КРАТКОСРОЧНЫЙ ПРОГНОЗ ЧАСОВОГО ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ С УЧЕТОМ ПОГОДЫ ДЛЯ СУБЪЕКТОВ РФ

Ф.Л.Быков, ***ФБГУ «Гидрометцентр России»***

В.А.Гордин, ***НИУ ВШЭ***

Москва, 2019г

Данные о часовом потреблении электроэнергии субъектами РФ

Данные АО «АТС» («Администратор Торговой Системы оптового рынка электроэнергии») о потреблении в 66 субъектах РФ [1]
http://www.atsenergo.ru/results/market/fact_region

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 также использование данных с остальных синоптических станций соответствующего региона.

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

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

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

Даем прогнозы потребления электроэнергии, используя соответствующие метеорологические прогнозы. Эталон – при фактической погоде.

Комплексный прогноз погоды на сайте http://method.meteorf.ru/ansambl/ansambl.html

Список городов

Прогноз порывов ветра | Фактические порывы ветра

А Б В Г Д Е Ж З И Й К Л М Н О П Р С Т У Х Ц Ч Ш Э Ю Я

А

Адлер | Аддан | Амдерма
Анадырь | Анучино | Апатиты
Арзамас | Архангельск | Астрахань
Астраханка | Армавир | Ачинск
Аян

Б

Балашов | Барабинск | Барнаул
Белгород | Бийск | Билибино
Биробиджан | Благовещенск | Бологое
Братск | Брянск | Бугульма
Бугуруслан | Бузулук

В

Великие Луки | Верхоянск | Вилюйск
Витим | Владивосток | Владикавказ
Владимир | Волгоград | Вологда
Волоколамск | Воркута | Воронеж
Выкса

Г

Гагарин | Глазов | Грозный
Гусь-Хрустальный

Д

Дальнереченск | Двинской Березник | Дербент
Дмитров | Диксон

Е

Евпатория | Екатеринбург | Елабуга
Елатьма | Елец | Ельня
Енисейск | Ербогачен | Ефремов

Ж

Жиганск | Жиздра

З

Зея | Зима | Златоуст
Змеиногорск

И

Иваново | Игарка | Ижевск
Иркутск | Ишим

Й

Йошкар-Ола

К

Казань | Калининград | Калач
Калуга | Камышин | Каргополь
Кемерово | Керчь | Кемь
Кингисепп | Кисловодск | Киренск
Киров | Клин | Ключи
Коломна | Комсомольск-на-Амуре | Корсаков
Кострома | Котлас | Красная Поляна
Краснодар | Красноярск | Красный Холм
Кудымкар | Курган | Курск
Кызыл | Кызыл-Озек

Л

Ленск | Липецк | Лиски
Лодейное поле

М

Магадан | Магнитогорск | Майкоп
Малоярославец | Малые Кармакулы | Мамакан
Махачкала | Мезень | Минеральные воды
Минусинск | Мирный | Можайск
Моздок | Мончегорск | Москва
Мурманск | Мценск

Н

Назрань | Нальчик | Нарьян-Мар
Находка | Нефтеюганск | Невельск
Нерчинск | Нижний Тагил | Нижневартовск
Нижний Новгород | Николаевск-на-Амуре | Новгород
Новокузнецк | Новороссийск | Новосибирск
Новый Иерусалим | Новый Уренгой | Норильск
Ноглики | Ноябрьск | Нюрба
Няндома

О

Оймякон | Оленёк | Олекминск
Омск | Опочка | Орёл
Оренбург | Орск | Осташков
Охотск | Оха

П

Павловский Посад | Певек | Пенза
Переславль-Залесский | Пермь | Петрозаводск
Петропавловск-Камчатский | Печора | Пограничный
Поронайск | Псков | Пушкино
Пушкинские Горы

Р

Рославль | Ростов | Ростов-на-Дону
Ряжск

С

Салехард | Самара | С.-Петербург
Саранск | Саратов | Севастополь
Сеймчан | Серпухов | Симферополь
Смоленск | Советск | Советская гавань
Сочи | Средне-Колымск | Ставрополь
Старый Оскол | Сургут | Сухиничи
Сыктывкар

Т

Таганрог | Тайшет | Тамбов
Тверь | Терней | Тикси
Тихвин | Тобольск | Тольятти
Томск | Торопец | Туапсе
Тула | Тулун | Тура
Тында | Тюмень

У

Улан-Удэ | Ульяновск | Урюпинск
Усть-Илимск | Усть-Ишим | Усть-Кут
Уфа | Ухта | Уэлен

Ф

Феодосия

Х

Хабаровск | Хакасская | Ханты-Мансийск
Хатанга

Ц

Цимлянск

Ч

Чайковский | Чебоксары | Челябинск
Чердынь | Череповец | Черкесск
Черняховск | Чистополь | Чита

Ш

Шарья

Э

Элиста

Ю

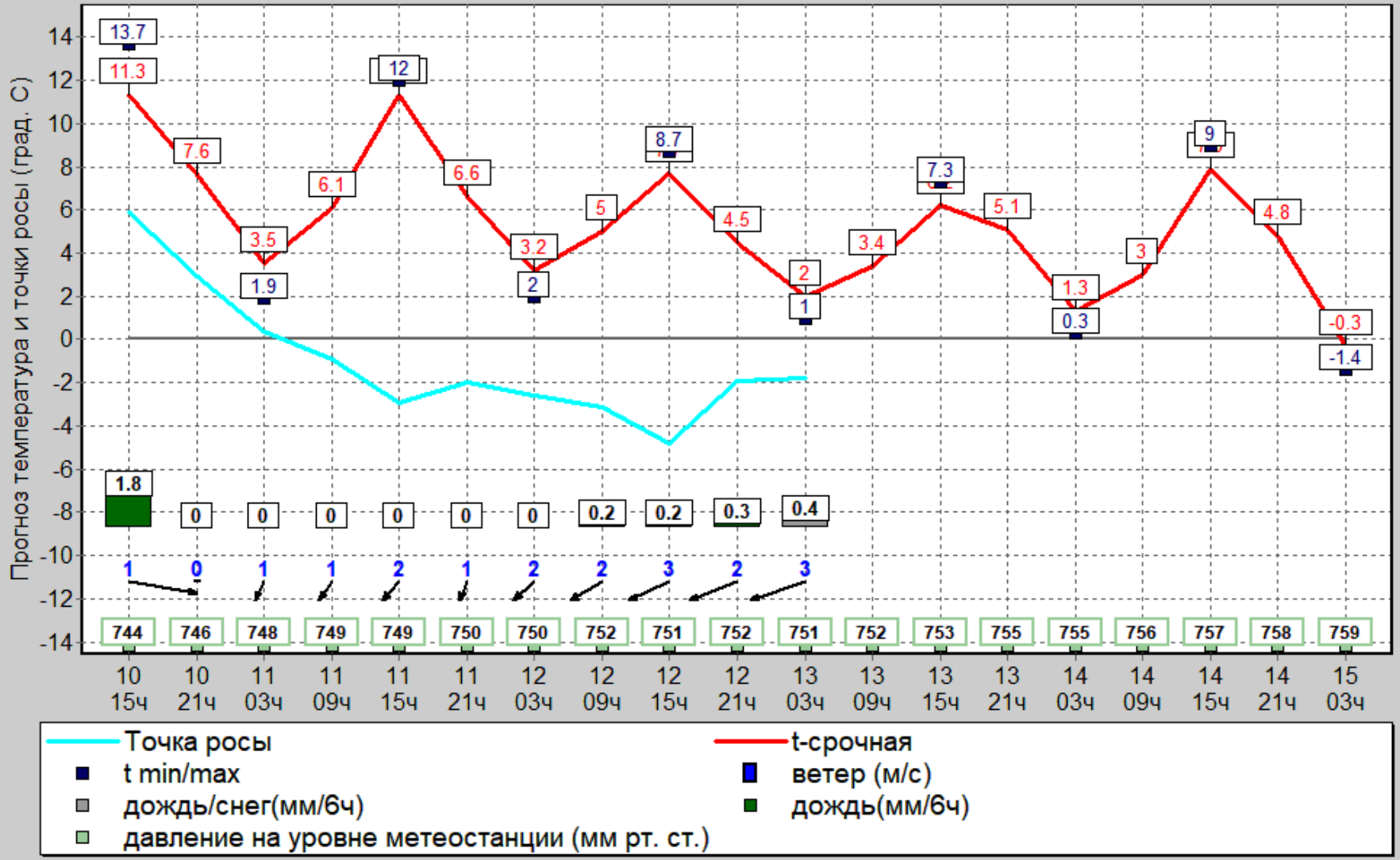
Южно-Курильск | Южно-Сахалинск | Юрьев-Польский

