

Обоснование и выбор направлений исследований, исследование экзогенных факторов и их влияния на объемы спроса на грузовые железнодорожные перевозки

Руководитель: чл.-корр. РАН, проф. К. В. Рудаков

Вычислительный центр им. А. А. Дородницына
Российской академии наук

Соглашение номер 14.604.21.0041

Решается проблема повышения эффективности транспортировки грузов. Для решения задачи выполняется прогноз потребностей у заказчиков РЖД в узлах погрузки/разгрузки с учетом временных интервалов доставки.

Цель проекта: повышение качества прогнозирования загруженности железнодорожных узлов.

Требуется сформировать прогноз отправления/погрузки грузов в заданном периоде:

- | | |
|---------------------------|---------------------------|
| 1) на месяц посуточно; | 4) на год ежемесячно; |
| 2) на месяц подекадно; | 5) на год поквартально; |
| 3) на квартал ежемесячно; | 6) на период больше года; |

и прогноз отправления/погрузки грузов с разложением:

- 1) по группам грузов;
- 2) по родам подвижного состава;
- 3) по комбинированному разложению, учитывающему перечисленные варианты.

Описание данных для блока вагонов:

Дата погрузки	Станция отправления	Станция назначения	Кол-во вагонов	Код груза	Род вагона	Суммарный вес груза, т.
2007-01-01	020108	932902	1	1	216	56
2007-01-01	032105	840109	1	19	040	63
2007-01-01	035508	843408	2	3	070	120

- код груза: нефть и нефтепродукты, сахар, продукты перемола и т.д.,
- род вагонов: полувагоны, крытые вагоны, цистерны, платформы, прочие.

- выполнение аналитического обзора научно-технической, нормативной, методической литературы,
- обоснование и выбор направлений дальнейших исследований,
- исследование экзогенных факторов и их влияния на объемы спроса на грузовые железнодорожные перевозки.

В частности, были поставлены следующие задачи:

1. Выбрать базовую модель прогнозирования.
2. Рассмотреть проблему обнаружения причинно-следственных связей в разнородных временных рядах и предложить способ ее решения на основе современных методов или их развития.
3. Провести исследования предметной области, включая основные процессы в области организации грузовых железнодорожных перевозок и возможные источники данных объемов спроса на грузовые железнодорожные перевозки и экзогенных факторов в историческом и текущем периоде.

1.1 Выполнение аналитического обзора

современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР

Cortez P., et al. *Evolving time series forecasting ARMA models* // Journal of Heuristics, 2004. Vol. 10(4). pp. 419–429

Shi X., Ren Q. *The Advanced Nonparametric Model for Short-Term Traffic Volume Forecasting* // Proceedings of the 11th International Conference of Chinese Transportation Professionals, China, August 2011. Pp. 1442–1453.

Arroyo J., Maté C. *Forecasting histogram time series with k-nearest neighbors methods* // International J. of Forecasting, 2009. Vol. 25(1), pp. 192–207.

Y. Chen, S.L. Bressler, M. Ding. *Frequency decomposition of conditional Granger causality and application to multivariate neural field potential data* // Journal of Neuroscience Methods 150:228–237 (2006).

R. B. Kline, *Principles and Practice of Structural Equation Modeling*. New York: Guilford. 2005.

G. Sugihara et al. *Detecting Causality in Complex Ecosystems* // Science, 338-6106(2012), 496–500.

1. 2 Обоснование и выбор направлений дальнейших исследований

с целью разработки математической модели прогнозирования объемов спроса на грузовые железнодорожные перевозки, учитывающей влияние экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки.

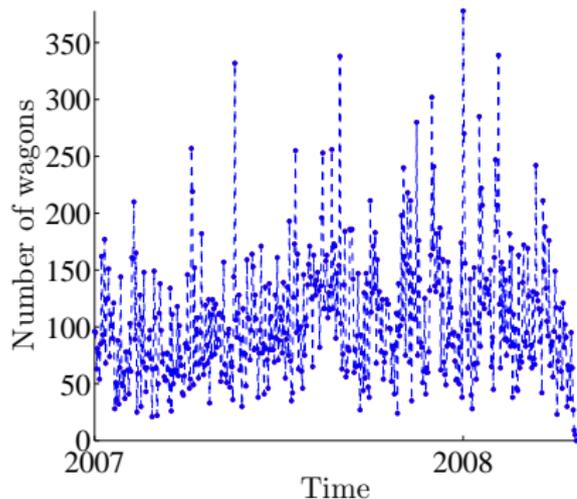
1. Критерий точности.
2. Критерий согласованности.
3. Критерий учета экзогенных факторов.
4. Физические ограничения, связанные со спецификой задачи.

