

Методы краткосрочного прогнозирования метеоданных по данным наблюдений сети метеостанций с использованием алгоритмов машинного обучения

Сергей Стрижак

ИСП РАН

02.07.2021

Семинар Крупного проекта ННГУ: Надежный и логически
прозрачный искусственный интеллект

Содержание

- Актуальность и постановка задачи с метеоданными и сетью метеостанций
- Постановка задачи для ветроэнергетики
- Обзор научных коллективов
- Решение тестовой задачи
- Базовый предиктор для прогноза погоды
- Постановка задачи с сетью метеостанций
- Первые результаты
- Заключение

Использование метеостанций

Типы станций: обслуживаемые, необслуживаемые, автоматические, метеорологические, агрометеорологические, гидрологические, авиационные метеорологические

Сбор данных: средняя скорость ветра, максимальная скорость ветра, направление ветра, давление, температуру, количество осадков, облачность, др. Всего 14 параметров.
Кратность записи данных – 1 раз в 3 часа
Кратность записи в 15 минут, 1 час – дополнительная услуга от Росгидромет

Назначение станций: прогноз погоды, предсказания паводков и наводнений, планирование агрокультур, использование в энергетике (ветер, солнце), коммунальные и дорожные работы, оценка загрязнений воздуха и водоемов

Для работы с **сетью метеостанций** необходимо использовать алгоритмы теории графов, методы оптимизации, ANN



Цель и задачи работы

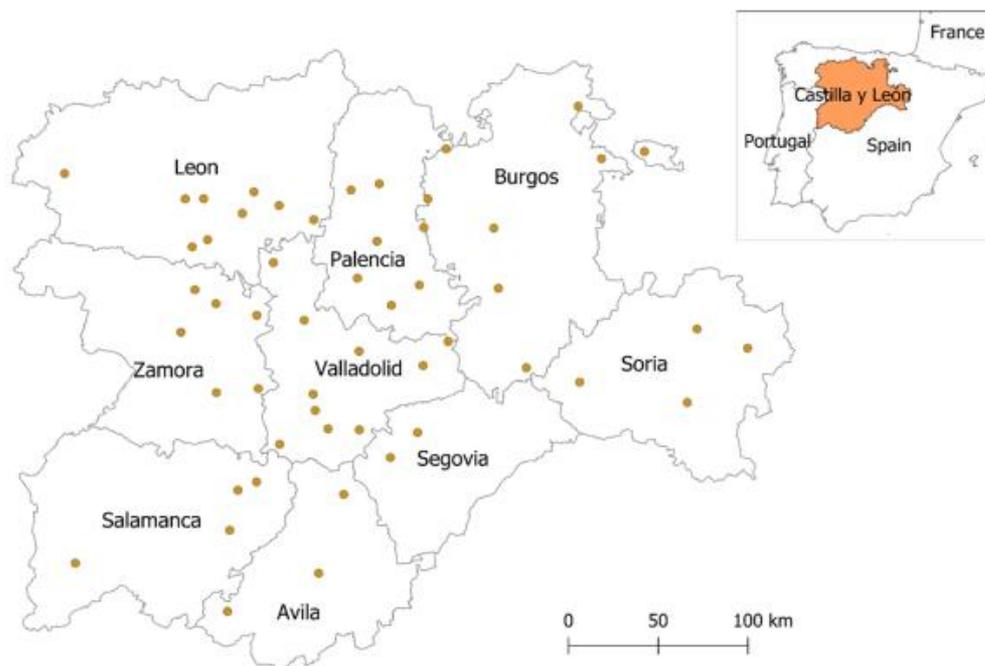
- **Цель:** Разработать ансамбль нейронных сетей для предсказания метеоданных в произвольной географической точке
- **Задача 1:** собрать необходимые метеоданные
- **Задача 2:** разработать архитектуру нейронных сетей для предсказания на 1-3 дней, обучить ANN и оценить ошибку
- **Задача 3:** применить данную методику для задач ветроэнергетики
- **Задача 4:** подготовить публикации в журналы Q1,Q2 в 2022

Научная группа ИСП РАН: Даниил Рязанов, Александр Иванов, Константин Кошелёв, Булат Кашфутдинов, Сергей Стрижак

Обзор научных коллективов.

Сеть метеостанций в Испании, регион Castilly y Leon

Fig. 1 Meteorological station network of the InfoRiego program of the Agrarian Technology Institute of Castilla y León



Рассматриваются **53** метеостанций. Разработана модель виртуальных метеостанций.

Использованы алгоритмы: Distance weighting, inverse-squared distance weighting, multilinear regression, and random forest regression.

Данные взяты за 1 год (07.2017-06.2018) + (X,Y,Z) из базы данных Аграрного технологического института Catilla

Результат: IDSW and the ANN with the softsign activation functions were the most precise approach to perform the interpolations and to use in the VWS as preferred methods for the Castilla y Leo'n territory

V.M. Franco et al. Neural Computing and Applications (2020) 32:12801–12812

Сеть метеостанций в Дании и в Нидерландах

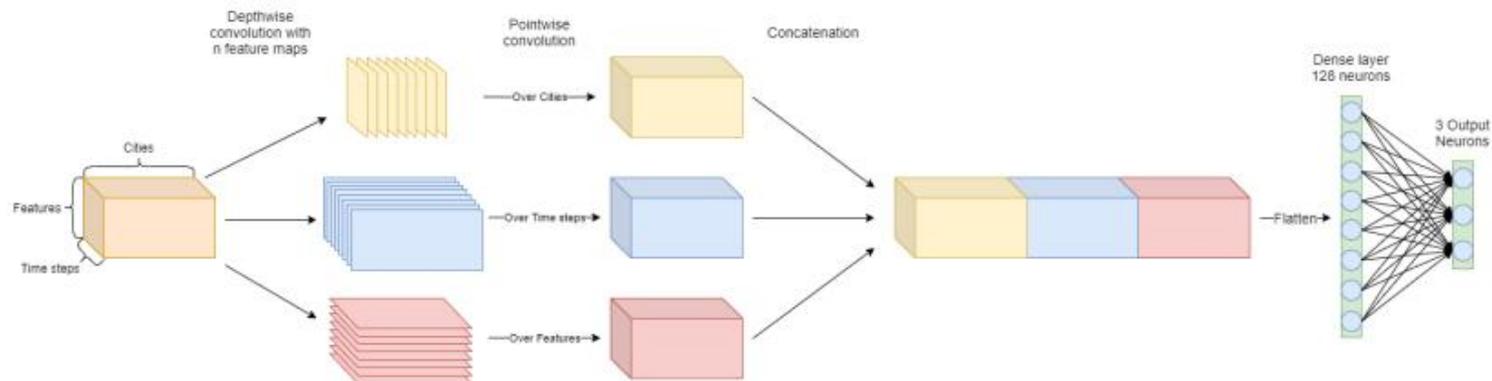


Fig. 1: Using depthwise-separable convolutions to go over different dimensions of the input tensor. The resulting three tensors are concatenated along the city-dimension and then flattened and fed into a dense layer with 128 hidden neurons. Lastly, a dense layer with three neurons calculates the final output.

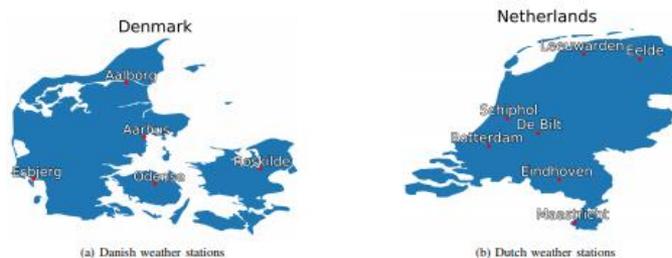


Fig. 3: The cities where the wind data was taken: (a) Aalborg, Aarhus, Esbjerg, Odense and Roskilde. (b) Schiphol, De Bilt, Leeuwarden, Eelde, Rotterdam, Eindhoven, Maastricht.

TABLE II: The average MAEs and MSEs of different models over target cities in Denmark dataset.

Model	MAE				MSE			
	6h	12h	18h	24h	6h	12h	18h	24h
Persistence	1.649	2.210	2.309	2.313	4.908	7.929	8.702	8.811
2D	1.304	1.746	1.930	2.004	2.824	5.088	6.120	6.610
2D+Attention	1.313	1.715	1.905	1.950	2.885	4.896	5.933	6.201
2D+Upscaling	1.307	1.723	1.858	1.985	2.826	4.931	5.639	6.474
3D	1.311	1.677	1.908	1.957	2.855	4.595	5.958	6.238
Multidimensional	1.302	1.706	1.873	1.925	2.804	4.779	5.773	6.066

TABLE III: The average MAEs and MSEs of different models over target cities in the Netherlands dataset.