Я

Якутск | Ялта | Ярославль

193.7.160.230/.../ST_PETERBU... | Центрально-Черноземное УГМС | Республика Бел

Начальная дата расчета прогноза 10.04.19 - 03 час местного времени
Пункт Москва (27612) 55.83°с.ш. 37.62°в.д.; высота 156 м



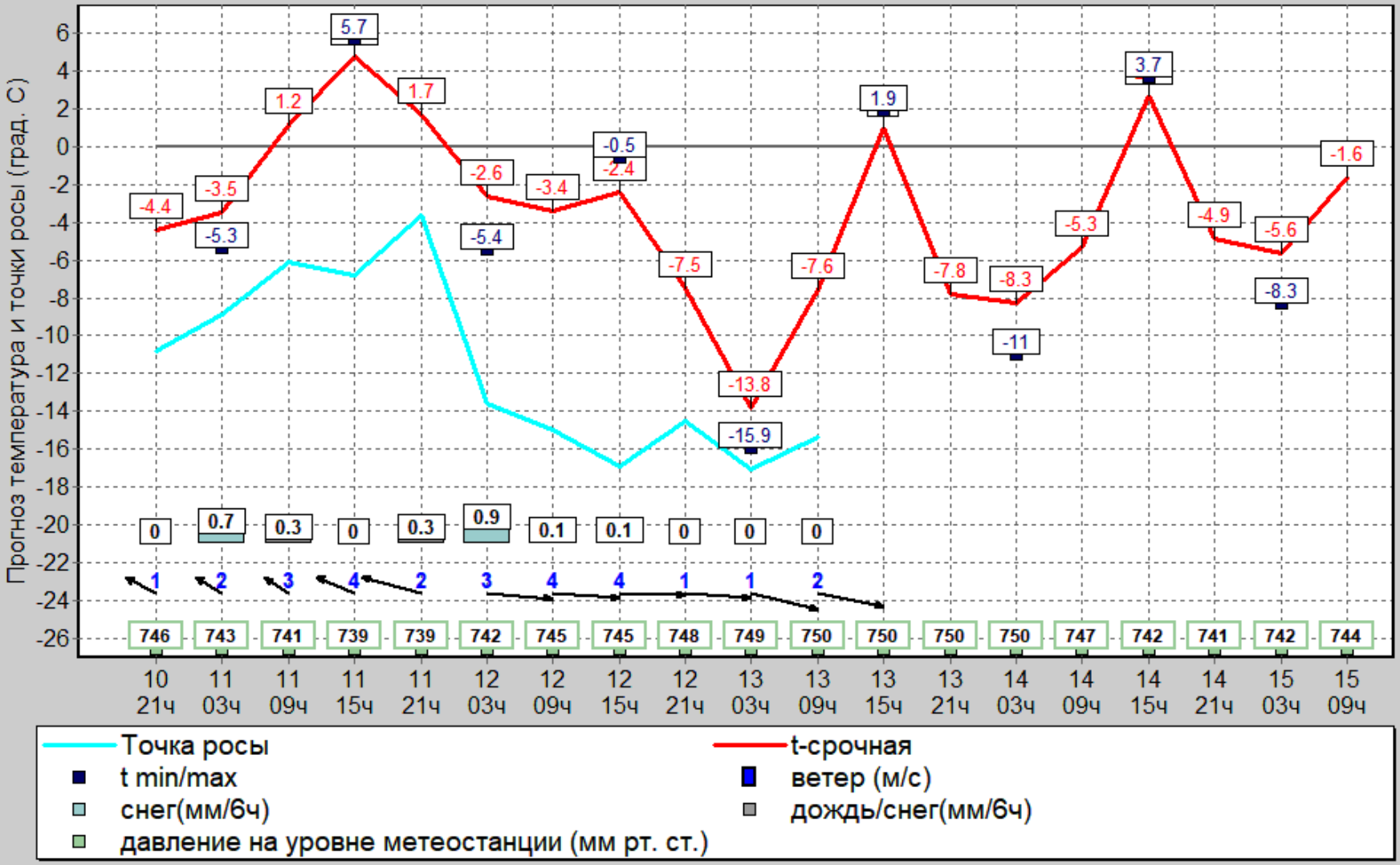
Прошедшая погода

Оценки прогноза температуры и осадков

Сравнительная оценка

© Методический кабинет Гидрометцентра России

Начальная дата расчета прогноза 10.04.19 - 09 час местного времени
Пункт Якутск (24959) 62.02°с.ш. 129.72°в.д.; высота 101 м