Название алгоритма	Сумма баллов	Среднее время работы на ряде
SSA	198	менее секунды
Model Selection	92	менее секунды
Local Forecasting	–	менее секунды
ARMA	462	менее секунды
Subseries Producing Superposition	295	35 секунд
Kernel Smoothing	–	120 секунд
Exponential Smoothing	–	менее секунды

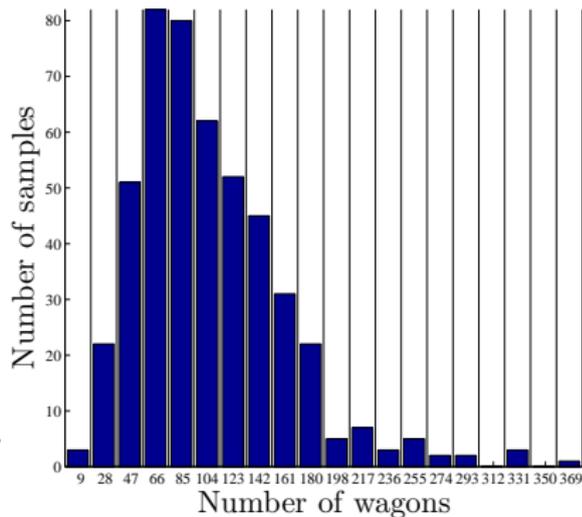
Схема системы прогнозирования



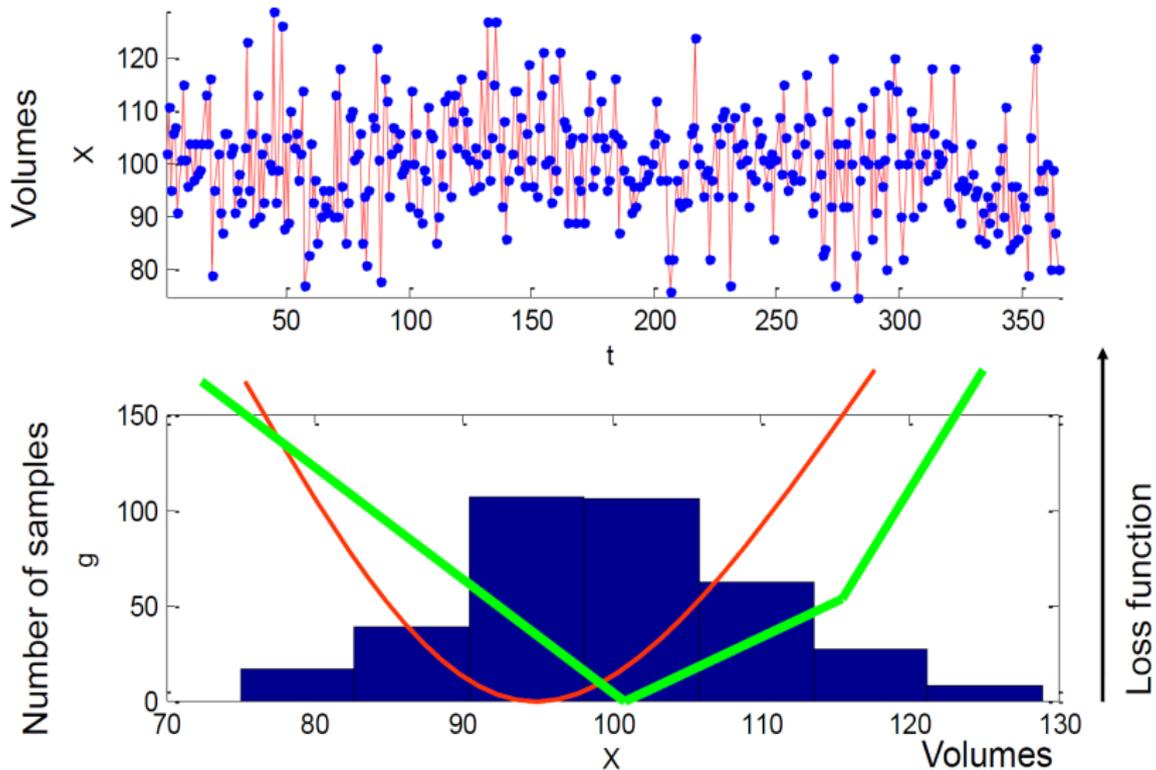
Пример распределения данных: прибытие вагонов с нефтью и нефтепродуктами



