



**Федеральное государственное бюджетное учреждение
«Гидрометеорологический научно-исследовательский центр Российской
Федерации»**



Быков Филипп Леонидович

Диссертация на соискание
ученой степени кандидата физико-математических наук

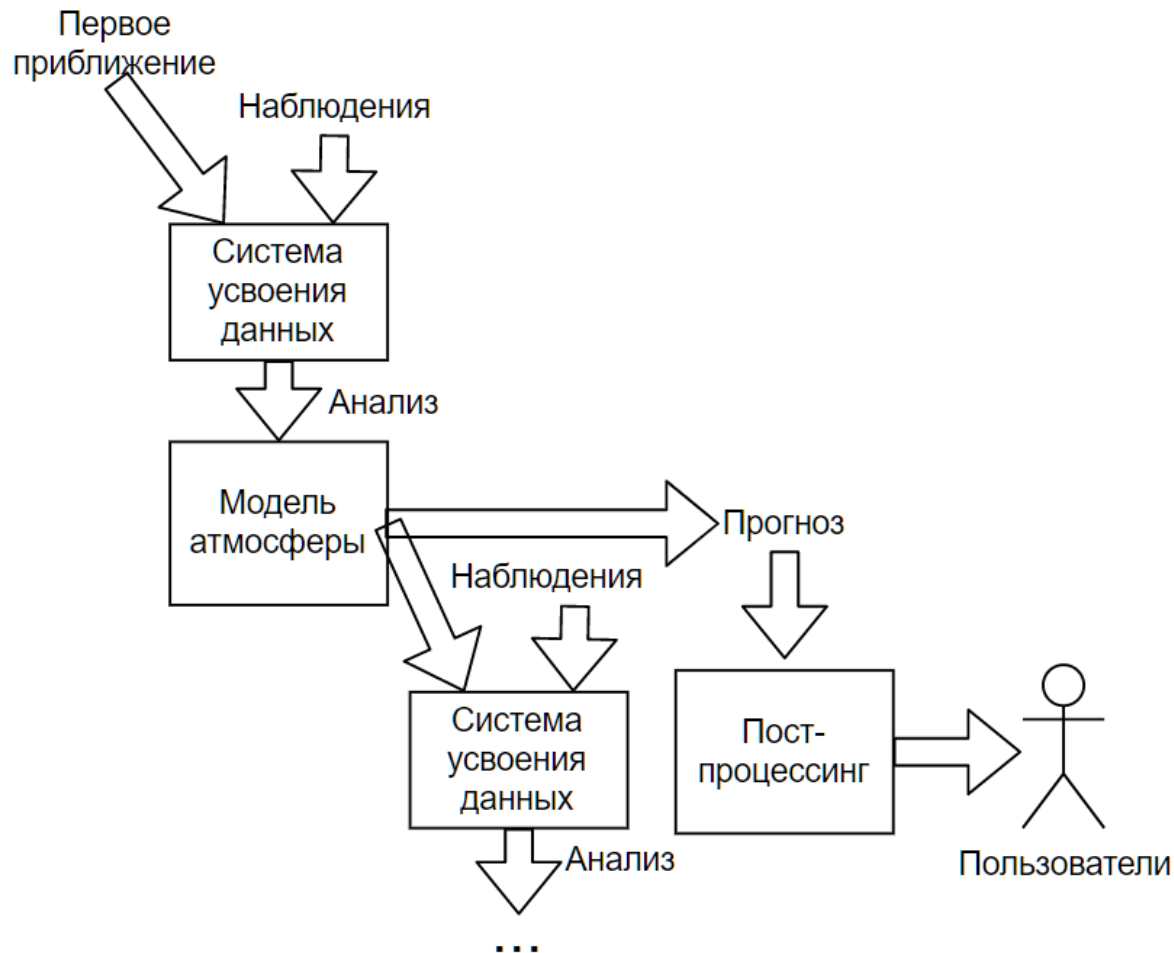
***Нейросетевые методы пост-процессинга полей
приземных метеорологических параметров***

Научный руководитель

д. ф.-м. н. В.А. Гордин

Москва 2021

Система прогноза погоды



Задачи пост-процессинга:

1. Коррекция (уточнение) прогноза;
2. Прогноз дополнительных параметров.

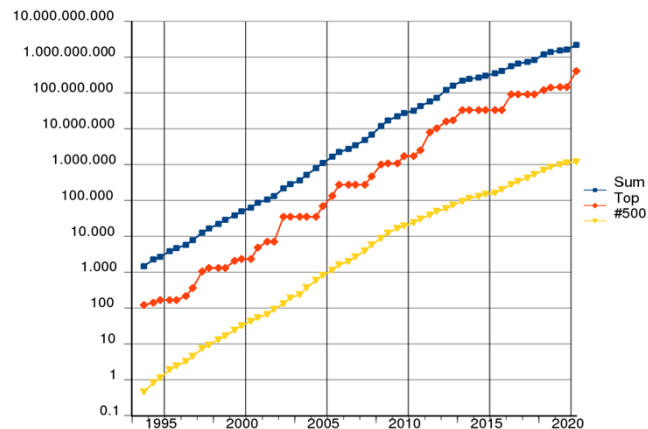
Подходы:

1. Точного прогноза;
2. Model output statistic.

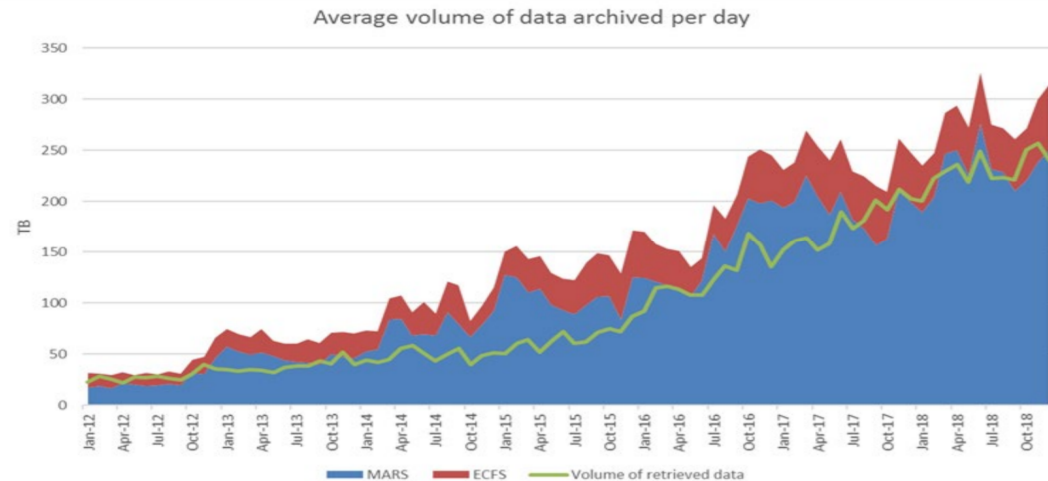
Будем решать задачу коррекции. Это возможно только в подходе MOS.

Актуальность темы

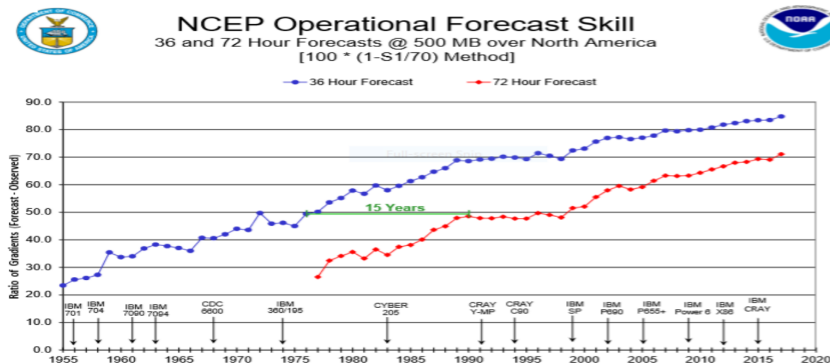
Производительность top500 растет
в 100 раз за 10 лет



Объем суточных данных в ECMWF вырос за 5 лет в 5 раз



Точность прогнозов NCEP (США) растет на сутки за 10 лет



Анализ растущего количества данных при
одновременном уменьшении
погрешностей прогноза
всё более сложен для человека.
Необходимы новые методы!

Цель исследования

Разработка нового автоматизированного комплекса нейросетевых методов пост-процессинга прогнозов приземных метеорологических параметров одной или нескольких различных численных моделей с учетом измерений на синоптических станциях за L последних суток.

Требования к комплексу методов:

- был применим **в случаях частичного отсутствия** прогностической информации и (или) данных наблюдений за период L ;
- давал прогнозы **и в точках синоптических станций, и в виде полей на модельной сетке**;
- обеспечивал качество прогнозов, заметно превосходящее качество прогнозов, откорректированных оперативными методами;
- был применим в оперативном режиме на доступных вычислительных системах.

Задачи исследования

Сформулированы математически и решены 3 задачи:

- Квазилинейная систематическая коррекция (комплексификация) прогнозов (глава 2)
учитывает известные смещения предыдущих прогнозов той же заблаговременности на рассматриваемой станции;
- Дополнительная коррекция с учетом предыдущих прогнозов (глава 3)
в том числе прогнозов большей заблаговременности на тот же срок;
- Интерполяция поправок из точек синоптических станций в точки модельной сетки (глава 4)
с оценкой неоднородностей поправок с помощью предикторов неоднородности.

Глава 1. Данные и методы

Приземные метеорологические параметры

Рассматриваются следующие метеопараметры:

- Температура на высоте 2 м T ($^{\circ}\text{C}$)
- Точка росы на высоте 2 м Td ($^{\circ}\text{C}$)
- Давление, приведенное к уровню моря $PMSL$ (гПа)
- Компоненты скорости ветра на высоте 10 м U и V (м/с)
- Модуль скорости ветра на высоте 10 м S (м/с)
- Скорость порывов ветра на высоте 10 м G (м/с)



Пусть t – начальный срок, τ – заблаговременность. Обозначим набор из этих 7 параметров согласно i -й модели как $\vec{\Psi}_i(t, \tau, \vec{x})$

При оптимизации всех параметров, кроме порывов ветра, использовалась функция потерь Хьюбера:

$$e_{huber}(X_{fact}, X_{pred}) = \begin{cases} c^{-1} (X_{pred} - X_{fact})^2 & \Leftarrow |X_{pred} - X_{fact}| < c; \\ 2|X_{pred} - X_{fact}| - c & \Leftarrow |X_{pred} - X_{fact}| \geq c, \end{cases}$$

где $c = 2^{\circ}\text{C} / 2 \text{ гПа} / 2 \text{ м/с}$.

