Water quality forecasting

Aida Usmanova

71%

of Earth is covered by water

3%

is drinkable

1%

is readily available

Larissa

What causes water pollution?

Rapid urban development

Building roads, houses, industries.

Agriculture

Chemicals and pesticides used to protect crops

Oil spills

Oil leaks from vehicles and mechanic trades

Radioactive waste

Sideways of nuclear energy production

How to determine water quality?

Biological Oxygen Demand

tells how much oxygen is being removed from water

Ammonia

is nutrient for the plant growth

Dissolved Oxygen

Rivers cannot have too much oxygen in general

Nitrate

predominates in unpolluted waters

Suspended solids

could be organic or inorganic. High concentrations are attributed to human influence

The Southern Bug River Basin dataset

(https://www.kaggle.com/vbmokin/wq-southern-bug-river-01052021)

- 22 stations
- Observations from 2000 to 2021
- Parameters:
  - BOD
  - NH4
  - O2
  - NO3, NO2
  - Suspended solids
  - SO4
  - CL

Water Quality Index

1. $$W_i = \frac{w_i}{\Sigma_1^n w_i}$$

2. $$q_i = \frac{C_i}{S_i} * 100$$

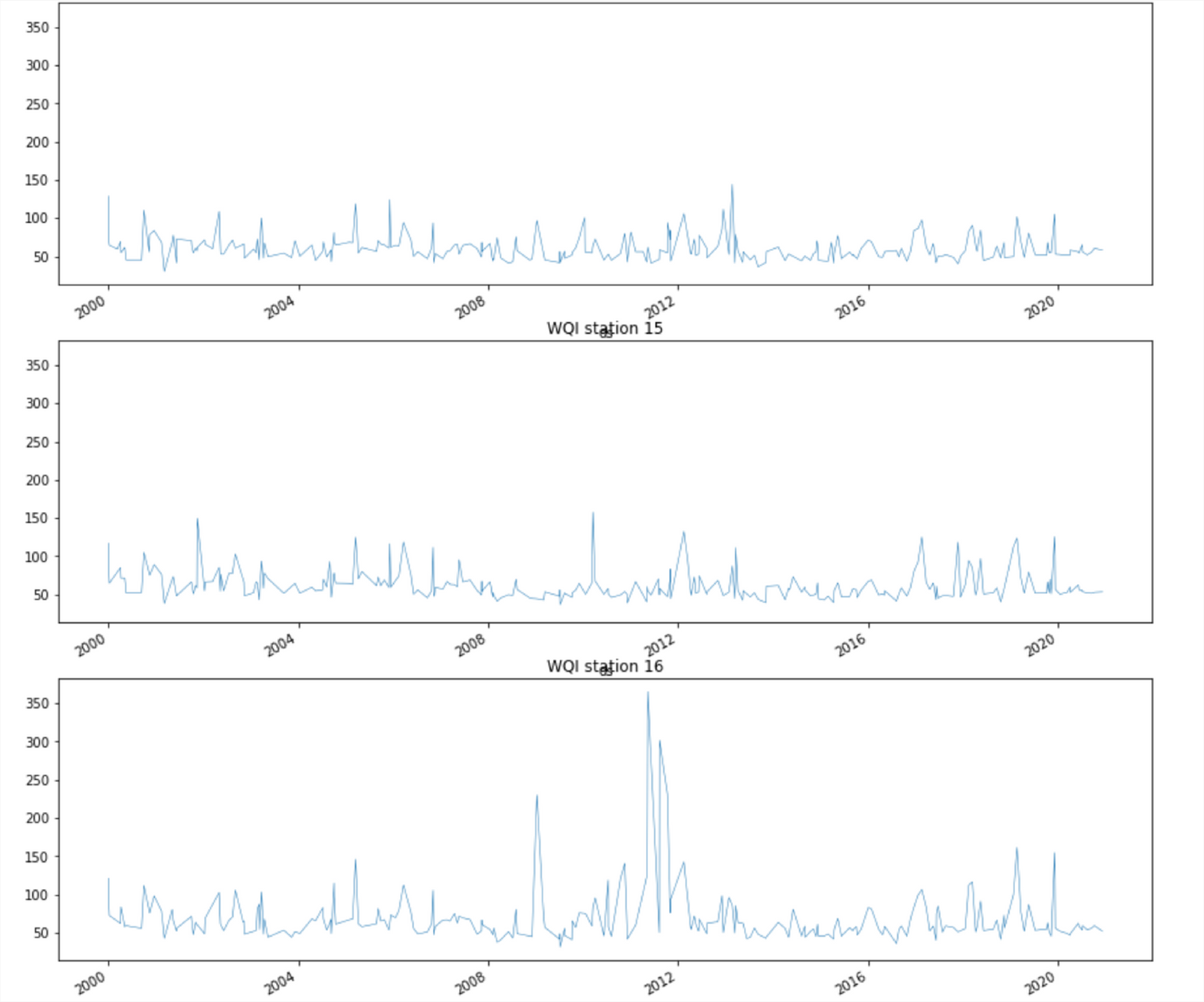
3. $$SI_i = W_i * q_i$$
$$WQI = \Sigma_1^n SI_i$$