Model	MAE				MSE			
	1h	2h	3h	4h	1h	2h	3h	4h
Persistence	9.55	11.34	12.90	14.37	183.61	246.95	310.38	375.36
2D	8.11	9.17	10.15	11.12	116.89	149.01	181.78	218.49
2D+Attention	8.08	9.10	10.11	11.00	115.96	147.75	180.66	213.23
2D+Upscaling	8.16	9.07	10.14	10.85	117.80	147.21	182.44	208.96
3D	8.17	9.26	10.15	10.93	118.35	151.51	181.35	211.19
Multidimensional	8.12	9.05	9.95	10.94	116.78	144.51	174.07	208.73

2D, 3D CNN нейронные сети

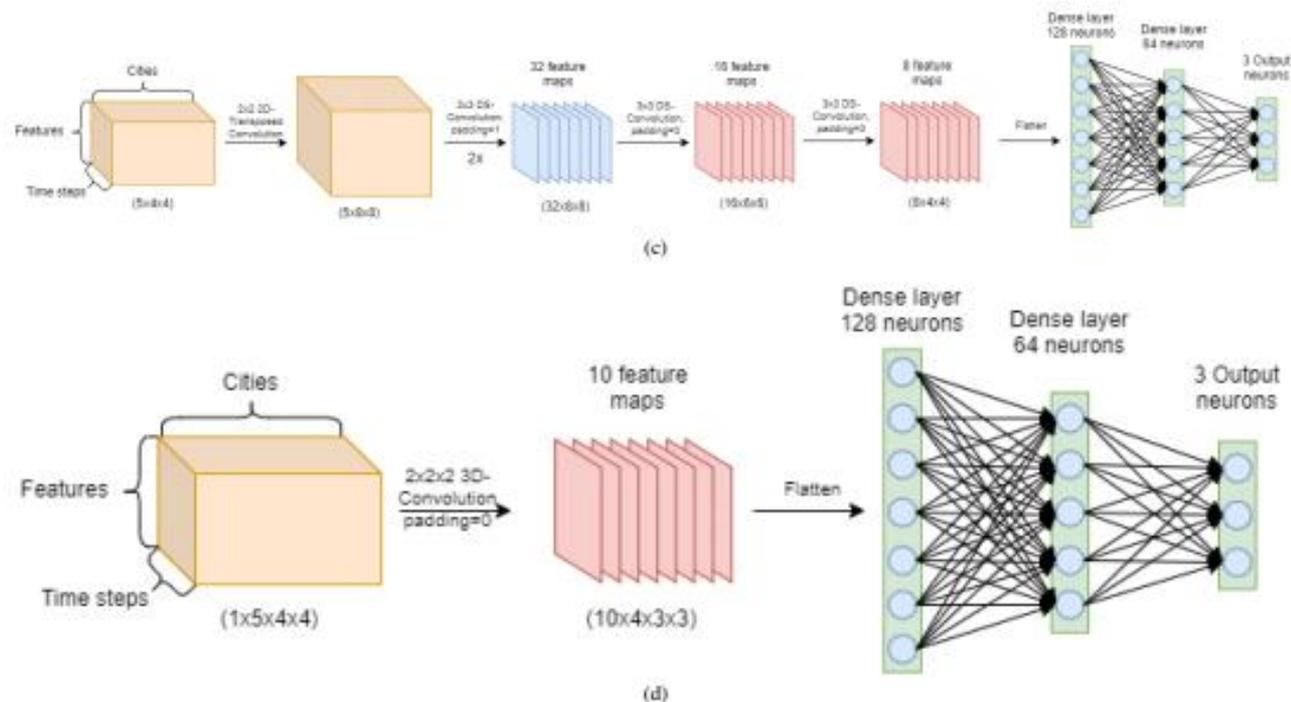


Fig. 2: The model architecture of the models that were tested. (a) 2D convolutional neural network. (b) 2D convolutional neural network with attention. (c) 2D convolutional neural network with transposed convolution (upsampling). (d) 3D convolutional neural network. The input dimensions and number of output neurons are adjusted to the dataset. Here, the models for the DK dataset are shown.

- 1) Denmark dataset is hourly measurements of temperature, pressure, wind speed and wind direction of five cities in **Denmark from 2000-2010**. The dataset is split into training and validation (years 2000-2009) and test set (year 2010).
- 2) the Royal Netherlands Meteorological Institute. It contains hourly weather measurements from 7 cities in the Netherlands from **January 1, 2011 until March 29, 2020** resulting in 81.000 data points. Прогноз метеоданных на 6,12,24 часов вперед

Прогноз температуры воздуха в Китае по данным GFS модели, сети метеостанций и данным с метеоспутника

- **1,812** observation stations in China
- Использованы данные GFS 3 h precipitable water vapor (GFS PWV), relative humidity (GFS RH).
- Forecast length of the GFS data (GFS PWV and GFS RH) used was 3 h per day, and there were eight periods of data per day (i.e., 00, 03, 06, 09, 12, 15, 18, and 21 UTC).
- Chinese geostationary meteorological satellite Fengyun-4A (FY-4A).
- Использованы 6 ML алгоритмов для предсказания T_{air} : MLR, GBTD, KNN, RF, XGB, and DNN
- China Meteorological Data
- Service Center (CMDC) (<http://data.cma.cn/>)

Результат: RMSE of the XGB model was 2.376°C, the RMSE using satellite data combined with numerically modeled T_{air} data reached 1.946°C

$$R^2 = \frac{[\sum (T_{ea} - \bar{T}_{ea})(T_{oa} - \bar{T}_{oa})]^2}{\sum (T_{ea} - \bar{T}_{ea})^2 \sum (T_{oa} - \bar{T}_{oa})^2}, \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{ea} - T_{oa})^2}{N}}, \quad (2)$$

$$bias = \frac{\sum_{i=1}^N (T_{ea} - T_{oa})}{N}, \quad (3)$$

$$MSE = \frac{\sum_{i=1}^N (T_{ea} - T_{oa})^2}{N}$$

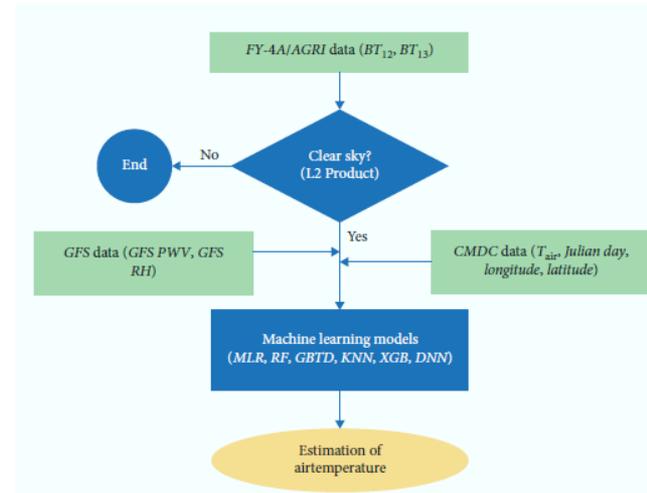
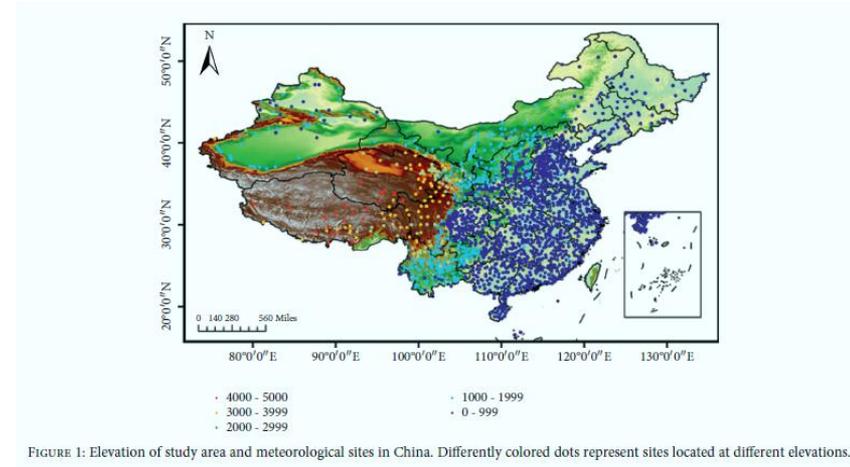


FIGURE 3: Flowchart of the T_{air} estimation model in this study.

Оптимальное проектирование сети метеостанций в Германии.

группа Prof. Jürgen Kurths (Potsdam Institute for Climate Impact Research)

$$k_i = \sum_{j=1}^N A_{i,j}, \quad \text{- the number of connections linked to the node directly}$$

$$B_i = \sum_{i \neq j \neq v \in \{V\}} \frac{\sigma_i(j,k)}{\sigma(j,k)}, \quad \text{- the betweenness centrality of any } i \text{ node}$$

$$DIL_i = k_i + \sum_{j=N_G(i)} I_{eij} \cdot \frac{k_i - 1}{k_i + k_j - 2}, \quad \text{- The degree and influence of line}$$

$$Bri_i = \sum_{j \notin N_G(i) \vee k \notin N_G(i)} \frac{\sigma_i(j,k)}{\sigma(j,k)} \quad \text{- the bridgeness measure}$$

$$WDB_i = B_i + I_i, \quad \text{- a combination of two network measures, degree and betweenness centrality}$$

$$I_i = \sum_{j=1}^{k_i} \frac{B_j \cdot (k_j - 1)}{(k_i + k_j - 2)}, \quad \text{- the cumulative effect of the influence or contribution of the directly connected nodes of } i$$

Data: 110 years at a daily resolution (1 January 1901 to 31 December 2010).

Результат: The new network measure, weighted degree–betweenness (WDB), combines the measures of degree and betweenness centralities is used for network in Germany.

A. Agarwal et al.: Optimal design of hydrometric station networks. Hydrol. Earth Syst. Sci., 24, 2235–2251, 2020

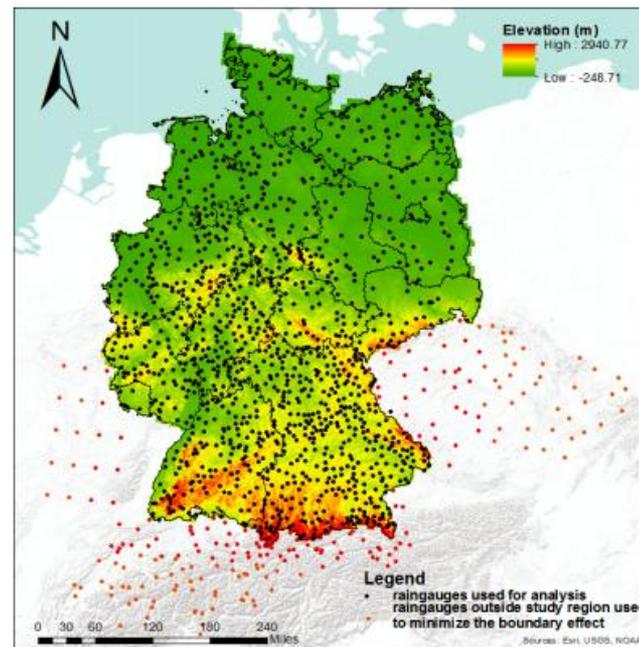
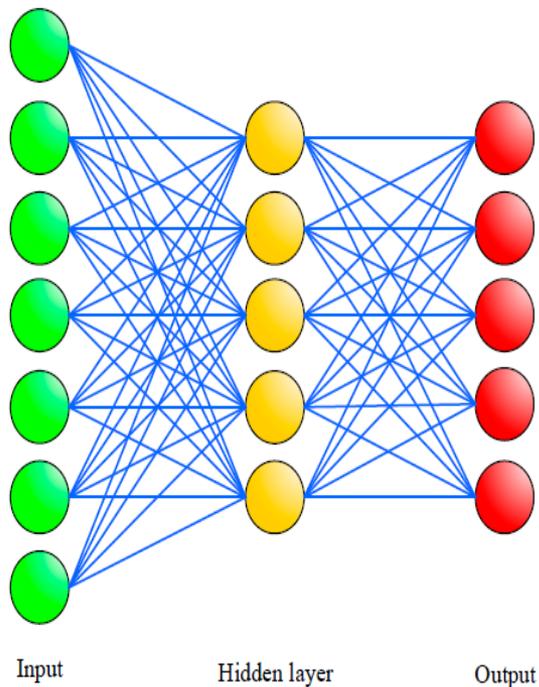


