# КРАТКОСРОЧНЫЙ ПРОГНОЗ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ С УЧЕТОМ ТЕМПЕРАТУРЫ ВОЗДУХА И КЛАСТЕРИЗАЦИЕЙ ПОТРЕБИТЕЛЕЙ

<u>Ф.Л.Быков</u>, В.А.Гордин ФГБУ «Гидрометцентр России», НИУ ВШЭ, факультет экономических наук

Москва, 2019г

#### Данные о часовом потреблении электроэнергии субъектами РФ

Данные AO «ATC» («Администратор Торговой Системы оптового рынка электроэнергии») о потреблении в 66 субъектах РФ [1] <a href="http://www.atsenergo.ru/results/market/fact\_region">http://www.atsenergo.ru/results/market/fact\_region</a>



3 субъекта: Амурская, Омская, Псковская области были отброшены, так как в них имеются плохо предсказуемые перетоки электроэнергии с соседними государствами.

Временное разрешение = 1 час. Данные с августа 2014 по июль 2016.

Итого N = 1105146 значений в  $N_r = 63$  субъектах (по ~17542 в каждом субъекте).

#### Метеорологические данные

Использовались данные с синоптических станций в административных центрах субъектов и метеорологические прогнозы для них. В дальнейшем возможно так же использование данных с остальных синоптических станций соответствующего региона.

Измерение температуры синоптическими станциями каждые 3 часа. Итого 367863 телеграмм (561 телеграмма отсутствует) от  $N_r$  = 63 станций (по ~5839 с каждой).

Архив метеорологических прогнозов по схеме комплексного прогноза [3] <a href="http://method.meteorf.ru/ansambl/ansambl.html">http://method.meteorf.ru/ansambl/ansambl.html</a> Разрешение прогнозов = 6 часов. Итого 183512 прогноза (700 прогнозов отсутствует).

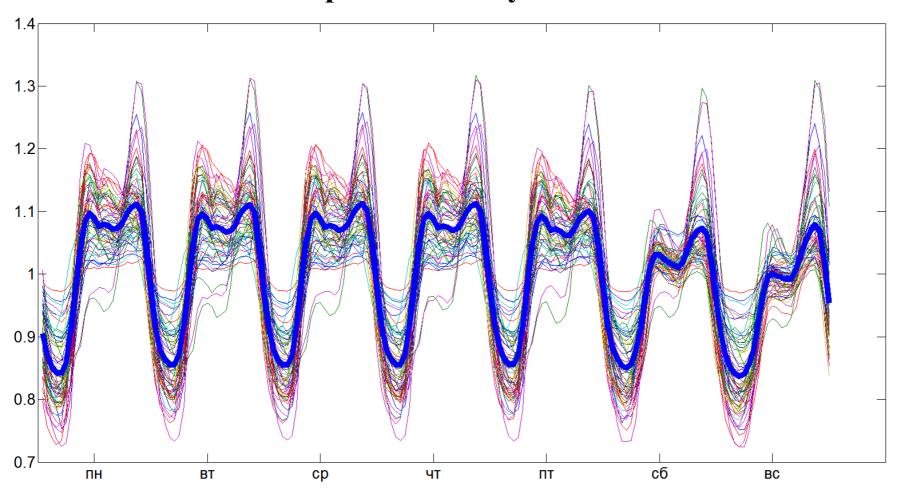
Метеорологические параметры мы кусочно-линейно интерполировали в те сроки, когда они отсутствовали.

Будем давать прогнозы потребления электроэнергии, используя соответствующие метеорологические прогнозы.

Метеорологические прогнозы от срока 00:00 ВСВ (03:00 мск) в Гидрометцентре готовы около 8:30 мск. Контракты на поставку электроэнергии на следующий день заключаются до 13:30 мск. Поэтому наибольший практический интерес представляют прогнозы часового потребления электроэнергии с заблаговременностью 11-37ч и соответствующие им метеорологические прогнозы с заблаговременностью 18-48ч.

Далее мы используем метеорологические прогнозы с заблаговременностью 24-48ч, когда говорим о прогнозе на 1-е сутки, и с заблаговременностью 48-72ч, когда говорим о прогнозе на 2-е сутки.

#### Зависимость потребления в субъектах РФ от часа недели



Синим — среднее по 63 субъектам РФ. Часовое потребление в субъектах меняется поразному в зависимости от дня недели и (местного) времени суток. Видно, что есть субъекты-жаворонки и субъекты-совы, динамика в выходные в разных субъектах сильно отличается. 1 — среднее часовое потребление в субъекте РФ

#### Метод главных компонент

Допустим, есть некоторый набор векторов в M-мерном пространстве:  ${}^x\!\!ri$ , где  $r\!=\!1,...,N$  — номер вектора,  $i\!=\!1,...,M$  — размерность. Рассмотрим собственные числа и вектора ковариационной матрицы  $A_{ij}\!=\!\frac{1}{N}\!\sum_r\!x_{ri}x_{rj}$ . Выберем подпространство,

натянутое на несколько единичных собственных векторов  $\vec{e}_k$ , k=1,...,K < M, соответствующих наибольшим собственным значениям  $d_k$  матрицы A. Тогда среднеквадратичная погрешность приближения исходных векторов системой из собственных векторов равна

$$\frac{1}{N} \sum_{r} \left\| \vec{x}_r - \sum_{k=1}^{K} (x_r, e_k) \vec{e}_k \right\|^2 = d_{K+1} + \dots + d_M,$$

причем именно такой выбор базисных векторов  $\vec{e}_k$  обеспечивает наименьшее возможное значение правой части.