Прошедшая погода | Оценки прогноза температуры и осадков | Сравнительная оценка

© Методический кабинет Гидрометцентра России

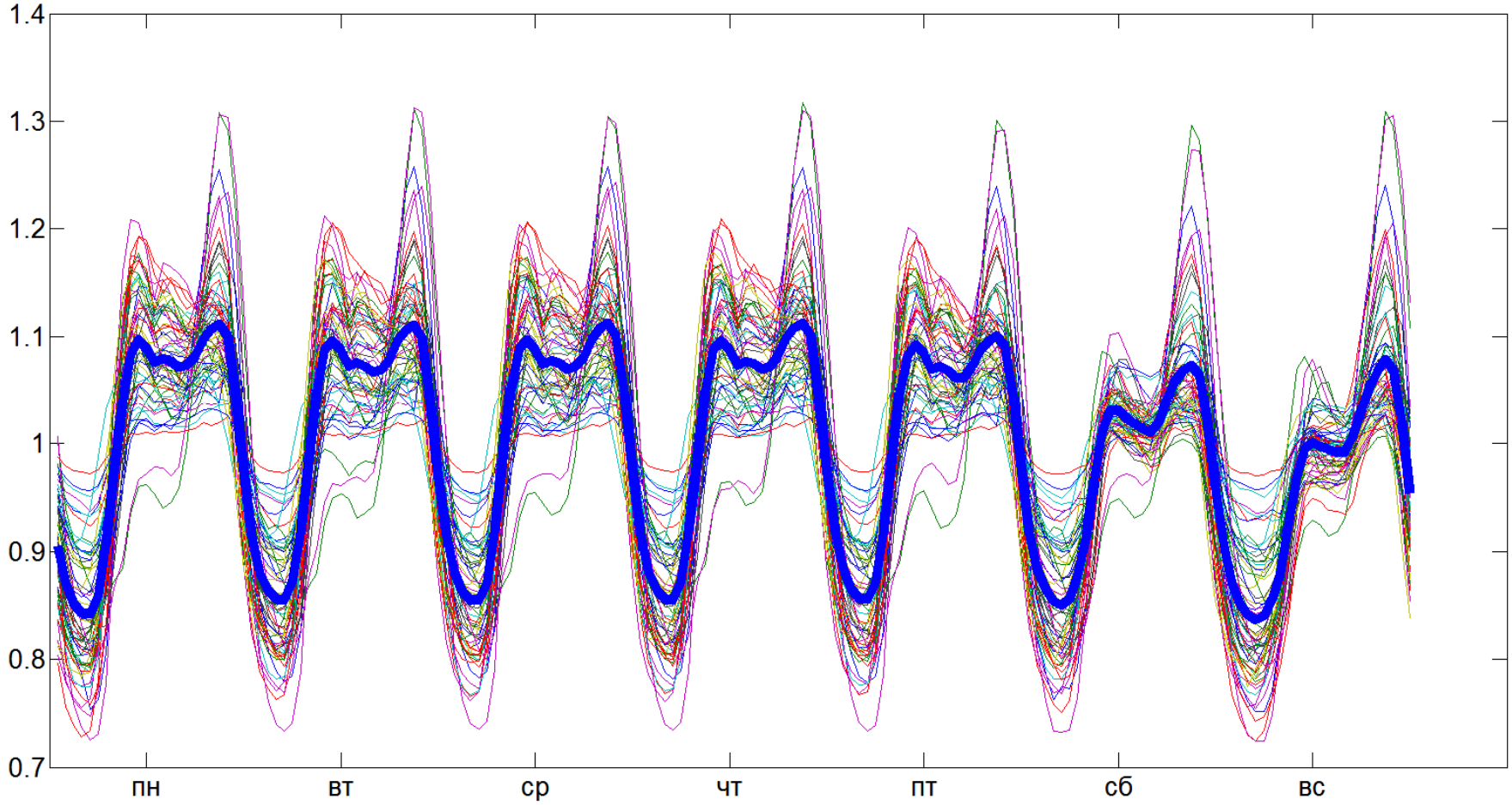
Технологические сроки

Метеорологические прогнозы по нашей схеме от срока 00:00 ВСВ (03:00 мск) в Гидрометцентре РФ 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.

У нас есть и другие начальные сроки, но здесь они практически менее интересны.

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.

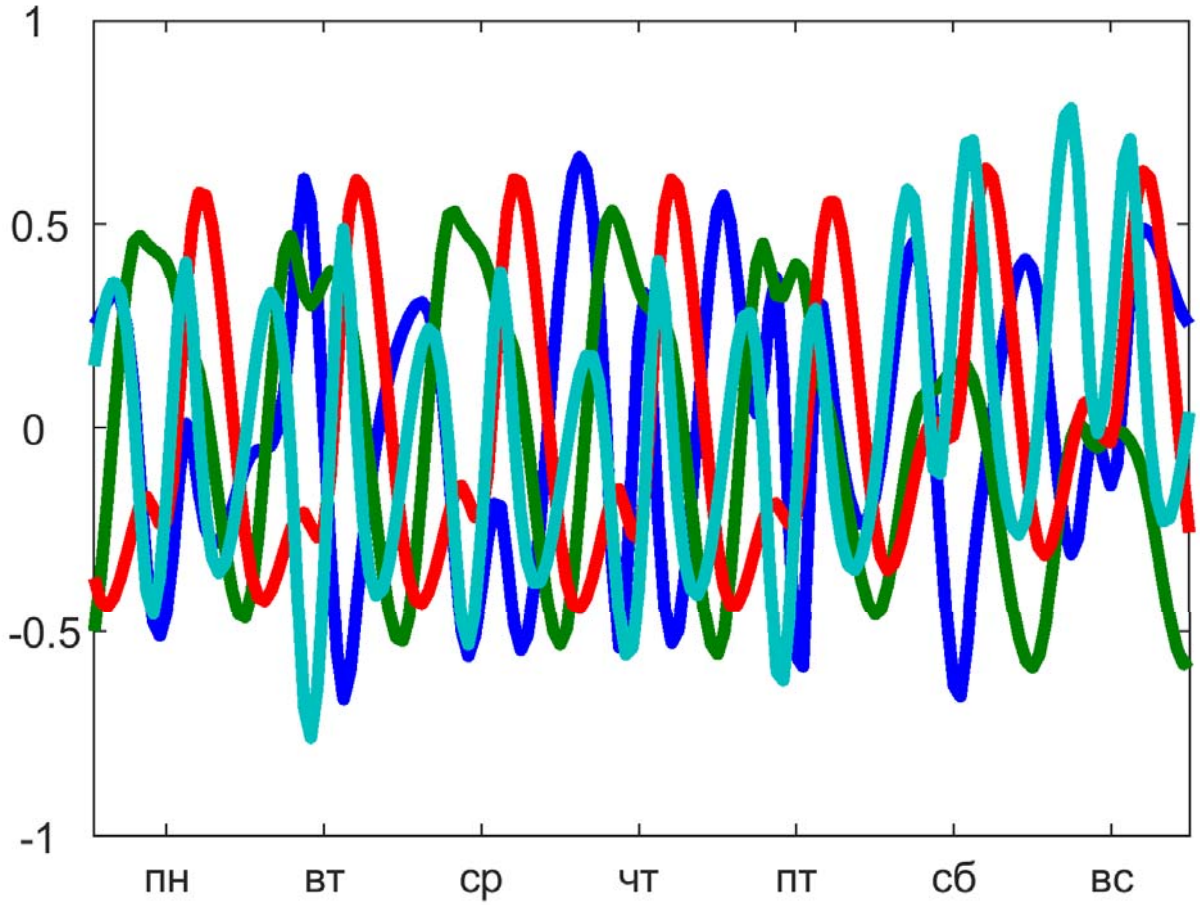
Зависимость потребления в субъектах РФ от часа недели



Синим – среднее по 63 субъектам РФ. Часовое потребление в субъектах меняется по-разному в зависимости от дня недели и (местного) времени суток. Есть «субъекты-жаворонки» и «субъекты-совы», динамика в выходные в разных субъектах сильно отличается. Здесь 1 – среднее часовое потребление в каждом субъекте РФ отдельно.

Зависимости первых базисных (естественных ортогональных) функций $H_k\left(t_w\right)$ от часа недели

$$H_0 \equiv 1, H_1 - синяя, H_2 - зеленая, H_3 - красная, H_4 - циан$$



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_w
3. Температура воздуха $T(t)$ и она же 12 часов назад $T(t-12)$

МНК оценку по этим 3 предикторам - потребление электроэнергии обозначим $B_r(t)$.

$$W_r(t) \approx B_r(t) = \sum_{k=1}^{4} h_{rk} H_k(t_w) + \alpha_{r1} + \alpha_{r2} W_r(t-168) + \ldots$$

$$+\alpha_{r3} T_r(t) + \alpha_{r4}\left[T_r(t)\right]^2 + \alpha_{r5}\left[T_r(t)\right]^3 + \alpha_{r6} T_r(t-12) + \alpha_{r7}\left[T_r(t-12)\right]^2$$

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

Учет тенденции изменения потребления

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 Δt - заблаговременность:

$$W_r\left(t+\Delta t\right) \approx D_r\left(t,\Delta t\right) = B_r\left(t+\Delta t\right)\frac{\sum_{t'=t}^{t-23} B_r\left(t'\right)W_r\left(t'\right)}{\sum_{t'=t}^{t-23}\left(B_r\left(t'\right)\right)^2}$$