(a) Временной ряд



(b) Гистограмма



Определим прогноз \hat{x} случайной величины из распределения p_x

$$\hat{x} = \operatorname{argmin}_{x \in \mathbb{R}} \int L(x, y) p_x(y) dy.$$

для некоторой функции потерь $L(x, y)$, например:

- $L(x, y) = (y - x)^2$,
- $L(x, y) = |y - x|$.

По временному ряду \mathbf{x} построим гистограмму

$$\mathcal{H} = \{(y_k, g_k)\}_{k=1}^K,$$

K — число интервалов $[y_k^{\min}, y_k^{\max}]$ со средним значением y_k ,
 g_k — высота столбца гистограммы на интервале y_k .

Тогда прогнозируемое значение x_{T+h} ряда \mathbf{x} находится как

$$\hat{x} = \operatorname{argmin}_{x \in \{y_1, \dots, y_K\}} \sum_{k=1}^K g_k L(x, y_k) \in \{y_1, \dots, y_K\}.$$

Прогнозирование выполняется непараметрическим методом. Используется функция ошибки, учитывающая особенности каждой станции. При построении прогноза учитывается сезонность и внешние факторы.

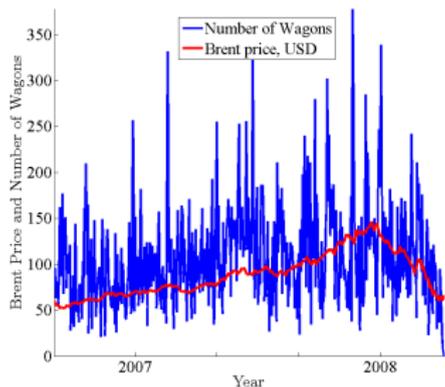
Код станции, усл.	ARMA, неделя	HIST, неделя	ARMA, месяц	HIST, месяц
1	2.76	1.07	2.41	1.02
2	5.02	1.02	4.76	0.93
3	2.64	1.24	NaN	1.06
4	4.02	1.93	3.18	1.57
5	2.20	1.56	1.94	1.23

Результаты прогнозирования объемов погрузок по фиксированной станции на неделю и месяц. В таблице указано отклонение от истинного значения в процентах. Сравниваются базовый алгоритм ARMA и предлагаемый алгоритм HIST.

1.3 Влияние экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки

Задача повышения точности прогноза решается путем введения в прогностическую модель следующих факторов:

- сезонности и данных о погоде (для учета перевозок сельскохозяйственной продукции),
- биржевых цен на перевозимые грузы на российских и зарубежных биржах.



На графике синей линией показано суммарное количество вагонов на фиксированной станции, красной линией — цена за баррель нефти. Цена за баррель является фактором, используемым для прогнозирования объема поставок.

1.3.1 Экспертный анализ

значимости и характера влияния экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки

Выделен набор экзогенных факторов, которые с точки зрения экспертов оказывают влияние на объем грузоперевозок:

Вид фактора, влияющего на объем грузоперевозок	Группы грузов и отрасли, на которые оказывается влияние	Степень и характер влияния
Мировые и внутренние цены на соответствующие товары.	Нефть и нефтепродукты, черные металлы, цветные металлы, удобрения, уголь.	На экспортные перевозки влияние сильное. Связь бывает как прямой, так и обратной.
Курс рубля к доллару.	Грузы, отправляемые на экспорт (нефть и нефтепродукты, металлы, уголь).	Степень влияния для экспортных перевозок зачастую высокая.
Сезонность производства природно-климатического характера.	Зерно, овощи, бахчевые культуры.	Степень влияния высокая. Динамика перевозки связана со сбором урожая.
...