Чем более доминируют наибольшие собственные числа над остальными, тем точнее наше приближение меньшим количеством векторов.

Далее будем использовать метод главных компонент для существенного сокращения количества констант, зависящих от субъекта. Однако при этом возникают константы, не зависящие от субъекта и описывающие вектора  $\vec{e}_k$ . Зато эти константы определяются по существенно большей статистике. Количество используемых коэффициентов сокращается с NM до K(N-1+M). У нас N - количество векторов, M - размерность, K - количество компонент, используемых для описания. Если N < M, то матрица A будет вырождена и будут существовать фиктивные соотношения между векторами  $\vec{x}_r$ , что сделает оценки собственных векторов  $\vec{e}_k$  неустойчивыми.

Подобное упрощение позволит в дальнейшем определять коэффициенты для новых субъектов по существенно меньшей статистике: вектора  $\vec{e}_k$  определять заново не нужно. Например, будет достаточно месячной статистики потребления какого-либо потребителя для определения всех коэффициентов, описывающих его потребление вместо годовой.

В неделе 168 часов — слишком много, чтобы оценивать зависимость потребления от часа недели  $t_w = 24t_d + t_h$  независимо для каждого часа. Будем ее описывать периодическими кубическими сплайнами с ежедневными узлами 4, 10, 12, 16, 18, 20 ч местного времени (это минимумы и максимумы суточного потребления). Всего 6\*7=42 узла. В этом 42-мерном подпространстве сплайнов 168-мерного пространства зависимостей от часа недели  $t_w$  выберем ортонормированный базис  $P_k$ , k=1,...,42:

$$\sum_{t_{w}=1}^{168} P_{k}\left(t_{w}\right) P_{q}\left(t_{w}\right) = \delta_{k}^{q}$$

Для каждого субъекта r найдем сплайн  $S_r$  на этой сетке с 42 узлами, обеспечивающий наилучшее среднеквадратическое приближение  $W_r(t) \approx S_r(t_W)$  и разложим его по базису  $P_k$ . Раз базис  $P_k$  ортонормированный, то при использовании разложений получим

$$S_r(t_w) = \sum_{k=1}^{42} s_{rk} P_k(t_w), \quad s_{rk} = (S_r, P_k),$$

В силу ортонормированности базиса расстояние в подпространстве сплайнов равно расстоянию в пространстве коэффициентов

$$\left\| S_r(t_w) - S_p(t_w) \right\|^2 = \sum_{t_w=1}^{168} \left( S_r(t_w) - S_p(t_w) \right)^2 = \sum_{k=1}^{42} \left( s_{rk} - s_{pk} \right)^2.$$

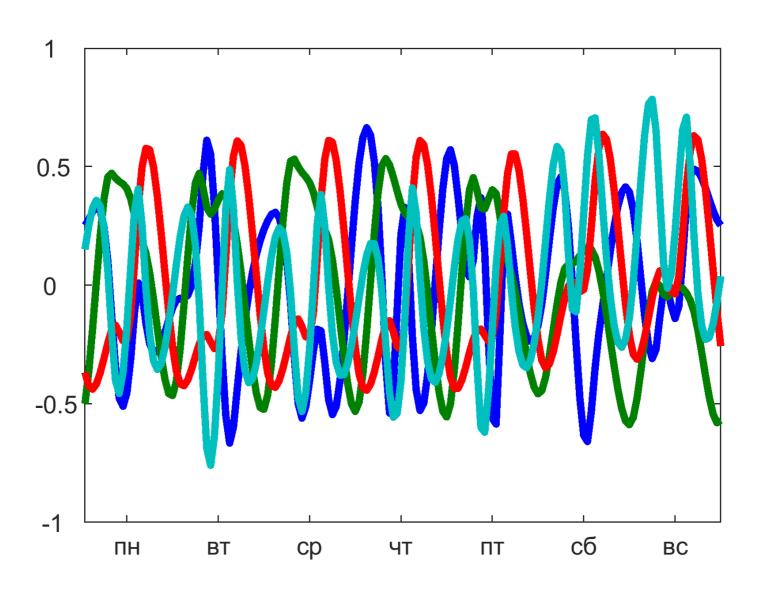
Применим метод главных компонент к коэффициентам разложения  $S_{rk}$ , получаем, что четыре компоненты  $H_k(t_w)$ , k=1,...,4 описывают  $c_W=99,5\%$  вариации сплайнов  $S_r(t_w)$ , то есть вместо 42 коэффициентов для каждого субъекта можно определять лишь 4 коэффициента, потеряв в среднем лишь  $1-c_W=0,5\%$  точности:

$$A_{r}(t_{w}) = \sum_{k=1}^{4} h_{rk} H_{k}(t_{w})$$

$$\sum_{r} \left\| S_{r}(t_{w}) - A_{r}(t_{w}) \right\|^{2} = (1 - c_{w}) \sum_{r} \left\| S_{r}(t_{w}) \right\|^{2}.$$

При этом количество коэффициентов, используемых для описания зависимости от дня недели и времени суток, сократится с  $42N_r = 2646$  до  $4\binom{N_r + 41}{r} = 416$ , где  $N_r = 63$  - количество субъектов.