Критерий прогноза порывов ветра

Согласно коду КН-01 SYNOP информация о порывах ветра G может не передаваться, если порывы не более 10 м/с. При оптимизации будем использовать:

$$\tilde{G}_{fact} = \begin{cases} G_{fact} & \Leftarrow \text{синоптическая станция передала } G_{fact}; \\ \min(1.9 \cdot |\vec{U}_{fact}|, 10.5 \text{ м/с}) & \Leftarrow \text{иначе,} \end{cases}$$

где число 1.9 получено регрессией G_{fact} по $|\vec{U}_{fact}|$.

Для оценки качества прогноза порывов используем критерий Багрова – Хайдке [Багров, 1959]

$$HSS(g) = 2 \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{31}a_{23} + a_{13}a_{32}} \rightarrow \max$$

Для применимости градиентного спуска

перейдем к сглаженной частотной матрице:

$$a_{lk\delta} = \sum_{j=1}^M \sigma_l \left[\delta(G_{pred,j} - g) \right] \sigma_k \left[\delta(\tilde{G}_{fact,j} - g) \right],$$

где M – количество прогнозов,

$$\delta^{-1} = 1 \text{ м/с}, \sigma_1(x) = \sigma(x), \sigma_2(x) = 1 - \sigma(x), \sigma_3(x) \equiv 1.$$

Итоговый критерий

$$G_{loss}(\tilde{G}_{fact}, G_{pred}) = - \sum_{g \in \{12, 18, 24 \text{ м/с} \}} HSS_{\delta}(g) \rightarrow \min$$

Частотная матрица	Наблюда- лось $\geq g$	Наблюда- лось $< g$	Всего
Прогнозиро- валось $\geq g$	a11	a21	a31
Прогнозиро- валось $< g$	a12	a22	a32
Всего	a13	a23	a33

Прогностические данные

Модель	Cosmo Ru6-ENA	PLAV- 2006	ICON-De	GFS	IFS	UKMO	JMA
Организация	Гидрометцентр России		DWD (Герма- ния)	NCEP (США)	ECMWF	MetOffice (Велико- британия)	JMA (Япония)
Шаг сетки	6.6 км	75км	0,25 ⁰	0,5 ⁰	0,5 ⁰	1 ⁰ 2,5 ⁰	1,25 ⁰
Заблаговре- менность, шаг	120 ч	120 ч	168 ч	180 ч	168 ч	48 ч 120 ч	72 ч
	3 ч	3 ч	6 ч	3 ч	6 ч	6 ч 12 ч	6 ч
Используемые поля	T, Td, U, V, P, G	T, Td, U, V, P	T, U, V, P, G	T, Td, U, V, P, G	T, U, V, P, G	T, Td, U, V, P	T, Td, U, V, P
Пар прогноз- факт в год*	6.3×10^7	6.8×10^7	2.2×10^7	8.8×10^7	3.7×10^7	1.5×10^7	2.2×10^7
Надежность	97.7%	86.0%	99.2%	99.5%	98.5%	99.7%	97.5%

*Рассматривались заблаговременности до 144 ч

Данные доступны благодаря А.Н.Багрову, И.И.Жабиной, И.И.Кулаковой, А.Ю.Недачиной, Ю.А.Степанову

Валидация и тестирование

Валидация гиперпараметров всегда проводилась на независимом периоде в 365 суток.

Тестирование проводилось в оперативном режиме (на сегодняшних прогнозах).

Обучение	Тестирование
04/2017-03/2019	04/2019-09/2020
Скользющий последний год (1 раз в 2 недели)	10/2020-01/2021
Скользящие 2 последних года (1 раз в 2 недели)	Начиная с 02/2021

Глава 2. Систематическая коррекция и комплексификация

Систематическая коррекция (часть 1 главы 2)

Пусть известны наблюдения на синоптической станции $X_{fact}(t)$ за параметром $X=T, Td, U, V, S, G$. А значит и смещения соответствующих прогнозов

$$BIAS_X(t - \delta t, \tau) = X_{fact}(t - \delta t + \tau) - X(t - \delta t, \tau)$$

Тогда можно применить систематическую коррекцию: $\sum_{\delta t=\lfloor \tau \rfloor}^L \exp(w_{\delta t}) BIAS_X(t - \delta t, \tau)$

$$X_{sc}(t, \tau) = X(t, \tau) + \frac{\sum_{\delta t=\lfloor \tau \rfloor}^L \exp(w_{\delta t}) BIAS_X(t - \delta t, \tau)}{\sum_{\delta t=\lfloor \tau \rfloor}^L \exp(w_{\delta t})}, \quad (2.1)$$

где L – период обучения систематической коррекции (использовалось $L=35$ дней). Коррекция (2.1) применяется, если доступны данные наблюдений за не менее 25 из 35 дней.

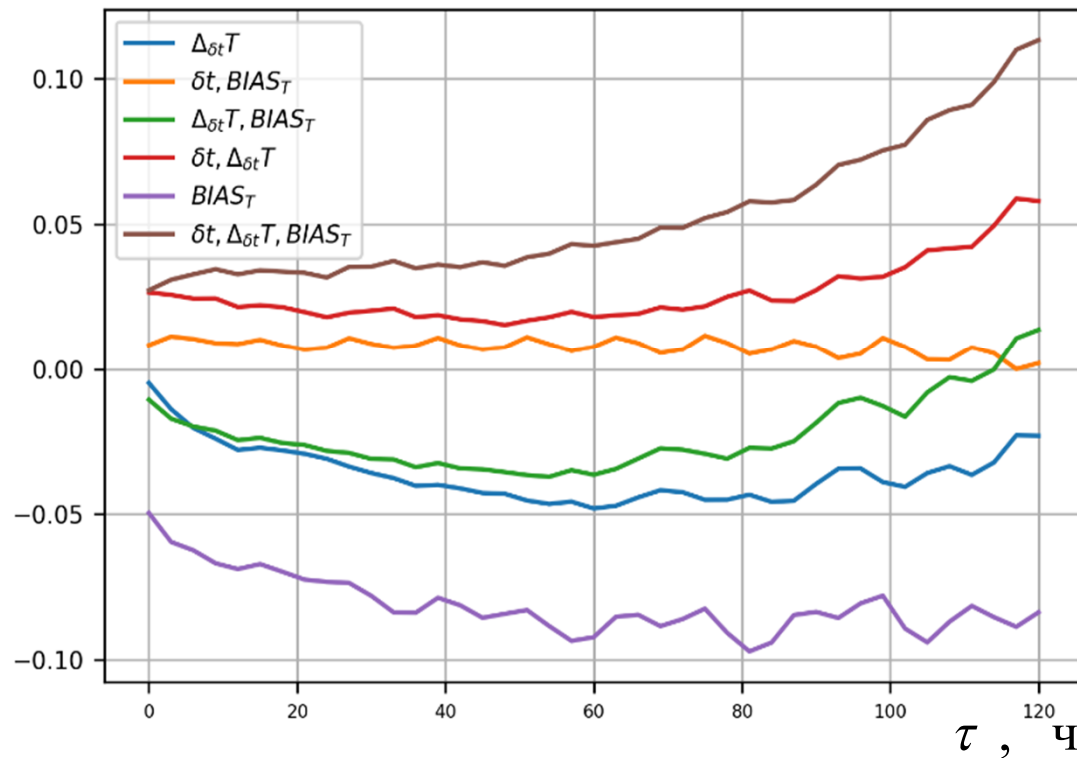
Классический подход: $w_{\delta t} \equiv 0$ или $w_{\delta t} = -\lambda \delta t, \lambda = const$

Предлагаем: вычислять веса $w_{\delta t}$ нейронной сетью:

$$w_{\delta t} = w_{NeuralNet}(\delta t, \tau, \vec{\Psi}(t, \tau) - \vec{\Psi}(t - \delta t, \tau), BIAS_X(t - \delta t, \tau))$$

Если веса $w_{\delta t}$ зависят от смещений $BIAS_X(t - \delta t, \tau)$, то (2.1) квазилинейна (далее квазилинейная систематическая коррекция, КСК)

Сравнение линейной и квазилинейной систематической коррекции



Выигрыш в RMS (больше - лучше) в зависимости от заблаговременности при систематической коррекции T_{2m} с различными наборами аргументов нейронной сети $w_{\delta t}$ по сравнению с зависящей только от сдвига по времени

$$w_{\delta t} = w(\delta t)$$

Квазилинейная КСК более эффективна, чем **линейная**

$$\Delta_{\delta t} T = T_{2m}(t, \tau) - T_{2m}(t - \delta t, \tau)$$

Комплексификация (часть 2 главы 2)

Рассмотрим несколько прогнозов, погрешности которых $X_{fact}(t+\tau) - X_i(t, \tau)$ независимы.

Тогда их оптимальной линейной комбинацией является

$$X_{comp}(t, \tau, I(t, \tau)) = \sum_{i \in I(t, \tau)} \frac{X_i(t, \tau) + \mu_i(t, \tau)}{\sigma_i^2(t, \tau)} \bigg/ \sum_{i \in I(t, \tau)} \sigma_i^{-2}(t, \tau),$$

где $I(t, \tau)$ – список моделей, $\mu_i(t, \tau)$ и $\sigma_i^2(t, \tau)$ – среднее и дисперсия погрешностей i -й модели.

