

# Case 3. Forecasting volumes of Russian railways freight transportation

**Keep a hierarchical structure of time series without losing performance**

Forecast with hierarchical aggregation of

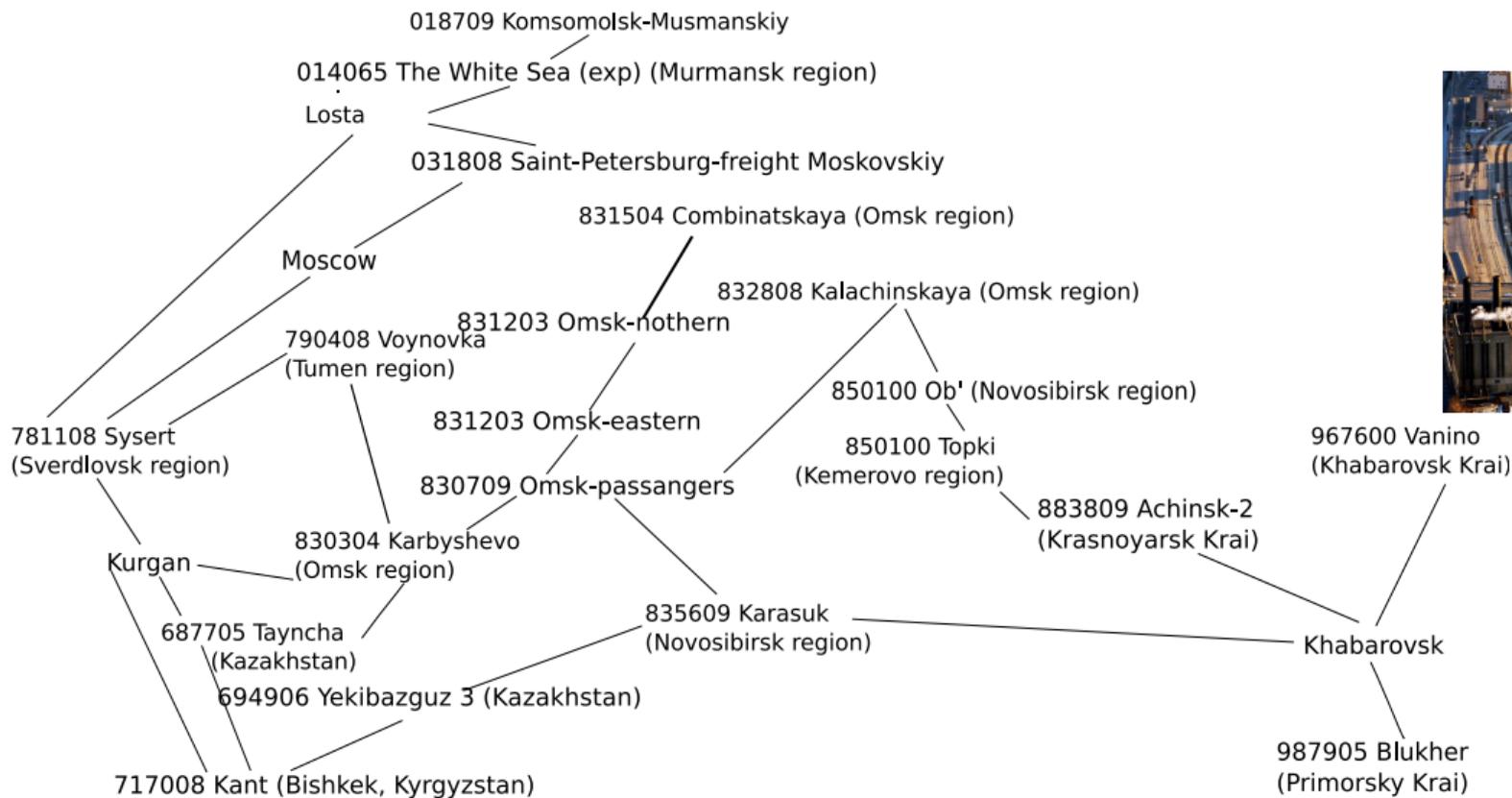
- ▶ types of freight in
- ▶ stations, regions, and roads,
- ▶ for a day, week, month, and quarter,
- ▶ counting all combinations above.