Parameter	Standard (mg/l)	Weight	Relative weight
BOD	0 - 5	3	0.1
Ammonia	0 - 0.5	4	0.133
Dissolved oxygen	6.5 - 12	4	0.133
Nitrate	0 - 18	5	0.167
Nitrite	0 - 0.1	3	0.1
Chloride	250 - 1000	3	0.1
Sulphate	200 - 400	4	0.133
Suspended solids	0 - 600	4	0.133
		$\sum w = 30$	$\sum W = 1.00$

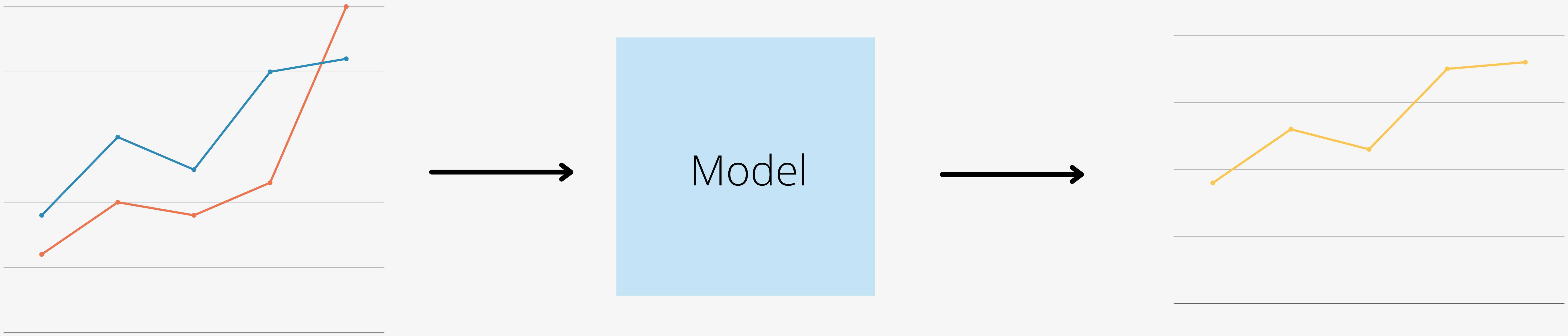
[1] Batabyal A. K., Chakraborty S. Hydrogeochemistry and water quality index in the assessment of groundwater quality for drinking uses

Water classification based on WQI:

- < 50 - Excellent
- 50 - 100 - Good
- 100 - 200 - Poor
- 200 - 300 - Very poor
- > 300 - Unsuitable for drinking