1.3.2 Выявление причинно-следственных связей в разнородных временных рядах объемов спроса на грузовые железнодорожные перевозки и экзогенных факторов

Решение о достоверности экспертного высказывания о зависимости объемов погрузок от внешних факторов принимается на основе теста Грейнджера и анализа перекрестных отображений исследуемых временных рядов.

Исследуемые ряды	Высказывание эксперта	Оценка достоверности	Решение о наличии связи
«Нефть и нефтепродукты» — «Цены на нефть»	Связь сильная	0.99	Положительное
«Сахар» — «Цены на сахар»	Связь сильная	0.97	Положительное
«Продукты перемола» — «Цены на пшеницу»	Связь сильная	0.46	Отрицательное

1.3.3 Разработка методов оценки достоверности

экспертных высказываний о влиянии экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки

Экспертным высказываниям о влиянии внешних факторов $S = \{s_1, \dots, s_M\}$ на прогнозируемые ряды x присваивается оценка достоверности $\mu(x, s)$.

Тест Грейнджера: решение о наличии связи принимаем при больших значениях достоверности

$$\mu(x, s) \sim \frac{ESS_x - ESS_{x,s}}{ESS_{x,s}} \frac{T - k_{x,s}}{k_{x,s} - k_x}, \quad ESS = \sum_{t=1}^T (\hat{x}_t - x_t)^2,$$

$k_{x,s}, k_x$ — порядок расширенной $\hat{x} = \hat{x}(x, s)$ и базовой $\hat{x} = \hat{x}(x)$ моделей прогнозирования.

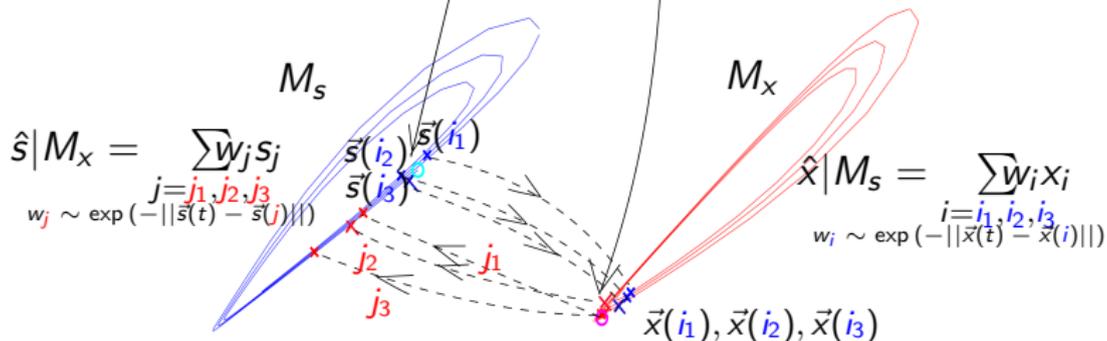
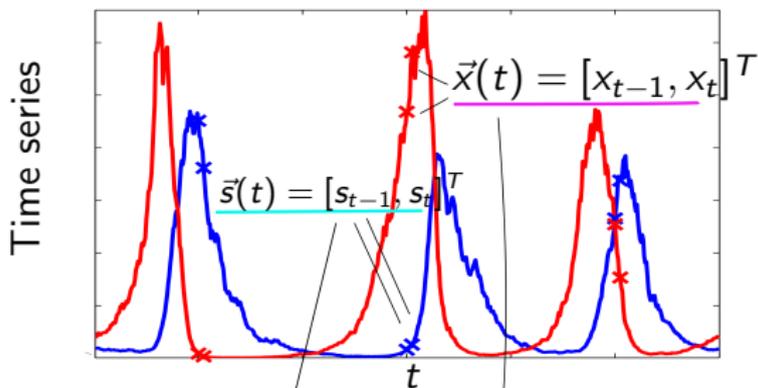
Метод ССМ: решение о наличии связи принимаем при больших значениях достоверности

$$\mu(x, s) \sim \rho(\hat{x}, x_{T+1}) = \frac{1}{\sigma_x \sigma_{\hat{x}}} E(\hat{x} - E\hat{x})(x_{T+1} - E x_{T+1}),$$

где

$$M_s = \{\vec{s} \mid \vec{s}(t) = [s_{(t-D)}, \dots, s_t]^T\},$$
$$\hat{x} \mid M_s = \sum_{d=1}^{D+1} w_d x_{i_d}, \quad w_d = \frac{u_d}{\sum_d u_d}, \quad u_d = \exp\left(-\frac{\|\vec{x}(T+1), \vec{x}(i_d)\|}{\|\vec{x}(T+1), \vec{x}(i_1)\|}\right).$$