Оценим μ_i и σ_i^2 как взвешенное среднее по последним смещениям моделей: $\mu_i(t, \tau) = R\{BIAS_{X_i}(t, \tau)\}$,

$$\sigma_i^2(t, \tau) = \alpha(\tau)^2 + R\{BIAS_{X_i}(t, \tau)^2\} - (\mu_i(t, \tau))^2, R\{B(t)\} = \sum_{\delta t = \lfloor \tau \rfloor}^L \exp[w_{\delta t}] B(t - \delta t) \bigg/ \sum_{\delta t = \lfloor \tau \rfloor}^L \exp[w_{\delta t}],$$

где $\alpha(\tau)$ – регуляризация. Обучение комплексного прогноза заключается в оптимизации весов $w_{\delta t}$, вычисляемых с использованием нейронной сети:

$$w_{\delta t} = \left[\sum_{i \in I(t, \tau)} \rho_{it} \right]^{-1} \sum_{i \in I(t, \tau)} \rho_{it} w_{NeuralNet}(\vec{\theta}, \delta t, \tau, \vec{\Psi}_i(t, \tau) - \vec{\Psi}_i(t - \delta t, \tau), BIAS_{X_i}(t - \delta t, \tau), i)$$

Одновременно оптимизируются веса ρ_{it} моделей в $w_{\delta t}$, регуляризации α_τ и параметры $\vec{\theta}$ нейронной сети $w_{NeuralNet}$:

$$\sum_{t, \tau, \vec{x}} e(X_{comp}, X_{fact}) \rightarrow \min_{\vec{\theta}, \vec{\alpha}, \vec{\rho}}$$

Дополнительная коррекция (глава 3)

Предикторы (всего 123 штуки) при коррекции прогноза с заблаговременностью 24 ч :

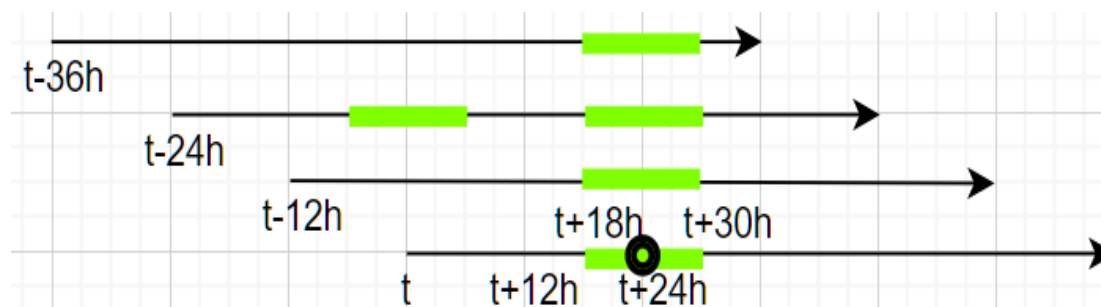
1. Прогнозы (всех рассматриваемых параметров) от **начальных сроков и с заблаговременностями** согласно рис.

2. Заблаговременность

3. Синус широты

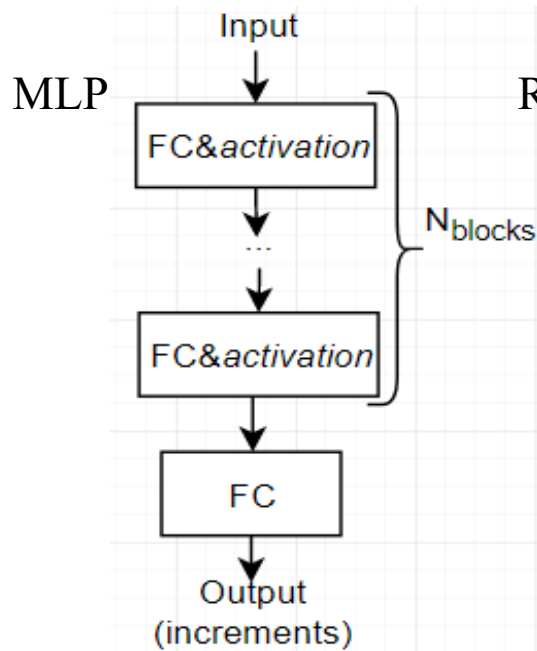
4. Синус (местного) времени суток

5. Высота Солнца над горизонтом

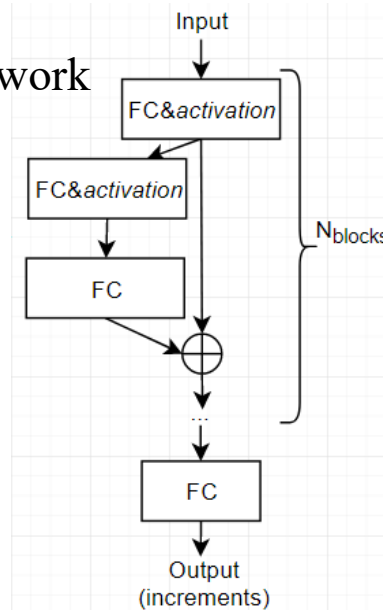


Для больших заблаговременностей кол-во предикторов сокращается: нет предыдущих прогнозов на тот же срок

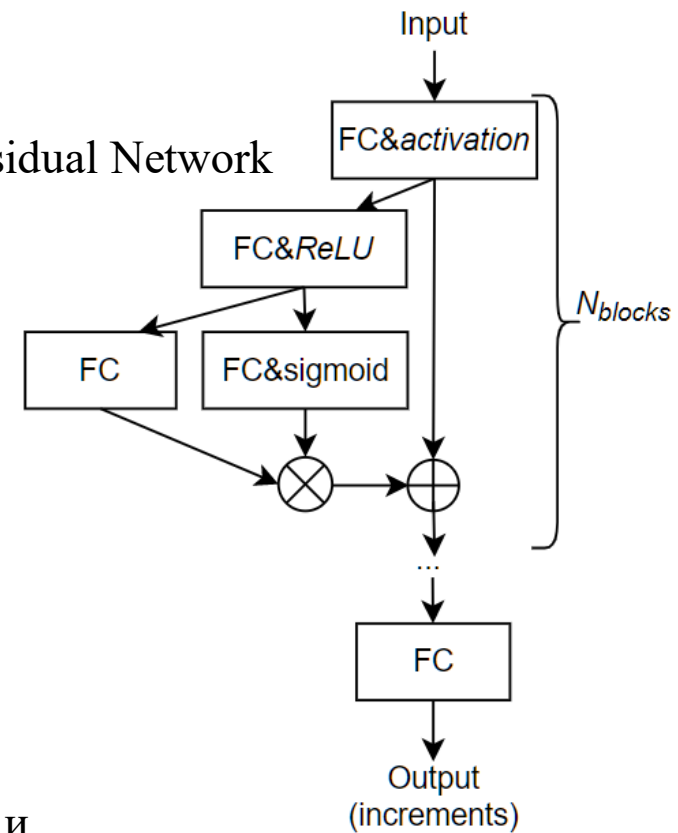
Оптимизация архитектуры



Residual Network



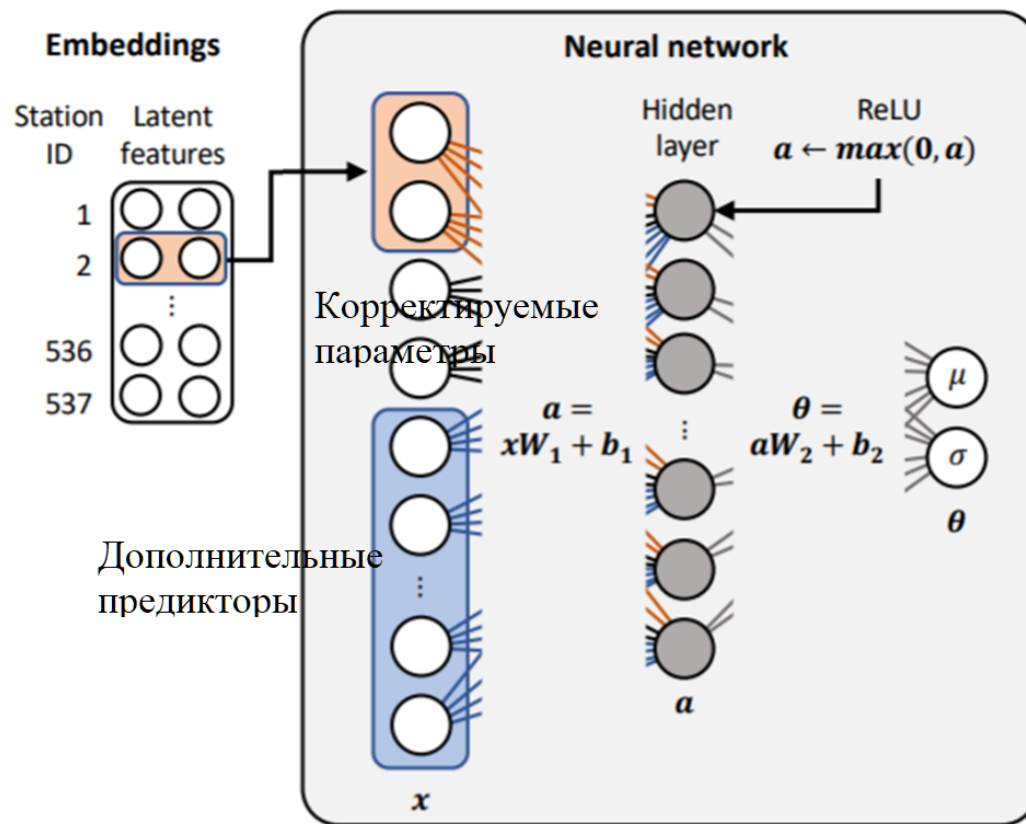
Gated Residual Network



Без систематической коррекции: MLP 64-32 с ReLU активациями

С коррекцией: GRN 24-16 с tanh активациями при коррекции T, Td, U и
GRN 12-8 с tanh активациями при коррекции давления P