Data preparation



WQI of stations 14 and 15

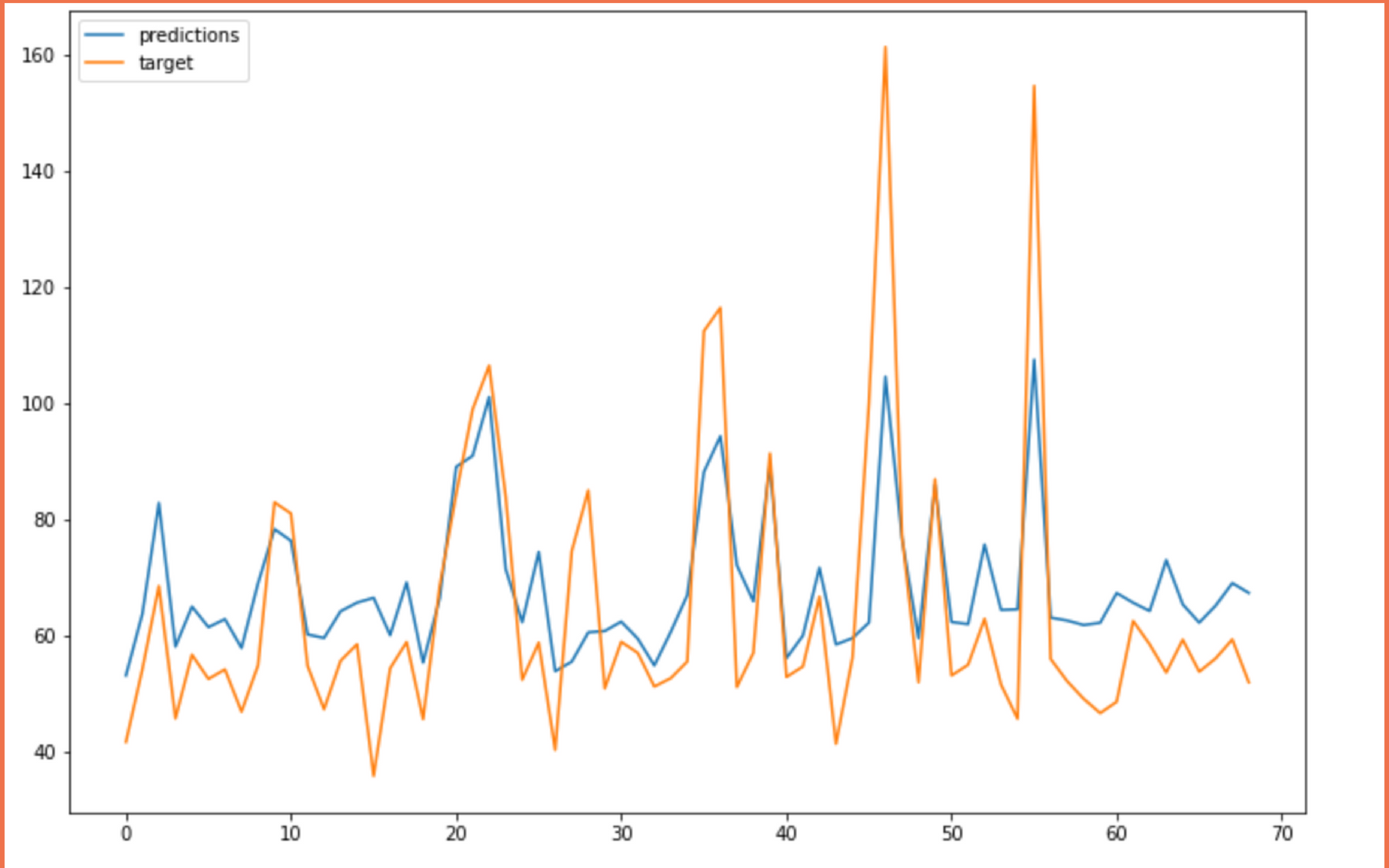
WQI of station 16

Train size: 164
Test size: 69

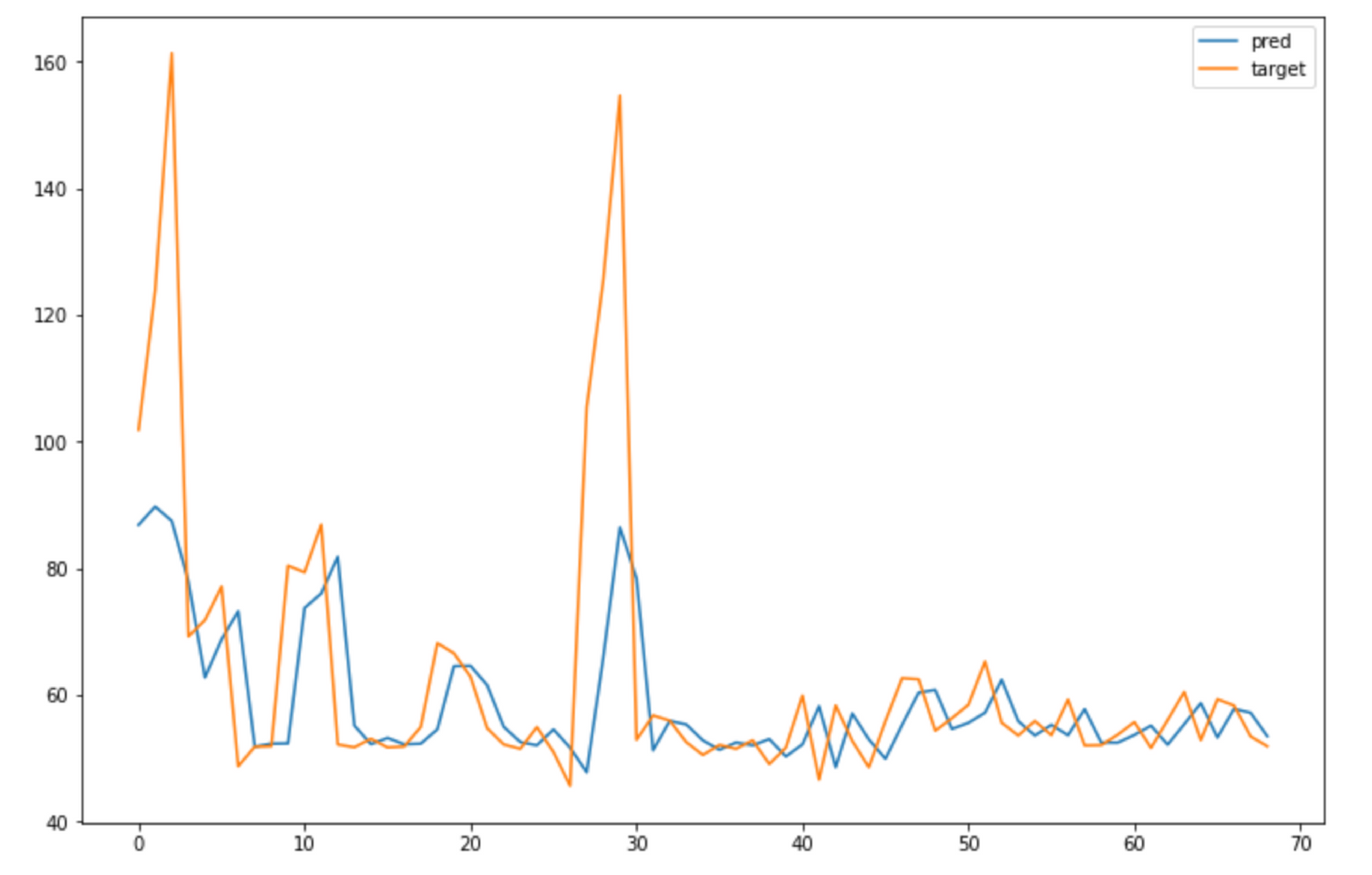
Linear regression

RMSE = 15.37

R2 score = 0.58



LSTM



RMSE = 18.20

R2 score = 0.37

Model	RMSE	R2	Adjusted R2
Linear Regression	15.37	0.58	0.56
LSTM	18.20	0.37	0.35

Results

- Benefit from a longer sequence
- Try more parameters
- LSTM doesn't suit time series forecasting

[2] Gers F. A., Eck D., Schmidhuber J. Applying LSTM to time series predictable through time-window approaches

Thank you for attention!

[1] Batabyal A. K., Chakraborty S. Hydrogeochemistry and water quality index in the assessment of groundwater quality for drinking uses //Water Environment Research. – 2015. – T. 87. – №. 7. – C. 607-617.

[2] Gers F. A., Eck D., Schmidhuber J. Applying LSTM to time series predictable through time-window approaches //Neural Nets WIRN Vietri-01. – Springer, London, 2002. – C. 193-200.