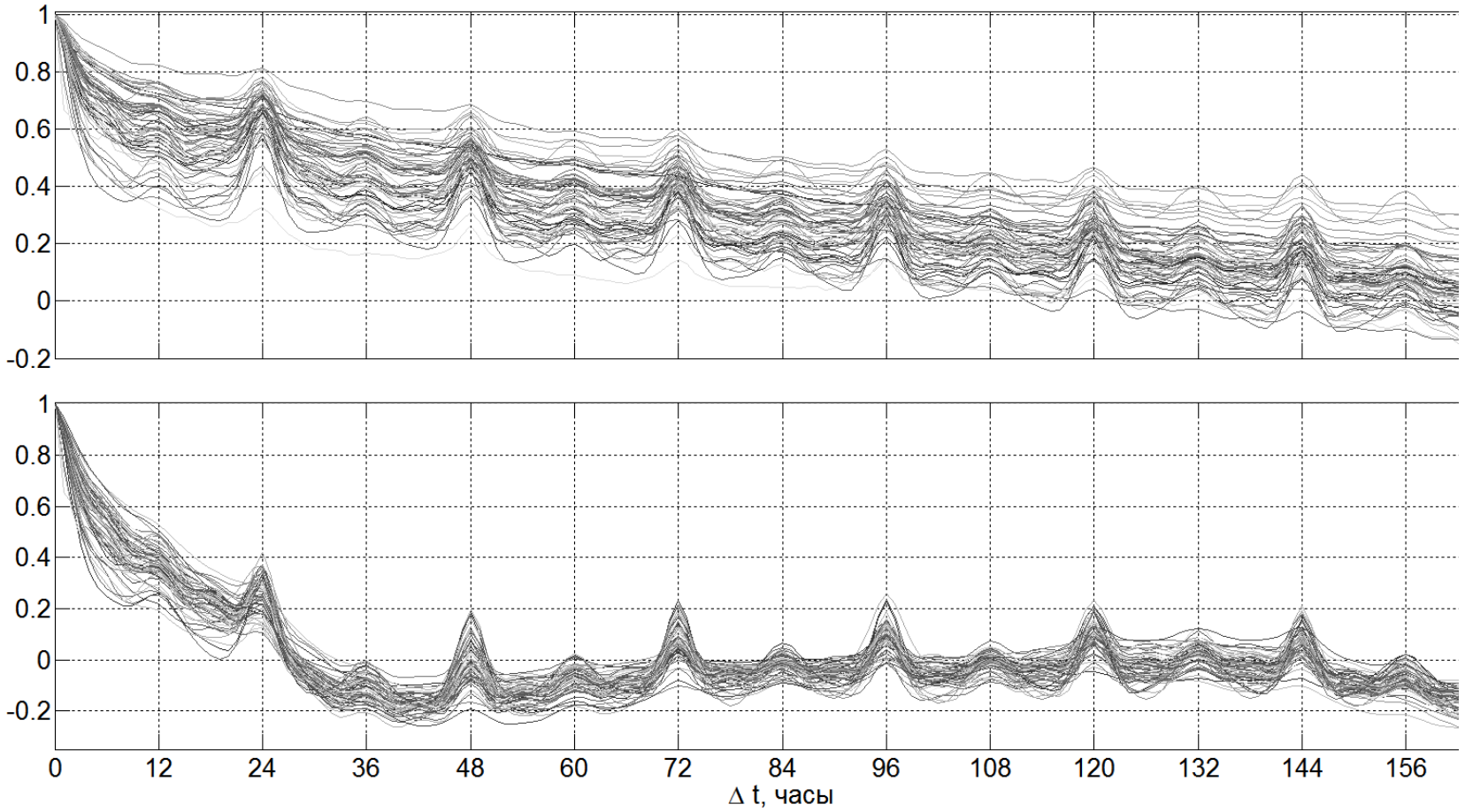
Корреляционные функции ошибок для субъектов РФ

КФ

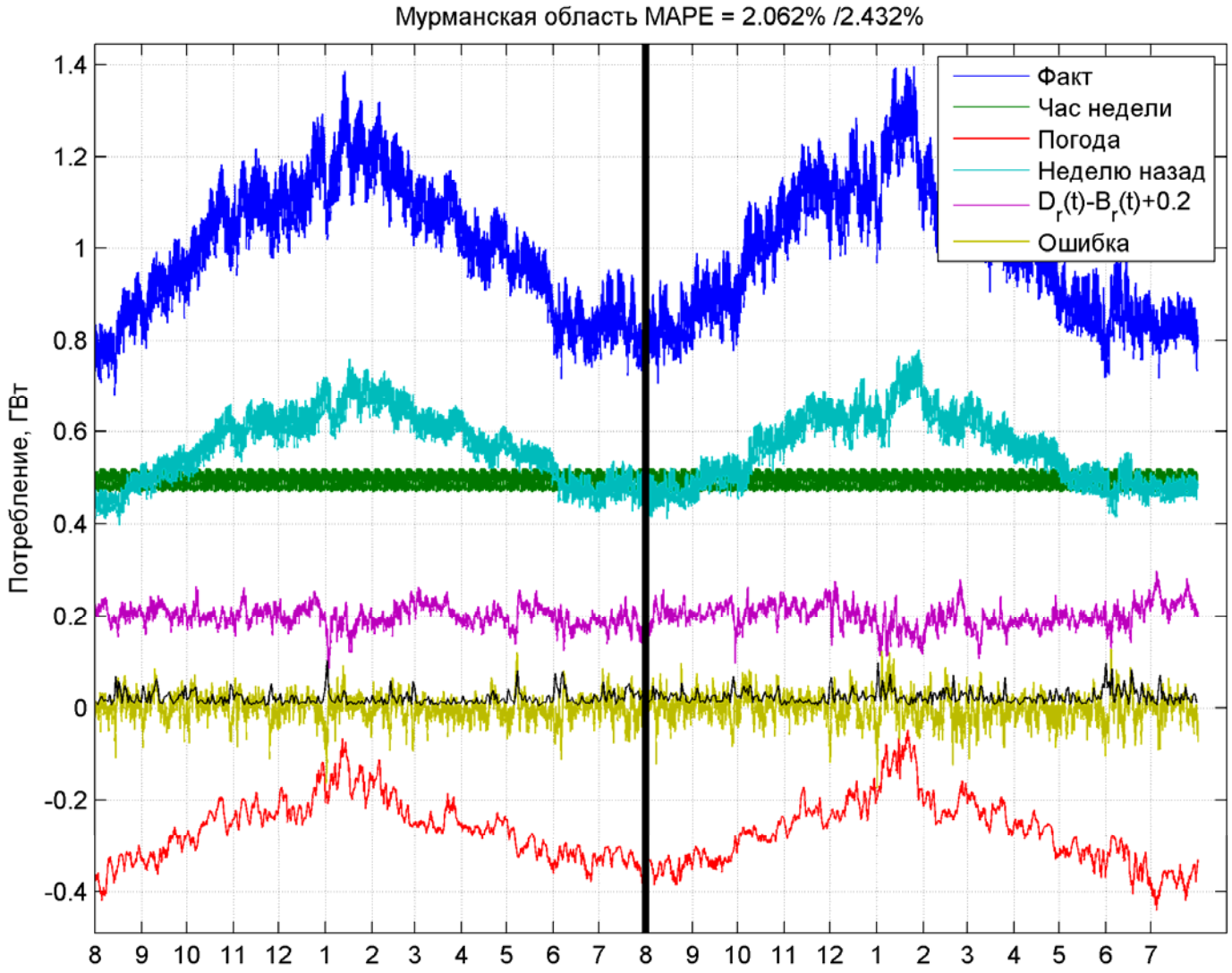
$W_r(t) - B_r(t)$

КФ

$W_r(t) - D_r(t, \Delta t)$



Компоненты прогноза для Мурманской области



Некоторые способы оценки качества прогноза

Пусть $W(t_i)-$ фактическое потребление, $F(t_i)-$ какой-то прогноз потребления, тогда