# Зависимости базисных функций $H_k(t_w)$ от часа недели



#### Факторы, влияющие на потребление электроэнергии

Предсказываем потребление W(t), используя следующие предикторы:

- 1. Фактическое часовое потребление ровно неделю назад W(t-168)
- 2. Местный час недели  $t_{_{\mathcal{W}}}$
- 3. Температура воздуха T(t) и она же 12 часов назад T(t-12)

МНК оценку по этим 3 предикторам - потребление электроэнергии обозначим  $B_{r}(t)$ .

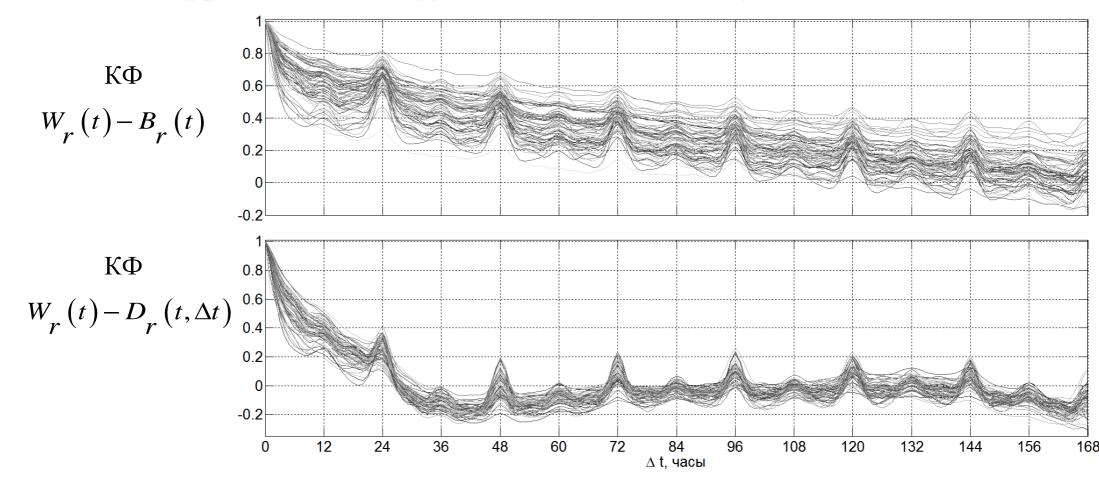
$$\begin{split} W_{r}(t) &\approx B_{r}(t) = \sum_{k=1}^{4} h_{rk} H_{k}(t_{w}) + \alpha_{r1} + \alpha_{r2} W_{r}(t-168) + \dots \\ &+ \alpha_{r3} T_{r}(t) + \alpha_{r4} \left[ T_{r}(t) \right]^{2} + \alpha_{r5} \left[ T_{r}(t) \right]^{3} + \alpha_{r6} T_{r}(t-12) + \alpha_{r7} \left[ T_{r}(t-12) \right]^{2} \end{split}$$

Число коэффициентов, зависящих от субъекта можно при помощи метода главных компонент уменьшить 11 до 5 практически без увеличения погрешности прогноза

Для устранения возможных ошибок прогноза, вызванных относительно медленно меняющимися, но пока не учтенными нами факторами, такими как сезон года, открытие, закрытие и модернизация производств, миграция населения и т. п., мы вводим ежесуточно мультипликативную поправку - вместо оценки  $B_r(t)$  используем оценку  $D_r(t, \Delta t)$ , где  $\Delta t$  - заблаговременность:

$$W_{r}(t + \Delta t) \approx D_{r}(t, \Delta t) = B_{r}(t + \Delta t) \frac{\sum_{t'=t}^{t-23} B_{r}(t') W_{r}(t')}{\sum_{t'=t}^{t-23} (B_{r}(t'))^{2}}$$

## Корреляционные функции ошибок для субъектов РФ



#### Некоторые оценки качества прогноза

Пусть  $W(t_i)$  – фактическое потребление,  $F(t_i)$  – прогноз потребления, тогда