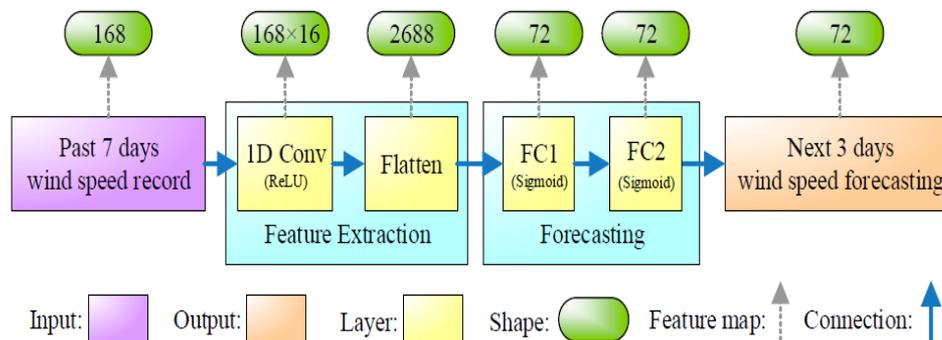
Figure 4. Location of rain stations in Germany and adjacent areas. Black dots indicate stations lying inside Germany that are used in the analysis. Red dots indicate stations outside of Germany that are used for network construction only in order to minimize the boundary effect. © Esri, USGS, NOAA.

1229 stations across Germany

Примеры использования нейронных сетей для предсказания скорости ветра



Multilayer Perceptron



WindNet Convolution Neural Network

Chiou-Jye Huang, Ping-Huan Kuo. A Short-Term Wind Speed Forecasting Model by Using Artificial Neural Networks with Stochastic Optimization for Renewable Energy Systems. *Energies*, 2018.

Результат прогноза скорости ветра на 3 дня

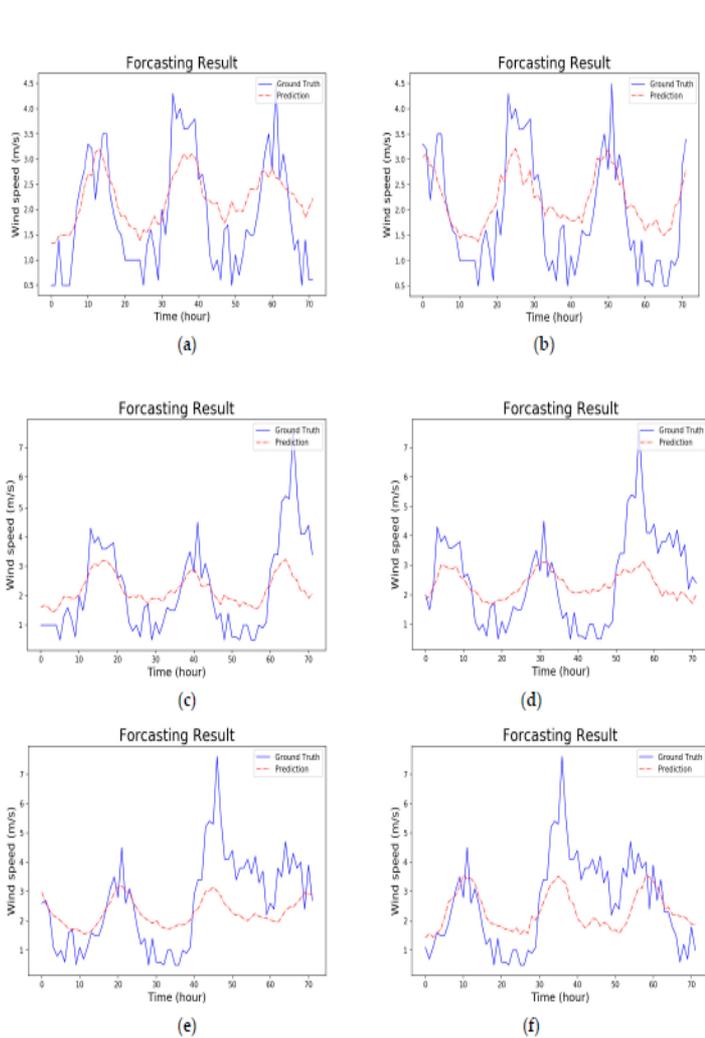


Figure 10. The forecasting results of random forest: (a) Partial results A; (b) Partial results B; (c) Partial results C; (d) Partial results D; (e) Partial results E; (f) Partial results F.

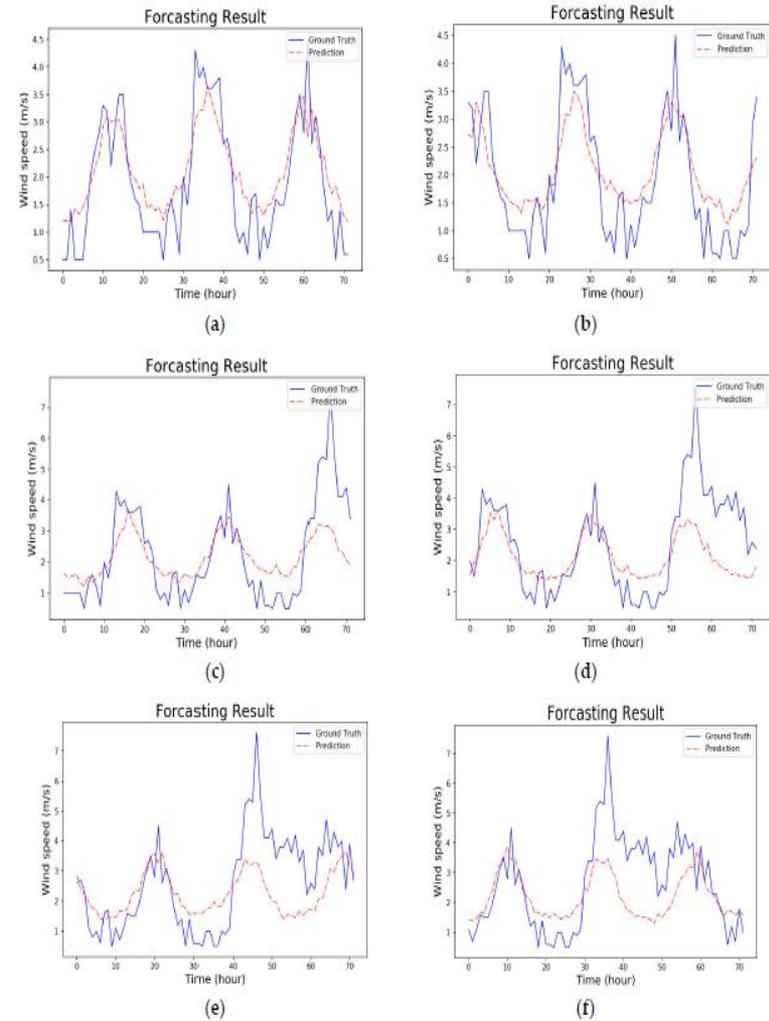


Figure 13. The forecasting results of the proposed WindNet: (a) Partial results A; (b) Partial results B; (c) Partial results C; (d) Partial results D; (e) Partial results E; (f) Partial results F.

Ветроэнергетика в РФ

Ветроэнергетика и новые ветропарки в РФ:

Ульяновская область (2018-2019)

Республика Саха, поселок Тикси (2018-2019)

Республика Адыгея (2019-2020)

Ставропольский край (с 2019)

Краснодарский край (2020)

Мурманская область (2021)

Основные участники рынка: РОСАТОМ (АО «НоваВинд»), ПАО «Фортум» / Группа «РОСНАНО», «Энел Россия».

В России создается новая индустрия, ежегодно строятся новые ветропарки, начато производство ветрогенераторов, углубляется локализация отдельных элементов ветроэлектрических установок (ВЭУ).

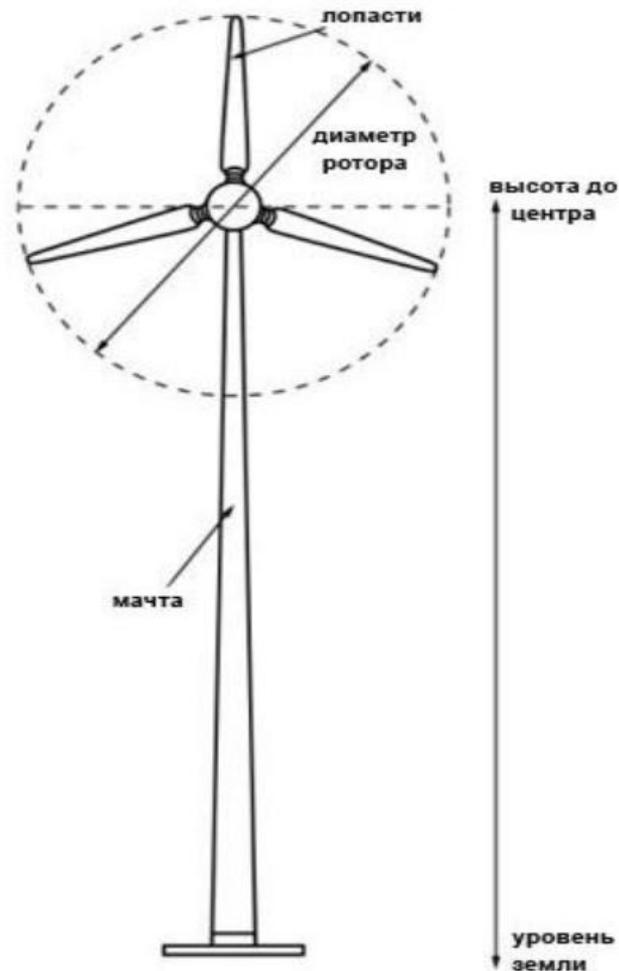
Потребность в разработке цифровых двойников, потребность **в метеоданных.**

Нужна информация (временные ряды) за **20-30 лет.**



Цифровой двойник ветропарка – максимально полное описание технической системы на основе компьютерных программ и математических моделей для решения задач конструирования, проектирования, эксплуатации промышленного объекта.