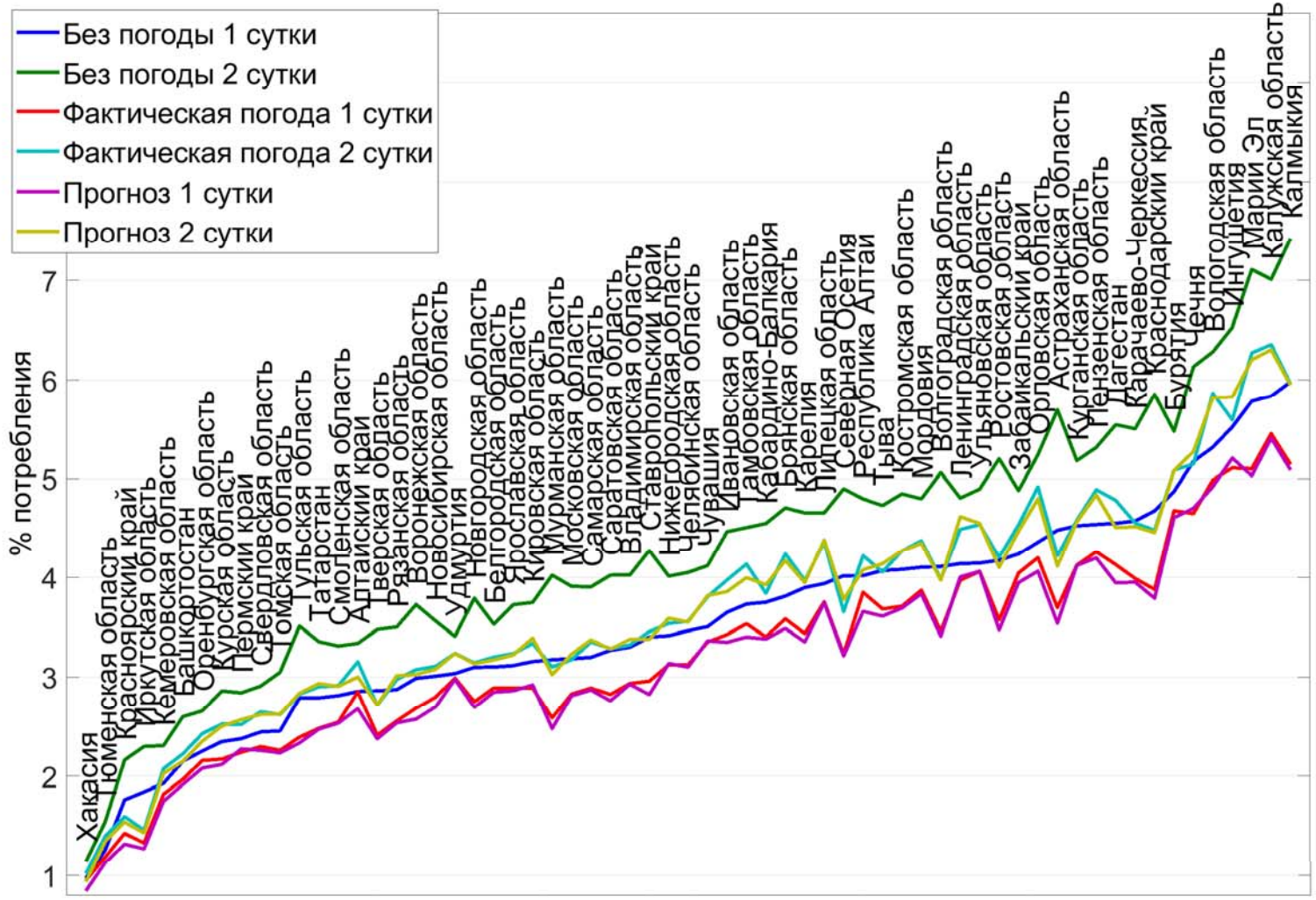
$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|W\left(t_i\right)-F\left(t_i\right)\right|}{W\left(t_i\right)}$$

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

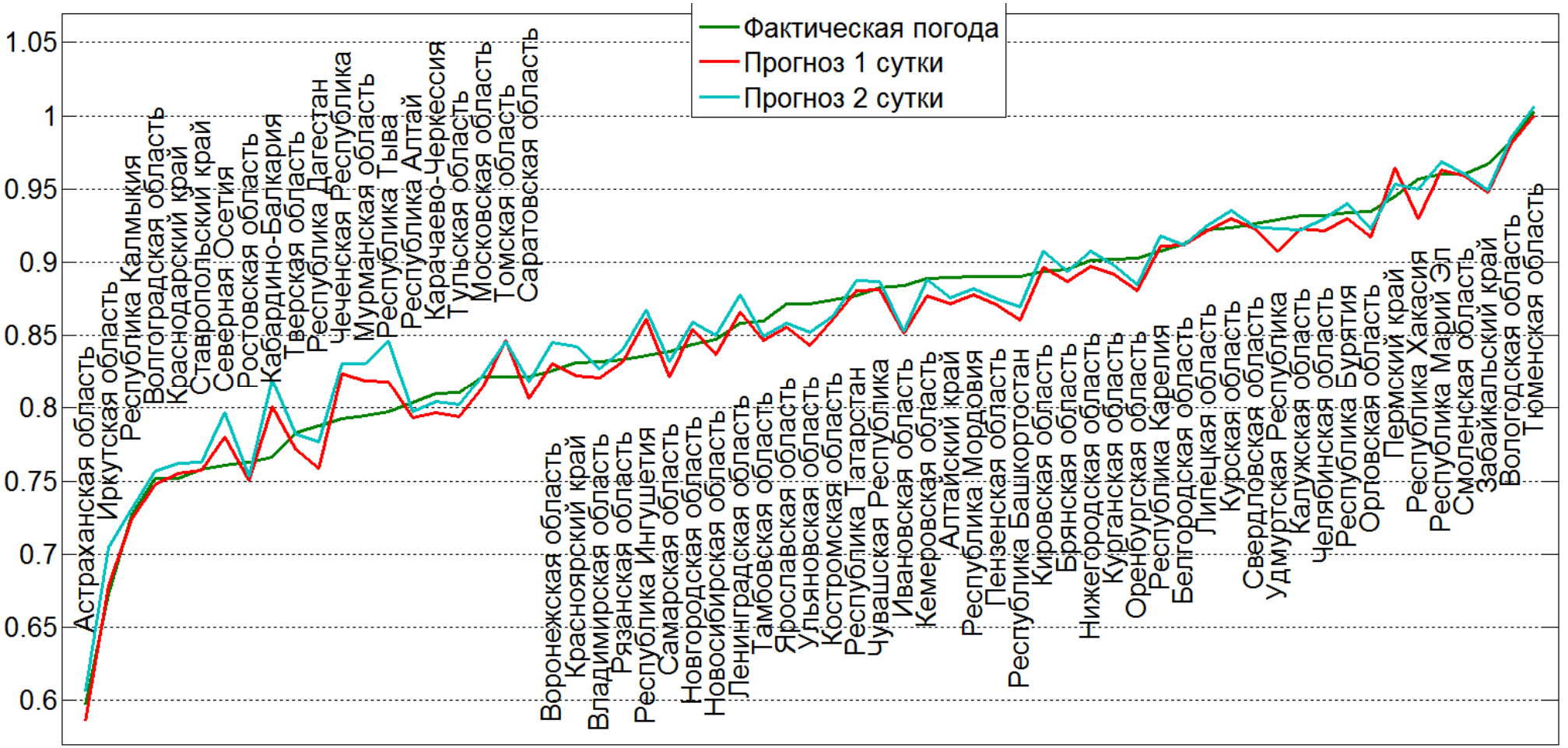
$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(W\left(t_i\right)-F\left(t_i\right)\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}W\left(t_i\right)}$$