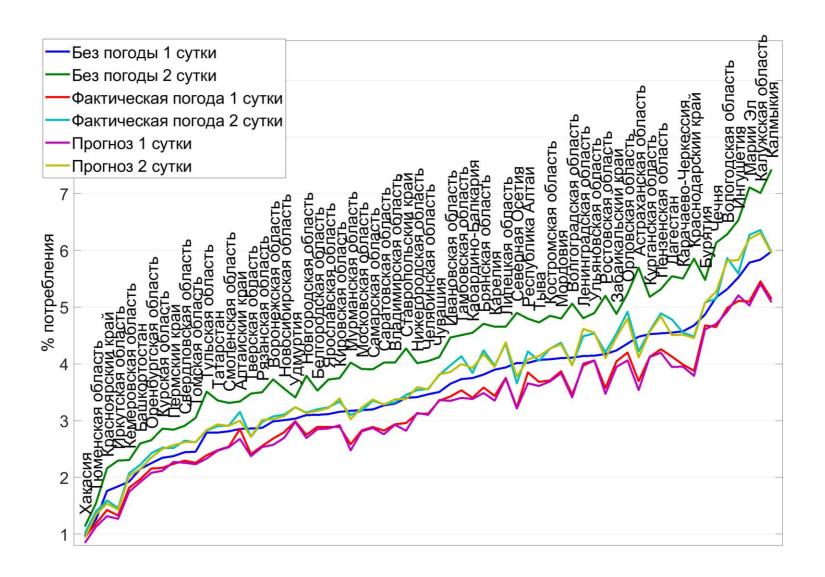
$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| W\left(t_{i}\right) - F\left(t_{i}\right) \right|}{W\left(t_{i}\right)}$$

$$NMAE = \frac{\sum_{i=1}^{N} \left| W\left(t_{i}\right) - F\left(t_{i}\right) \right|}{\sum_{i=1}^{N} W\left(t_{i}\right)}$$

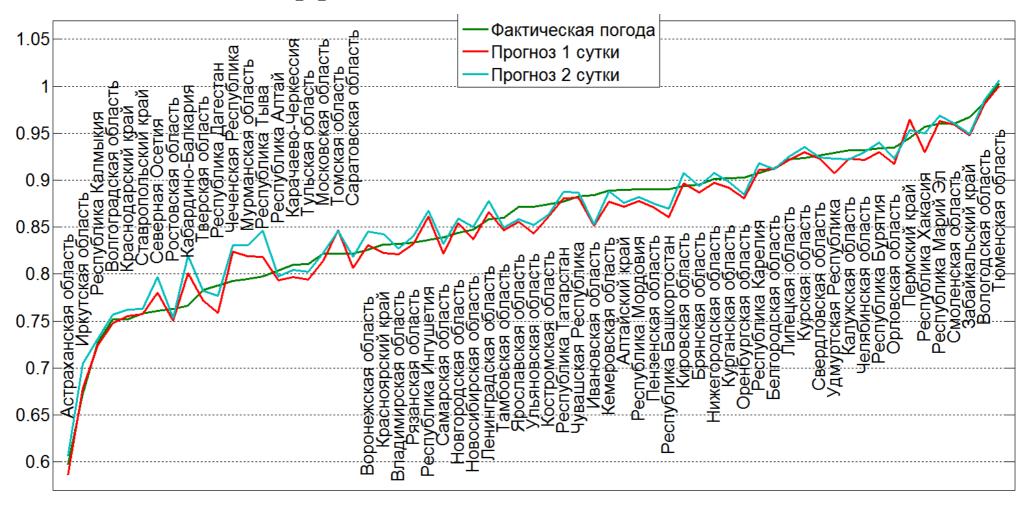
$$NRMSE = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(W\left(t_{i}\right) - F\left(t_{i}\right)\right)^{2}}}{\frac{1}{N} \sum_{i=1}^{N} W\left(t_{i}\right)}$$

Оптимизация происходила по NRMSE, далее приведены оценки MAPE

#### Результаты



#### Эффект от использования погоды



Отношение ошибок прогноза при использовании фактической и прогностической погоды в сравнении с её неиспользованием.

#### Сравнение результатов с методом Чучуевой

В диссертации [5] предложен метод прогноза потребления электроэнергии в объединенных энергосистемах (ОЭС) РФ методом выборки максимального правдоподобия и приведены оценки за 2011г. Просуммировав прогнозы по субъектам, можно получить соответствующие прогнозы и рассчитать оценки МАРЕ

ОЭС	Чучуева, 2011г Предлагаемый метод, 2015		
Урала	0.91%	1.01%	
Средней Волги	1.65%	1.77%	
Юга	1.83%	3.41%*	
Северо-Запада	1.67%	3.15%*	
Центра	1.53%	2.02%	
Сибири	1.65%	1.05%*	
Европейской части	1.12%	1.33%	

<sup>\*</sup>Границы ОЭС не совпадают с границами субъектов или мы не рассматривали некоторые из субъектов, входящих с ОЭС

## Данные о потреблении электроэнергии отдельным зданием

	«Forecasting Energy Consumption»	«Cold Start Energy Forecasting»	
	https://www.drivendata.org/competitio	https://www.drivendata.org/competition	
	<u>ns/51</u>	<u>s/55</u>	
Зданий всего/test	267/267	1383/625	
Объем train/test	7782405/1309176	621360/7529	
Шаг по времени	15мин/1час/1сут	1час/1сут/1неделя	
Заблаговременности	1сут-1неделя	1сут/1неделя/2недели	
Температура	Датчики + расстояние до датчика	1 значение	
Рабочие дни здания	Дни недели+гос.праздники	Дни недели	
Площадь здания	$B m^2$	Малая/средняя/большая/	
Температура внутри	18/20/22 °C	Низкая/средняя/высокая	
Место/участников	12/1034	17/1291	
Метрика	Взвешенная NRMSE	NMAE	

Далее речь пойдет о данных с конкурса «Cold Start Energy Forecasting»

#### Особенности набора данных

В «train» длина рядов данных была не более 28 суток. В «test» – 1-7 суток.