Индивидуальные параметры для станций



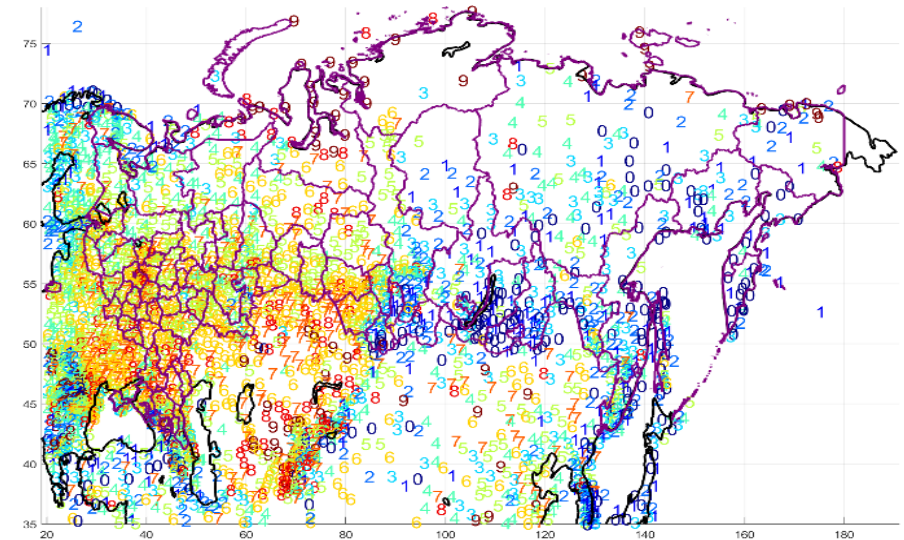
Raw ensemble		1.16	1.16
<i>Benchmark post-processing methods</i>			
EMOS-gl	Global EMOS	1.01	1.00
EMOS-loc	Local EMOS	0.90	0.90
EMOS-loc-bst	Local EMOS with boosting	0.85	0.80
QRF	Local quantile regression forest	0.95	0.81
<i>Neural network models</i>			
FCN	Fully connected network	1.01	1.01
FCN-aux	... with auxiliary predictors	0.92	0.91
FCN-emb	... with station embeddings	0.91	0.91
FCN-aux-emb	... with both of the above	0.88	0.87
NN-aux	1-hidden layer neural network with auxiliary predictors	0.90	0.86
NN-aux-emb	... and station embeddings	0.82	0.78

Коррекция 48 ч ансамблевых прогнозов $T2m$ для 537 станций в Германии [Rasp, Lerch, 2018]

Индивидуальные параметры

При коррекции оригинальных прогнозов COSMO-Ru индивидуальные параметры зависят от географического района:

Уменьшение RMS при коррекции с E индивидуальными параметрами:

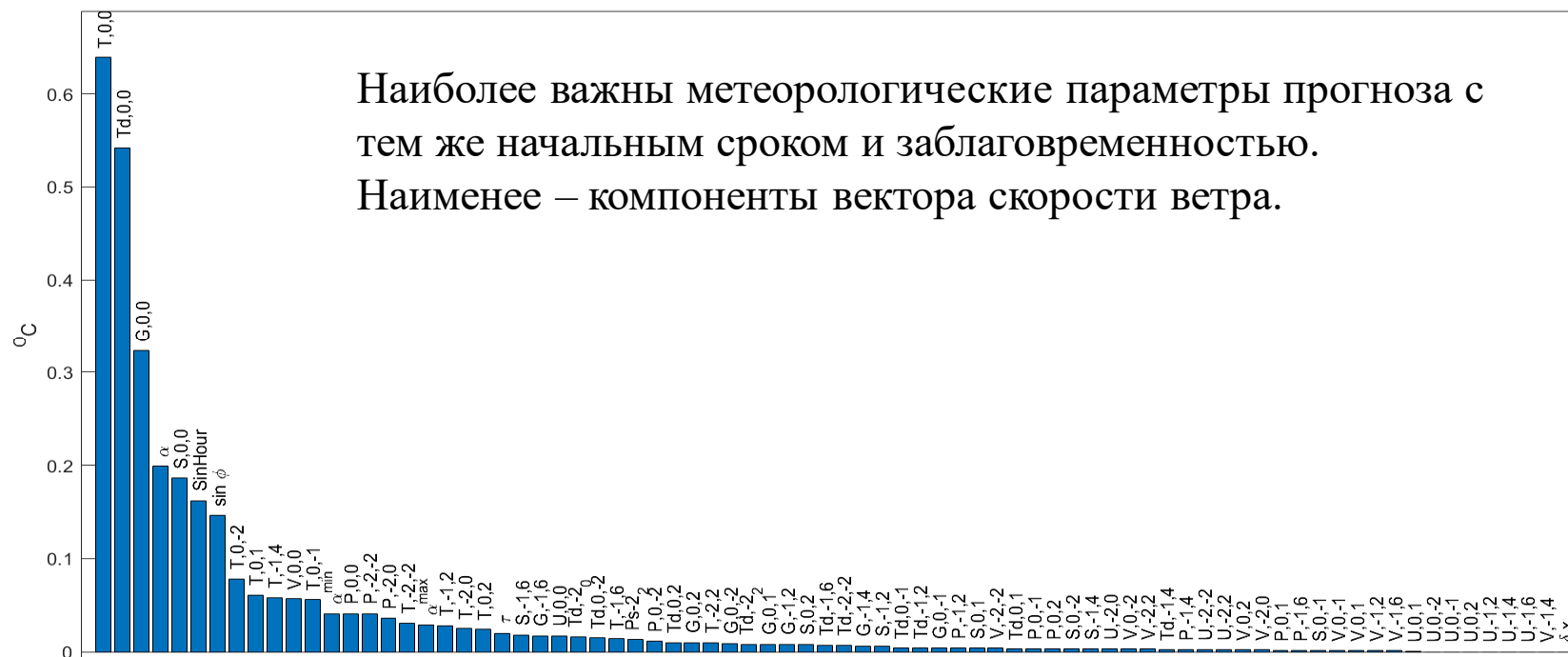


COSMO-Ru						COSMO-Ru+KCK				
E	0	1	2	3	4	0	1	2	3	4
$T, ^\circ\text{C}$	0.290	0.315	0.345	0.349	0.364	0.128	0.139	0.142	0.144	0.147
$Td, ^\circ\text{C}$	0.379	0.410	0.438	0.443	0.457	0.120	0.124	0.122	0.125	0.127
$P, \text{гПа}$	0.291	0.329	0.334	0.325	0.354	0.012	0.015	0.012	0.012	0.013
$U, \text{м/с}$	0.291	0.374	0.388	0.386	0.390	0.038	0.038	0.039	0.038	0.038

Индивидуальные параметры **неэффективны** после систематической коррекции

Оценка важности предикторов

Оценка важности (методом permutation importance) предикторов при коррекции T2m



Глава 4. Интерполяция поправок по горизонтальным координатам

Оптимальная интерполяция [Гандин, 1963]

Будем интерполировать поправки со станций $\vec{x}_i$ на модельную сетку.

Рассмотрим задачу интерполяции случайного поля $F(\vec{x}), \vec{x} \in \Omega \subset \mathbb{S}^2$ со средним $\mu(\vec{x})$, стандартным отклонением $\sigma(\vec{x})$ и нормированной корреляционной функцией (НКФ) $K(\vec{x}, \vec{y})$.

Предположим, что результат интерполяции в точку $\vec{y}$ вычисляется по линейной формуле

$$\hat{F}(\vec{y}) = \mu(\vec{y}) + w_x(\vec{y})^T [F_x - \mu_x], \quad (4.1)$$

Минимизируя MSE , интерполяционные веса в (4.1) вычисляются по формулам многомерной регрессии:

$$w_x(\vec{y}) = \sigma_x^{-1} \circ K_{xx}^{-1} K_{xy} \circ \sigma(\vec{y})$$

Используемая оценка НКФ K должна быть ядром симметричного положительно определенного оператора.

В 95% случаев сумма весов $\sum_{i \in I} w_i(\vec{y})$ лежит на отрезке $[0.98, 1.03]$, т.о. $\hat{F}$ зависит от относительных значений $\mu(\vec{x}) - \mu(\vec{y})$ и $\sigma(\vec{x}) / \sigma(\vec{y})$.

Погрешность интерполяции и ковариационная модель

Средняя погрешность в точке «пропущенной» j -й станции (leave-one-out):

$$L(\mu(\vec{x}), \sigma(\vec{x}), K(\vec{x}, \vec{y})) = \frac{1}{N} \sum_{j=1 \dots N} e(F(\vec{x}_j), \hat{F}_j(\vec{x}_j)), \quad (4.2)$$

где e – функция потерь, $\hat{F}_j$ – интерполяция без использования станции $\vec{x}_j$.

Функционал (4.2) зависит от относительных значений $\mu(\vec{x}) - \mu(\vec{y})$, $\sigma(\vec{x}) / \sigma(\vec{y})$

то есть от локальной изменчивости полей среднего и дисперсии.

Как правило предполагают, что поле $X(\vec{x})$ – гауссовское и оптимизируют логарифм правдоподобия гауссовской ковариационной модели

$$L_{gauss}(\mu(\vec{x}), \sigma(\vec{x}), K(\vec{x}, \vec{y})) = \frac{1}{2} \left[\bar{F}_x^T K_{xx}^{-1} \bar{F}_x + \ln \det K_{xx} + \dim K_{xx} \ln 2\pi \right], \bar{F}_x = \frac{F_x - \mu}{\sigma}. \quad (4.3)$$

Функционал (4.3) зависит и от абсолютных значений $\mu(\vec{x}), \sigma(\vec{x})$, то есть и от их крупномасштабной изменчивости. При оптимизации (4.3) достаточно рассматривать лишь пары точек, то есть матрицы K_{xx} размера 2×2 .

