4.27) и (4.28) выполняется как задача линейного программирования. В отличие от регрессионного анализа, в котором оцениваемые переменные могут иметь как положительные, так и отрицательные значения, в линейном программировании на оцениваемые параметры налагаются ограничения по знаку. Для снятия такого ограничения предлагается переход к решению двойственной задачи линейного программирования [39]. В рассматриваемой задаче восстановления матрицы корреспонденций оцениваемые корреспонденции имеют только положительные значения, знакопеременными являются остатки регрессии  $e_i = y_i -$

$\sum x_j a_{ij}$ . Сумма абсолютных значений ошибок  $e_i$  будет целевой функцией задачи линейного программирования. Применение аппарата линейного программирования в частном случае заметно упрощается по сравнению с регрессионным анализом в общем случае. Применение МНМ потребует лишь введения дополнительных переменных, позволяющих включить остатки регрессии в состав целевой функции линейного программирования.

С учетом введения фиктивных переменных оценка матрицы корреспонденций будет сводиться к определению вектора-столбца параметров [44]

$$(x_1, x_2, \dots, x_m, g_1, h_1, g_2, h_2, \dots, g_n, h_n),$$

минимизирующего линейную функцию

$$\sum_{i=1}^n (g_i + h_i), \quad (4.29)$$

или с учетом весовых коэффициентов

$$\sum_{i=1}^n w_i (g_i + h_i), \quad (4.30)$$

где  $g_i = \max(0, e_i)$ ,  $h_i = \min(-e_i, 0)$ .

Линейные ограничения задач (4.29) и (4.30) формулируются в виде равенств

$$A \mathbf{x} + \mathbf{g} - \mathbf{h} = \mathbf{y}. \quad (4.31)$$

Ограничения на знак переменных и весовых коэффициентов следующие:

$$x \geq 0, g \geq 0, h \geq 0, w \geq 0. \quad (4.32)$$

Далее вектор-столбец оцениваемых параметров  $(x_1, x_2, \dots, x_m, g_1, h_1, g_2, h_2, \dots, g_n, h_n)$  размерностью  $(m+2n) \times 1$  удобнее рассматривать как  $(x_1, x_2, \dots, x_m, x_{m+1}, \dots, x_{m+2n})$ . В матричной форме задачи (4.29) и (4.30), кроме вектора-столбца оцениваемых параметров, представлены векторами

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}$$

и измененной добавлением фиктивных переменных матрицей  $A_2$  размерности  $n \times (m+2n)$

$$A_2 = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & 1 & -1 & 0 & 0 & \dots & \dots & 0 & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & 0 & 1 & -1 & \dots & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & 0 & 0 & \dots & \dots & 1 & -1 \end{bmatrix} \quad (4.33)$$

или

$$A_2 = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} & 1 & 0 & \dots & 0 & -1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & 1 & \dots & 0 & 0 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & 0 & 1 & 0 & 0 & \dots & -1 \end{bmatrix}. \quad (4.34)$$

При этом выражение (4.34) проще, чем выражение (4.33), применять для формирования вектора  $x$  (используется MATLAB):

```
e = y - a*x; % current vector of errors
for i=1:n
    if e(i)>=0
        x(m+i)= e(i);
        x(m+n+i)=0;
    else
        x(m+i)=0;
        x(m+n+i)= abs(e(i));
    end
end
```

и формирования самой матрицы  $A_2$ :

```
A2=A+repmat(eye(n),1,2)
```

В соответствии с процедурой вектор ошибок  $e$  определяется с учетом результатов оценки вектора корреспонденций  $x$  на преды-

дущей итерации. В зависимости от знака ошибки  $e_i$  задаются элементы вектора  $\mathbf{x}^2$ , начиная с индекса  $m+1$ . Таким образом, задачи (4.29) и (4.30) формулируются в следующем виде:

$$\min \sum_{i=m+1}^{m+2n} x_i, \quad (4.35)$$

$$\min \sum_{i=m+1}^{m+2n} w_i x_i. \quad (4.36)$$

при этом ограничения (4.31) и (4.32) меняются на условия

$$A\mathbf{x} = \mathbf{y}; \quad x \geq 0, w \geq 0.$$

Для применения рассмотренных выше методов ВВП [см. формулу (4.24)] и МНМ [выражения (4.35), (4.36)] оценки матрицы корреспонденций необходимо использовать некоторое начальное значение оцениваемых параметров  $x_{ij}^0$ .

Вектор начальных оценок корреспонденций  $\mathbf{x}^0$  можно получить на основе какой-либо предварительной обработки данных интенсивности движения. При выборе метода предварительной оценки корреспонденций можно учитывать, что в ПДП, проектах ОДД и реконструкции УДС объектом проектирования часто являются ограниченные по размерам фрагменты УДС. Влияние расстояния на распределение корреспонденций в выделяемых участках сетей не имеет такого значения, как в случаях больших сетей (генеральный план, КТС, КСОД города в целом). В этой связи в рассматриваемых задачах допустимо предположение, что корреспонденции между “нагрузочными” вершинами пропорциональны входящим и выходящим потокам этих вершин. В соответствии с этим для оценки начальных значений корреспонденций автором использовалась простейшая форма гравитационной модели

$$x_{ij}^0 = a_i b_j A_i B_j, \quad (4.37)$$

где  $x_{ij}^0$  – количество транспортных средств, следующих из вершины  $i$  в вершину  $j$ ;  $a_i$  – интенсивность движения из  $i$ ;  $b_j$  – интенсивность движения  $j$ ;  $A_i, B_j$  – коэффициенты балансировки по столбцам и строкам матрицы.

Использование формулы (4.37) не исключает возможности учета влияния факторов расстояния или времени на распределение корреспонденций. В случаях, когда сеть, для которой проводится оценка матрицы, занимает значительную территорию (город в целом или городские районы крупных и крупнейших городов), вместо упрощенной гравитационной модели следует использовать гравитационную модель общего вида:

$$x_{ij}^0 = a_i b_j A_i B_j f(c_{ij}). \quad (4.38)$$

Функция распределения поездок  $f(c_{ij})$  входит в состав стандартных исходных данных для прогнозирования матрицы корреспонденций в составе генерального плана или КТС. Поэтому при предварительном применении модели (4.38) для получения начального приближения  $x^0$  дополнительно требуются лишь использованные ранее данные прогнозирования матрицы корреспонденций в составе действующих генерального плана или КТС.

#### **4.5. Предварительная оценка исходных данных и выявление грубых ошибок**

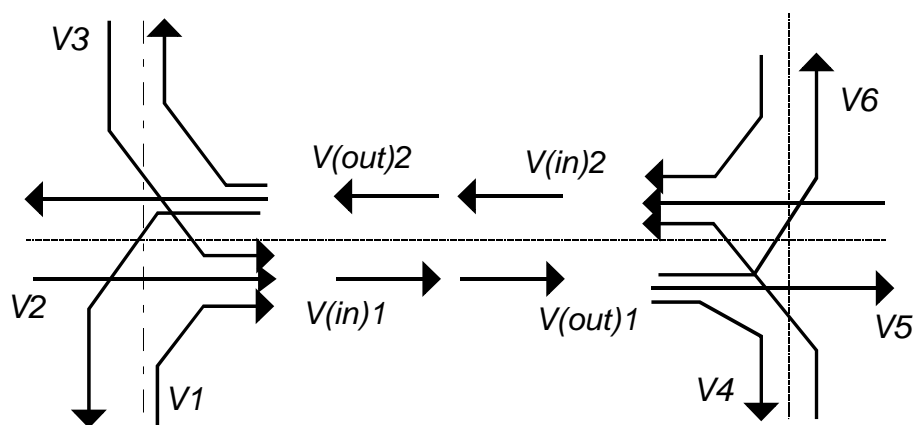
Предварительная обработка данных должна включать оценку их точности в целом и процедуру выявления недопустимых грубых ошибок. Для этих целей автором предлагается формализованное описание УДС в следующем виде (рис. 4.7):

входящие на перегон  $F(in)_i$  и выходящие  $F(out)_i$  с перегона потоки образуют пары, при этом на перегонах с односторонним движением получается одна пара сравниваемых значений, а на перегонах с двухсторонним движением – две;

данные замеров интенсивности движения образуют матрицу размерностью  $n \times 2$ , где  $n$  – количество пар сравниваемых значений  $V(out)_i, V(in)_i$ ;