Влияние скорости ветра на мощность ветрогенератора



$$P = 0,5 \cdot V^3 \cdot S \cdot \rho \cdot \xi$$

V - скорость ветра (м/с)

S - площадь отметания (кв.м)

ρ - плотность воздуха (куб.м)

ξ - коэффициент погрешности

Источники климатических данных

1. <https://rp5.ru/>

- a. Место предсказания
- b. Язык интерфейса
- c. Единицы измерения
- d. Возможность загрузки данных
- e. Различные крупные города которые возможно нам понадобятся
- f. Координаты выбранного места
- g. Выбор интересующих данных
- h. Контакты

Путь -> Выбор места (город)-> интересующие интервал предсказания (предыдущие дни)->возможность скачать данные

2. <https://www.gismeteo.ru/>

- a. Место предсказания
- b. Необходимые параметры (скорость, температура и тд.)
- c. Координаты выбранного места
- d. Возможность распечатки
- e. Диапазон дат (только вперёд) до 10 дней
- f. Контакты

Путь -> Выбор места (город)-> интересующие интервал предсказания (только вперед)->возможность распечатать (сводку данных)

3. <https://www.meteoprog.ua/>

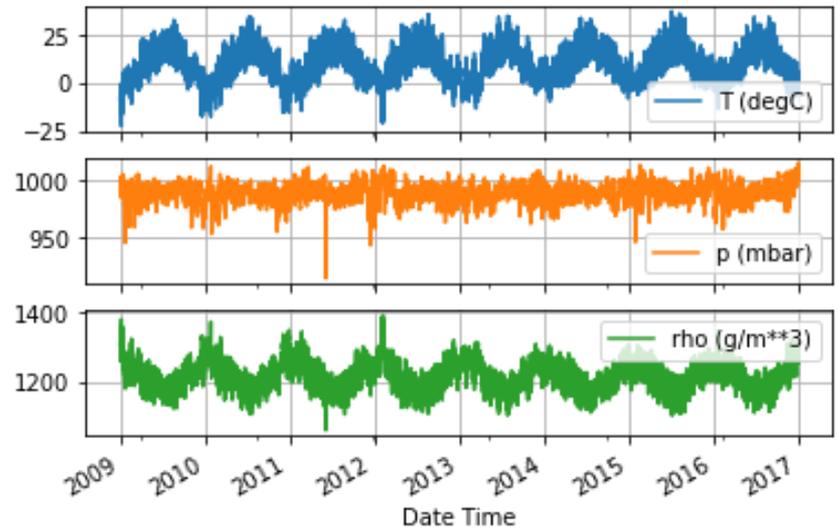
- a. Место предсказания
- b. Возможность укрупнить регион (алушта - весь крым)
- c. Вывод графиков (в принципе бесполезно, но для общей динамики норм)
- d. Выводится общая карта с самыми основными природными признаками
- e. Различные погодные новости (пример: ураган где то)
- f. Различные крупные города которые возможно нам понадобятся
- g. Контакты

Путь -> Выбор места (город)-> интересующие интервал предсказания (только вперед)->возможность распечатать (сводку данных)

Нигде нет ошибки предсказания

Тестовый пример ИСП РАН - обработка метеоданных для известной метеостанции в Jupyter-Lab

- Используется [временной ряд наблюдений за погодными условиями](#), которые записываются [Институтом Биохимии Макса Планка](#).
- Набор данных содержит 14 различных признаков таких как температура воздуха, атмосферное давление и влажность.
- Данные собираются каждые 10 минут начиная с 2003 года. Для простоты мы будем использовать только данные собранные в период с 2009 по 2016.
- Данный пример является введением в практику прогнозирования для временных рядов с использованием инструментария TensorFlow.
- Рассматривается несколько моделей, включая сверточные нейронные (CNN) сети и рекуррентные нейронные сети (RNN).
- Эта тема представлена в двух главных разделах с подразделами:
- Предсказание следующего временного шага:
 - - По одному признаку.
 - - По всем признакам.
- Предсказание многих шагов по времени:
 - - Однократный: сразу все предсказывая за раз.
 - - Авторегрессия: предсказание с исходной гипотезой о текущем значении

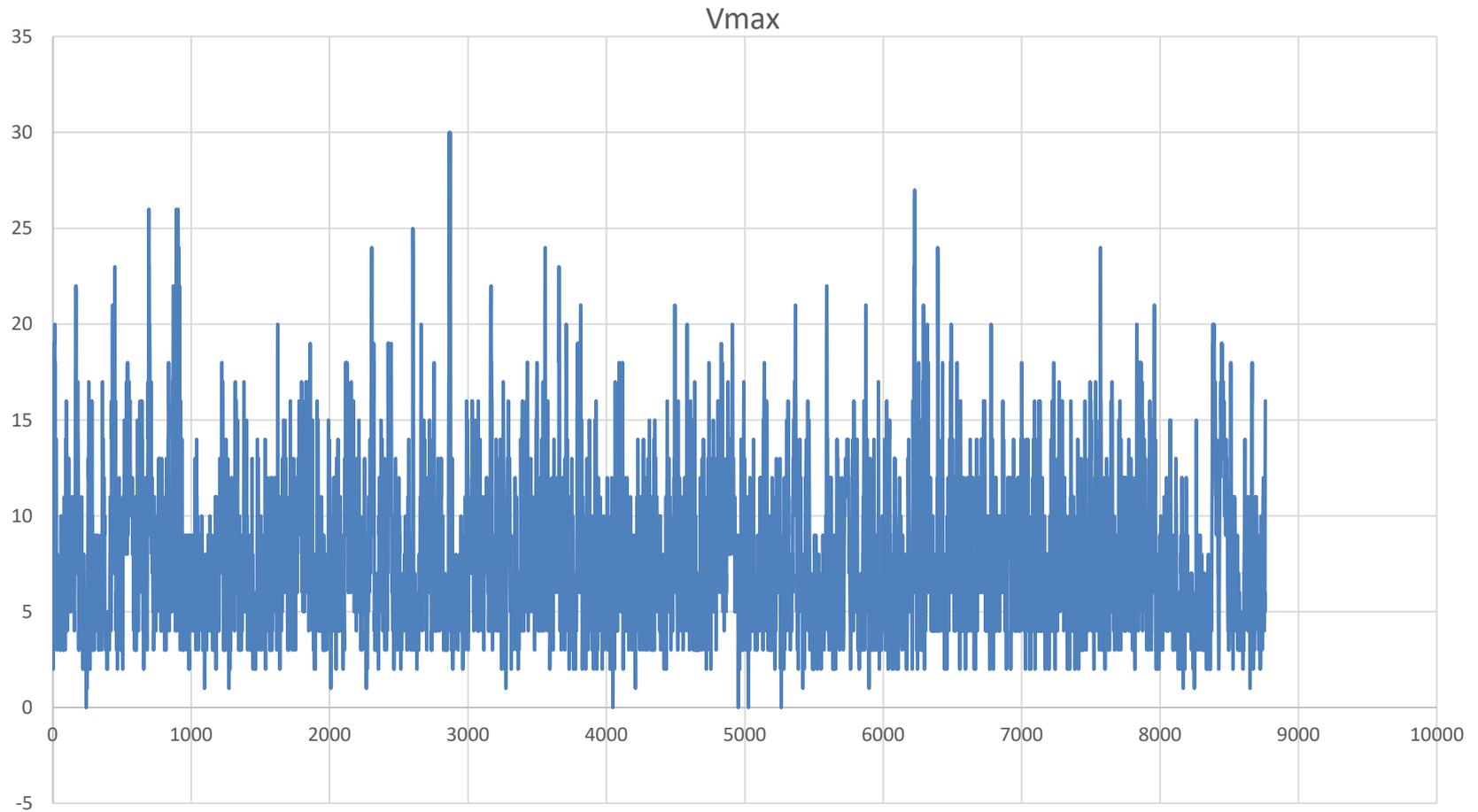


Для обучения, валидации и тестирования данные подготавливаются и разбиваются на 3 группы: 70%, 20%, 10%

Dataset за 2017-2019 с метеостанции 34949 г. Ставрополь

Number	Year	Month	Date	Hour	Direction	Vmean	Vmax	Temperature	Pressure
34949	2017	1	1	0	185	1	9	-4,9	1020,4
34949	2017	1	1	3	0	0	2	-5,9	1019
34949	2017	1	1	6	0	0	2	-6,8	1019,5
34949	2017	1	1	9	288	5	7	1,3	1017,2
34949	2017	1	1	12	295	4	8	-0,2	1017,5
34949	2017	1	1	15	282	7	11	-0,1	1018,2
34949	2017	1	1	18	266	12	18	-0,3	1018,2
34949	2017	1	1	21	281	8	18	-0,2	1018,3
34949	2017	1	2	0	269	11	15	0,6	1018
34949	2017	1	2	3	277	11	19	0,8	1017,3
34949	2017	1	2	6	267	13	19	1,4	1017,4
34949	2017	1	2	9	274	12	19	2	1016,4
34949	2017	1	2	12	273	12	19	2,2	1015,5
34949	2017	1	2	15	253	10	19	2,2	1016,3
34949	2017	1	2	18	270	13	20	2,1	1016,2
34949	2017	1	2	21	250	6	17	1,3	1017,3
34949	2017	1	3	0	263	10	15	1,4	1016,9
34949	2017	1	3	3	269	8	18	1,8	1016,3
...									
34949	2019	12	30	12	85	5	9	-0,4	1027,2
34949	2019	12	30	15	112	3	6	-1,2	1027,7
34949	2019	12	30	18	137	3	5	-1	1026,5
34949	2019	12	30	21	120	3	5	-1,3	1024,8
34949	2019	12	31	0	124	2	4	-1,4	1023,5
34949	2019	12	31	3	191	2	5	-1	1021,7
34949	2019	12	31	6	229	3	6	-0,4	1021,4
34949	2019	12	31	9	259	2	5	2,6	1018,9
34949	2019	12	31	12	268	5	7	2,7	1016,8
34949	2019	12	31	15	261	8	11	0,8	1016,4
34949	2019	12	31	18	260	10	14	1,9	1014,8
34949	2019	12	31	21	241	6	16	1,1	1014,1