Теорема Мерсера [1909] и выравнивающее отображение (kernel trick)

Пусть $\Omega \subset \mathbb{R}^n$ – связный компакт. Если $K(\vec{x}, \vec{y}), \vec{x}, \vec{y} \in \Omega$ – ядро симметричного положительно определенного оператора представимо в виде почти всюду сходящегося ряда

$$K(\vec{x}, \vec{y}) = \sum_{j=1}^{\infty} \lambda_j e_j(\vec{x}) e_j(\vec{y})$$

где e_j – ортонормированный базис, $\lambda_j > 0$

$g(\vec{x}) = (e_1(\vec{x})\sqrt{\lambda_1}, e_2(\vec{x})\sqrt{\lambda_2}, \dots)$ – выравнивающее отображение в $H = L^2(\Omega)$ [Aronszajn, 1950]

Если $g: \Omega \rightarrow H$, то $\langle g(\vec{x}), g(\vec{y}) \rangle_H$ – ядро симметричного положительно определенного оператора

Если $K(\vec{x}, \vec{y})$ – нормир. корреляционная ф-я, то g переводит поле F в однородное и изотропное:

$$K(\vec{x}, \vec{y}) = \langle g(\vec{x}), g(\vec{y}) \rangle_H = 1 - \frac{1}{2} \|g(\vec{x}) - g(\vec{y})\|_H^2.$$

Если K – липшицево и $\Omega = \mathbb{S}^{n-1} \Rightarrow \exists \delta > 0$: последовательность $j^{1+1/(n-1)+\delta} \lambda_j$ ограничена [Ferreira, 2008], то есть g хорошо приближается отображением с конечномерным образом.

Оптимизация квазилинейной неоднородной оптимальной интерполяции (КНОИ)

Предположим, что выравнивающее отображение представимо в виде графика нейронной сети $\varpi(\Theta(t, \vec{x}))$, принимающей на вход предикторы неоднородности $\Theta(t, \vec{x})$:

$$g(\vec{x}) = (\vec{x}, \varpi(\Theta(t, \vec{x})))$$

тогда КФ вычисляется как $K(\vec{x}, \vec{y}) = K_g(|g(\vec{x}) - g(\vec{y})|, \vec{v}) = K_g\left(\sqrt{|\vec{x} - \vec{y}|^2 + |\varpi(\Theta(t, \vec{x})) - \varpi(\Theta(t, \vec{y}))|^2}, \vec{v}\right)$

Будем минимизировать погрешность интерполяции $L(\mu(\vec{x}), \sigma(\vec{x}), K(\vec{x}, \vec{y}))$, оптимизируя параметры нейронных сетей $\varpi(\Theta(t, \vec{x})), \mu(\Theta(t, \vec{x})), \sigma(\Theta(t, \vec{x}))$ и 4 параметра $\vec{v} = (\varepsilon, \beta, R_1, R_2)$ семейства НКФ (I_0 – индикатор 0):

$$K_g(r, \vec{v}) = \varepsilon I_0(r) + (1 - \varepsilon) \left[\beta \exp\left(-r^2/2R_2^2\right) + (1 - \beta)(1 + r/R_1) \exp(-r/R_1) \right]$$

Чтобы отказаться от линейности интерполяции по известным значениям поля F , добавим в список аргументов нейронных сетей величину $\tilde{F}(\vec{x})$ – результат интерполяции в предположениях однородности, изотропности и линейности: $\varpi(\vec{x}) = \varpi(\Theta(t, \vec{x}), \tilde{F}(\vec{x}))$

Вычисление интерполяционных весов

При вычислении интерполяционных коэффициентов w_x необходимо решать СЛАУ $K_{xx} w_x = K_{xy}$

При обратном распространении ошибки необходимо вычислять градиенты:

$$\nabla_{K_{xy}} e = K_{xx}^{-1} \nabla_{w_x} e, \nabla_{K_{xx}} e = -w_x \nabla_{K_{xy}}^T e \quad (4)$$

Градиенты (4) ограничены $2 \|K_{xx}^{-1}\|^2 \max_i |F(\vec{x}_i)|$. Если матрица K_{xx} плохо обусловлена, то обучение (сходимость градиентного спуска) будет очень медленной.

Если выполнено $\|w_x\| = \|K_{xx}^{-1} K_{xy}\| < C$, то градиенты (4) не более $2C \|K_{xx}^{-1}\| \max_i |F(\vec{x}_i)|$.

Будем вычислять интерполяционные веса по модифицированной формуле:

$$\hat{w}_i(\vec{y}) = \begin{cases} w_i(\vec{y}), & \text{если } \|w_x\| < C = 1.5; \\ \frac{K(\vec{x}_i, \vec{y})}{\mathcal{G}_i} \bigg/ \max \left(1, \sum_{i \in I(\vec{y})} \frac{K(\vec{x}_i, \vec{y})}{\mathcal{G}_i} \right), & \text{иначе,} \end{cases}$$

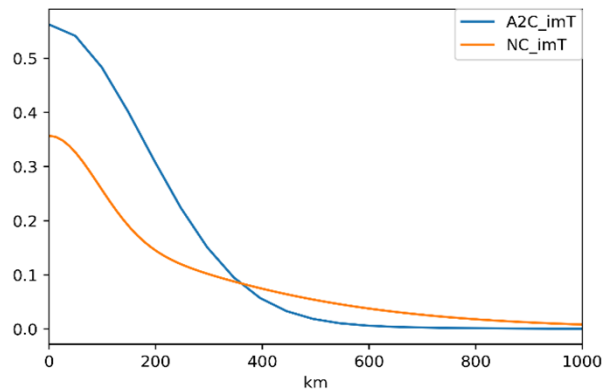
где $\mathcal{G}_i = \sum_{k \in I(\vec{y})} K(\vec{x}_k, \vec{x}_i) > 1$ – плотность наблюдений в окрестности i -й станции.

Оптимальная интерполяция (ОИ)	Квазилинейная неоднородная ОИ (КНОИ)
Однородность и изотропность	Неоднородности описываются предикторами
Поле – гауссовское	Прямая оптимизация погрешности
Линейность оператора интерполяции	Квазилинейность оператора интерполяции

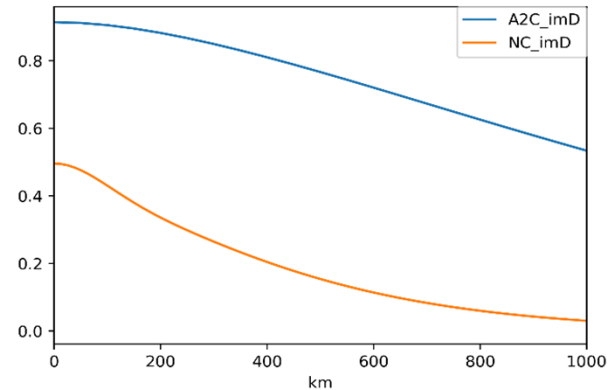
Используемые предикторы неоднородности

- Первое приближение
- Высота над уровнем моря
- Синус времени года
- Высота Солнца над горизонтом
- Результат интерполяции (в предположениях однородности, изотропности и линейности)