для статистической оценки качества выборок значений интенсивности движения и выявления выбросов используются значения разностей пар, то есть

$$d_i = V(out)_i - V(in)_i. \quad (4.39)$$

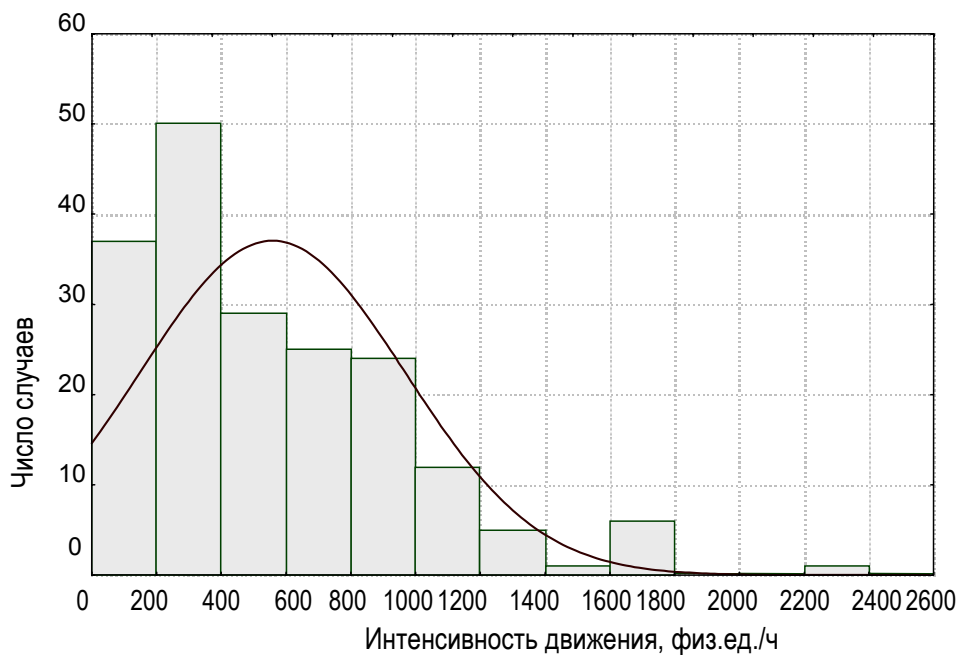


**Рис. 4.7.** Представление перегона улицы в модели оценки ошибок подсчетов интенсивности движения:  $V(in)_1, V(in)_2$  – входящие на перегон потоки;  $V(out)_1, V(out)_2$  – выходящие с перегона потоки;  $V1, V2, V3$  – потоки, полученные в результате замеров на предыдущем перекрестке и образующие поток  $V(in)_1$ ;  $V4, V5, V6$  – потоки, полученные в результате замеров на следующем перекрестке и образующие поток  $V(out)_1$ ;  $d_1 = V(out)_1 - V(in)_1$  – ошибка подсчетов интенсивности движения на перегоне для 1-го из направлений движения

Таким образом, сравниваются две выборки, одна из которых состоит из значений интенсивности движения потоков, входящих на перегоны  $V(in)_i$ ; вторая – выходящих  $V(out)_i$ , где  $i = 1, 2, \dots, n$ ,  $n$  – количество пар потоков данной сети. Поскольку этот случай можно рассматривать как сравнение двух  $n$ -мерных векторов, то формально перечень мер или критериев близости двух выборок очень широк (евклидова норма, расстояние Махолобиса, коэффициент

корреляции, критерий Стьюдента и т.д.). Вместе с тем свойства выборок случайных величин  $V(in)$  и  $V(out)$  требуют применения определенных статистических критериев.

Такие характеристики УДС, как значения интенсивности движения ТС, интенсивности движения пешеходов, имеют, как правило, распределение с так называемой положительной асимметрией, у которого мода  $\tilde{x}$  находится левее среднего значения  $\bar{x}$ . Это подтверждают результаты обследований как отдельных участков УДС, так и магистральной УДС целого города (рис. 4.8 и табл. 4.1).



**Рис. 4.8.** Гистограмма значений интенсивности движения (физ.ед./ч) на перегонах УДС Иркутска (1995 г.) и попытка аппроксимации нормальным распределением

**Таблица 4.1**

Статистика распределения значений интенсивности движения на перегонах УДС Иркутска (1995 г.)

| Статистика            | Значение статистики | Статистика        | Значение статистики |
|-----------------------|---------------------|-------------------|---------------------|
| Минимальное значение  | 46                  | Мода $\tilde{x}$  | 464,5000            |
| Максимальное значение | 2276                | Среднее $\bar{x}$ | 558,5368            |
| Вариационный размах   | 2230                |                   |                     |



Предложенное ранее формализованное описание УДС (см. рис. 4.7) позволяет применять метод парных сравнений, который не требует нормального распределения  $V(out)_i$  и  $V(in)_i$  и основан на процедуре, использующей разности пар (4.39). В общих чертах суть этого метода заключается в оценке близости двух упорядоченных выборок, элементы которых образуют пары. Оценка корректности использования данных интенсивности движения сводится к проверке гипотезы, что связанные выборки  $V(out)_i$  и  $V(in)_i$  принадлежат к одной генеральной совокупности.

Поэтому для предварительной оценки качества данных предлагается использовать проверку значимости среднего разности пар с применением  $t$ -критерия Стьюдента, который признается в статистической литературе оптимальным критерием метода парных сравнений [20]. Проверка статистической значимости разности пар проводится по формуле

$$\epsilon = \frac{\bar{d}}{S_d} = \frac{(\sum d_i)/n}{\sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}}}, \quad (4.40)$$

при этом число степеней свободы для критерия Стьюдента  $\nu = n - 1$ .

Проверяется нуль-гипотеза  $\mu_d=0$  (где  $\mu_d$  – истинное среднее значение разности пар), которая принимается, если рассчитанное значение критерия Стьюдента  $t$  меньше критического или равно ему  $t_{(n;\alpha)}$ . Как отмечается в специальной литературе [20], проверка связанных выборок с помощью  $t$ -критерия не имеет строгих допущений; переменные могут значительно отклоняться от нормального распределения; достаточно, чтобы их разности распределялись по нормальному закону.

Кроме проверки близости выборок (4.40) представлялось необходимым оценивать ошибки наблюдений с использованием тех показателей, которые будут в дальнейшем применяться для оценки

ошибок, возникающих при восстановлении матриц корреспонденций. С этой целью были выбраны следующие показатели:

средняя ошибка

$$\bar{d} = \sum_i d_i / n; \quad (4.41)$$

средняя абсолютная ошибка

$$\bar{d}_{abs} = \sum_i |d_i| / n; \quad (4.42)$$

отношение средней абсолютной ошибки к среднему значению интенсивности движения на перегоне в одном направлении

$$E = \bar{d}_{abs} / \bar{V}. \quad (4.43)$$

Показатели “средняя абсолютная ошибка” (4.42) и “отношение средней абсолютной ошибки к среднему значению корреспонденций” (4.43) часто используются для оценки точности восстановления матриц корреспонденций [113,115]. Кроме того, показатель  $\bar{d}_{abs}$  аналогичен такому критерию, как среднее абсолютное отклонение (САО) от среднего значения [20]

$$CAO = \frac{\sum |x_i - \bar{x}|}{n},$$

который рекомендуют применять в случаях загрязненных выборок малого объема.

Для анализа качества обследований интенсивности движения и ошибок, возникающих при сведении таких данных в одну выборку, автором использовались материалы обследований:

Санкт-Петербурга в 1977–1978 и 1985–1988 гг.;

Иркутска в 1995 г. и центральной части Иркутска в 1998 г.

При выполнении этого анализа предстояло доказать с формальных позиций корректность использования для восстановления матриц корреспонденций данных разных замеров интенсивности движения, сводимых в одну выборку (см. рис. 4.5 и 4.6). Оценка проводилась с использованием статистик, представленных в табл. 4.2. Как показали результаты анализа (табл. 4.3), средняя ошибка  $\bar{d}$

[формула (4.41)] не несет объективной информации о точности исходных данных; значительно информативнее статистики  $\bar{d}_{abs}$  и  $\bar{d}_{abs}/\bar{V}$ .

В соответствии с имевшимися у авторов данными максимальные абсолютные значения ошибок достигают 450 – 470 авт./ч (в одном направлении). Диапазон значений средней абсолютной ошибки составил  $\bar{d}_{abs}=53,8 - 190,6$ , а значений отношения  $\bar{d}_{abs}/\bar{V}$  – от 0,082 до 0,391 (см. табл.4.4). Средняя абсолютная ошибка составила 115,3 авт./ч, то есть 19% интенсивности движения.

Проведенный статистический анализ материалов различных обследований показывает, что сведение замеров интенсивности движения в единые выборки для оценки матриц корреспонденций является корректным с позиций статистики. Все рассмотренные в данной работе выборки входящих и выходящих потоков оценены как относящиеся к одним и тем же генеральным совокупностям. Коэффициенты корреляции между выборками  $V(out)_i$  и  $V(in)_i$ , как правило, превышали значение 0,9.