Оптимизацию мы проводили по *NRMSE*, далее приведены оценки *MAPE*

Ошибки прогноза для разных регионов



Эффект от использования знания погоды или ее прогноза



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

Задача прогноза потребления электроэнергии отдельным зданием

Данные компании **Schneider Electric**	«**Forecasting Energy Consumption**» https://www.drivendata.org/competitions/51	«**Cold Start Energy Forecasting**» https://www.drivendata.org/competitions/55
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	В $м^2$	Малая/средняя/большая/ …
Температура внутри	18/20/22 ^{O}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.

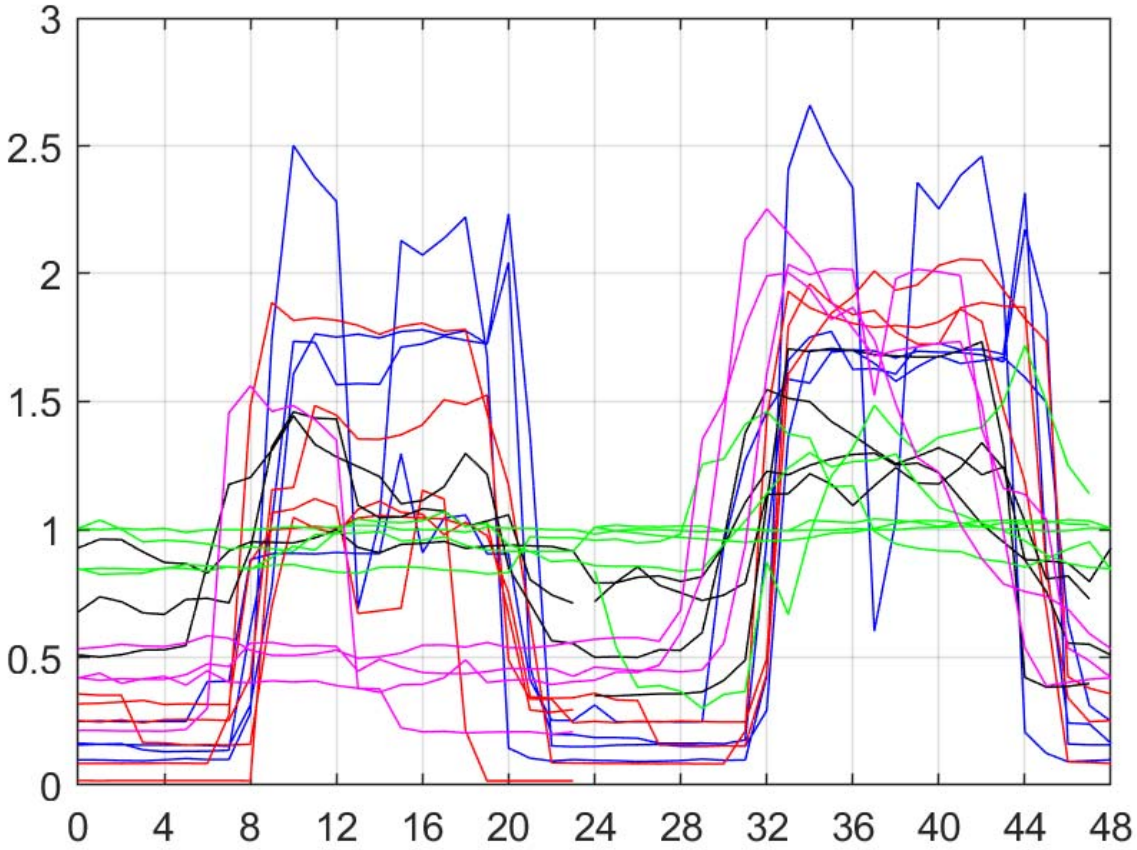
В данных о температуре были пропуски. Они было заполнены средне климатическими значениями, посчитанными на предоставленном архиве.

Метрика $NMAE\left(x_i, y_i\right) = \frac{1}{N}\sum_i \frac{1}{\overline{y}_i}\sum_{t=t_{i1}}^{t_{i2}}\left|x_i\left(t\right) - y_i\left(t\right)\right|$, где x_i - прогноз, y_i - факт, $\overline{y}_i$ - среднее фактическое потребление *i*-го здания(не известно), N – количество слагаемых в суммах.

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

Режим работы зданий отражается на его электропотреблении

Примеры энергопотребления:



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

Кластеризация методом k-средних

Кластеризация – построение оптимального разбиения множества векторов x_i на непересекающиеся кластеры I_k. Алгоритм кластеризации [7] *k*-средних минимизирует среднее расстояние точек кластеров I_k от центра кластера $\vec{\mu}_k$:

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

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 – разбиение Вороного $I_k=\left\{i\middle|\left\|\vec{x}_i-\vec{\mu}_k\right\|_{L_2}=\arg\min_k\left\|\vec{x}_i-\vec{\mu}_k\right\|_{L_2}\right\}$

2. – Выберем наиболее оптимальные центры кластеров - центры масс: $\vec{\mu}_k=\frac{1}{\#I_k}\sum_{i\in I_k}\vec{x}_i$

Алгоритм *k*-средних всегда сходится, но не всегда находит глобальный минимум (1)

Если данных для обучения 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:

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

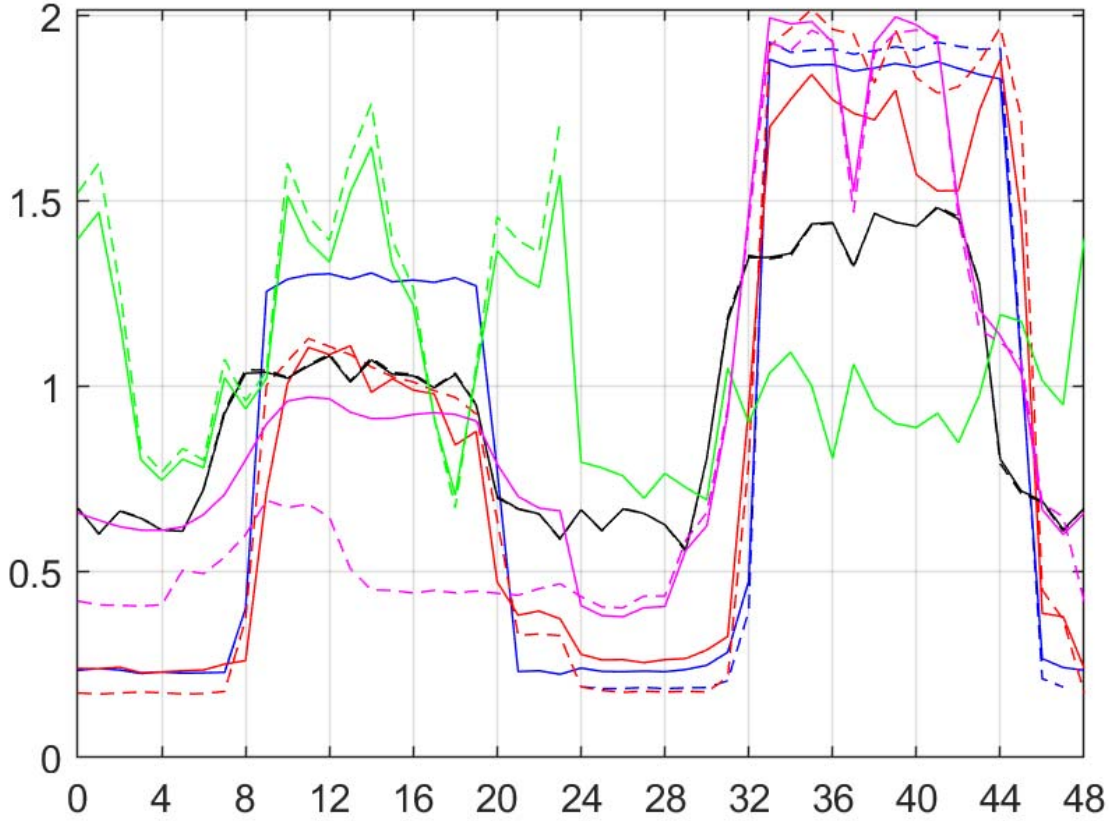
$$\sum_{i\in I_k}\left[Q_i(t)-M_i\left(s(t),w(t),T_{air}(t)-T_{internal}(t)\right)\right]^2+ \qquad (2)$$