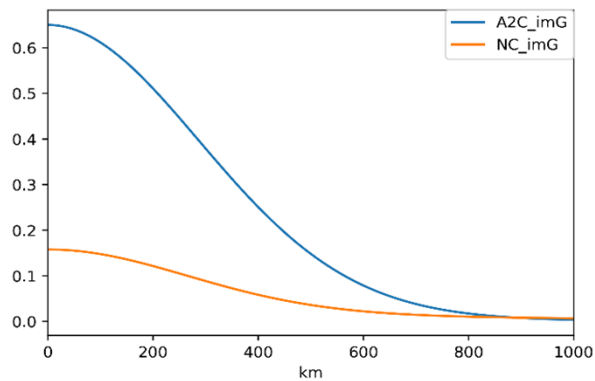
Корреляционные функции



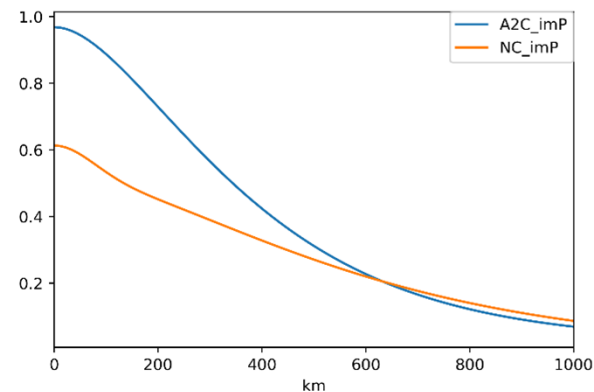
T2m



Td2m



U10m



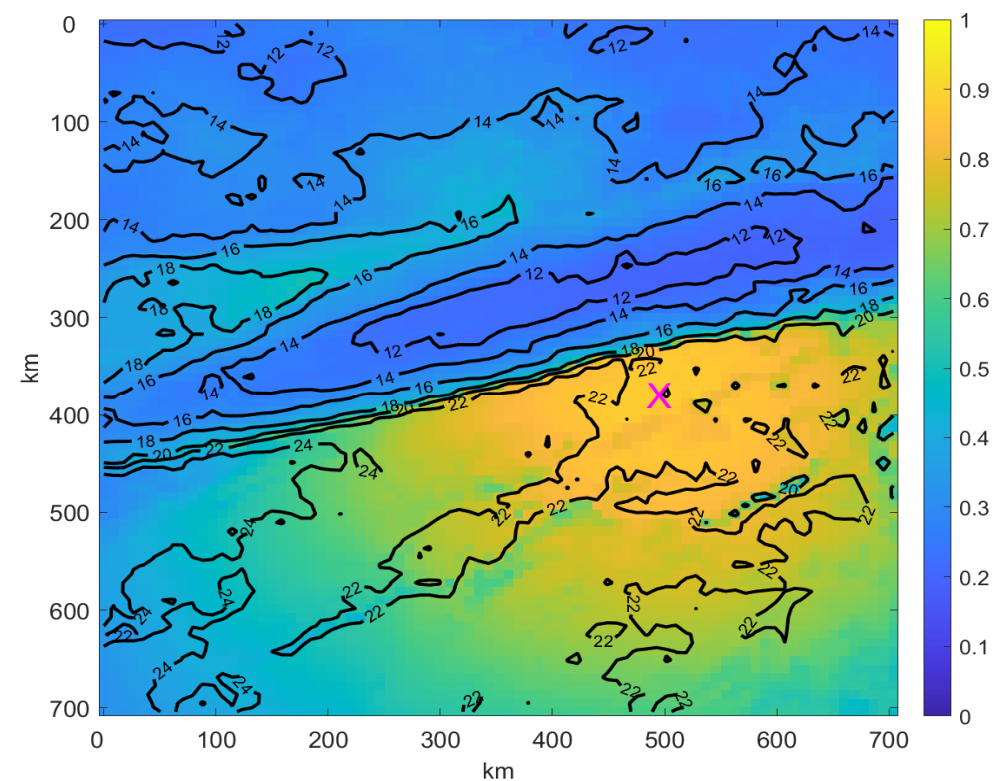
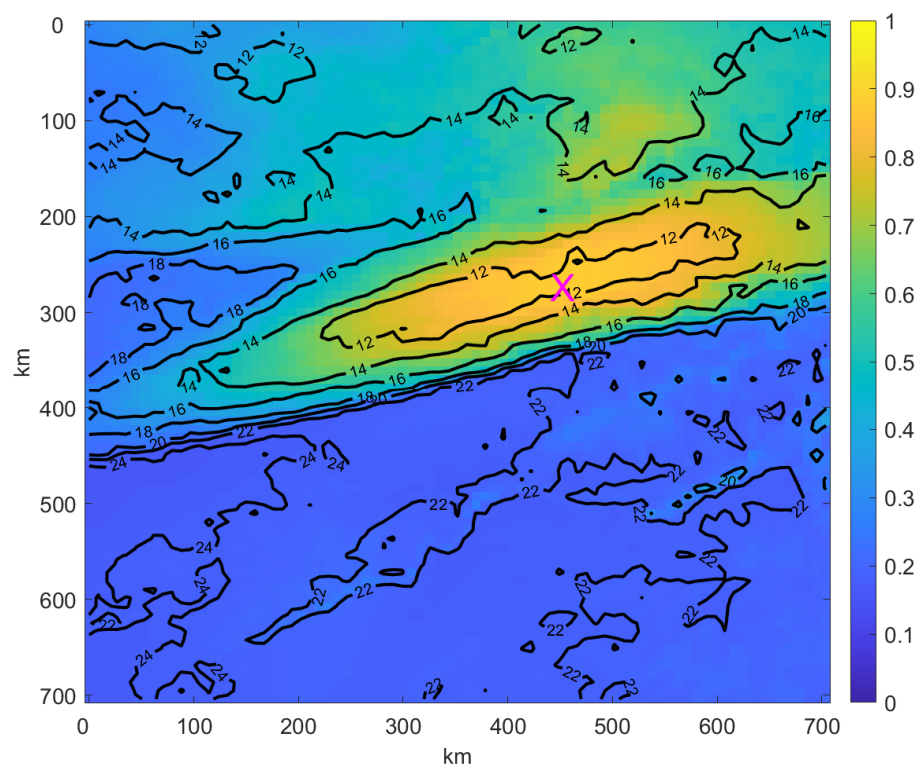
PMSL

КНОИ позволяет объяснить большую часть вариации, чем **ОИ**.

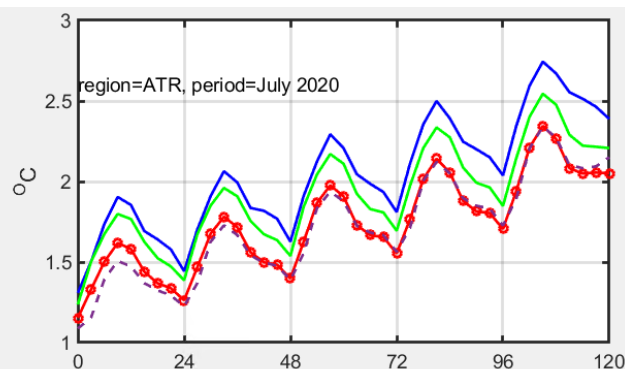
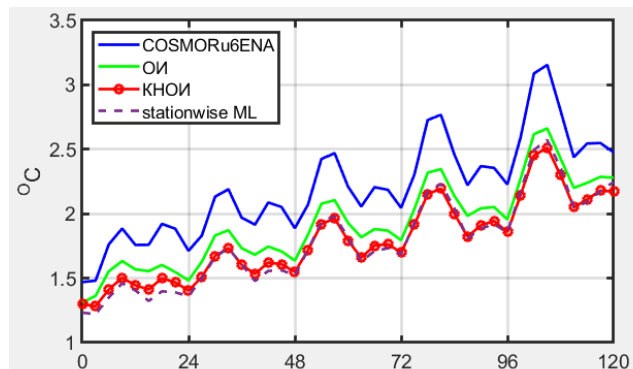
Для **КНОИ** по оси ординат – расстояние в расширенном пространстве.

Пример корреляционной функции

Значения КФ $K(\vec{x}, \vec{y})$ поля поправок к температуре $T2m$ при отмеченном $\times$ и $\vec{y}$, пробегающей по области вблизи атмосферного фронта. Изолинии – поле температуры из модели COSMO-RuBy



Сравнение качества интерполяции

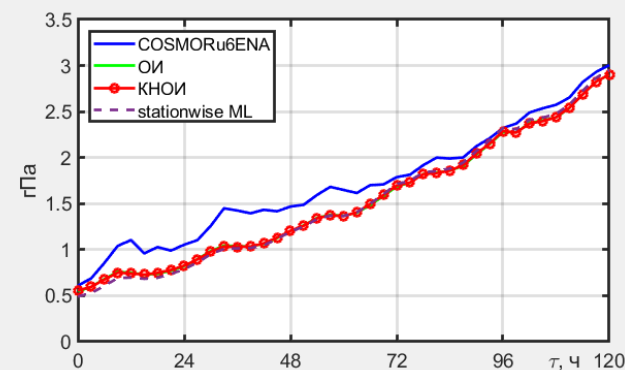
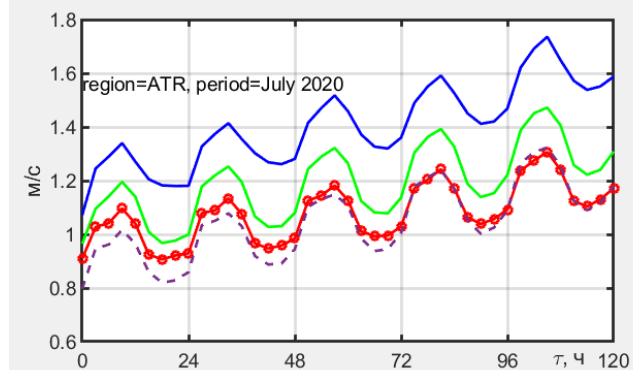


За июль 2020 г в АТР.

stationwise ML — пост-процессинг, использующий наблюдения на данной станции. **ОИ** и **КНОИ** — сравнение на «пропущенной» станции.

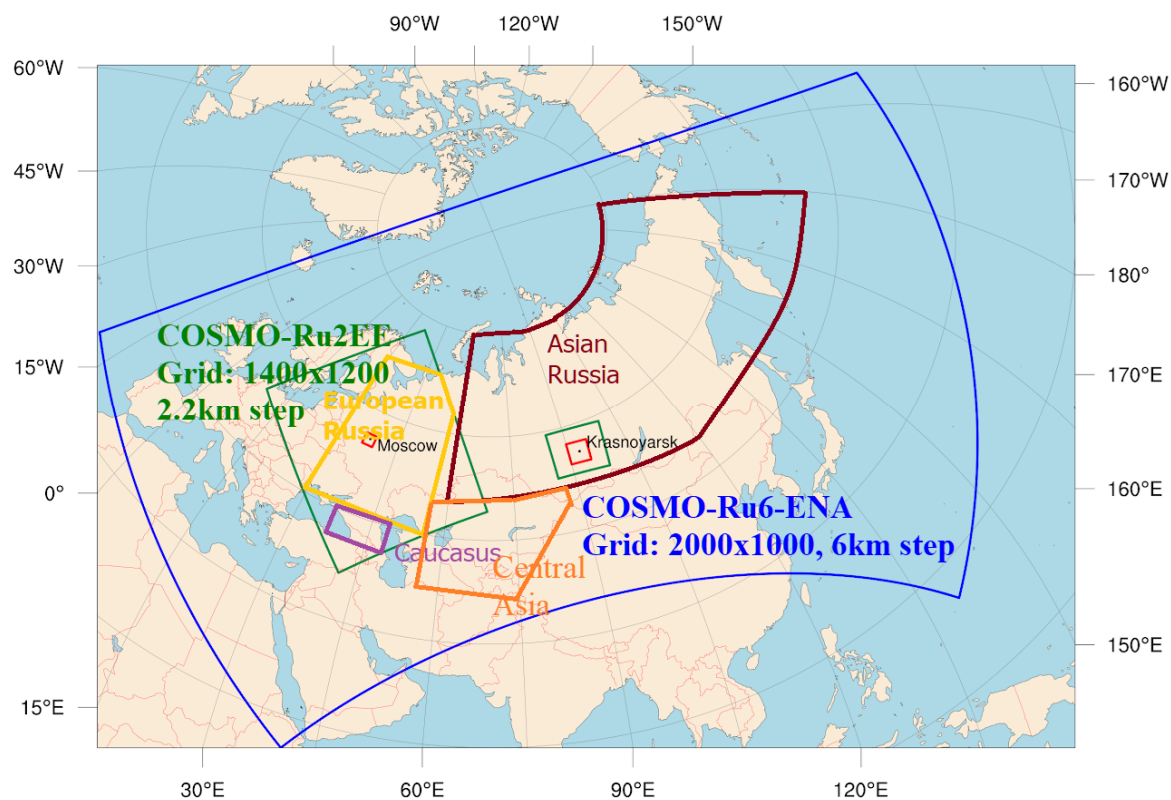
Предположение однородности и изотропности поправок состоятельно:

1. Для давления *PMSL*;
2. На равнинных территориях в зимний период.