Satisfy the conditions:

- ▶ minimize error,
- ▶ incorporate important external factors,
- ▶ respect hierarchical structure,
- ▶ do not exceed physical bounds of forecast values.



# The railroad map counts $\sim 78$ regions, $\sim 4000$ stations, and $\sim 100$ rail-yards



## Задача

Построить прогнозы семейства временных рядов, связанных в иерархическую многоуровневую структуру и описывающих объемы погрузки ряда грузов в заданных узлах РЖД с разным уровнем детализации.

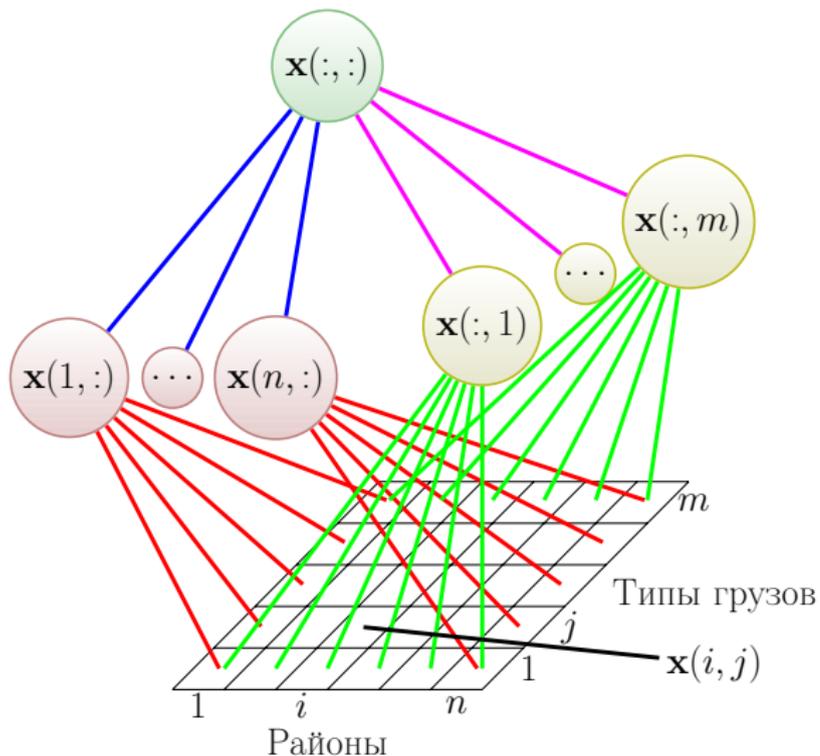
## Требования к модели

- прогнозы должны быть точны — обеспечивать минимально возможное значение заданной функции потерь;
- прогнозы должны удовлетворять физическим ограничениям — лежать в заданном интервале для каждого временного ряда;
- прогнозы должны удовлетворять условию согласованности (структуре иерархии).

## Проблема согласования прогнозов

Прогнозы, полученные для каждого временного ряда независимо, могут не удовлетворять структуре иерархии, т. е. не быть *согласованными*.