Для обучения были даны данные о часовом потреблении их и прогнозировали. В «test» нужно было получить в т. ч. и данные о суточном/недельном потреблении. Далее будем говорить только о часовом, т.к. суточное и недельное можно получить суммированием.

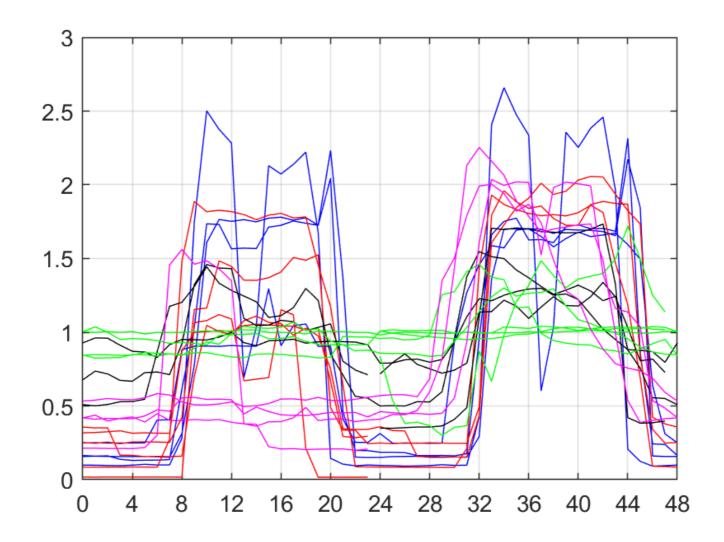
В данных о температуре были пропуски. Они было заполнены средне климатическими значениями, посчитанными на предоставленном архиве.

Метрика 
$$NMAE(x_i, y_i) = \frac{1}{N} \sum_{i} \frac{1}{\overline{y}_i} \sum_{t=t_{i1}}^{t_{i2}} |x_i(t) - y_i(t)|$$
, где  $x_i$  - прогноз,  $y_i$  - факт,

 $\overline{y}_i$  - среднее фактическое потребление (не известно), N – количество слагаемых в сумме.

Далее будем рассматривать только нормированное потребление  $Q_i\left(t\right) = W_i\left(t\right) / \bar{W}_i\left(t\right)$  и оптимизировать среднюю абсолютную погрешность

#### Различный режим работы зданий отражается на электропотреблении



Четыре кластера потребителей (цвет): зеленые работают всегда, синие и черные только днем, пурпурные — только по рабочим. По оси x: 0-24ч — нерабочий день 24-48ч — рабочий день.

#### Кластеризация методом к-средних

Алгоритм кластеризации [7] k-средних минимизирует среднее расстояние точек кластеров  $I_k$  от центра кластера  $\vec{\mu}_k$ :

$$\sum_{k=1}^{K} \sum_{i \in I_k} \|\vec{x}_i - \vec{\mu}_k\|_{L_2} \to \min_{I_k, \vec{\mu}_k}$$

Выберем случайные  $\vec{\mu}_k = \vec{x}_{i_k}$ . Процесс оптимизации состоит из последовательного применения двух шагов:

1. Выберем наиболее оптимальное разбиение для данных центров кластеров  $I_k = \left\{i \big| \big\| \vec{x}_i - \vec{\mu}_k \big\|_{L_2} = \arg\min_k \big\| \vec{x}_i - \vec{\mu}_k \big\|_{L_2} \right\}$ 

2. Выберем наиболее оптимальные центры кластеров - центр масс векторов кластера:

$$\vec{\mu}_k = \frac{1}{\#I_k} \sum_{i \in I_k} \vec{x}_i$$

Алгоритм k-средних всегда сходится, но не всегда дает оптимальное решение. Задача кластеризации NP-сложная [8]

Если данных относительно мало, то разумно предположить, что потребление зданий из одного кластера одинаковым образом зависит от температуры. Для потребителей, которые при кластеризации попали в один кластер, совместно уточняем зависимости потребления электроэнергии  $M_i$  от времени суток и градиента температуры:

$$M_{i}(s(t), w(t), dT(t)) = M_{i1}(s(t), w(t)) + c_{k1}dT(t)^{2} + c_{k2}dT(t)^{4}$$