Анализ качества исходных данных был необходим еще с позиций определения требований к точности оценки матрицы корреспонденций, которые должны применяться на практике. Признаком сходимости итерационной оценки можно будет принять абсолютные величины погрешности вычислений корреспонденций

$$|e_i| = \left| x_i^n - x_i^{n-1} \right|, \quad (4.44)$$

где  $x_i^n$  – оценка  $i$  корреспонденции на  $n$ -итерации;  $x_i^{n-1}$  – оценка  $i$  корреспонденции на предыдущей  $n-1$ -итерации.

Значения допустимой невязки оценок корреспонденций (4.44), получаемые при последовательных итерациях, должны назначаться с учетом качества реальных исходных данных. Данные табл. 4.3 и 4.4 позволяют утверждать, что для практических целей допустима точность оценки матрицы корреспонденций, при которой невязка может составлять 10–20 авт./ч.

Таблица 4.2

Точность обследований интенсивности движения  
в Иркутске в 1995 г.

|  |                              |          |
|--|------------------------------|----------|
| Количество перегонов   |                              | 114      |
| Сравнение выборок $V(out)$ и $V(in)$   |                              |          |
| <p>Диаграмма разброса значений входящих <math>V(out)_i</math> и выходящих <math>V(in)_i</math> потоков</p> |                              |          |
| Средние значения   | $\bar{V}(out)$               | 625,912  |
|  | $\bar{V}(in)$                | 629,096  |
| Значение критерия Стьюдента  | Вычисленное $t$              | 0,266    |
|  | Критическое $t_{(113;0.05)}$ | 1,984    |
|  | Статистическая надежность    | 0,821    |
| Коэффициент корреляции $R$   |                              | 0,927    |
| Анализ ошибок $d_i = V(out)_i - V(in)_i$   |                              |          |
| Интенсивность движения на перегонах<br>в одном направлении $V$ , физ.ед./ч                                 | максимальная $V_{max}$       | 1799,00  |
|  | средняя $V$                  | 627,5044 |
|  | минимальная $V_{min}$        | 21,000   |
| Средняя ошибка $\bar{d}$   |                              | 3.1842   |
| Абсолютная ошибка $\bar{d}_{abs}$  | максимальная $d_{abs}^{max}$ | 478,000  |
|  | средняя $\bar{d}_{abs}$      | 102,693  |
|  | минимальная $d_{abs}^{min}$  | 0,000    |
| Отношение $\bar{d}_{abs}/\bar{V}$  |                              | 0,614    |

Таблица 4.3

## Показатели точности обследований интенсивности движения

| Фрагмент сети, год обследования                 | Средняя ошибка $\bar{d}$ | Средняя абсолютная ошибка $\bar{d}_{abs}$ | Отношение средней абсолютной ошибки к средней интенсивности $\bar{d}_{abs}/\bar{V}$ |
|---|--------------------------|---|---|
| Санкт-Петербург                                 |                          |   |   |
| Невский пр., 1977                               | 5,5                      | 116,1                                     | 0,276   |
| Московский пр., 1977                            | -20,4                    | 60,4                                      | 0,082   |
| Октябрьский (Адмиралтейский) р-н, 1979          | 14,4                     | 59,2                                      | 0,085   |
| Невский р-н, 1978 – 1979                        | 4,2                      | 114,9                                     | 0,273   |
| Невский р-н, 1984 – 1985                        | -12,9                    | 142,7                                     | 0,391   |
| Кировский р-н, 1984 – 1985                      | -7,4                     | 53,8                                      | 0,223   |
| Наб. Фонтанки, 1985                             | -24,5                    | 78,3                                      | 0,105   |
| Московский пр., 1987                            | -36,3                    | 190,6                                     | 0,176   |
| Московский р-н, 1987                            | 29,4                     | 141,6                                     | 0,185   |
| Иркутск   |                          |   |   |
| Ул. Ленина, 1995                                | 67,4                     | 141,7                                     | 0,155   |
| Ул. К. Маркса, 1995                             | -15,3                    | 178,2                                     | 0,224   |
| Центр Иркутска, 1995                            | -41,                     | 134,5                                     | 0,234   |
| Ул. Сергеева – Маяковского, 1995<br>16.00-17.00 | 4,8                      | 105,9                                     | 0,119   |
| Ул. Сергеева – Маяковского, 1995<br>17.00-18.00 | 18,4                     | 108,9                                     | 0,117   |
| УДС Иркутска в целом, 1995                      | 3,2                      | 102,7                                     | 0,164   |

Таблица 4.4

## Средние показатели точности обследований интенсивности движения

| Оценки                  | Среднее значение | Доверительный интервал |        | Значение    |              |
|-------------------------|------------------|------------------------|--------|-------------|--------------|
|                         |                  | -95,0%                 | +95,0% | минимальное | максимальное |
| $\bar{d}_{abs}$         | 115,3            | 92,5                   | 138,1  | 53,8        | 190,6        |
| $\bar{d}_{abs}/\bar{V}$ | 0,187            | 0,140                  | 0,234  | 0,082       | 0,391        |

Правила исключения выделяющихся одного или нескольких наблюдений давно рассматриваются в статистике и предложено

большое количество критериев отбраковки. В литературе, посвященной робастным оценкам, рекомендован ряд процедур исключения резко выделяющихся наблюдений, основанных на использовании критериев Смирнова-Граббса, Граббса, Титьена-Мура и т.д. [18,20,54,61,64]. Подчеркивается, что по сравнению с другими критериями отбраковки они имеют высокую мощность.

Процедура, базирующаяся на применении критерия Смирнова-Граббса, в рассматриваемом случае сводится к определению статистики  $T_i$ , то есть

$$T_i = \frac{d_i - \bar{d}}{S_d}, \quad (4.45)$$

где  $\bar{d} = \sum_{i=1}^n \frac{d_i}{n}$  – среднее значение ошибки;  $S_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$  – среднеквадратическое отклонение;  $n$  – количество наблюдений (то есть пар).

Отбраковка наблюдения производится, если при заданной значимости  $\alpha$  расчетное значение  $T$  превосходит критическое  $T_\alpha$ .

Одна из часто предлагаемых при выполнении регрессионного анализа процедур исключения наблюдений, содержащих грубые ошибки, использует так называемые студентизированные остатки, то есть нормированные отклонения [18,54]. В рассматриваемом случае эта статистика рассчитывается в соответствии с формулой

$$\epsilon = \frac{d}{S_d} = \frac{d}{\sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}}}. \quad (4.46)$$

Согласно этому методу парами, содержащими грубые ошибки, считаются такие, у которых  $z$  превосходит некоторое заранее заданное значение  $z_\alpha$ . Первоначально предполагалось, что при тестировании рассматриваемой процедуры для отбраковки будут ис-

пользоваться уровнем значимости  $\alpha = 5\%$  и соответствующее ему  $z_{0,05} = 1,960$  [20].

Тестирование критериев выявления грубых ошибок позволяет утверждать, что предпочтительнее применение статистики  $z$  (нормированного отклонения). Критерий Смирнова – Граббса (4.45) оказался мало чувствительным к ошибкам в рассматриваемых автором данным. Лишь однажды при анализе ошибок данных обследований интенсивности движения в Иркутске в 1994 – 1995 гг. был отмечен случай, когда у одного из наблюдений вычисленное значение  $T$  превышало критическое.

