

**БИЛЬФЕЛЬД Н.В., СОБАКИН И.А., БЕЛЯНИН Н.А.
ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ДОРОЖНО-ТРАНСПОРТНЫХ
ПРОИСШЕСТВИЙ В ПЕРМСКОМ КРАЕ НА ОСНОВЕ
МАТЕМАТИЧЕСКИХ МНОГОФАКТОРНЫХ МОДЕЛЕЙ**

УДК 629.331:004.94, ВАК 05.13.18, ГРНТИ 28.17.19

Прогнозирование количества дорожно-транспортных происшествий в Пермском крае на основе математических многофакторных моделей

Prediction of the quantity of road-transport accidents in the Perm Region on the basis of mathematical multifactor models

Бильфельд Н. В., Собакин И. А.,
Белянин Н. А.

N. V. Bilfeld, I. A. Sobakin,
N. A. Belyanin

Березниковский филиал Пермского национального исследовательского политехнического университета, г. Березники

Berezniki branch of the Perm National Research Polytechnic University, Berezenky

В данной статье рассмотрены основные причины большого количества ДТП на дорогах России. Обоснована актуальность уменьшения количества ДТП. Приведена статистика ДТП в Пермском крае за последние 10 лет. Проанализированы факторы, влияющие на количество ДТП. Построена авторегрессионная модель. Произведен прогноз значений на три года.

The main reasons for the large number of accidents on the roads of Russia are discussed. relevance of reducing the number of accidents is justified. The statistics of accidents in the Perm region over the past 10 years is reviewed. The factors affecting the number of accidents is analyzed. The autoregressive model is constructed. The forecast of values for three years is made.

Ключевые слова: Дорожно-транспортное происшествие (ДТП), правила дорожного движения (ПДД), транспортное средство (ТС), моделирование, линейно-многофакторные модели (ЛММ), авторегрессионные модели (АвРМ), прогнозирование.

Keywords: Road traffic accident (DTP), rules of the road (traffic rules) vehicle (TS), modeling, linear-multifactor models (LMM), autoregressive models (AvRM), forecasting.

Введение

В 1885 году Карл Бенц закончил первый в мире автомобиль с двигателем внутреннего сгорания, который получил название «*Motorwagen*». С того дня прошло больше 100 лет, на сегодняшний день автомобиль является неотъемлемой частью нашей жизни. Однако несмотря на все свои плюсы,

автомобиль является опасным для человека: загрязнение окружающей среды, шум, но самая главная проблема — это дорожно-транспортные происшествия (далее ДТП). Согласно своду правил дорожного движения [1], ДТП трактуется как событие, возникшее в процессе движения по дороге транспортного средства и с его участием, при котором погибли или ранены люди, повреждены транспортные средства, сооружения, грузы либо причинен иной материальный ущерб. По статистике [2] в Пермском крае ежегодно происходит около 3500 ДТП, в которых погибают в среднем около 400 человек и около 4200 человек получают травмы или ранения. По данным Росстата [3], количество ДТП в Пермском крае ежегодно растет.

Большое количество автотранспорта на дорогах, нарушение скоростных режимов и банальное незнание/игнорирование правил дорожного движения как водителями, так и пешеходами, – все это приводит к тому, что дороги становятся опасным местом, на которых создаются конфликтные ситуации, которые, в большинстве случаев, приводят к ДТП.

Целью данной исследовательской работы является исследование влияния социально-экономических факторов на количество ДТП в Пермском крае, а также поиск возможных путей, направленных на снижения количества аварий.

Методика

Одним из методов исследования социально-экономических систем является построение математических моделей [**Ошибка! Источник ссылки не найден.**]. В основном, при моделировании систем используются:

- линейно-многочленные модели (ЛММ) – позволяет дать количественную оценку влияния набора различных неуправляемых факторов на некоторое выходное значение [4].
- модели в пространстве состояний (МПрС) – применяются для описания многосвязных систем. Выходное значение такой системы рассчитывается по конечному набору входных факторов, которые называются состоянием [5].
- трендовые модели (ТрМ) – основная цель данных моделей показать общую тенденцию развития модели. Трендовые модели делятся на подтверждающие тренд и предупреждающие его смену [6].
- авторегрессионные модели (АВРМ) – широко используются для описания стационарных случайных процессов. В этих моделях текущее значение зависит от предыдущих значений этого же ряда [**Ошибка! Источник ссылки не найден.**].
- Наиболее подходящими для задач прогнозирования считаются факторные модели, потому что помимо приближения данные модели позволяют исследовать возможное влияние на объект путем изменения входных факторов.