# Условие согласованности прогнозов



$$x_t(:, :) = \sum_{i=1}^n x_t(i, :);$$

$$x_t(:, :) = \sum_{j=1}^m x_t(:, j);$$

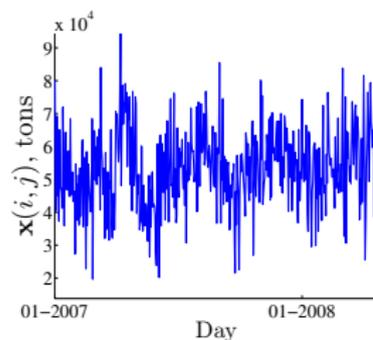
$$x_t(i, :) = \sum_{j=1}^m x_t(i, j),$$

$$i = 1, \dots, n;$$

$$x_t(:, j) = \sum_{i=1}^n x_t(i, j),$$

$$j = 1, \dots, m;$$

$$t = 1, \dots, T.$$



## Алгоритмы согласования прогнозов

- 1 Albert B. Schwarzkopf, Richard J. Tersine, John S. Morris *Top-down versus bottom-up forecasting strategies*. The International Journal Of Production Research, 26(11):1833–1843, 1988.
- 2 Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang. *Optimal combination forecasts for hierarchical time series*. Computational Statistics and Data Analysis, 55(9):2579–2589, 2011.

## Базовые публикации

- 1 Tim Van Erven and Jairo Cugliari. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. 2013.
- 2 М. М. Стенина, В. В. Стрижов. *Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования*. Системы и средства информатики, 24(2):21–34, 2014.

Срез иерархии, вектор независимых и вектор согласованных прогнозов:

$$\chi_t = \begin{pmatrix} x_t(:, :) \\ \dots \\ x_t(n, 1) \\ \dots \\ x_t(n, m) \end{pmatrix}, \quad \hat{\chi} = \begin{pmatrix} \hat{x}(:, :) \\ \dots \\ \hat{x}(n, 1) \\ \dots \\ \hat{x}(n, m) \end{pmatrix}, \quad \hat{\varphi} = \begin{pmatrix} \hat{y}(:, :) \\ \dots \\ \hat{y}(n, 1) \\ \dots \\ \hat{y}(n, m) \end{pmatrix}.$$

Условие согласованности  $\mathbf{S}\chi_t = \mathbf{0}$ ,  $t = 1, \dots, T$ ,

где  $\mathbf{S}$  — матрица связей, имеет размер  $(2 + n + m) \times (1 + n + m + nm)$  и записывается в виде

$$\mathbf{S} = \left( \begin{array}{cccc|cccc|cccc|c|cccc}
-1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
-1 & 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\hline
0 & -1 & \dots & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\
\dots & & \dots & & & \dots & & & & \dots & & & & & \dots & \\
0 & 0 & \dots & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \\
\hline
0 & 0 & \dots & 0 & -1 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\
\dots & & \dots & & & \dots & & & & \dots & & & & & \dots & \\
0 & 0 & \dots & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 1
\end{array} \right).$$

# Обозначения: ограничения и функция потерь

Множество векторов, удовлетворяющих условию согласованности

$$\mathcal{A} = \{\boldsymbol{\chi} \in \mathbb{R}^d \mid \mathbf{S}\boldsymbol{\chi} = \mathbf{0}\}.$$

$$\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_{T+1}, \hat{\boldsymbol{\varphi}} \in \mathcal{A}, \quad \hat{\boldsymbol{\chi}} \notin \mathcal{A}.$$

Прогнозы и исторические значения временных рядов удовлетворяют физическим ограничениям

$$\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_{T+1}, \hat{\boldsymbol{\chi}}, \hat{\boldsymbol{\varphi}} \in \mathcal{B},$$

$$\mathcal{B} = \{\boldsymbol{\chi} \in \mathbb{R}^d \mid \chi(i) \in [A_i, B_i], \text{ для всех } i = 1, \dots, d\},$$

$$A_i, B_i \in [-\infty; +\infty] \text{ для всех } i = 1, \dots, d.$$

Для задачи прогнозирования объемов погрузки

$$A_i = 0, \quad B_i = +\infty, \quad i = 1, \dots, d.$$

Качество прогнозов оценивается с помощью функции потерь, которая зависит от вектора прогнозов и среза иерархии в момент времени  $(T + 1)$

$$l_h(\boldsymbol{\chi}_{T+1}, \hat{\boldsymbol{\chi}}).$$

## Дано

Матрица связей  $\mathbf{S}$ , множества  $\mathcal{A}$ ,  $\mathcal{B}$  и вектор независимых прогнозов  $\hat{\chi}$

$$\hat{\chi} \notin \mathcal{A}, \quad \hat{\chi} \in \mathcal{B}.$$

## Требуется построить

вектор согласованных прогнозов  $\hat{\varphi}$ , который удовлетворяет следующим требованиям:

- $\hat{\varphi} \in \mathcal{A}$ ,  $\mathcal{A} = \{\chi \in \mathbb{R}^d \mid \mathbf{S}\chi = \mathbf{0}\}$  — согласованность;
- $\hat{\varphi} \in \mathcal{B}$  — физические ограничения;
- $l_h(\chi_{T+1}, \hat{\varphi}) \leq l_h(\chi_{T+1}, \hat{\chi})$  для любого среза действительных значений  $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$  — качество.

## Цель

Сравнить качество прогнозов, полученных предложенными алгоритмами согласования, с качеством независимых прогнозов и согласованных прогнозов, полученных при помощи существующих алгоритмов согласования, для различных типов иерархических структур.

## Данные

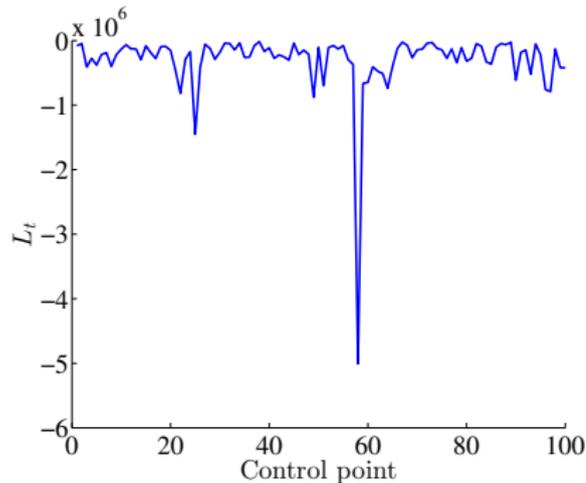
- **Трехуровневая иерархия:** данные о посуточной загруженности узлов РЖД. 37 типов грузов, 98 ЖД веток.
- **Двухуровневая иерархия:** данные о почасовом потреблении электроэнергии в 20 регионах Канады (Global Energy Forecasting Competition 2012).

Для согласования прогнозов  $H = 100$  последних точек истории решалась оптимизационная задача  $\hat{\varphi} = \arg \min_{\chi \in A \cap B} \|\chi - \hat{\chi}\|_2^2$ .

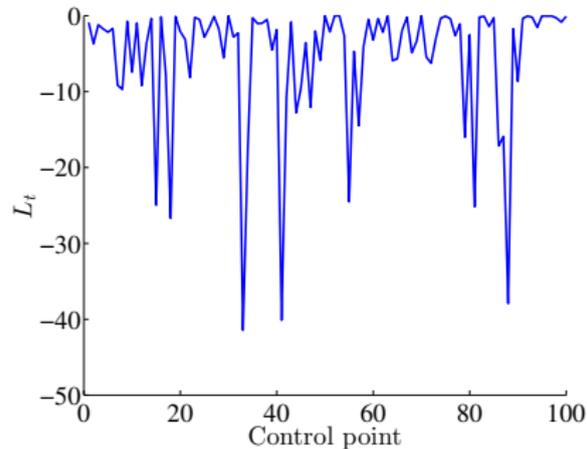
Изображена величина

$$L_t = \|\chi_t - \hat{\varphi}\|_2^2 - \|\chi_t - \hat{\chi}\|_2^2, \quad t = (T - H + 1), \dots, T.$$

Во всех контрольных точках потери уменьшились.