В дальнейшем для выявления грубых ошибок авторы использовали нормированное отклонение  $z$  [формула (4.46)]. Следует отметить, что использование уровня значимости  $\alpha = 5\%$  не всегда обеспечивало эффективность процедур выявления грубых ошибок. Особенностью исходных данных, которые исследовал автор, являлось то, что при принятии доверительной вероятности  $\beta = 95\%$  (соответственно уровень значимости  $\alpha = 5\%$ ) они лишь в редких случаях классифицировались как содержащие грубые ошибки. Результаты тестов позволяют рекомендовать задаваться доверительной вероятностью  $\beta$  в зависимости от объема выборки:

сеть до 50 перегонов,  $\beta = 80\%$ ;

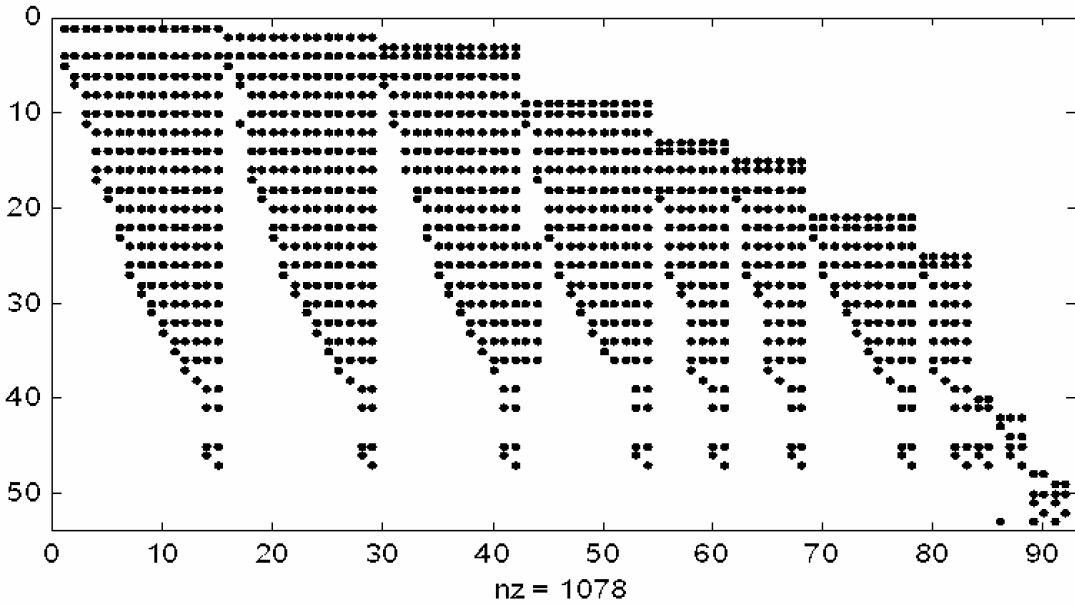
сеть более 50 перегонов,  $\beta = 90\%$ .

#### **4.6. Результаты тестирования методов робастного оценивания матриц корреспонденций**

Тестирование методов робастной оценки матриц корреспонденций выполнялось с использованием библиотек операций с матрицами и оптимизации пакета MATLAB.

Следует сразу отметить, что метод вариационно- взвешенных приближений (ВВП) [см. формулу (4.24)], который при использо-

вании пакета MATLAB реализуется в матрично-векторной форме как  $x = \text{inv}(A' * W * A) * A' * W * Y$ , явился не пригодным для решения рассматриваемой задачи.



**Рис. 4.9.** Структура матрицы инцидентности для графа, отображающего улицу одностороннего движения (ул. К. Маркса в Иркутске)

Как указывалось ранее, при оценке матриц корреспонденций предлагается рассматривать отдельно каждое из направлений движения (см. рис. 4.1 и 4.3), что позволяет использовать распределение потоков “все или ничего”. Но при этом получаются матрицы  $A$  инцидентности с коллинеарными столбцами, что можно показать на реальном примере (рис. 4.9). Получение матриц инцидентности со структурой, аналогичной структуре матрицы, показанной на рис. 4.9, будет неизбежным в случаях, когда оцениваются матрицы корреспонденций “вдоль маршрута движения”. Исходные матрицы инцидентности, построенные для реальных УДС, получаются разреженными и плохо обусловленными из-за наличия коррелированных столбцов и строк. Плохая обусловленность матрицы  $A$  приводит к тому, что небольшие изменения ее элементов, в свою очередь, при-



водят к значительным ошибкам матричных вычислений и вектора оценок  $\mathbf{x} = (A'WA)^{-1} A'WY$ .

В этой связи можно утверждать, что матрицы инцидентностей, которые необходимо использовать в задаче (4.24), будут иметь плохую обусловленность. Соответственно, те методы регрессионного анализа, которые используют нормальные уравнения (4.21) или операции вычисления обратных матриц [см. формулы (4.24),(4.25)], не пригодны для оценки корреспонденций “вдоль маршрутов движения”.

Тестирование вариантов метода наименьших модулей (МНМ) выполнялось с помощью библиотеки Optimization Toolbox пакета MATLAB. Минимизация целевых функций (4.35) и (4.36) рассматривается как задача линейного программирования со смешанными ограничениями, в которой целевая функция

$$\min \mathbf{w}^T \mathbf{x} \quad (4.47)$$

линейные ограничения

$$A2 \mathbf{x} = \mathbf{y}, \quad (4.48)$$

при этом матрица  $A2$  получается на основе матрицы инцидентности  $A$  [см. формулу (4.35)]; двухсторонние ограничения

$$\mathbf{x}^{lb} \leq \mathbf{x} \leq \mathbf{x}^{ub}. \quad (4.49)$$

Поэтому исходные данные включали:

$A$  – матрицу инцидентности ранга  $n \times m$ , задающую принадлежность корреспонденций дугам графа сети (ее элементы  $a_{ij}$  имеют значение 0 или 1); если корреспонденция  $j$  проходит по дуге  $i$ , то  $a_{ij} = 1$ , в противном случае  $a_{ij} = 0$ , где  $i = 1, 2, \dots, n$ ,  $n$  – количество дуг графа УДС, на которых задана интенсивность движения,  $j = 1, 2, \dots, m$ ,  $m$  – количество оцениваемых корреспонденций .

$\mathbf{y}$  – вектор-столбец значений интенсивности движения на дугах сети;

$\mathbf{x}0$  – начальное приближение значений корреспонденций – исходный вектор-столбец значений, который может быть задан:

в виде старой матрицы корреспонденций, преобразованной в вектор;

значениями корреспонденций, полученными в результате экспертной оценки;

в виде существующей матрицы корреспонденций, восстановленной с гравитационной модели;

$lb$  – коэффициент, задающий нижние допустимые значения корреспонденций  $x^{lb}$ ,  $x^{lb} = x lb$ ;

$ub$  – коэффициент, задающий верхние допустимые значения корреспонденций,  $x^{ub}$ ,  $x^{ub} = x ub$ ;

$div$  – коэффициент, задающий верхние допустимые значения искусственных переменных  $x^{ub}$ ,  $x^{ub}_i = x_i/div$ ;

$v$  – показатель степени весовых коэффициентов  $w_i = |e_i|^{v-2}$ .

Соответственно, выходные данные включали следующие векторы-столбцы:

$x$  – восстановленные значения корреспонденций;

$b$  – значения интенсивности движения на дугах, рассчитанные на основе восстановленных корреспонденций  $b = Ax$ ;

$e$  – остатки (ошибки) регрессии – разность между наблюдаемыми и рассчитанными значениями интенсивности движения на дугах  $e = y - Ax$ .

Двухсторонние ограничения задачи линейного программирования (4.49) могут использоваться как средство управления сходимостью итераций и точности оценки. Получаемый в результате решения задачи линейного программирования (4.47) – (4.49) вектор-столбец  $x^2$  размером  $m+2n$  состоит из корреспонденций  $x$ , которые составляют его элементы  $x_1, x_2, \dots, x_m$ , и из искусственных переменных  $x_{m+1}, x_{m+2}, \dots, x_{m+2n}$ , одновременно являющихся той частью вектора  $x^2$ , которая входит в состав целевой функции (4.47). Поэтому значения вектора нижних ограничений принимаются  $x^{lb}_i = 0$ , а векторы верхних ограничений  $x^{ub}$  назначаются:

$x_i^{ub} = k x_i$ ,  $k > 1$ , где  $x_i$  – значение, полученное на предыдущей итерации;

$x_{m+j}^{ub} = (1/div) x_j$  и  $x_{m+n+j}^{ub} = (1/div) x_j$ , где  $x_j$  – модуль остатка регрессии  $e_j$ , полученный на предыдущей итерации,  $div$  – коэффициент,  $div > 1$ .

Тестирование вариантов метода наименьших модулей (МНМ) было выполнено авторами на основе следующих данных:

искусственные данные без ошибок;

искусственные данные с внесенными грубыми ошибками;

данные реальных обследований.

Задачей первого этапа тестирования являлась оценка точности методов. Поэтому тест выполнялся по следующей схеме:

задавались искусственная матрица корреспонденций  $x$ , преобразованная в вектор, и соответствующие этой матрице точные значения интенсивности движения на дугах графа  $y = A x$ ;

проводилось восстановление матрицы корреспонденций и выполнялась оценка точности этого восстановления с применением целого ряда статистик.