$$+\alpha\sum_{i\in I_k}\left\|\frac{\partial M_i}{\partial s}\right\|^2_{L_2}+\beta\sum_{i\in I_k}\left\|\frac{\partial^2 M_i}{\partial s^2}\right\|^2_{L_2}+\frac{\gamma}{N_k}\sum_{i,j\in I_k}\left\|M_{i1}-M_{j1}\right\|^2_{L_2}\to\min_{M_i,i\in I_k,c_{k1},c_{k2}},$$

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, α,β,γ – параметры регуляризации по Тихонову, [9], $\|\cdot\|_{L_2}$ – норма в L_2.

Функции M_i ищутся в конечно-параметрическом семействе, что позволяет свести задачу (2) к задаче квадратичного программирования с разреженной ковариационной матрицей, для которой существуют быстрые и экономичные методы решения, [10].

Примеры оценки потребления до и после совместной (по кластеру) оптимизации (2)

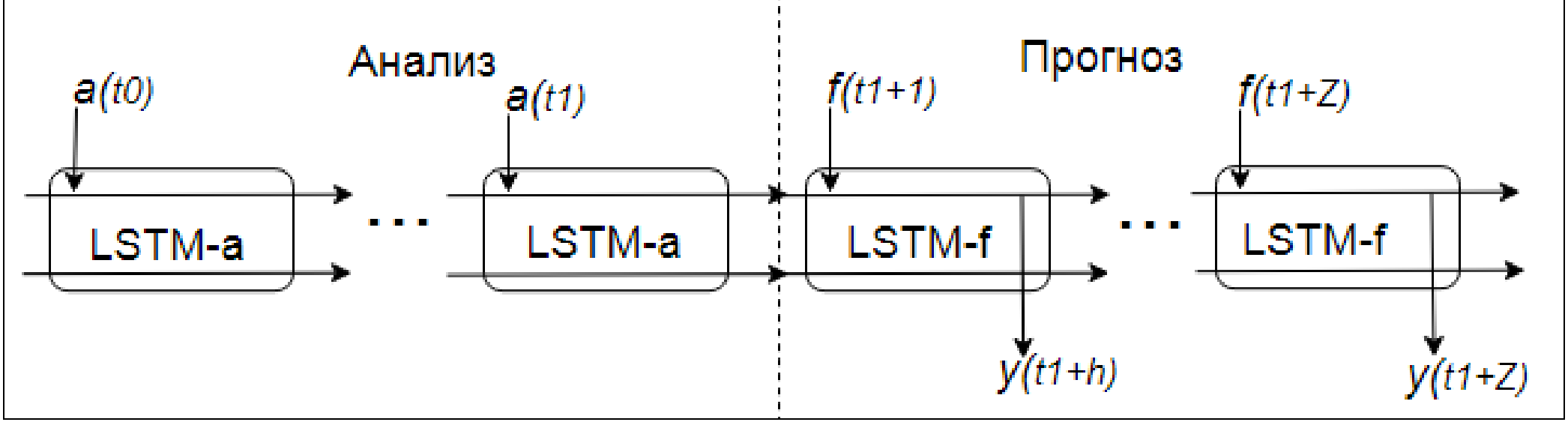


Предварительная оценка M_{i1} (сплошные) и среднее потребление от времени суток (пунктир). Тут для зеленого потребителя не было данных о потреблении в рабочие дни, а у синего – в нерабочие

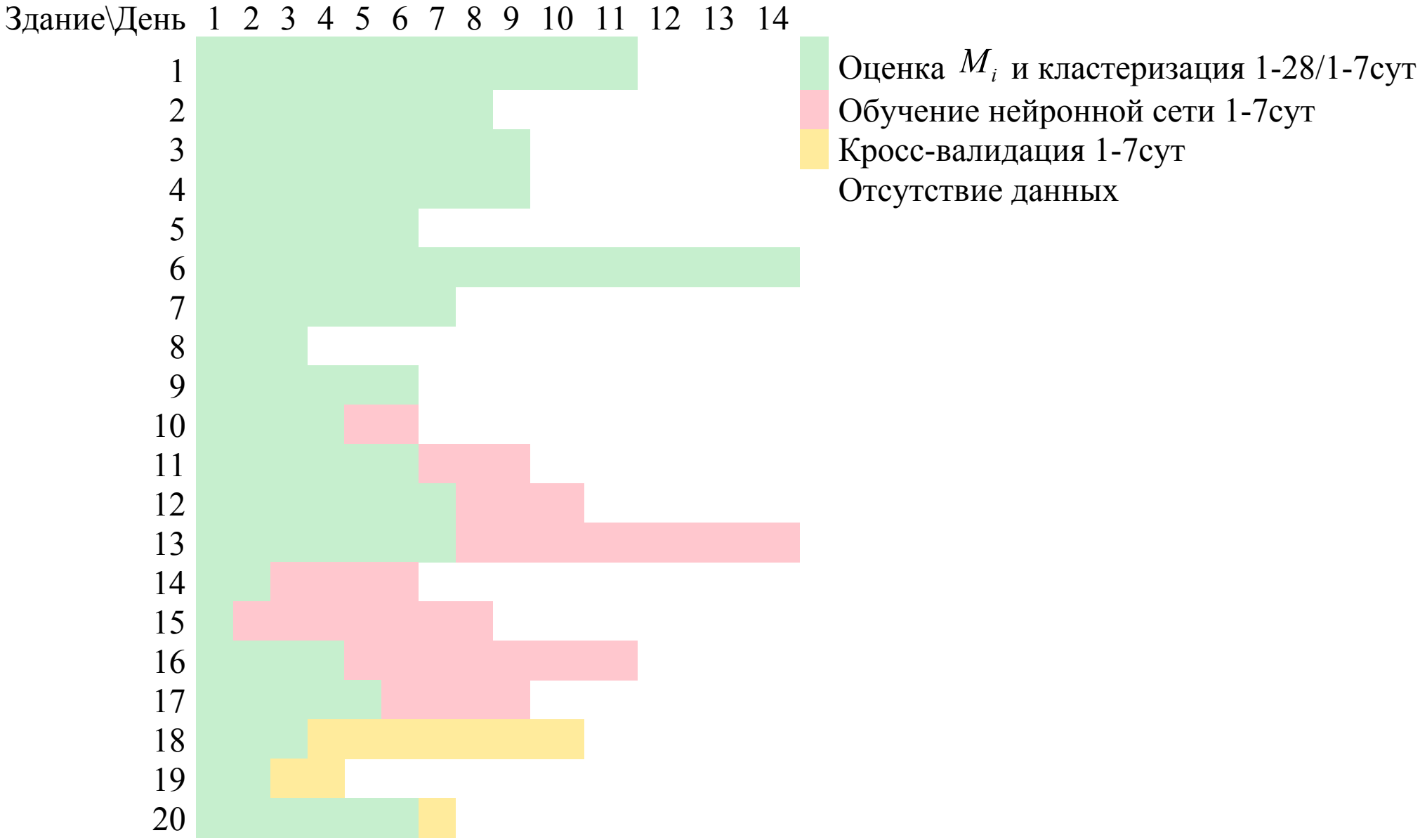
Схема анализ-прогноз с помощью нейронной сети типа LSTM [11][12]