Прогнозирование по модели предполагает следующие действия:

1. Определение факторов и их значение; выбор значений реакции системы;
2. Нормирование полученных данных и проведение корреляционного анализа;
3. Выбор подходящей модели и ее построение;

4. Анализ прогнозных свойств модели;
5. Прогнозирование и проверка результатов прогнозирования.

Прогнозирование ДТП

Для построения прогнозной модели в качестве критерия выберем y – количество ДТП в Пермском крае (шт.). В качестве входных факторов были выбраны: x_1 – количество дорожных знаков и светофоров(шт.), x_2 – количество пешеходов (чел.), x_3 – количество пьяных водителей (млн. чел.), x_4 – количество транспортных средств(шт.), x_5 – количество женщин за рулем (млн. чел.), x_6 – количество мужчин за рулем (млн. чел.). Статические данные (Таблица 1) были взяты с сайта Росстата [7], где находятся в открытом доступе.

Таблица 1 – Исходные данные

Год	2010	2011	2012	2013	2014	2015	2016	2017	2018
y	2459	2600	2520	2700	2980	3126	3560	3500	3721
x_1	11845	13611	19567	17017	15593	12873	18962	21569	22359
x_2	2635275	2633550	2631073	2634461	2636154	2637032	2634409	2632097	2623122
x_3	11845	12252	12843	13600	16517	15222	15632	16431	15892
x_4	81781	87149	89884	91275	107432	132319	239943	242106	247449
x_5	27260	29015	29994	30425	35810	44100	79981	80702	82483
x_6	54620	58100	59990	69000	71630	89230	160000	161500	165000

Проведем нормирование (

Таблица) вышеуказанных данных для исключения влияния размерности по формуле:

$$y_{\text{норм.}} = \frac{y(t) - \min_t(y(t))}{\max_t(y(t)) - \min_t(y(t))},$$

где $\min_t y(t)$ – минимум среди набора значений реакции системы, а $\max_t y(t)$ – максимум из набора. Аналогичным образом нормируются факторы.

Проведем корреляционный анализ для исключения факторов с высокой взаимной корреляцией, а также факторов с низкой корреляцией с реакцией системы из ЛММ:

$$r_{xy} = \frac{\sum((x_i - \bar{x}) \cdot (y_i - \bar{y}))}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}},$$

где x_i – значения переменной x , y_i – значения переменной y , \bar{y} – среднее значение критерия y , \bar{x} – среднее значение фактора x , рассчитывающееся по формуле:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k,$$

Анализ корреляции факторов показал, что из числа выбранных можно исключить фактор «Количество женщин за рулем» (x_6) имеющий высокую корреляцию с x_4 (Количество ТС) и «Количество мужчин за рулем» (x_5).

Таблица 2 – Нормированные данные

	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	-0,7389	0,527	0,7347	0,735	0,7427
x_2		1	-0,1636	-0,5256	-0,5227	-0,5206
x_3			1	0,7249	0,7249	0,7266
x_4				1	1	0,9986
x_5					1	0,9986
y	0,6525	-0,454	0,8499	0,9593	0,9593	0,9601

Построим линейную многофакторную модель (ЛММ) количества случившихся ДТП в Пермском крае вида:

$$y(\vec{x}) = a_0 + \sum_{i=1}^n b_i * x_i(t),$$

где a_0 – независимый коэффициент, b_i – коэффициент влияния $x_i(t)$ на $y(t)$, $x_i(t)$ – значение фактора.

Поиск коэффициентов производился методом наименьших квадратов:

$$S = \sum_{t=1}^n (y_m - y_э)^2 \rightarrow \min,$$

где y_m – значение реакции системы, $y_э$ – экспериментальное значение.

В результате получили коэффициенты линейной многофакторной модели: $a_0 = 0,2554$, $a_1 = -0,2556$, $a_2 = -0,2228$, $a_3 = 0,3895$, $a_4 = 0,6489$. Квадратичная погрешность аппроксимации ЛММ составила $S = 0,0153$.