Оценка точности включала:

парное сравнение заданных  $x_i$  и оцененных  $x_i$  значений корреспонденций с использованием статистик

$$d_i = x_i - \hat{x}_i; \quad \bar{d} = \sum_i d_i / m; \quad \bar{d}_{abs} = \sum_i |d_i| / m; \quad D = \bar{d}_{abs} / \bar{x}; \quad (4.50)$$

парное сравнение заданных  $y_i$  и оцененных  $y_i$  значений потоков на дугах (т.е. анализ остатков регрессии  $e = y - A^*x$ ) с использованием статистик

$$e_i = y_i - \hat{y}_i; \quad \bar{e} = \sum_i e_i / n; \quad \bar{e}_{abs} = \sum_i |e_i| / n; \quad E = \bar{e}_{abs} / \bar{y}. \quad (4.51)$$

Задачей второго этапа тестирования вариантов МНМ являлась оценка их сходимости при наличии грубых ошибок. Представляет-

ся, что наиболее точно исследовать сформулированную задачу можно при проведении численного эксперимента, когда:

задаются точные значения корреспонденций  $x_i$  и потоков  $y_i$ , соответствующих этим корреспонденциям, т.е.  $y = A x$ ;

задаются ошибки исходных данных (т.е. значения  $y_i^*$  с внесенными ошибками);

восстанавливается матрица корреспонденций  $x_i$  при заданных с ошибками значениях потоков  $y_i^*$ ;

восстановленные значения корреспонденций  $x_i$  сравниваются с заданными точными значениями  $x_i$  с использованием статистик (4.50);

Проводимый в такой форме численный эксперимент позволит установить:

влияние ошибок исходных данных (т.е. значений интенсивности движения) на сходимость простого МНМ и взвешенного МНМ;

точность восстановления матрицы корреспонденций при использовании простого МНМ и взвешенного МНМ в случае ошибок исходных данных;

предельные допустимые значения ошибок исходных данных, позволяющие выполнять восстановление матрицы корреспонденций.

На третьем этапе тестирования выполнялось сравнение простого и взвешенного МНМ на примерах реальных данных с целью:

установить влияние верхних и нижних ограничений оцениваемых переменных на сходимость и точность оценки;

сравнить сходимость и точность простого и взвешенного МНМ при одних и тех же верхних и нижних ограничениях оцениваемых переменных.

На последнем этапе тестирования точность восстановления корреспонденций уже оценивалась только парным сравнением за-

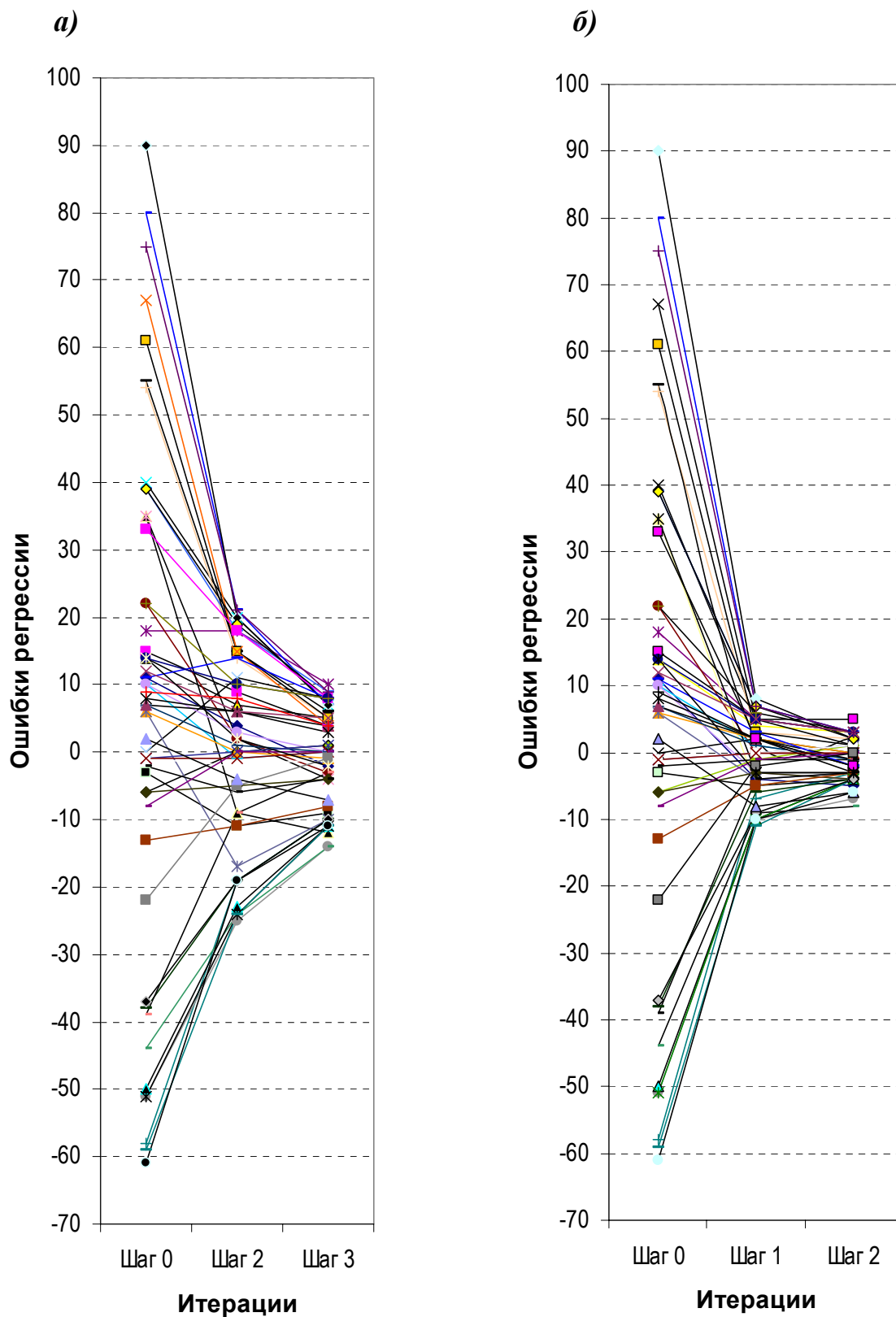
данных  $y_i$  и оцененных  $y_i$  значений потоков на дугах (т.е. анализом остатков регрессии  $e = y - A*x$ ) с использованием статистик (4.51).

Тестирование вариантов МНМ показало, что метод отличается хорошей устойчивостью при плохо обусловленных матрицах и матрицах неполного ранга. Это особенно важно для практического использования метода, поскольку позволяет применять его для случаев, когда число ненулевых элементов матрицы корреспонденций (т.е. количество оцениваемых корреспонденций) превышает количество дуг графа УДС, на которых известны значения интенсивности движения.

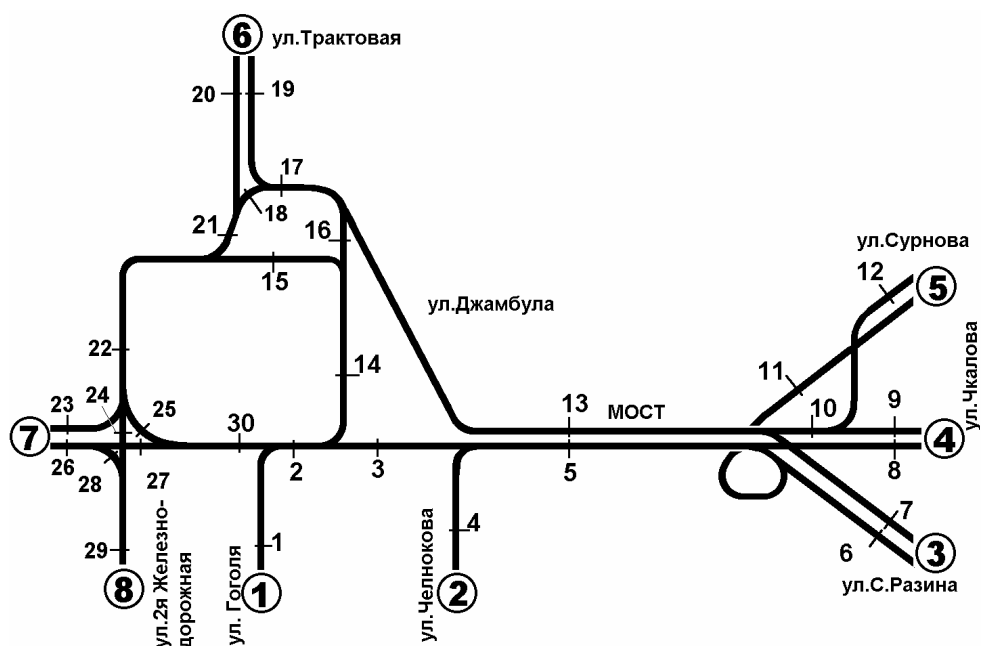
По результатам тестирования с применением искусственных загрязненных данных и реальных данных простой МНМ имеет более высокую скорость схождения по сравнению со взвешенным МНМ. В частности, это можно показать в виде графика “итерации – остатки регрессии” (рис. 4.10); нулевой итерацией (шаг 0) принимается определение вектора  $x_0$  с использованием гравитационной модели (4.37).