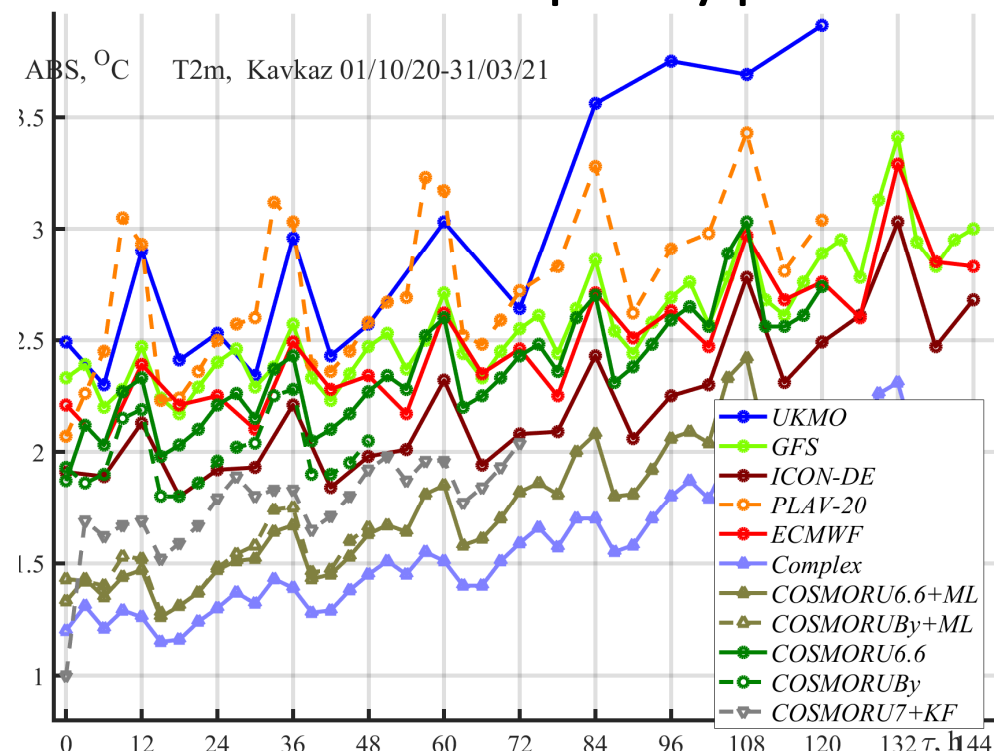
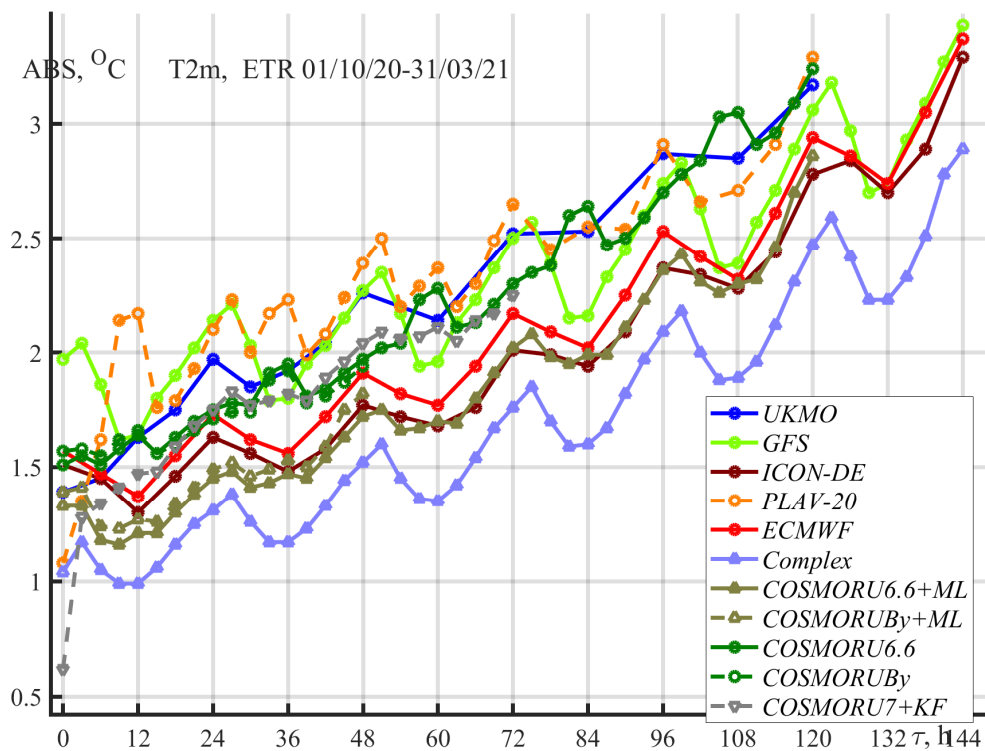


Глава 5. Сравнительные оценки качества прогнозов

Регионы для оценок и области COSMO-Ru



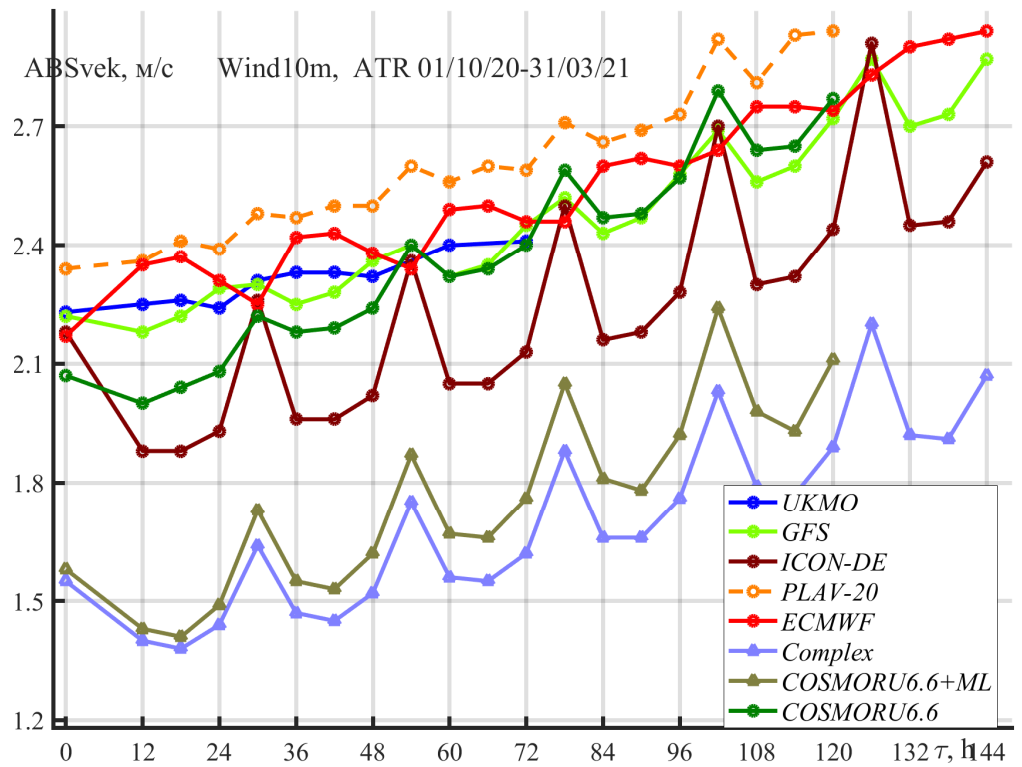
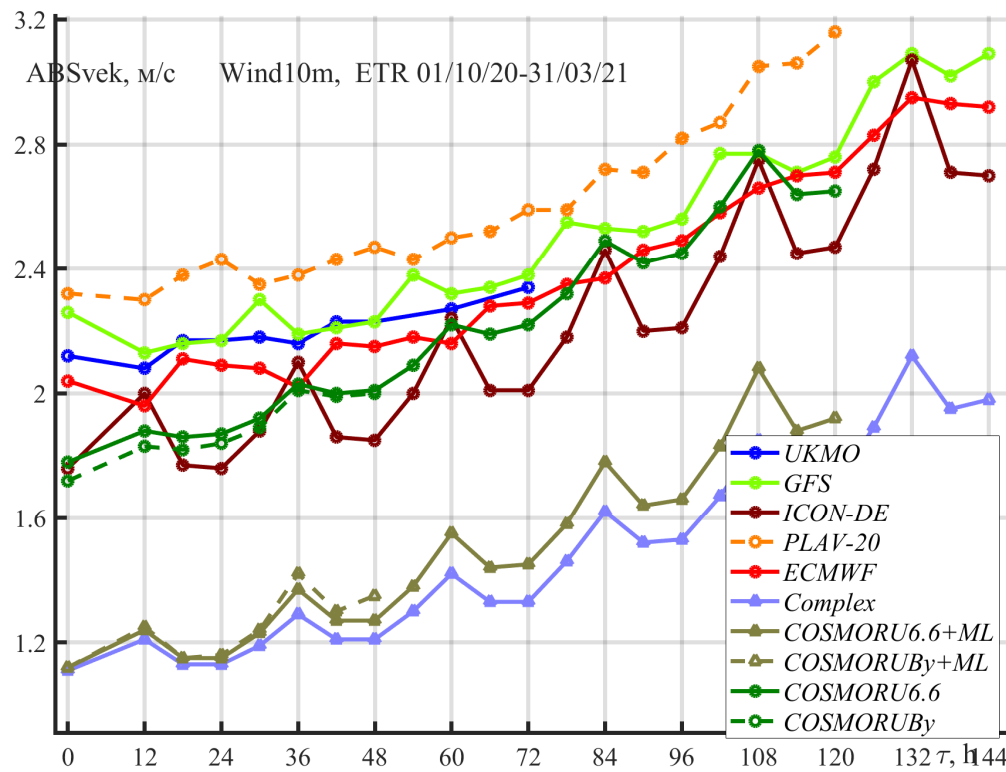
Оценки качества прогнозов температуры



Средняя абсолютная ошибка в зависимости от заблаговременности прогноза

COSMORU7+KF – результат коррекции фильтром Калмана [Алферов Ю.В.,Климова Е.Г.,2020]

Оценки качества прогнозов скорости ветра



Средняя абсолютная векторная ошибка в зависимости от заблаговременности прогноза

Оценки прогнозов порывов ветра более 18 м/с

GFS	Pr=17%	HSS=0.21	ICON-De	Pr=34%	HSS=0.34
76	200	276	156	303	459
375	105554	105929	302	106029	106331
451	105754	106205	458	106332	106790
COSMO-Ru6ENA	Pr=34%	HSS=0.35	Ru6ENA+корр	Pr=47%	HSS=0.43
153	261	414	210	306	516
291	100081	100381	234	100036	100270
444	100342	100786	444	100342	100786
COSMO-RuBy	Pr=26%	HSS=0.35	RuBy+корр	Pr=37%	HSS=0.42
115	103	218	164	178	342
328	99068	99396	279	98993	99272
443	99171	99614	443	99171	99614
Комплексный прогноз	Pr=56%	HSS=0.41	Оценки по ЕТР за период 1 октября 2020 – 31 марта 2021 Pr – предупрежденность HSS – критерий Багрова – Хайдке		
253	506	759			
200	104934	105134			
453	105440	105893			

$U = 100\%$, если ошибка менее 3 °С, иначе 0%.