Согласно полученным коэффициентам, фактор x_4 является наиболее значащим, т.к. количество транспортных средств напрямую влияет на количество ДТП. Однако значение коэффициента a_2 у фактора x_2 получилось неверным: при увеличении количества пешеходов, количество ДТП снижается. Довольно часто ДТП происходят по вине или с участием пешеходов, не соблюдающих ПДД, но цифры говорят об обратном.

Результаты построения ЛММ оказались неоднозначными, поэтому стоит проверить другие распространенные модели на возможность прогнозирования.

Построим трендовую модель (Рисунок) по формуле:

$$y(t) = a + b * t,$$

где a – свободный коэффициент, b – коэффициент ежегодного изменения реакции системы.

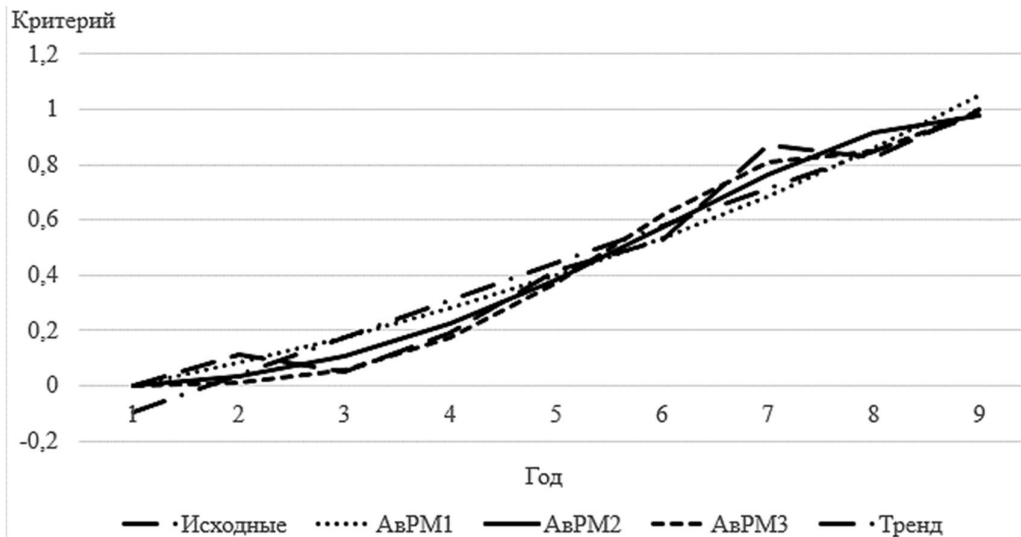


Рисунок 1 – Исходные данные, линия тренда и АвРМ

С помощью мастера поиска решений *MS Excel*, были получены следующие коэффициенты: $a = -0,2338$, $b = 0,1354$. По графику видно, что линия тренда неплохо аппроксимирует наши данные, но данный вид модели не позволяет исследовать управление.

Построим авторегрессионные модели 1-го, 2-го и 3-го порядка (Рисунок) по следующей формуле:

$$y(t_i) = a + \sum_j^N a_j y(t_{i-j}),$$

где, a – независимый коэффициент, a_j – коэффициенты влияния $i-j$ -го расчетного значения критерия системы, N – порядок.

Проведя расчеты с помощью мастера поиска решений *MS Excel*, были получены следующие коэффициенты:

- 1-го порядка $a_0 = 0,0828$, $a_1 = 1,1288$ и квадратичную погрешность $S = 0,0638$ (Линия АвРМ1 Рисунок);
- 2-го порядка $a_0 = 0,032$, $a_1 = 2,2793$, $a_2 = -1,4961$ и квадратичную погрешность $S=0,03401$ (Линия АвРМ2 Рисунок);
- 3-го порядка $a_0 = 0,0109$, $a_1 = 4,2407$, $a_2 = -7,475$, $a_3 = 5,5145$ и квадратичную погрешность $S = 0,0255$ (Линия АвРМ3 Рисунок);

По графику видно (Рисунок), что АвРМ 2-го и 3-го порядка неплохо аппроксимирует нашу модель. Погрешности приближения небольшие, однако данные модели не объясняют влияние внешних факторов на модель, поэтому они подходят для данной работы.