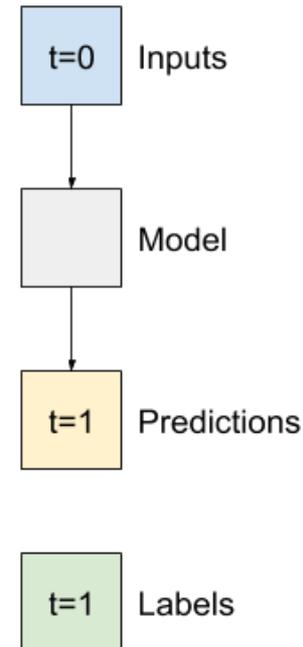
Если имеется порыв ветра, то
 $V_{\max} - V_{\text{mean}} > 5.4 \text{ м/с}$



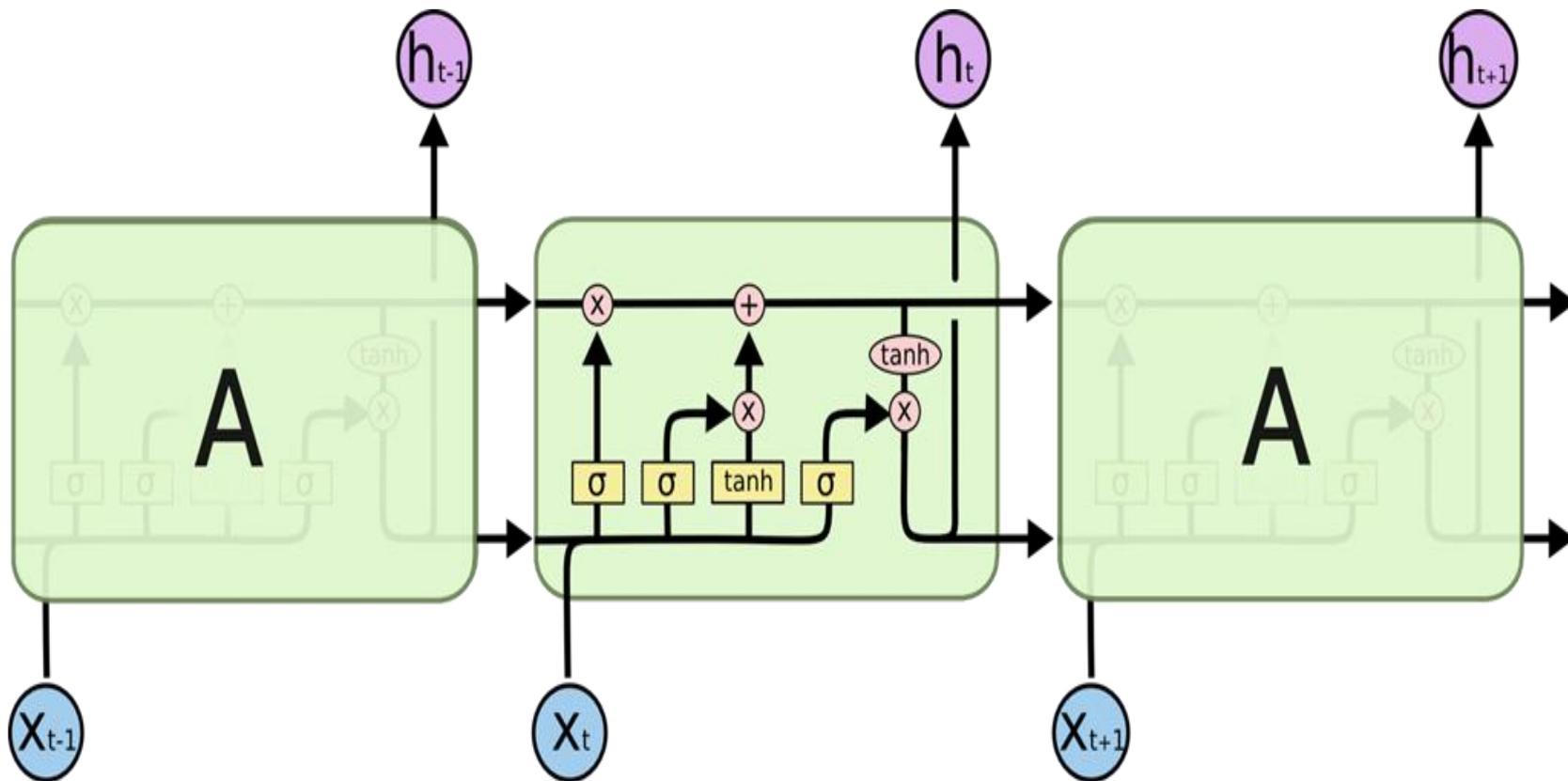
Простейшая модель прогноза

- Простейшая предсказательная модель предсказывает один признак на следующем временном шаге просто копируя значение на текущем временном шаге.

- Реализована на Python в виде отдельной функции
- Dataset – метео данные по станции в г. Ставрополь
- Точность прогноза – 77 %
- Точность прогноза менее этого значения не имеет смысла



Прогноз метеоданных для станции в г. Ставрополь. Исходные данные - Dataset 2017-2019.



Выбрана нейронная сеть LSTM

Состоит из четырех взаимодействующих слоев

<https://habr.com/ru/company/wunderfund/blog/331310/>

Автор оригинала: Christopher Olah

Результаты обучения

- Для обучения нейросети был использован датасет с метеоданными из г. Ставрополь за 3 годы
- Данные были нормализованы
- Для предсказания данных было выбрано окно в 3 дня
- Основная метрика - корень среднеквадратичной ошибки RSME
- Среднее отклонение предсказанного результата от фактического результата около 2,5

Параметры нейросети

Model: "sequential_6"

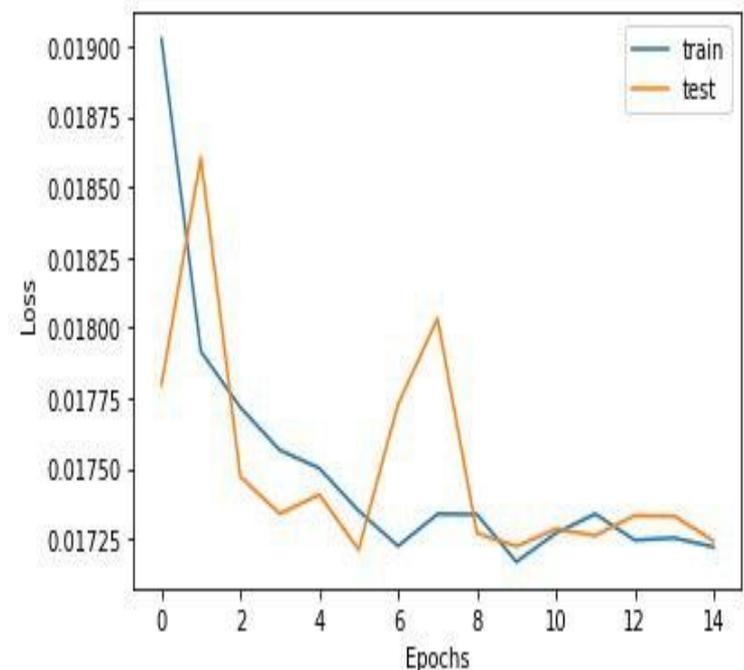
Layer (type)	Output Shape	Param #
lstm_10 (LSTM)	(None, 30, 150)	91200
dropout_15 (Dropout)	(None, 30, 150)	0
lstm_11 (LSTM)	(None, 50)	40200
dropout_16 (Dropout)	(None, 50)	0
dense_10 (Dense)	(None, 100)	5100
dropout_17 (Dropout)	(None, 100)	0
dense_11 (Dense)	(None, 1)	101