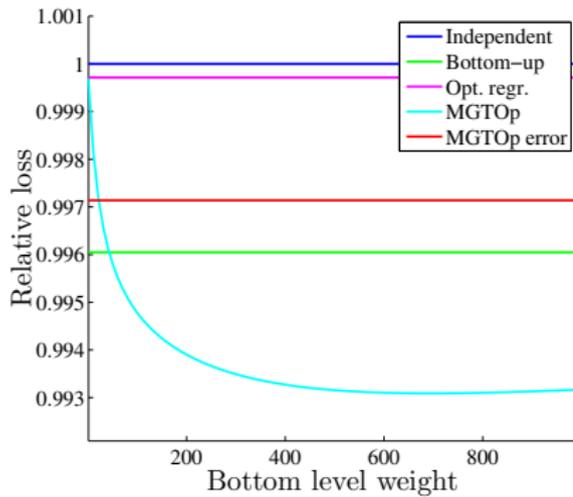
Для РЖД



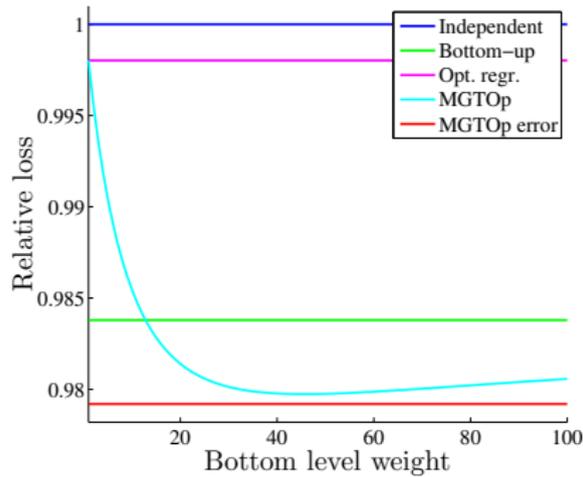
Для электроэнергии

$$\text{Relative loss}(\text{algorithm}) = \frac{\sum_{t=T-H+1}^T \|\chi_t - \hat{\varphi}_{\text{algorithm}}\|_2^2}{\sum_{t=T-H+1}^T \|\chi_t - \hat{\chi}\|_2^2}.$$

$$\hat{\varphi}_{\text{теор.игр.}} = \arg \min_{\chi \in \mathcal{A} \cap \mathcal{B}} l_r(\chi, \hat{\chi}), \quad l_r(\chi, \hat{\chi}) = \sum_{i=1}^d w_i (\chi(i) - \hat{\chi}(i))^2.$$



Для РЖД



Для электроэнергии

Функция потерь  $l_h(\chi_t, \hat{\chi}) = \|\chi_t - \hat{\chi}\|_2^2$ .

Средние потери прогнозирования отгрузки в узлах РЖД,  $\times 10^8$

Уровень иерархии	Независимые прогнозы	Восходящее согласование <sup>1</sup>	Оптимальная регрессия <sup>2</sup>	Модиф. теор.-игр. согл. (веса 700)
Вся иерархия	10.038	9.999	10.035	9.969
Верхний уровень	2.858	2.868	2.856	2.840
Средний уровень, ветки	2.549	2.486	2.545	2.487
Средний уровень, грузы	2.338	2.351	2.340	2.348
Нижний уровень	2.294	2.294	2.294	2.294

Средние потери прогнозирования потребления электроэнергии

Уровень иерархии	Независимые прогнозы	Восходящее согласование	Оптимальная регрессия	Модиф. теор.-игр. согл. (погрешности)
Вся иерархия	2727	2683	2722	2670
Верхний уровень	2083	2039	2076	2029
Нижний уровень	644	644	646	642

<sup>1</sup> Albert B. Schwarzkopf, Richard J. Tersine, John S. Morris *Top-down versus bottom-up forecasting strategies*. The International Journal Of Production Research, 26(11):1833—1843, 1988.

<sup>2</sup> Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang. *Optimal combination forecasts for hierarchical time series*. Computational Statistics and Data Analysis, 55(9):2579—2589, 2011.