Анализ ошибок оценки корреспонденций и потоков, невязок оценок корреспонденций по итерациям показал, что необходимая для практических задач точность обеспечивается на 2-й итерации. В ряде случаев тестов сильного загрязнения увеличение числа итераций приводило к снижению качества оценки. Поэтому для практических целей можно рекомендовать процедуру восстановления матрицы всего из двух итераций с использованием простого или комбинированного МНМ (взвешенный МНМ на 2-й итерации).

Примером могут служить результаты тестов, выполнявшихся для участка УДС на подходах к мосту в Иркутске (рис. 4.11 – 4.14, табл. 4.5 – 4.7), для которого в дальнейшем были оценены матрицы корреспонденций и распределения потоков по реальным данным. Графики на рис. 4.12 – 4.14 показывают, как изменялись оценки корреспонденций и потоков по итерациям.



**Рис. 4.10.** Ошибки  $e_i = y_i - \hat{x}_i$  (остатки регрессии) на 0-й – 2-й итерациях: *а* – взвешенный МНМ; *б* – сочетание итераций (1-я – простой МНМ, 2-я – взвешенный МНМ)



**Рис. 4.11.** Улично-дорожная сеть лево- и правобережных подходов к старому мосту через Ангару в Иркутске:

①, ②, ..., ⑤ – “нагрузочные” вершины (истoki – стоки);  
 1, 2, ..., 30 – дуги, включенные в матрицу инцидентности  $A$

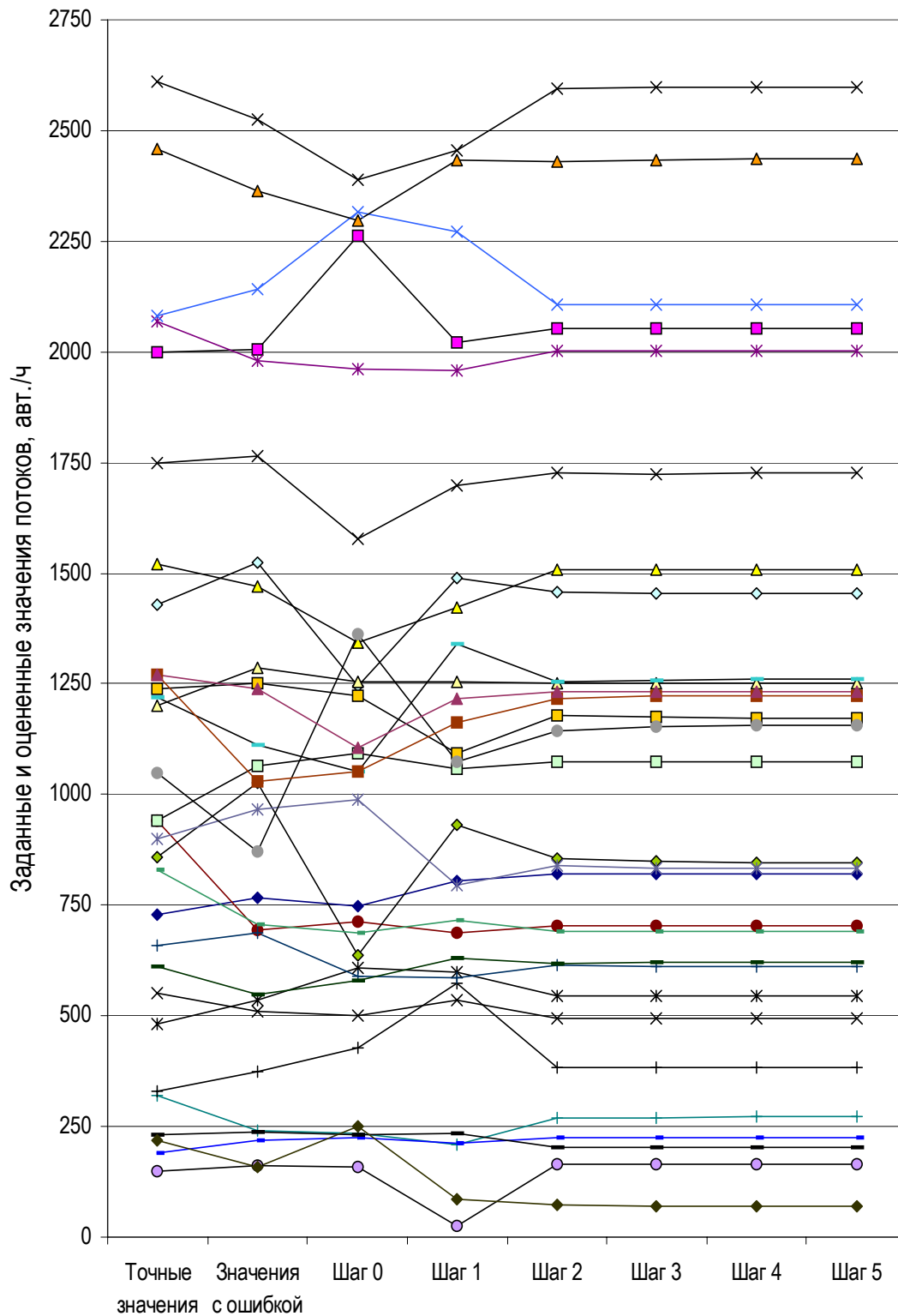
Хотя предлагаемый метод восстановления матриц ориентирован на загрязненные данные, он дает высокие показатели точности при хорошем качестве исходных данных, получаемых в результате специальных обследований интенсивности движения. В данном контексте под специальным обследованием понимается выполнение замеров интенсивности движения на исследуемом участке УДС за короткий период времени, что снижает вероятность возникновения грубых ошибок. В качестве примера можно привести результаты оценки матрицы корреспонденций на ул. К.Маркса в центре Иркутска (рис.4.15, табл. 4.8) по данным обследований 1998 г., выполненных за неделю. Использование процедуры восстановления матрицы простого и комбинированного МНМ (на 2-й итерации используется взвешенный МНМ) дало среднюю абсолютную ошибку соответственно 3,2 и 2,7 авт./ч. С учетом показателей точности восстановления матриц, приводимых в других исследованиях [113,115], этот результат можно оценивать как очень хороший.

**Таблица 4.5**

Парный анализ заданных и восстановленных значений потоков при оценке точности восстановления матрицы корреспонденций с использованием простого МНМ