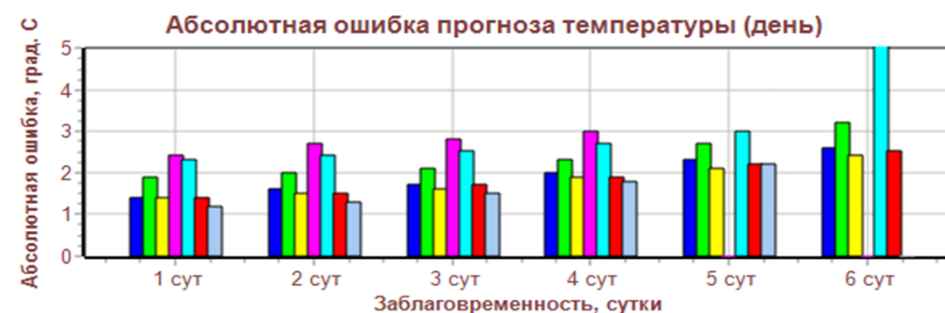
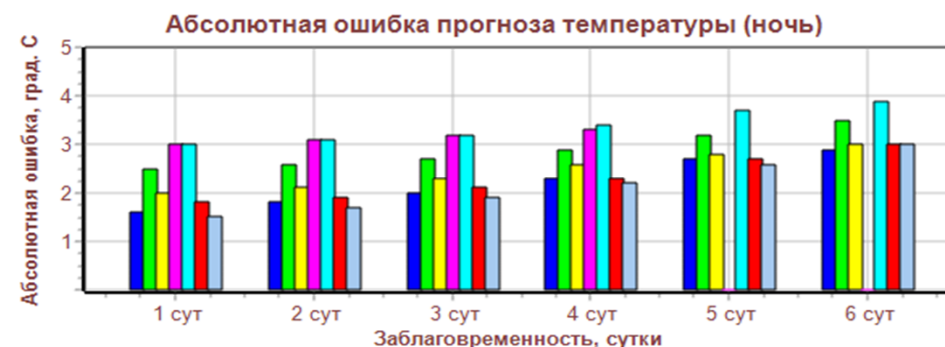
	1 сутки $U, \% / \text{MAE}, ^\circ\text{C}$	2 сутки $U, \% / \text{MAE}, ^\circ\text{C}$	3 сутки $U, \% / \text{MAE}, ^\circ\text{C}$
Минимальная ночная температура, ЕЧР			
Комплексный прогноз	96.0/1.17	95.1/1.31	92.7/1.46
РЭП	94.6/1.30	92.0/1.53	89.9/1.64
Синоптики УГМС	94.8/1.37	92.3/1.57	89.8/1.74
Максимальная дневная температура, ЕЧР			
Комплексный прогноз	97.8/1.00	96.7/1.14	94.0/1.35
РЭП	96.6/1.14	94.9/1.32	92.9/1.45
Синоптики УГМС	96.3/1.24	94.3/1.40	92.3/1.53
Минимальная ночная температура, АЧР			
Комплексный прогноз	92.8/1.39	90.6/1.57	89.0/1.74
РЭП	90.4/1.58	88.5/1.70	86.1/1.86
Синоптики УГМС	89.3/1.66	86.3/1.85	83.5/2.05
Максимальная дневная температура, АЧР			
Комплексный прогноз	96.6/1.15	94.8/1.28	93.4/1.42
РЭП	94.6/1.33	93.2/1.48	91.0/1.58
Синоптики УГМС	93.4/1.46	90.4/1.6	88.2/1.76

Сравнение с интернет-сайтами (Н.А.Светлова)

Сравнительная оценка прогнозов температуры воздуха,
опубликованных на различных ИНТЕРНЕТ-сайтах
по Европейской территории России (33 города*)
Период осреднения: 01.10.2020 - 31.03.2021



Сравнительная оценка прогнозов температуры воздуха,
опубликованных на различных ИНТЕРНЕТ-сайтах
по Азиатской территории России (20 городов**)
Период осреднения: 01.10.2020 - 31.03.2021



Результаты

Комплекс предложенных методов пост-процессинга прогнозов приземных метеорологических параметров одной или нескольких различных численных моделей:

- 1) Учитывает последние (за период длительностью 35 дней) измерения на синоптических станциях;
- 2) Применим в случаях частичного отсутствия прогностической информации и в случаях отсутствия (отбраковки алгоритмом контроля качества) части данных наблюдений;
- 3) Выдаёт прогнозы как в точках синоптических станций, так и в виде полей на сетке;
- 4) Существенно превосходит по качеству прогнозов прогнозы, рассчитанные линейными методами;
- 5) Применим в оперативном режиме счета на существующей инфраструктуре (для подготовки прогнозов в виде полей на модельной сетке высокого разрешения необходимо использование GPU).

Защищаемые положения

- Новый нейросетевой метод квазилинейной систематической коррекции (КСК) прогнозов приземных метеорологических параметров. КСК применяет к известным смещениям модели оператор с коэффициентами, вычисляемыми нейронной сетью;
- Новый нейросетевой квазилинейный комплексный прогноз приземных метеорологических параметров;
- Новый нейросетевой метод квазилинейной неоднородной оптимальной интерполяции (КНОИ) приземных метеорологических параметров. КНОИ эквивалентен оптимальной интерполяции с корреляционной функцией (КФ), зависящей от расстояния в евклидовом пространстве большей размерности. Вложение в пространство большей размерности вычисляется нейронной сетью и учитывает предикторы неоднородности (в том числе прогностические поля). Учет неоднородности позволяет объяснить большую часть вариации интерполируемого поля;
- При настройке параметров КФ напрямую минимизируется погрешность интерполяции в точку «пропущенной» станции без каких-либо предположений о статистическом распределении интерполируемого поля.

Дальнейшие исследования

Усовершенствования:

1. Добавление дополнительных предикторов (feature engineering);
2. Поиск более совершенных архитектур нейронных сетей;
3. Применение к другим метеорологическим полям (в т.ч. трехмерным);
4. Замена используемых предположений на более общие.

Приложения:

- А. Форсинг (nudging) численной модели найденными поправками;
- В. Использование найденных неоднородных КФ в системе усвоения данных;
- С. Применение метода КНОИ для усвоения косвенных данных дистанционного зондирования: возможна совместная оптимизация КФ и оператора наблюдений.

Спасибо за внимание!

Ф.Л.Быков был поддержан грантом № 20-04-021 в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2020– 2021 гг.

Полносвязный слой нейронной сети

Полносвязный слой [Rosenblatt, 1958] с N нейронами применяет линейный оператор с N -мерным образом, а затем поэлементно заранее заданную функцию активации $f(x)$:

$$F(\vec{x}, W) = f(W\vec{x}).$$

Любая непрерывная функция G на n -мерном кубе может быть представлена в виде суперпозиции функций одного переменного

$$G(y_1, \dots, y_n) = \sum_{q=1}^{2^{n+1}} \chi_q \left[\sum_{p=1}^n \psi_{pq}(y_p) \right],$$

где χ_q непрерывны, а ψ_{pq} непрерывны и монотонны [Колмогоров, 1957].

Следствие: G можно приблизить трехслойным персептроном с функциями активации $\max(0, x)$.

Если функция активации $f(x)$ – сигмоидальная, то такая непрерывная монотонная функция, что, $\lim_{x \rightarrow -\infty} f(x) = 0$ и $\lim_{x \rightarrow +\infty} f(x) = 1$. Тогда G можно приблизить двухслойным персептроном с $f(x)$ [Cybenko, 1989].

Обратное распространение ошибки

Пусть $\vec{X}$ – матрицы (тензоры) аргументов; $\vec{\theta}$ – матрицы (тензоры) параметров.

Разобьём модель $\{\vec{Y}_{pred}\} = F(\{\vec{X}\}, \{\vec{\theta}\})$ на блоки (операторы)

$$\{\vec{Y}_k\} = F_k(\{\vec{X}_k\}, \{\vec{\theta}_k\}), k = 1 \dots K, \text{ где } \{\vec{X}_1\} \subseteq \{\vec{X}\}, \{\vec{X}_k\} \subseteq \{\vec{X}\} \cup \bigcup_{s=1}^{k-1} \{\vec{Y}_s\}$$

Тогда для обучения методом градиентного спуска достаточно вычислить градиенты, идя в обратном порядке (метод обратного распространения ошибки):

$$\nabla_{\vec{X}_K} e = \nabla_{\vec{X}_K} F_K \circ \nabla_{F_K} e, \nabla_{\vec{\theta}_K} e = \nabla_{\vec{\theta}_K} F_K \circ \nabla_{F_K} e, \dots, \nabla_{\vec{\theta}_1} e = \nabla_{\vec{\theta}_1} F_1 \circ \nabla_{F_1} e$$

Так можно обучать любые модели методом градиентного спуска. Для сходимости градиентного спуска необходима ограниченность градиентов $\nabla_{\vec{\theta}} e$ глобальной константой.

Автоматическое дифференцирование

Пакет PyTorch строит в памяти граф вычислений, который затем используется для вычисления производных.

Пример:

Вычисление корреляции:

Снизу – программа. Справа – граф вычислений.

```
xn = x-x.mean()
```

```
yn = y-y.mean()
```

```
corr = (xn*yn).mean()/torch.sqrt((xn*xn).mean()*(yn*yn).mean())
```