Состоит из двух цепочек одинаковых блоков: блоков анализа и прогноза

Блок анализа принимает на вход вектор a_t, состоящий из:	Блок прогноза принимает на вход вектор f_t, состоящий из:
Фактическое потребление $Q_i(t+s)$, оценка потребления $M_i(t+s)$, градиент температуры $dT_i(t+s)$ при $s=-3,\ldots,0$ Индикаторы дней недели Время суток Индикаторы кластеров потребителей	Оценка потребления $M_i(t+s)$, градиент температуры $dT_i(t+s)$ при $s=-2,\ldots,2$ Индикаторы дней недели Время суток Индикаторы кластеров потребителей



Кросс-валидация (CV)



Результаты

Оценка M_i	Формула	Периодов*	Кластер.	Погрешность кросс-валлидации	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

*во всех случаях число параметров нейронных сетей лежало в пределах 4300-4600

**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 зданий - 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 http://www.atsenergo.ru/results/market/fact_region
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 http://method.meteorf.ru/ansambl/ansambl.html
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. – С. 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., Koutnık 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 https://www.drivendata.org/competitions/51
14. Competition: Power Laws: Cold Start Energy Forecasting https://www.drivendata.org/competitions/55

Спасибо за внимание!

Работа подготовлена в ходе проведения исследования (грант № 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»

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 MAPE

ОЭС	Чучуева, 2011г	Предлагаемый метод, 2015г
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%

*Границы ОЭС не совпадают с границами субъектов или мы не рассматривали некоторые из субъектов, входящих с ОЭС

Сеть с памятью LSTM

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

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

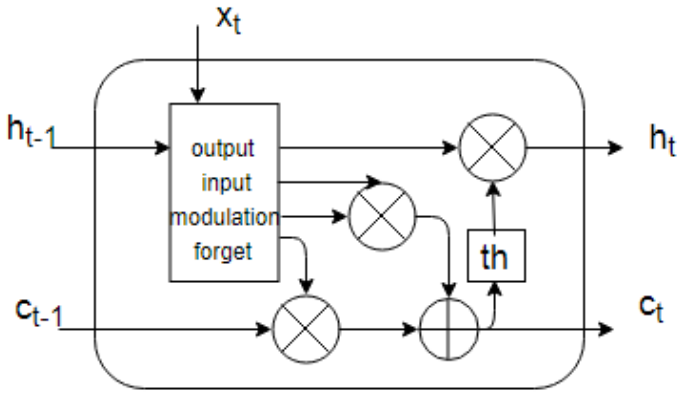
Входящий вентиль $i_t = \sigma\left(W_{ix} x_t + W_{ih} h_{t-1} + b_i\right)$

Модуляция $m_t = W_{mx} x_t + W_{mh} h_{t-1} + b_c$

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

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

Результат $h_t = \tanh(c_t) \circ o_t$



Градиент функции потерь затухает в соответствии с забывающим вентилем f_t:

$$\frac{\partial e_t}{\partial c_t} = f_{t+1} \circ \frac{\partial e_t}{\partial c_{t+1}} + o_t \circ \tanh'\left(c_t\right) \circ \frac{\partial e_t}{\partial h_t}$$

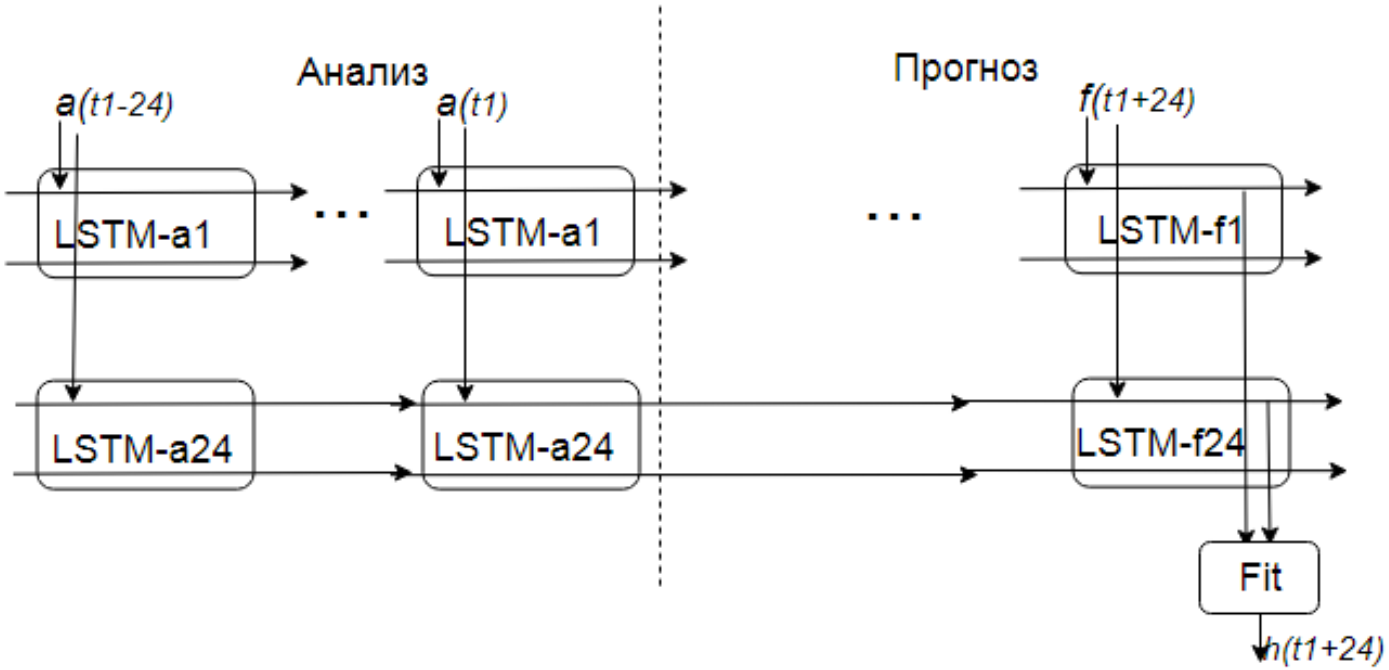
Можно научиться воспроизводить дальние зависимости

Инициализация b_f очень важна: временной масштаб процесса

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 $D(t) = \max\left(0, h_{i1}(t)\right)$
2. С учетом оценки M_i, то есть $C(t) = \max\left(0, \max\left(0, h_{i1}(t)\right) M_i(t) + h_{i2}(t)\right)$

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 «Adam с моментами Нестерова» [16], со скоростью обучения $\eta = 0,01$. Существенное увеличение η 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