| Дуга (см. рис.4.11)                | Значение потока на дуге |           | Оценка потоков |        |        |        |        |        |
|------------------------------------|-------------------------|-----------|----------------|--------|--------|--------|--------|--------|
|                                    | точное                  | с ошибкой | Итерации       |        |        |        |        |        |
|                                    |                         |           | 0              | 1      | 2      | 3      | 4      | 5      |
| Y1                                 | 730                     | 767       | 749            | 805    | 822    | 822    | 822    | 822    |
| Y2                                 | 2000                    | 2005      | 2263           | 2022   | 2054   | 2054   | 2054   | 2054   |
| Y3                                 | 1520                    | 1469      | 1342           | 1424   | 1508   | 1508   | 1508   | 1508   |
| Y4                                 | 550                     | 511       | 499            | 535    | 493    | 493    | 493    | 493    |
| Y5                                 | 2070                    | 1980      | 1961           | 1959   | 2001   | 2001   | 2001   | 2001   |
| Y6                                 | 940                     | 695       | 712            | 688    | 702    | 702    | 702    | 702    |
| Y7                                 | 320                     | 241       | 236            | 208    | 269    | 270    | 271    | 271    |
| Y8                                 | 190                     | 219       | 225            | 212    | 226    | 226    | 226    | 226    |
| Y9                                 | 230                     | 238       | 231            | 235    | 204    | 203    | 203    | 203    |
| Y10                                | 1430                    | 1525      | 1246           | 1489   | 1457   | 1455   | 1455   | 1455   |
| Y11                                | 940                     | 1066      | 1092           | 1059   | 1073   | 1073   | 1073   | 1073   |
| Y12                                | 1200                    | 1287      | 1256           | 1254   | 1253   | 1252   | 1252   | 1252   |
| Y13                                | 1750                    | 1766      | 1579           | 1697   | 1726   | 1725   | 1726   | 1726   |
| Y14                                | 480                     | 536       | 609            | 598    | 546    | 546    | 546    | 546    |
| Y15                                | 150                     | 161       | 159            | 24     | 164    | 164    | 164    | 164    |
| Y16                                | 330                     | 375       | 428            | 574    | 382    | 382    | 382    | 382    |
| Y17                                | 2080                    | 2141      | 2316           | 2271   | 2108   | 2107   | 2108   | 2108   |
| Y18                                | 1220                    | 1113      | 1053           | 1339   | 1254   | 1259   | 1262   | 1262   |
| Y19                                | 860                     | 1028      | 638            | 932    | 854    | 848    | 846    | 846    |
| Y20                                | 1240                    | 1252      | 1222           | 1093   | 1177   | 1174   | 1173   | 1173   |
| Y21                                | 2460                    | 2365      | 2297           | 2432   | 2431   | 2433   | 2435   | 2435   |
| Y22                                | 2610                    | 2526      | 2390           | 2456   | 2595   | 2597   | 2599   | 2599   |
| Y23                                | 900                     | 966       | 990            | 794    | 838    | 834    | 833    | 833    |
| Y24                                | 1050                    | 871       | 1363           | 1075   | 1143   | 1152   | 1155   | 1155   |
| Y25                                | 660                     | 689       | 588            | 587    | 614    | 611    | 611    | 611    |
| Y26                                | 830                     | 707       | 689            | 717    | 691    | 691    | 690    | 690    |
| Y27                                | 610                     | 549       | 580            | 630    | 618    | 621    | 621    | 621    |
| Y28                                | 220                     | 158       | 249            | 87     | 73     | 70     | 69     | 69     |
| Y29                                | 1270                    | 1029      | 1053           | 1162   | 1216   | 1222   | 1224   | 1224   |
| Y30                                | 1270                    | 1238      | 1105           | 1217   | 1232   | 1232   | 1232   | 1232   |
| Корреляция $R$                     | 0,989                   | 0,975     | 0,986          | 0,994  | 0,994  | 0,994  | 0,994  | 0,994  |
| Средняя ошибка $\bar{e}$           | 21,233                  | 33,000    | 17,833         | 12,867 | 12,767 | 12,767 | 12,467 | 12,467 |
| Стандартное отклонение $S_e$       | 18,065                  | 27,840    | 20,679         | 14,053 | 14,191 | 14,254 | 14,254 | 14,254 |
| Средняя абсолютная $\bar{e}_{abs}$ | 77,967                  | 128,333   | 94,167         | 58,867 | 59,633 | 59,800 | 59,800 | 59,800 |
| $t$ -критерий Стьюдента:           |                         |           |                |        |        |        |        |        |
| критическое значение               | $t_{(29;0,05)} = 2,045$ |           |                |        |        |        |        |        |
| вычисленное значение               | 1,175                   | 1,185     | 0,862          | 0,916  | 0,900  | 0,875  | 0,875  | 0,875  |





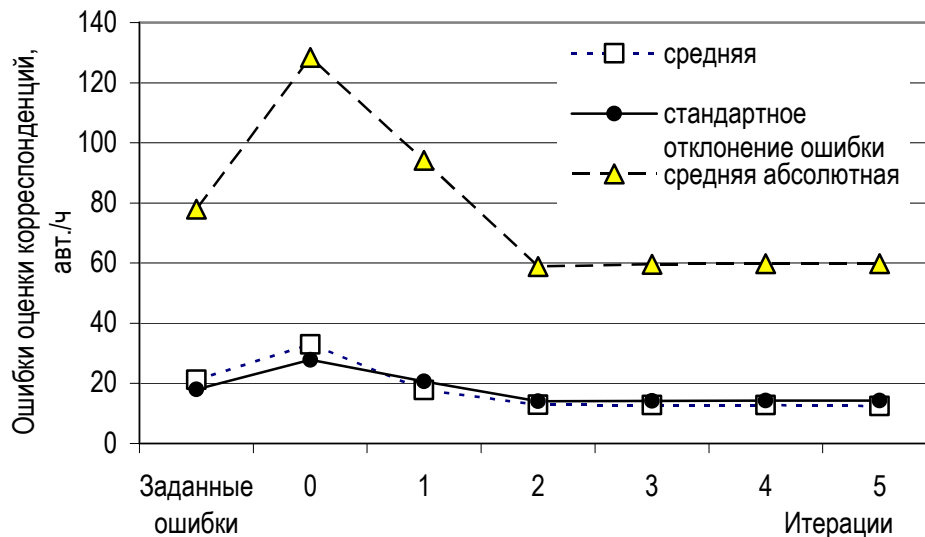
**Рис. 4.12.** Точные значения потоков, значения потоков с внесенными ошибками и восстановленные значения потоков по итерациям при использовании простого МНМ (см. табл. 4.5)

**Таблица 4.6**

**Оценка точности восстановления матрицы корреспонденций  
на основе парного анализа заданных  
и восстановленных значений корреспонденций**

| Точные значения<br>корреспонденций        |     | Оценка корреспонденций  |         |         |         |         |         |
|---|-----|-------------------------|---------|---------|---------|---------|---------|
|   |     | Итерации                |         |         |         |         |         |
|   |     | 0                       | 1       | 2       | 3       | 4       | 5       |
| X1  | 190 | 143                     | 157     | 160     | 160     | 160     | 160     |
| X2  | 40  | 45                      | 59      | 62      | 62      | 62      | 62      |
| X3  | 190 | 221                     | 235     | 238     | 238     | 238     | 238     |
| X4  | 160 | 197                     | 198     | 198     | 198     | 198     | 198     |
| X5  | 150 | 161                     | 164     | 164     | 164     | 164     | 164     |
| X6  | 250 | 179                     | 174     | 173     | 173     | 173     | 173     |
| X7  | 50  | 57                      | 52      | 51      | 51      | 51      | 51      |
| X8  | 250 | 275                     | 270     | 269     | 269     | 269     | 269     |
| X9  | 100 | 89                      | 50      | 42      | 40      | 40      | 40      |
| X10                                       | 90  | 73                      | 62      | 60      | 60      | 60      | 60      |
| X11                                       | 130 | 79                      | 152     | 167     | 170     | 171     | 171     |
| X12                                       | 70  | 88                      | 32      | 21      | 19      | 18      | 18      |
| X13                                       | 70  | 72                      | 44      | 38      | 37      | 37      | 37      |
| X14                                       | 90  | 78                      | 134     | 145     | 147     | 148     | 148     |
| X15                                       | 360 | 476                     | 420     | 409     | 407     | 406     | 406     |
| X16                                       | 350 | 390                     | 362     | 356     | 355     | 355     | 355     |
| X17                                       | 490 | 421                     | 477     | 488     | 490     | 491     | 491     |
| X18                                       | 300 | 242                     | 221     | 217     | 216     | 216     | 216     |
| X19                                       | 60  | 76                      | 55      | 51      | 50      | 50      | 50      |
| X20                                       | 300 | 371                     | 350     | 346     | 345     | 345     | 345     |
| X21                                       | 240 | 270                     | 228     | 220     | 218     | 217     | 217     |
| X22                                       | 340 | 293                     | 335     | 343     | 345     | 345     | 345     |
| X23                                       | 200 | 131                     | 149     | 152     | 153     | 153     | 153     |
| X24                                       | 40  | 41                      | 59      | 62      | 63      | 63      | 63      |
| X25                                       | 200 | 199                     | 217     | 220     | 221     | 221     | 221     |
| X26                                       | 170 | 178                     | 183     | 184     | 184     | 184     | 184     |
| X27                                       | 220 | 158                     | 87      | 73      | 70      | 69      | 69      |
| Корреляция $R$                            |     | 0,929                   | 0,932   | 0,924   | 0,922   | 0,921   | 0,921   |
| Средняя ошибка $\bar{d}$                  |     | 3,593                   | 6,444   | 7,974   | 7,222   | 7,259   | 7,259   |
| Стандартное отклонение $S_{\bar{d}}$      |     | 45,481                  | 44,702  | 47,750  | 48,456  | 48,670  | 48,670  |
| Средняя абсолютная ошибка $\bar{d}_{abs}$ |     | 116,000                 | 133,000 | 147,000 | 150,000 | 151,000 | 151,000 |
| $t$ -критерий Стьюдента:                  |     |                         |         |         |         |         |         |
| критическое значение                      |     | $t_{(26;0,05)} = 2,056$ |         |         |         |         |         |
| вычисленное значение                      |     | 0,410                   | 0,749   | 0,770   | 0,774   | 0,775   | 0,775   |





**Рис. 4.14.** Оценка точности восстановления матрицы корреспонденций с использованием простого МНМ (к табл. 4.7, 4.8). Изменение ошибок восстановления значений потоков (остатков регрессии) по итерациям:

средняя  $\bar{e} = (\sum e_i)/n$ , где  $e_i = y_i - \hat{x}_i$ ;