Total params: 136,601

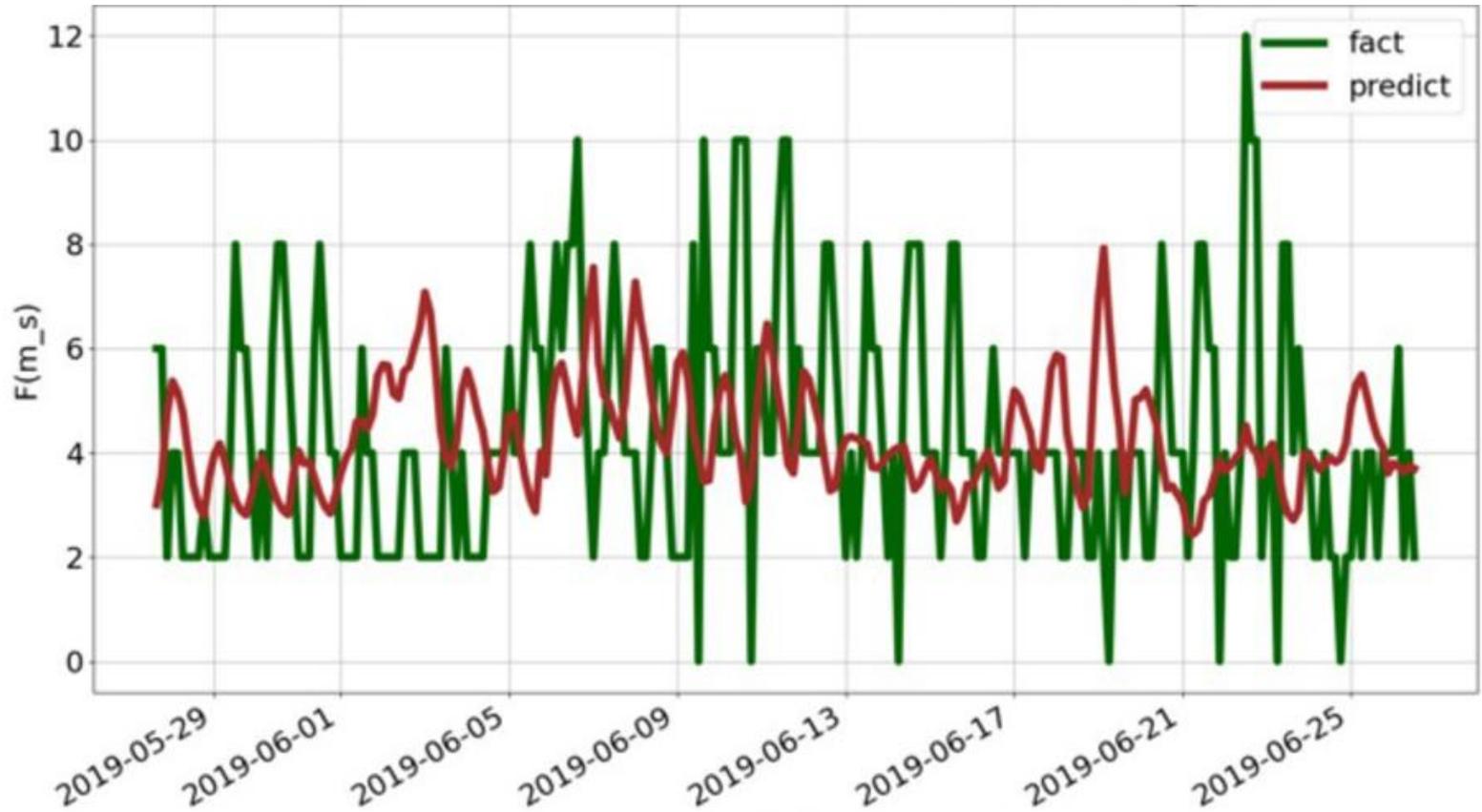
Trainable params: 136,601

Non-trainable params: 0

Параметры нейросети



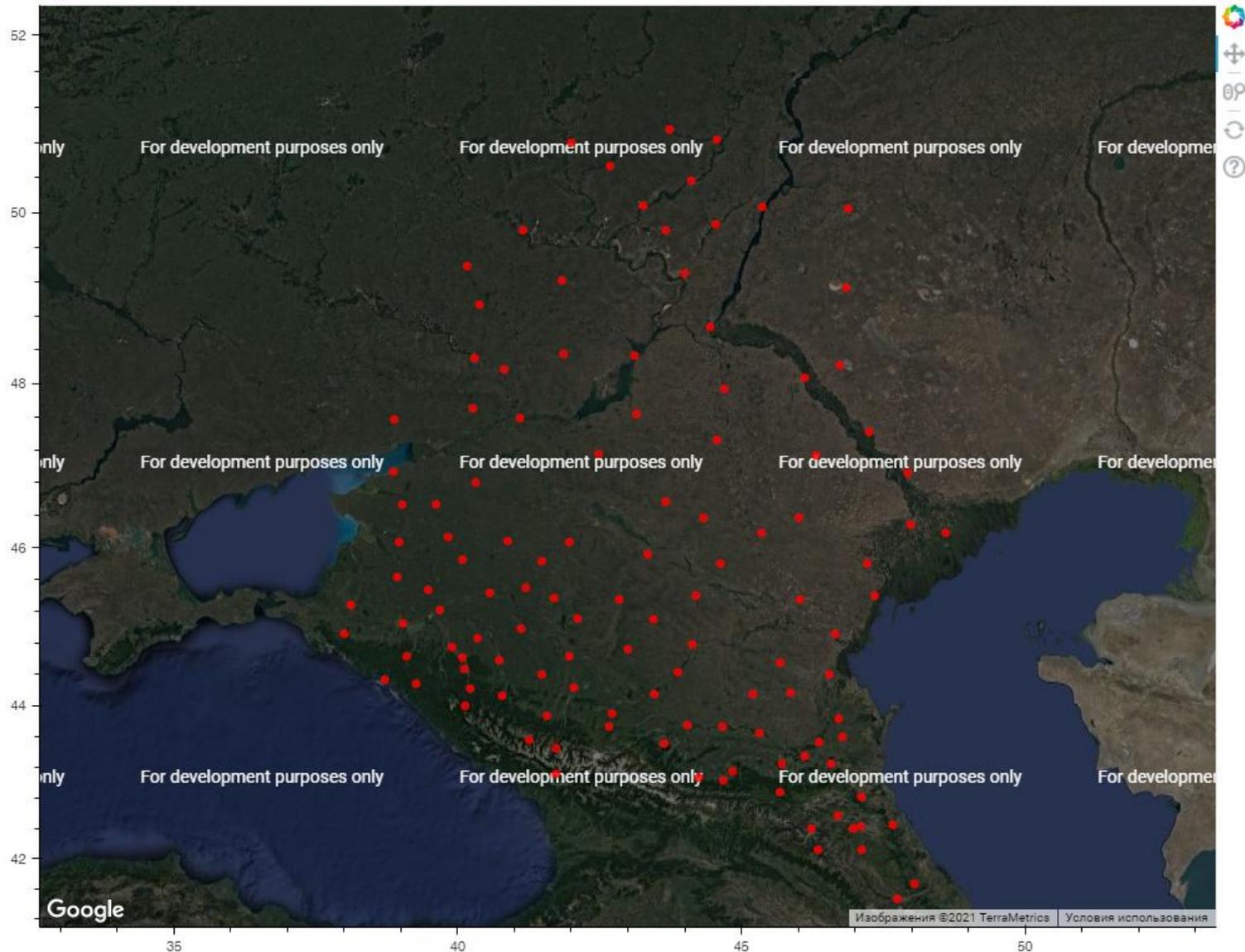
Результаты прогноза скорости ветра для станции в г. Ставрополь на 1 месяц
Фактические показатели и предикт для F (m_s)