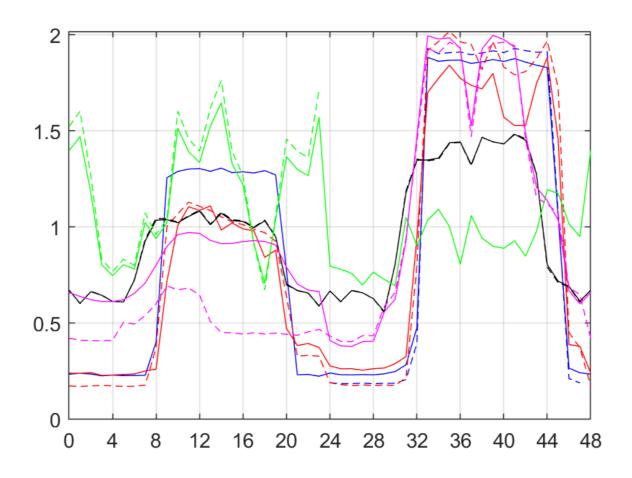
$$\sum_{i \in I_{k}} \left[ Q_{i}(t) - M_{i}(s(t), w(t), T_{air}(t) - T_{internal}(t)) \right]^{2} + (1)$$

$$+ \alpha \sum_{i \in I_{k}} \left\| \frac{\partial M_{i}}{\partial s} \right\|_{I_{2}}^{2} + \beta \sum_{i \in I_{k}} \left\| \frac{\partial^{2} M_{i}}{\partial s^{2}} \right\|_{I_{2}}^{2} + \frac{\gamma}{N_{k}} \sum_{i, j \in I_{k}} \left\| M_{i1} - M_{j1} \right\|_{L_{2}}^{2} \rightarrow \min_{M_{i}, i \in I_{k}, c_{k1}, c_{k2}},$$

где k=1,...,K — номер кластера потребителей, состоящего из  $N_k$  потребителей,  $T_{air}(t)$  — температура атмосферного воздуха,  $T_{internal}(t)$  — поддерживаемая в помещении температура,  $\alpha,\beta,\gamma$  — параметры регуляризации по Тихонову, [9],  $\|\cdot\|_{L_2}$  — норма в  $L_2$ .

Функции  $M_i$  ищутся в конечно-параметрическом семействе, что позволяет свести задачу (1) к задаче квадратичного программирования с разреженной ковариационной матрицей, для которой существуют быстрые и экономичные методы решения, [10].

#### До и после совместной оптимизации



Функции  $M_{i1}$  (сплошные) и среднее потребление от времени суток (пунктир) Тут для зеленого потребителя не было данных о потреблении в рабочие дни, а у синего – в нерабочие.

#### Схема анализ-прогноз с помощью нейронной сети типа LSTM [11][12]

Состоит из двух цепочек одинаковых блоков: блоков анализа и прогноза

Блок анализа принимает на вход вектор Блок прогноза принимает на  $a_t$ , состоящий из:

Фактическое потребление  $Q_i(t+s)$ , оценка | Оценка потребления  $M_i(t+s)$ , градиент потребления  $M_i(t+s)$ , температуры  $dT_i(t+s)$  при s = -3,...,0

Индикаторы дней недели

Время суток

Индикаторы кластеров потребителей

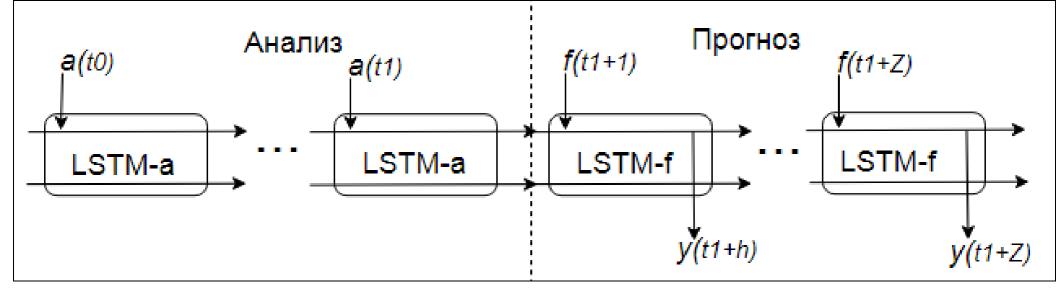
ВХОД вектор  $f_t$ , состоящий из:

градиент | температуры  $dT_i(t+s)$  при s = -2,...,2

Индикаторы дней недели

Время суток

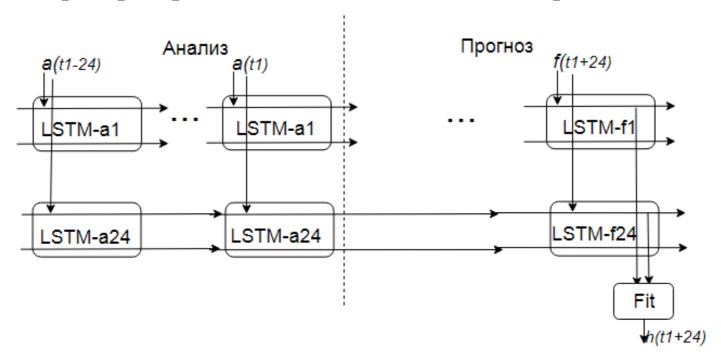
Индикаторы кластеров потребителей



#### Двупериодическая схема анализ-прогноз

Состоит из двух цепочек блоков анализ-прогноз:

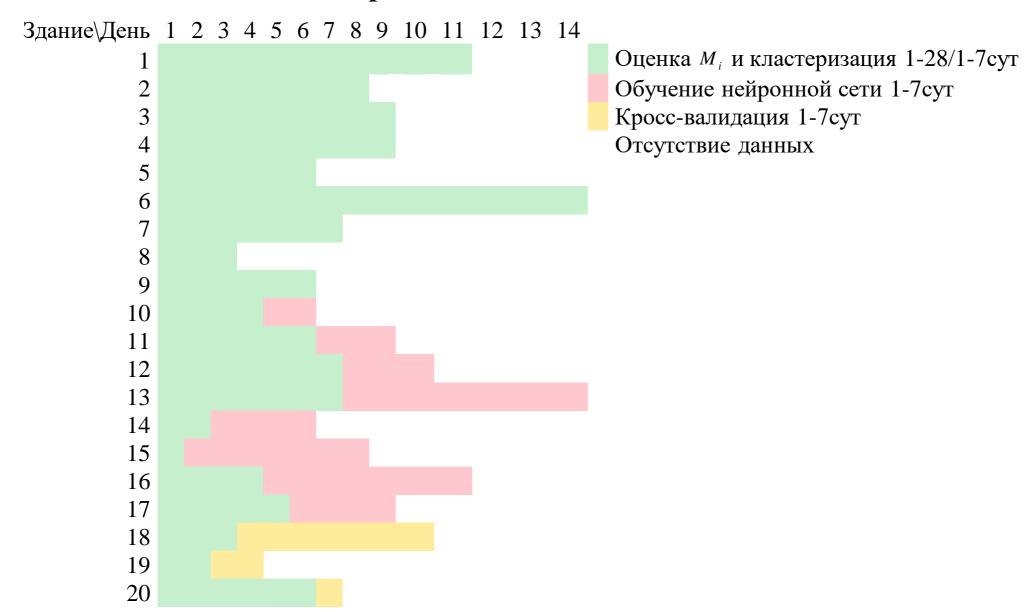
В первой сигнал распространяется с шагом 1 час, во второй – с шагом 24 часа



Итоговый прогноз потребления вычисляется:

- 1. Напрямую нейронной сетью  $D(t) = \max(0, h_{i1}(t))$
- 2. С учетом оценки  $M_i$ , то есть  $C(t) = \max(0, \max(0, h_{i1}(t))M_i(t) + h_{i2}(t))$

# Кросс-валидация (CV)



#### Результаты

$O$ ценка $M_i$	Формула	Периодов*	Кластер.	Погрешность на CV	Погрешность на test
Нет	D(t)	1	Нет	0.349-0.463	0.4948
Нет	D(t)	2	Нет	0.341-0.457	**
Да	D(t)	1	Нет	0.210-0.282	0.4020
Да	C(t)	1	Нет	0.149-0.197	**
Да	D(t)	1	Да	0.191-0.272	0.3666
Да	C(t)	1	Да	0.138-0.181	**
Да	D(t)	2	Нет	0.208-0.277	0.3853
Да	C(t)	2	Нет	0.145-0.193	0.3317
Да	D(t)	2	Да	0.190-0.269	0.3168
Да	C(t)	2	Да	0.134-0.177	0.3034

<sup>\*</sup>во всех случаях число параметров нейронных сетей лежало в пределах 4300-4600

<sup>\*\*</sup>после окончания конкурса невозможно получить оценку на test

#### Выводы

Для задачи прогноза потребления электроэнергии субъектами РФ был предложен экономичный метод, учитывающий прогноз температуры в административном центре субъекта. Параметры предложенного метода могут быть найдены при помощи решения простых систем линейных уравнений.

Величина улучшения от использования данных о погоде при прогнозе почасового потребления субъектов РФ эквивалентна уменьшению заблаговременности прогноза с 2 до 1 суток

В задаче прогноза потребления электроэнергии отдельным зданием были применены некоторые полезные техники, позволившие уменьшить погрешность прогноза:

- Кластеризация потребителей на 5-10%
- Прогнозирование отклонения от среднего, а не самого потребления на 5-50%
- Двупериодическая структура нейронной сети на 1-10%