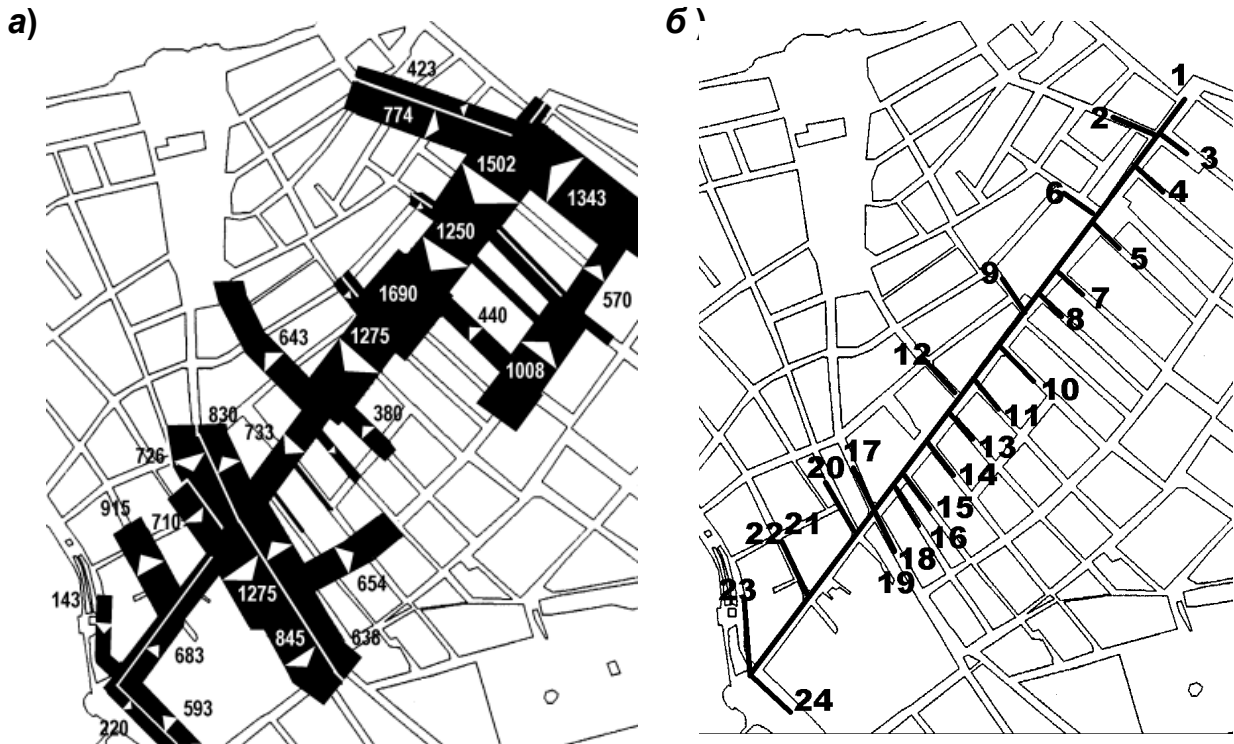
стандартное отклонение  $S_{\bar{e}} = \sqrt{\frac{\sum e_i^2 - (\sum e_i)^2/n}{n(n-1)}}$ ;

средняя абсолютная  $\bar{e}_{abs} = (\sum |e_i|)/n$

**Таблица 4.7**

Ошибки восстановления значений потоков  
(см. табл. 4.5 и рис. 4.12, 4.14)

| Итера-<br>ция    | Ошибки           |                   |                   | Абсолютные ошибки |                         |                   |
|------------------|------------------|-------------------|-------------------|-------------------|-------------------------|-------------------|
|                  | минималъ-<br>ная | средняя $\bar{e}$ | макси-<br>мальная | минималъ-<br>ная  | средняя $\bar{e}_{abs}$ | максималъ-<br>ная |
| 0                | -313,00          | 33,000            | 228,00            | 1,00              | 128,333                 | 313,00            |
| 1                | -244,00          | 17,833            | 252,00            | 5,00              | 94,167                  | 252,00            |
| 2                | -133,00          | 12,867            | 238,00            | 6,00              | 58,867                  | 238,00            |
| 3                | -133,00          | 12,767            | 238,00            | 11,00             | 59,633                  | 238,00            |
| 4                | -133,00          | 12,467            | 238,00            | 11,00             | 59,800                  | 238,00            |
| 5                | -133,00          | 12,467            | 238,00            | 11,00             | 59,800                  | 238,00            |
| Внесенные ошибки |                  |                   |                   |                   |                         |                   |
|                  | -168             | 21,333            | 245               | 5                 | 77,967                  | 245               |



**Рис.4.15.** Картограмма интенсивности движения в районе ул. К.Маркса (центр Иркутска) в вечерний час пик (а) и корреспондирующие вершины (б)

**Таблица 4.8**

Сравнение точности восстановления значений потоков с использованием взвешенного МНМ, простого МНМ и их комбинации

| МНМ        | Шаг | Ошибка      |              |                   | Абсолютная ошибка |                         |
|------------|-----|-------------|--------------|-------------------|-------------------|-------------------------|
|            |     | минимальная | максимальная | средняя $\bar{e}$ | максимальная      | средняя $\bar{e}_{abs}$ |
| Взвешенный | 0   | -61         | 90           | 6,981132          | 90                | 27,77358                |
|            | 1   | -25         | 21           | 0,716981          | 25                | 11,16981                |
|            | 2   | -14         | 10           | -0,54717          | 14                | 5,981132                |
| Простой    | 0   | -61         | 90           | 6,981132          | 90                | 27,77358                |
|            | 1   | -11         | 8            | -0,622641         | 11                | 4,433962                |
|            | 2   | -10         | 4            | -1,995600         | 8                 | 3,188679                |
| Комбинация | 0   | -61         | 90           | 6,981132          | 90                | 27,77358                |
|            | 1   | -11         | 8            | -0,622641         | 11                | 4,433962                |
|            | 2   | -8          | 5            | -1                | 8                 | 2,698113                |

Тестирование показало, что наибольший эффект на качество восстановления корреспонденций и значений потоков оказывают двухсторонние ограничения (4.49) корреспонденций  $x_i$   $i=1,2,\dots,m$  и искусственных переменных  $x_i$ ,  $i=m+1,m+2,\dots,m+2n$ . В соответствии с результатами тестирования рекомендуются следующие значения ограничений:

нижние ограничения оцениваемых переменных  $x_i^{lb}=kx_i$ ,  $k=0$ ;

верхние ограничения оцениваемых переменных  $x_i^{ub}=kx_i$ ,  $k \geq 2$ ;

нижние ограничения искусственных переменных  $x_i^{lb}=0$ ,  $i=m+1,m+2,\dots,m+2n$ ;

верхние ограничения искусственных переменных  $x_{m+j}^{ub} = (1/div) x_j$  и  $x_{m+n+j}^{ub} = (1/div) x_j$  рассчитывать для  $div \geq 30$ , где  $x_j$  – модуль остатка регрессии  $e_j$ , полученный на предыдущей итерации,  $n$  – количество дуг графа УДС, на которых заданы значения интенсивности движения;  $j=1,2,\dots,n$ .

Поскольку при восстановлении корреспонденций получаемые значения коэффициента корреляции очень высокие, предлагается следующее сочетание критериев, более чувствительных к результатам регрессионного анализа [см. формулы (4.40) и (4.51)]:

критерий Стьюдента для парного сравнения исходных и оцененных значений потоков;

средняя абсолютная разность пар замеренных и оцененных значений потоков  $\bar{e}_{abs}$ ;

отношение средней абсолютной ошибки к среднему значению интенсивности движения на перегоне в одном направлении  $E$ .

В настоящее время имеется большое количество математических пакетов, которые можно использовать для решения рассмотренной в данной главе задачи [74,139,144]. Выбор пакета MATLAB для решения задачи восстановления матриц корреспонденций вызван и тем, что:

пакет ориентирован на выполнение операций с векторами и матрицами, в том числе разряженными матрицами;

пакет применяется для решения сетевых задач [227], в том числе на его основе осуществляется разработка приложений для транспортных задач, например, программа **DelftOD Version 2.0** (Nanne J. van der Zijpp, Delft University of Technology <http://www.delftod.tudelft.nl>);

пакет достаточно известен и распространен в нашей стране, имеет сопровождение [144].

Предложенный метод полностью основан на применении стандартных математических библиотек, не требует специальных профессиональных знаний в области программирования, поэтому прост и доступен для использования. В дальнейшем авторы предполагают рассмотреть возможности применения метода робастного восстановления корреспонденций пассажирских (межостановочных матриц корреспонденций) и пешеходных потоков.