1.3.5 Построение кросс-корреляционных зависимостей временных рядов объемов спроса на грузовые железнодорожные перевозки и экзогенных факторов при прогнозе объема спроса на грузовые железнодорожные перевозки на отдельном железнодорожном узле



1.3.4 Расширение подхода Грейнджера

для обнаружения структуры связей между временными рядами экзогенных факторов и объемами спроса на грузовые железнодорожные перевозки

Обнаруженные для временных рядов x зависимости

$$x(t+1) = f(x(T), \dots, x(t-lag), s),$$

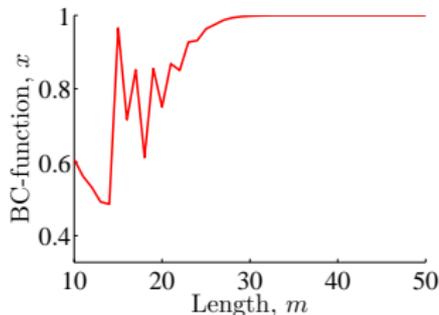
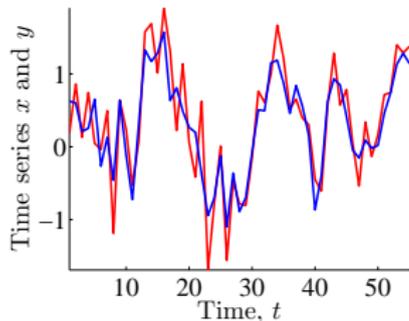
распространялись на близкие к ним временные ряды x' :

$$x'(t+1) = f(x'(t), \dots, x'(t-lag), s), \text{ при } BC(x, x') < \epsilon,$$

где

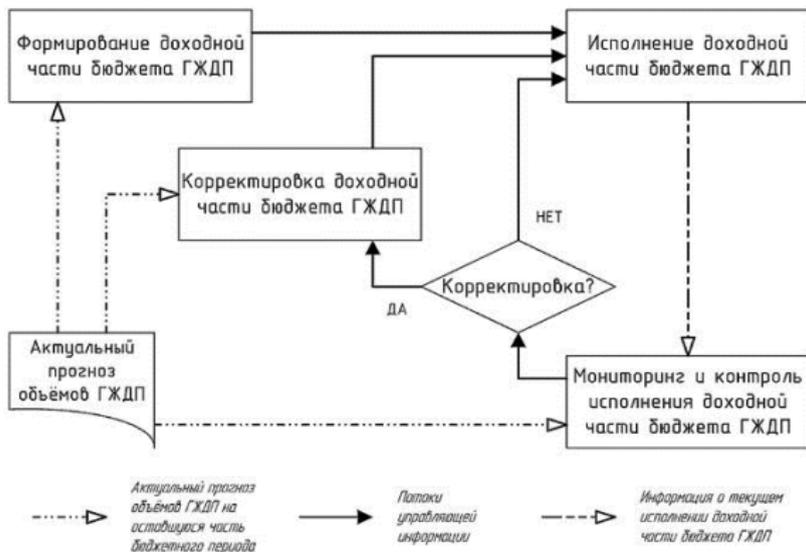
$$BC(x, x') = \frac{2}{1 + \exp(kCor(x, x'))} d(x, x'),$$

$$d(x, x') = \frac{1}{m} \sqrt{\sum_{t=1}^m (x(t) - x'(t))^2}.$$



1.4 Исследования предметной области

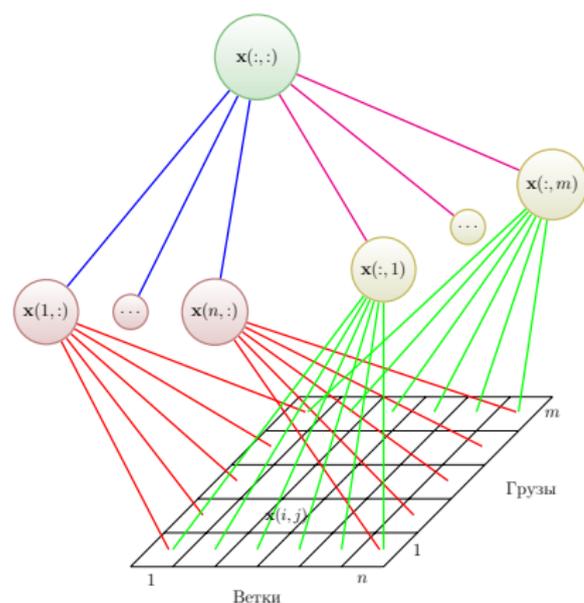
- Проанализированы процессы ИП
- Выявлены типы подвижного состава
- Проанализированы источники и структура данных