Построим модель в пространстве состояний по следующим уравнениям:

$$\begin{aligned} x'(n+1) &= a + b \cdot x'(n) \\ y(n) &= c + d \cdot x(n), \end{aligned}$$

где x' – вектор состояния, a – вектор функции перехода, b – матрица перехода, c и d – векторы функции выхода.

После проведения расчетов была получена квадратичная аппроксимация $S = 0,0701$. Значение получилось немного больше, чем у ЛММ. Чтобы окончательно определиться с выбором модели, необходимо сравнить

прогнозные свойства моделей. Одним из способов проверки моделей является метод постпрогноза, суть которого заключается в расчете реакции системы на несколько предыдущих лет. Последующее сравнение полученных данных с реальными значениями тех лет позволит количественно оценить прогнозные свойства рассчитанной модели.

Для ЛММ (линейной многофакторной модели) и МПС (модель в пространстве состояний), произведены расчеты постпрогнозов на 1, 2 и 3 года (Рис.2 ,3).

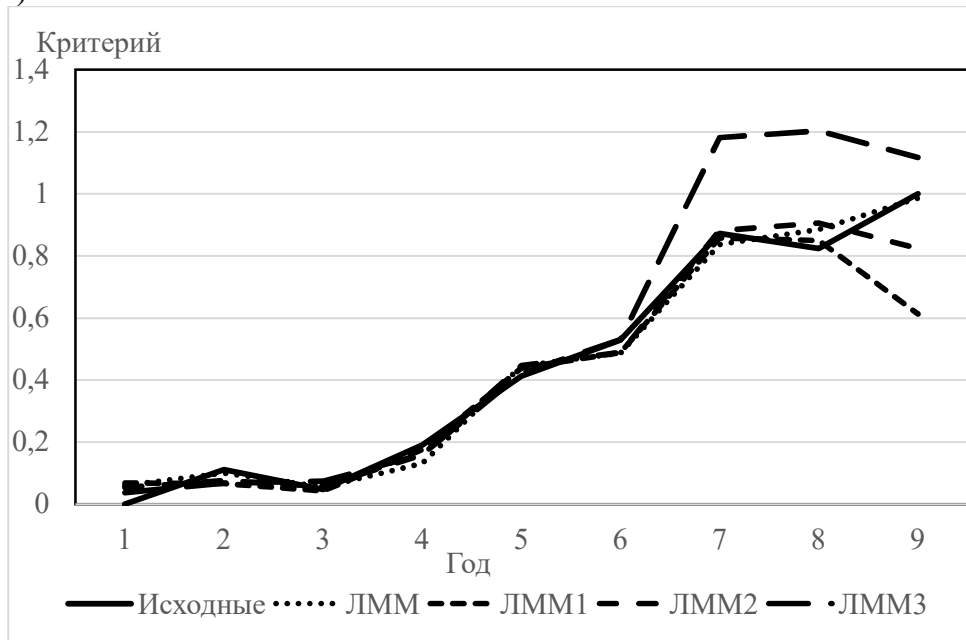


Рисунок 2 – Постпрогноз ЛММ

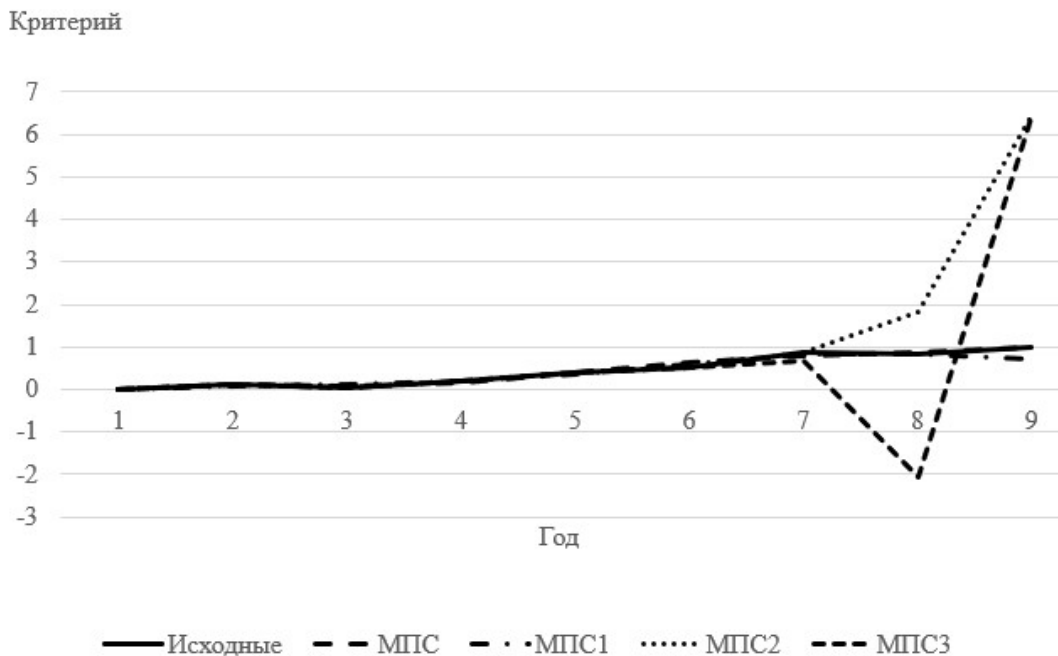


Рисунок 3 – Постпрогноз МПС

Для сравнения в таблицу (Таблица) были занесены абсолютные погрешности метода в зависимости от интервала времени:

Таблица 3 – Абсолютные погрешности моделей

	1 год	2 года	3 года
ЛММ	-0,38631	-0,17557	0,116956
МПС	-0,30487	5,42645615	5,38652598

Стоит отметить, что погрешность постпрогноза на длительное время у МПС в разы выше, чем у ЛММ. Следовательно, значения, спрогнозированные с помощью модели в пространстве состояний, будут в меньшей степени соответствовать реальным данным, в отличие от линейной многофакторной модели. Однако ЛММ также имеет большую погрешность во всех трех случаях и ни один график не показывает адекватной тенденции: постпрогноз на 2 и 3 года показывает падение количества ДТП, тогда как постпрогноз на 1 год – резкий рост количества ДТП.

Трудно сказать, что привело к такому результату, формулы, которые применялись при расчете, правильные, а сами расчеты были выполнены на компьютере. Вывод: наша модель не зависит от выбранных факторов.

Вернемся к авторегрессионным моделям. Рассмотрим прогнозные свойства авторегрессионной модели 1-го порядка с помощью постпрогноза.

Критерий

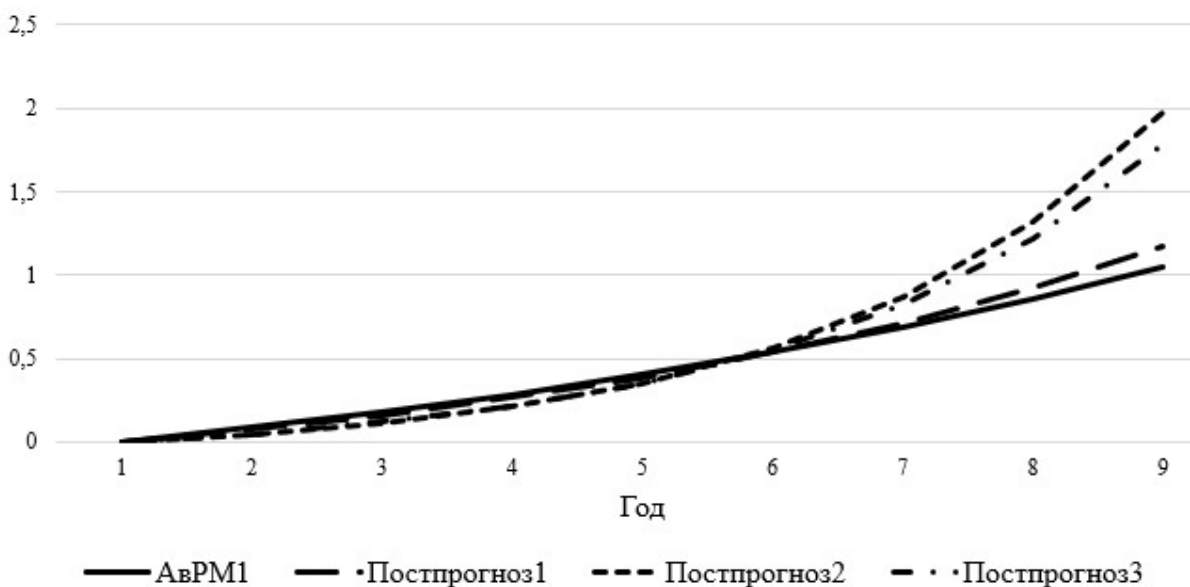


Рис. 1 Постпрогноз АвРМ

Погрешности постпрогноза были занесены в таблицу:

Таблица 4 – Погрешности постпрогноза АвРМ 1-го порядка

	1 год	2 года	3 года
АвРМ	0,0552	0,0136	0,0133

Значения погрешности для каждого года получились маленькими, графики сохраняют общую тенденцию роста, поэтому, для дальнейшего исследования, выберем именно эту модель.

Вывод

Спрогнозируем реакцию системы на следующие три года:

Таблица 5 – Спрогнозированные значения и их рост по сравнению с предыдущим годом

	2018	2019	2020	2021
Урасч.	1	1,269766835	1,516072931	1,794095
Рост, %	0	26,97668354	19,39774205	18,33834

Согласно спрогнозированным данным, количество ДТП на дорогах Пермского края будет расти с каждым годом примерно на 21%.

Список литературы

1. ПДД России 2019. Статистическая информация [Электронный ресурс] – Режим доступа: <http://www.pdd24.com>
2. [Показатели состояния дорожного движения – Официальный сайт Госавтоинспекции РФ \[Электронный ресурс\]](http://stat.gibdd.ru) – режим доступа: <http://stat.gibdd.ru>
3. Федеральная служба государственной статистики [Электронный ресурс] – Режим доступа: <http://www.gks.ru/>
4. Затонский А.В. Теоретический подход к управлению социально-техническими системами // Программные продукты и системы. 2008. № 1. С. 29-32.
5. Янченко Т.В., Затонский А.В. Определение оптимальной ранжировки частных критериев оценки краевого социального ресурса // Экономика и менеджмент систем управления. 2013. Т. 10. № 4. С. 99-104.
6. Затонский А.В., Сиротина Н.А., Янченко Т.В. Об аппроксимации факторов дифференциальной модели социально-экономической системы // Современные исследования социальных проблем (электронный научный журнал). 2012. № 11. С. 6.
7. Затонский А.В., Сиротина Н.А. Преимущества дифференциальной модели сложной экономической системы // Образование. Наука. Научные кадры. 2012. № 8. С. 98-102.
8. Иванова Е.В., Затонский А.В. Оценка и моделирование научно-исследовательской работы студентов как многоагентной системы // Современные наукоемкие технологии. 2009. № 7. С. 75-78.
9. Статистика ДТП в Пермском крае: [Электронный ресурс] – Режим доступа: <https://elibrary.ru/item.asp?id=32254694>

List of references

1. Traffic regulations of Russia 2019. Statistical information [Electronic resource] - access Mode: <http://www.pdd24.com>
2. Traffic indicators – Official website of the state traffic Inspectorate of the Russian Federation [Electronic resource] - access mode: <http://stat.gibdd.ru>
3. Federal state statistics service [Electronic resource] - access Mode: <http://www.gks.ru/>
4. Zatonkiy A. V. Theoretical approach to the management of socio-technical systems // Software products and systems. 2008. No. 1. Pp. 29-32.
5. Yanchenko T. V., Zatonkiy A.V. Determining the optimal ranking of private criteria for evaluating the regional social resource // Economics and management of management systems. 2013. Vol. 10. No. 4. Pp. 99-104.
6. Zatonkiy A.V., Sirotina N. A., Yanchenko T. V. on approximation of factors of the differential model of the socio-economic system // Modern studies of social problems (electronic scientific journal). 2012. No. 11. P. 6.
7. Zatonkiy A.V., Sirotina N. A. Advantages of a differential model of a complex economic system // Education. The science. Scientific personnel. 2012. No. 8. Pp. 98-102.
8. Ivanova E. V., Zatonkiy A.V. Assessment and modeling of students ' research work as a multi-agent system // Modern science-intensive technologies. 2009. No. 7. Pp. 75-78.
9. Accident statistics in the Perm region: [Electronic resource] - re-press access: <https://elibrary.ru/item.asp?id=32254694>