#### Литература

- 1. Объем суммарного фактического потребления электрической энергии в субъекте РФ <a href="http://www.atsenergo.ru/results/market/fact\_region">http://www.atsenergo.ru/results/market/fact\_region</a>
- 2. Быков Ф.Л., Гордин В.А. *Краткосрочный прогноз часового потребления электроэнергии с учетом погоды для субъектов РФ* // Известия РАН: Энергетика, 2017, № 5, с. 47—56.
- 3. Багров А.Н., Быков Ф.Л., Гордин В.А. Схема оперативного краткосрочного комплексного прогноза приземной температуры воздуха и влажности // Метеорология и гидрология, 2018, № 8, с. 5—18.
- 4. Комплексный прогноз по данным различных атмосферных моделей для городов России и Республики Беларусь <a href="http://method.meteorf.ru/ansambl/ansambl.html">http://method.meteorf.ru/ansambl/ansambl.html</a>
- 5. Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального правдоподобия. Дис. канд. техн. наук. М.: МГТУ, 2012. 153 с
- 6. Aggarwal S.K., Saini L.M., Kumar A. *Electricity price forecasting in deregulated markets: A review and evaluation* // Electrical Power and Energy Systems, 2009, vol. 31, pp. 13—22.
- 7. Lance G.N., Willams W.T. *A general theory of classificatory sorting strategies*. 1. *Hierarchical systems* // Computer Journal, 1967, № 9, pp. 373—380.
- 8. Mahajan M., Nimbhorkar P., Varadarajan K. *The planar k-means problem is NP-hard* //International Workshop on Algorithms and Computation. Springer, Berlin, Heidelberg, 2009. C. 274-285.
- 9. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1979
- 10. Gill P. E., Murray W., Saunders M. A., Wright M. H. *Sparse matrix methods in optimization* // SIAM Journal on Scientific and Statistical Computing, 1984, 5(3), 562-589.
- 11. Hochreiter S., Schmidhuber J. *Long short-term memory* // Neural computation, 1997, vol. 9, pp. 1735—1780.
- 12. Greff K., Srivastava R.K., Koutnik J., Steunebrink B.R. *Lstm: A search space odyssey* // IEEE transactions on neural networks and learning systems, 2017, №10, vol. 28.
- 13. Competition: Power Laws: Forecasting Energy Consumption <a href="https://www.drivendata.org/competitions/51">https://www.drivendata.org/competitions/51</a>
- 14. Competition: Power Laws: Cold Start Energy Forecasting <a href="https://www.drivendata.org/competitions/55">https://www.drivendata.org/competitions/55</a>

# Спасибо за внимание!

Работа подготовлена в ходе проведения исследования (№ 18-05-0011) в рамках Программы «Научный фонд Национального исследовательского университета "Высшая школа экономики" (НИУ ВШЭ)» в 2018 — 2019 гг. и в рамках государственной поддержки ведущих университетов Российской Федерации «5-100»

#### Сеть с памятью LSTM

Пусть  $\sigma(x) = (1 + e^{-x})^{-1}$ , тогда уравнения LSTM можно записать как

Забывающий вентиль 
$$f_t = \sigma (W_{fx} x_t + W_{fh} h_{t-1} + b_f)$$

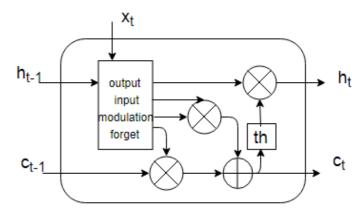
Входящий вентиль 
$$i_{t} = \sigma \left( W_{ix} x_{t} + W_{ih} h_{t-1} + b_{i} \right)$$

$$M$$
одуляция  $m_{\scriptscriptstyle t} = W_{\scriptscriptstyle mx} x_{\scriptscriptstyle t} + W_{\scriptscriptstyle mh} h_{\scriptscriptstyle t-1} + b_{\scriptscriptstyle c}$ 

Обновление состояния 
$$c_t = f_t \circ c_{t-1} + i_t \circ m_t$$

Выходящий вентиль 
$$o_t = \sigma (W_{ox} x_t + W_{oh} h_{t-1} + b_o)$$

$$P$$
езультат  $h_t = \tanh(c_t) \circ o_t$ 



Градиент функции потерь затухает в соответствии с забывающим вентилем  $f_t$ :

$$\frac{\partial e_{t}}{\partial c_{t}} = f_{t+1} \circ \frac{\partial e_{t}}{\partial c_{t+1}} + o_{t} \circ \tanh'(c_{t}) \circ \frac{\partial e_{t}}{\partial h_{t}}$$

Можно научиться воспроизводить дальние зависимости

Инициализация  $b_f$  очень важна: временной масштаб процесса

#### Технические подробности обучения нейронных сетей

Входные данные для нейронной сети предварительно нормировались так, чтобы значения каждого из предикторов принадлежали интервалу [-1,1]. Инициализация весов нейронной сети происходила случайным образом по методу Ксавье [15]. Подстройка весов нейронной сети осуществлялось методом градиентного спуска «Аdam с моментами Нестерова» [16], со скоростью обучения  $\eta = 0.01$ . Существенное увеличение  $\eta$  может привести к отсутствию сходимости, а уменьшение – к росту объёма вычислений без уменьшения итоговой погрешности.

Обучение состояло из 10000 шагов с minibatch [17], состоящими из 16 временных рядов длиной 2-14 суток: 1-7 суток для анализа и 1-7 суток для прогноза.

- 15. Glorot X., Bengio Y. *Understanding the difficulty of training deep feedforward neural networks* // In Proc. AISTATS, 2010, vol. 9. pp. 249-256
- 16. Dozat T. Incorporating Nesterov Momentum into Adam // ICLR Workshop. 2016. Vol. 1:2013-2016
- 17. Cotter A., Shamir O., Srebro N., Sridharan K. *Better mini-batch algorithms via accelerated gradient methods* // Advances in Neural Information Processing Systems, 2011, vol. 24, pp. 1647-1655