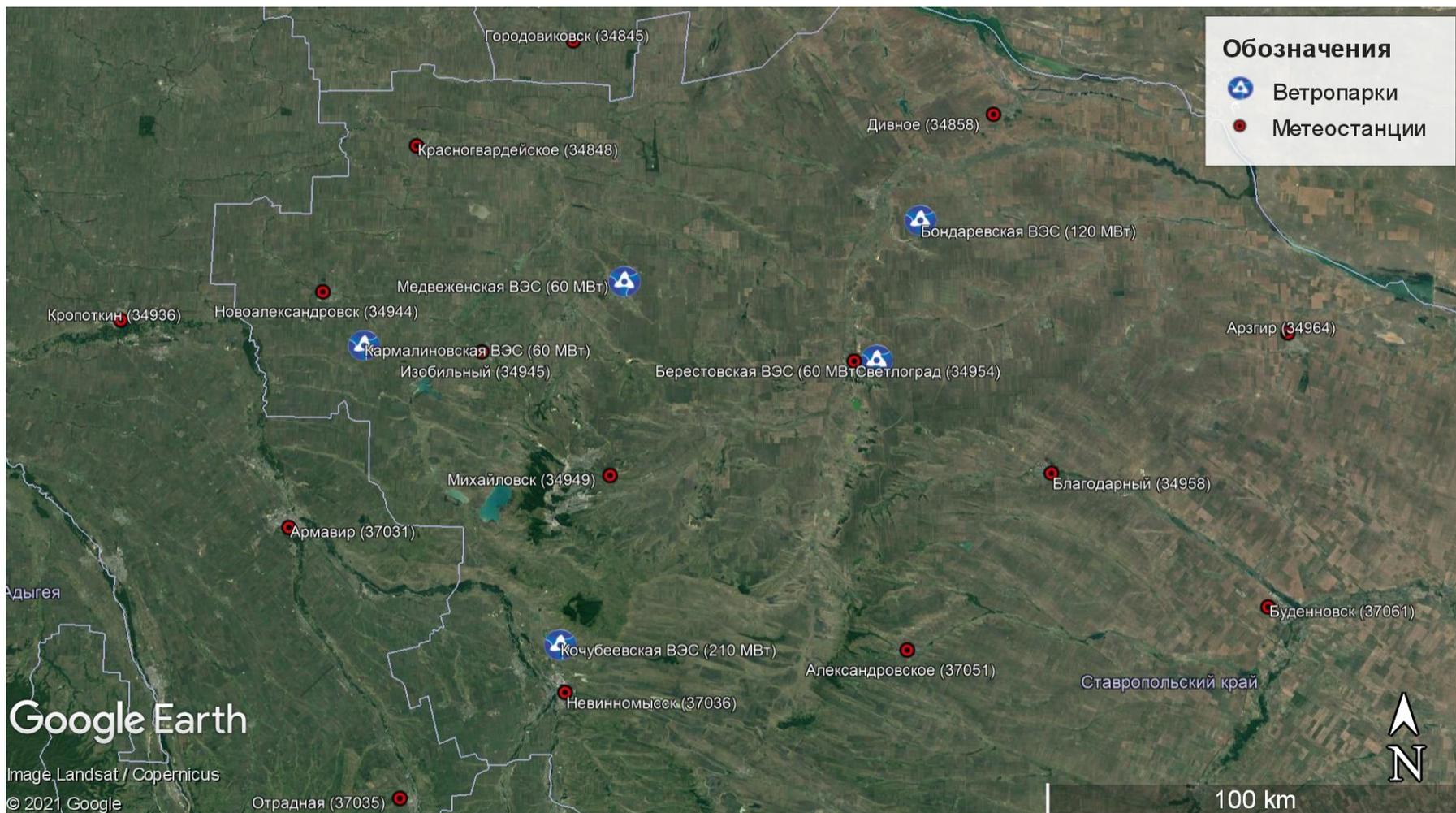
Выбрана нейронная сеть LSTM

Сеть доступных метеостанций в ФГБУ

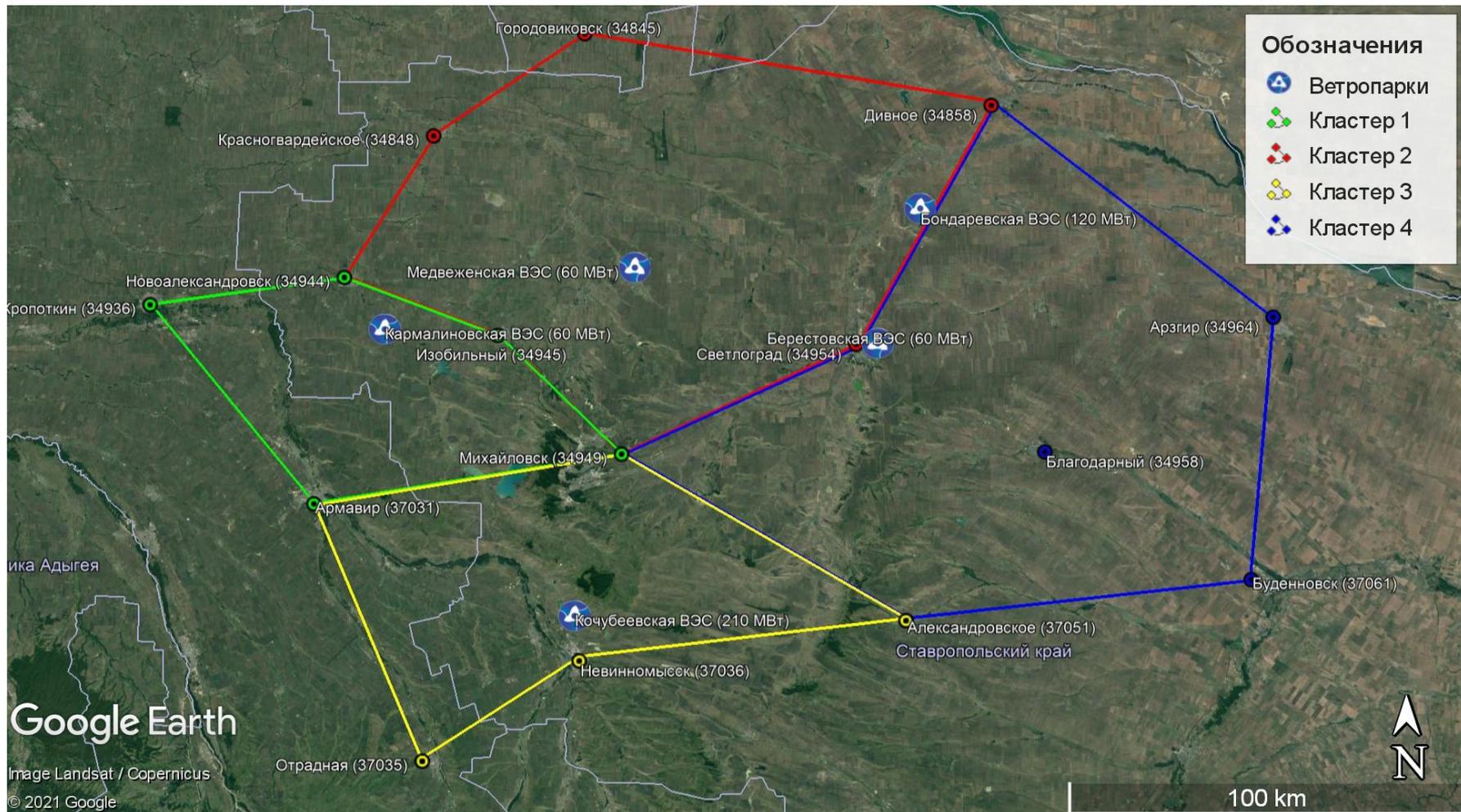
Северо-Кавказский УГМС (из 132 станций доступны 121 на rp5.ru)



Общий план метеостанций



Разбиение на кластеры



Рабочий процесс

PCA, POD

kNN

LSTM, 3D CNN, ансамбли

Подготовка данных
для 121 метео-
Станции за 10 лет

Формирование
мини-кластеров

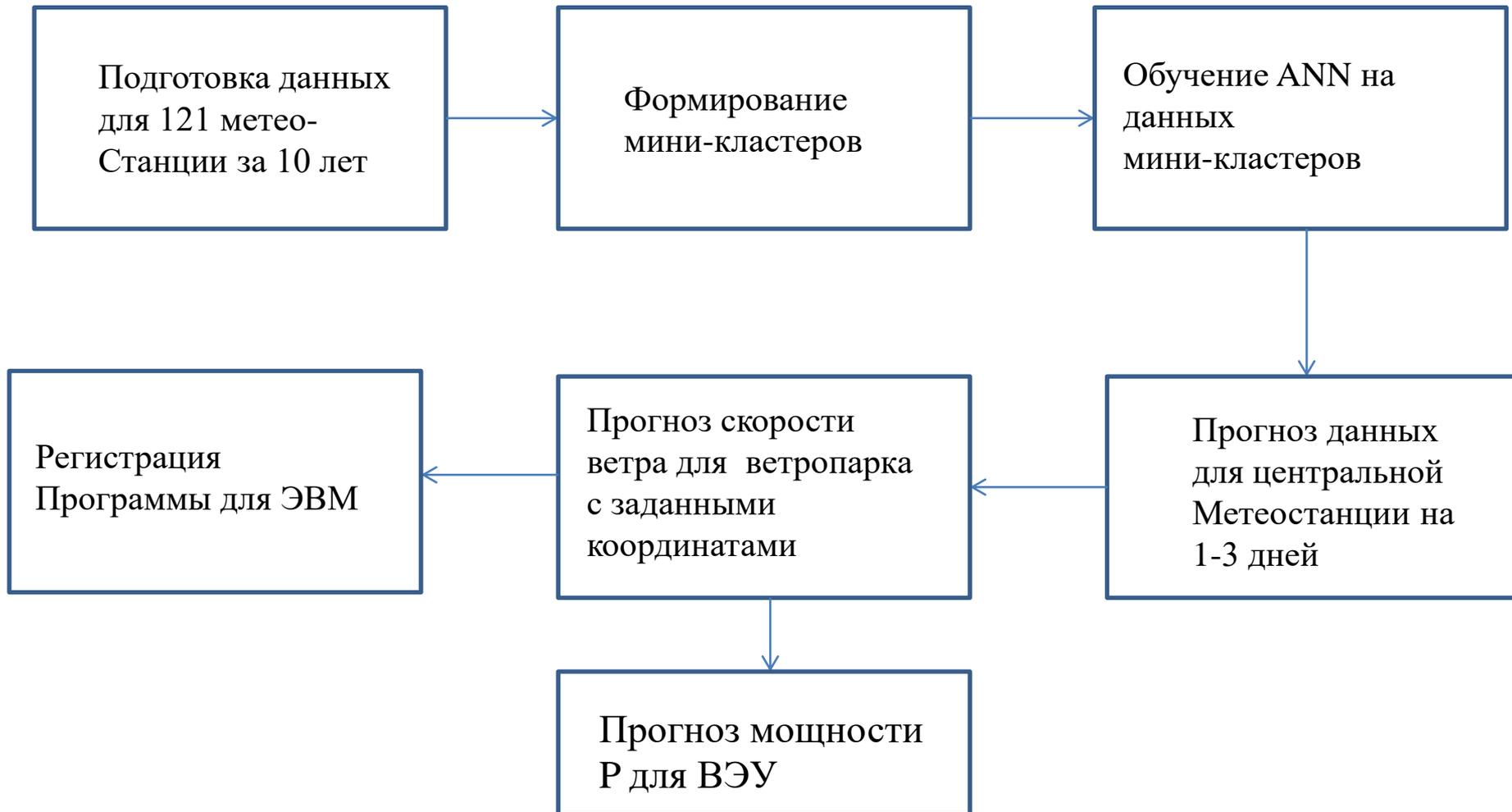
Обучение ANN на
данных
мини-кластеров

Регистрация
Программы для ЭВМ

Прогноз скорости
ветра для ветропарка
с заданными
координатами

Прогноз данных
для центральной
Метеостанции на
1-3 дней

Прогноз мощности
P для ВЭУ



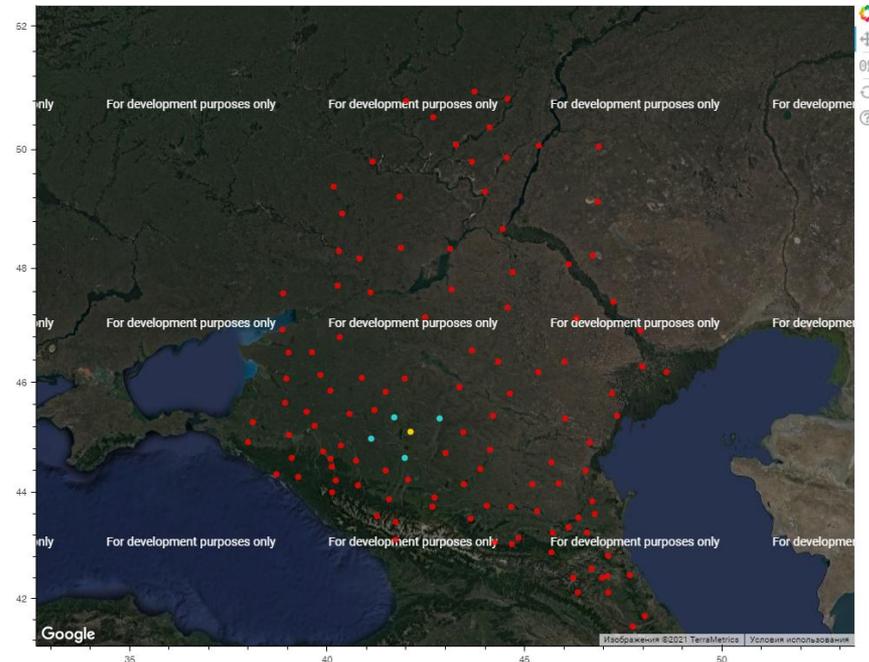
Работа с данными и первые результаты

- Подготовка данных для 121 метеостанций
- ПО: jupyter notebook, Python, pandas, scikit-learn 0.24.2
- Поиск ближайших k соседей (k=4) для центральной метеостанции
- Использование алгоритма kNN
- Sklearn.neighbors.BallTree.
- Работа с Google Earth Pro с использованием API
- Получена группа кластеров, сформирована база данных

[147]:

	Name	index	LAT	LON	ELEV	rp5	color	radLon	radLat
0	УРЮПИНСК	34240	50.800	42.000	110	1	red	0.733038	0.886627
1	ЕЛАНЬ	34253	50.950	43.733	132	1	red	0.763285	0.889245
2	НОВОАННИНСКИЙ	34254	50.533	42.683	98	1	red	0.744959	0.881967
3	РУДНЯ	34262	50.833	44.567	114	1	red	0.777841	0.887203
4	ДАНИЛОВКА	34267	50.367	44.117	102	1	red	0.769987	0.879070
...
116	КАСУМКЕНТ	37597	41.667	48.050	477	1	red	0.838631	0.727226
117	АХТЫ	37663	41.467	47.750	1015	1	red	0.833395	0.723736
118	Буденновск	37061	44.783	44.133	136	1	red	0.770266	0.781611
119	Дивное	34858	45.917	43.350	87	1	red	0.756600	0.801403
120	Михайловск	34949	45.109	42.113	452.9	1	red	0.735011	0.787301

121 rows × 9 columns



Заключение

- Подготовлен обучающий пример для немецкой метеостанции с использованием jupyter-lab
- Подготовлены 121 datasets за 10 лет для ФГБУ Северо-Кавказкого УГМС с частотой записи с периодичностью 3 часа. Данные взяты с сайта rp5.ru
- Планируем обучить модель в июле-августе 2021 для предсказания метеоданных на 1-3 дня вперед
- Для обучения модели используются ресурсы ННГУ, сервера с 8 GPU Nvidia A100
- Планируем рассмотреть задачу с прогнозом мощности ВЭУ Lagerway L100 P=2.5 МВт по данным прогноза скорости ветра в Ставропольском крае РФ

Дополнительные слайды

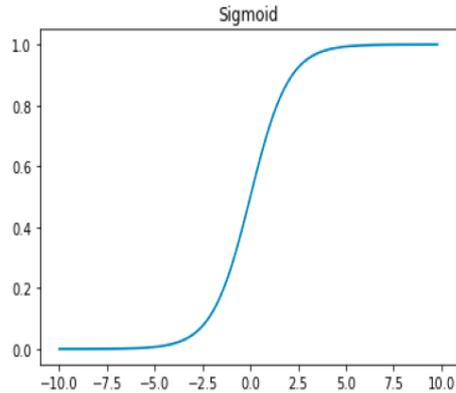
Новые ВЭС на юге РФ

<u>Березовская ВЭС</u> строится	220.00 МВт
<u>Кочубеевская ВЭС</u> в эксплуатации	210.00 МВт
<u>Адыгейская ВЭС</u> в эксплуатации	150.00 МВт
<u>Бондаревская ВЭС</u> строится	120.00 МВт
<u>Марченковская ВЭС</u> строится	120.00 МВт
<u>Ейская ВЭС</u> строится	72.00 МВт
<u>Кармалиновская ВЭС</u> строится	60.00 МВт
<u>Медвеженская ВЭС</u> проектируется	60.00 МВт
<u>Зимовниковская ВЭС</u> строится	0.00 МВт

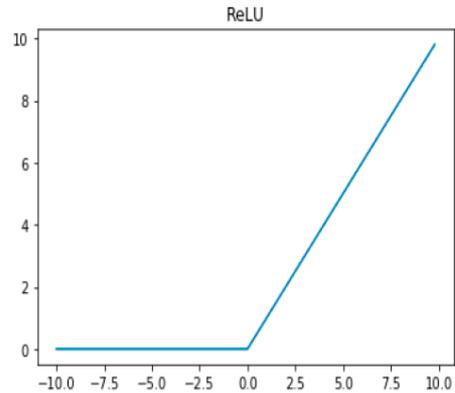
Математический аппарат для нейронной сети

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

$$\text{ReLU}(x) = \max(0, x) \quad (2)$$



(a)



(b)

$$g_t = \nabla_{\theta} f_t(\theta_{t-1}) \quad (3)$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (4)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (5)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (6)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (7)$$

$$\theta_t = \theta_{t-1} - \frac{\alpha \cdot \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (8)$$

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |y_n - \hat{y}_n| \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^N (y_n - \hat{y}_n)^2}{N}} \quad (10)$$

Использование ПО: TensorFlow, PyTorch, Scikit-learn, pandas

Прогноз кривой мощности ВЭУ

$$P_w = \frac{1}{2} \rho A v^3 \quad (1)$$

$$P_e = \eta C_p P_w \quad (2)$$

The maximum theoretical mechanical power that can be extracted by wind turbines is 0.5926 (the Betz limit)

Polynomial regression has been extensively used in the literature to estimate the power curve of wind turbines. This model can be considered as a standard extension of the linear regression $p_i = \beta_0 + \beta_1 v_i + \epsilon_i$, with a polynomial function

$$p_i = \beta_0 + \beta_1 v_i + \beta_2 v_i^2 + \dots + \beta_k v_i^k + \epsilon_i. \quad (3)$$

Model (3) can be written as

$$\mathbf{P} = \mathbf{V}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4)$$

where $\mathbf{P} = (p_1, p_2, \dots, p_n)^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$, and \mathbf{V} is a matrix with its i th row being defined as $\mathbf{V}_i = (1, v_i, v_i^2, \dots, v_i^k)$. We use the least squares method to estimate $\boldsymbol{\beta}$ by minimizing the residual sum of squares (RSS)

$$\text{RSS}(\boldsymbol{\beta}) = (\mathbf{P} - \mathbf{V}\boldsymbol{\beta})^\top (\mathbf{P} - \mathbf{V}\boldsymbol{\beta}). \quad (5)$$

Differentiating (5) with respect to $\boldsymbol{\beta}$, we solve

$$\frac{\partial \text{RSS}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{V}^\top (\mathbf{P} - \mathbf{V}\boldsymbol{\beta}) = 0 \quad (6)$$

where \mathbf{V}^\top stands for the transpose of \mathbf{V} and obtain

$$\hat{\boldsymbol{\beta}} = (\mathbf{V}^\top \mathbf{V})^{-1} \mathbf{V}^\top \mathbf{P}. \quad (7)$$

The ten-fold cross-validation

$$\text{CV}_{(10)} = \frac{1}{10} \sum_{i=1}^{10} \text{MSPE}_i \quad (8)$$

Прогноз $P(U)$ для ВЭУ Vestas Wind Systems A/S, Furlander AG

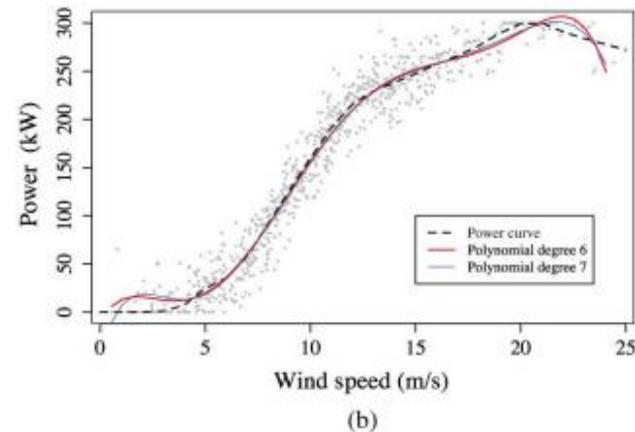
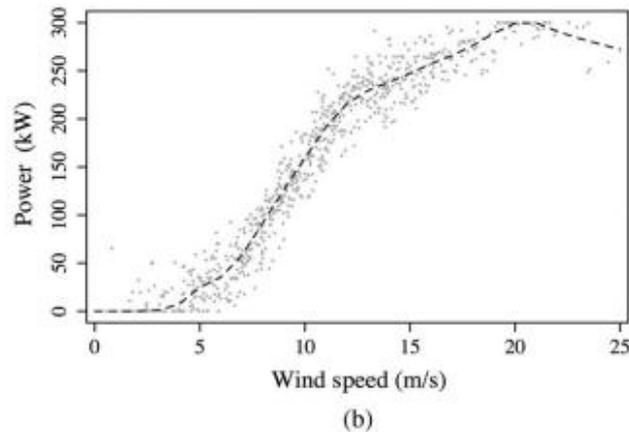
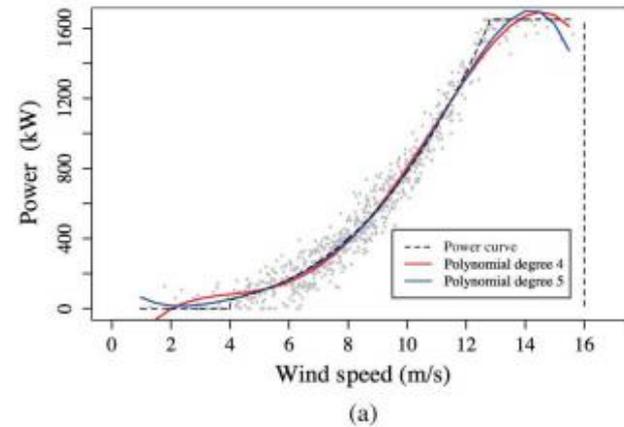
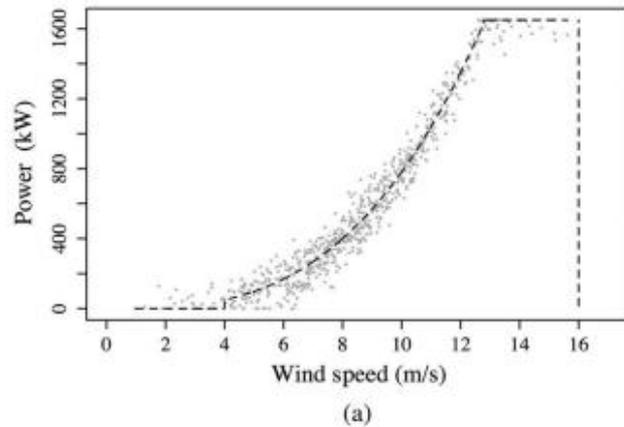


Fig. 1. The manufacturer wind turbine power curves for (a) turbine model V82 and (b) turbine model FL-255 with the scatter plots of 720 generated wind speed and hourly produced power for 1 month, assuming normally distributed noises about the manufacturer power curves with standard deviations $\sigma_\epsilon = 100$ and $\sigma_\epsilon = 20$, respectively.

Fig. 2. Fitted polynomial regression models for (a) turbine model V82 with degrees 4 and 5 and (b) turbine model FL-255 with degrees 6 and 7. The black-dashed curves are theoretical power curves from the manufacturers.