- Выявлен высокий интерес к теме исследования
- В основном иностранные работы
- Активная работа по защите результатов исследований по теме

1. Выполнен аналитический обзор современной научно-технической, нормативной, методической литературы, затрагивающей научно-техническую проблему, исследуемую в рамках НИР.
2. Проведен обоснованный выбор направлений дальнейших исследований с целью разработки математической модели прогнозирования объемов спроса на грузовые железнодорожные перевозки, учитывающей влияние экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки.
3. Выявлены и исследованы экзогенные факторы, влияющие на объемы спроса на грузовые железнодорожные перевозки с анализом значимости и характера их влияния. Разработаны методы оценки достоверности экспертных высказываний о влиянии экзогенных факторов на объемы спроса на грузовые железнодорожные перевозки.
4. Проведены исследования предметной области, описаны основные процессы в области организации грузовых железнодорожных перевозок, должны быть выявлены и исследованы возможные источники данных объемов спроса на грузовые железнодорожные перевозки и экзогенных факторов в историческом и текущем периоде.
5. Выполнены патентные исследования в соответствии с ГОСТ 15.011-96.

Прогнозируемые временные ряды могут разбиваться на наборы временных рядов, определяющих детализацию (по веткам, типам грузов и т.д.)



Условие

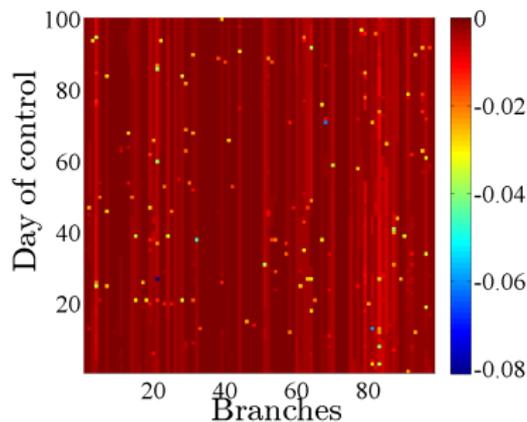
согласованности прогнозов
 $\hat{\chi} = [\hat{\chi}(:, 1), \hat{\chi}(1), \dots, \hat{\chi}(n)]^T :$

$$\hat{\chi}(:, j) = \sum_{j=1}^n \hat{\chi}(j).$$

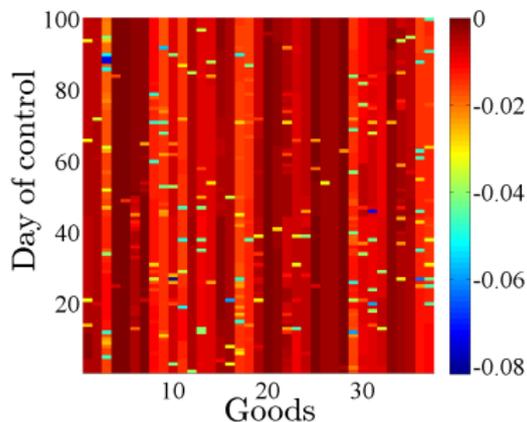
Необходимо построить согласованные прогнозы
 $\hat{\varphi} = [\hat{\varphi}(:, 1), \hat{\varphi}(1), \dots, \hat{\varphi}(n)]^T :$

$$\hat{\varphi} = \sum_{j=1}^n \hat{\chi}(j), \quad L(\hat{\chi}, \hat{\varphi}) \rightarrow \min .$$

Результат согласования прогнозов для 100 последних точек истории: суммарные потери во всех контрольных точках либо не увеличиваются, либо уменьшаются.



(f) Для каждой ветки



(g) Для каждого груза