Queen Mary
University of London
Science and Engineering

Interpretable Machine Learning for Dissolved Oxygen Forecasts

UK HEP Forum 2024
26/11/2024

Hankun He, Takuya Boehringer, Benjamin Schäfer,
Kate Heppell, Christian Beck

Motivation

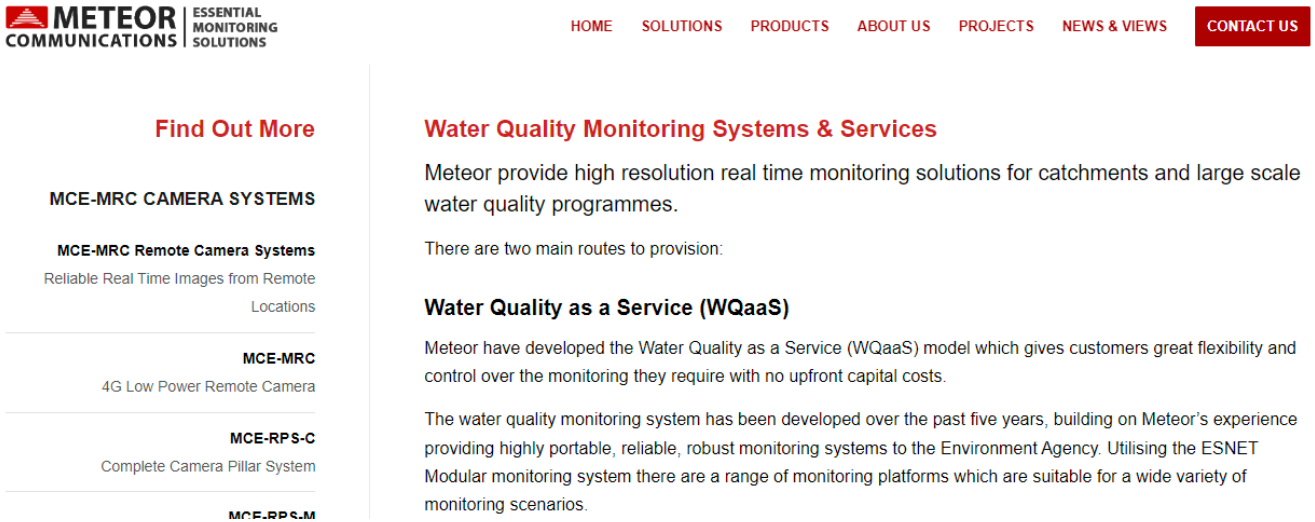
The ongoing sewage crisis in the River Thames has sparked deep concerns:

- Ecosystem
- Human health
- Economy

Image Credit: Maureen McLean/Alamy

Image Credit: ANGLING TRUST/PA

The data set considered



METEOR COMMUNICATIONS | ESSENTIAL MONITORING SOLUTIONS

HOME SOLUTIONS PRODUCTS ABOUT US PROJECTS NEWS & VIEWS CONTACT US

Find Out More

MCE-MRC CAMERA SYSTEMS

MCE-MRC Remote Camera Systems
Reliable Real Time Images from Remote Locations

MCE-MRC
4G Low Power Remote Camera

MCE-RPS-C
Complete Camera Pillar System

MCE-RPS-M

Water Quality Monitoring Systems & Services

Meteor provide high resolution real time monitoring solutions for catchments and large scale water quality programmes.

There are two main routes to provision:

Water Quality as a Service (WQaaS)

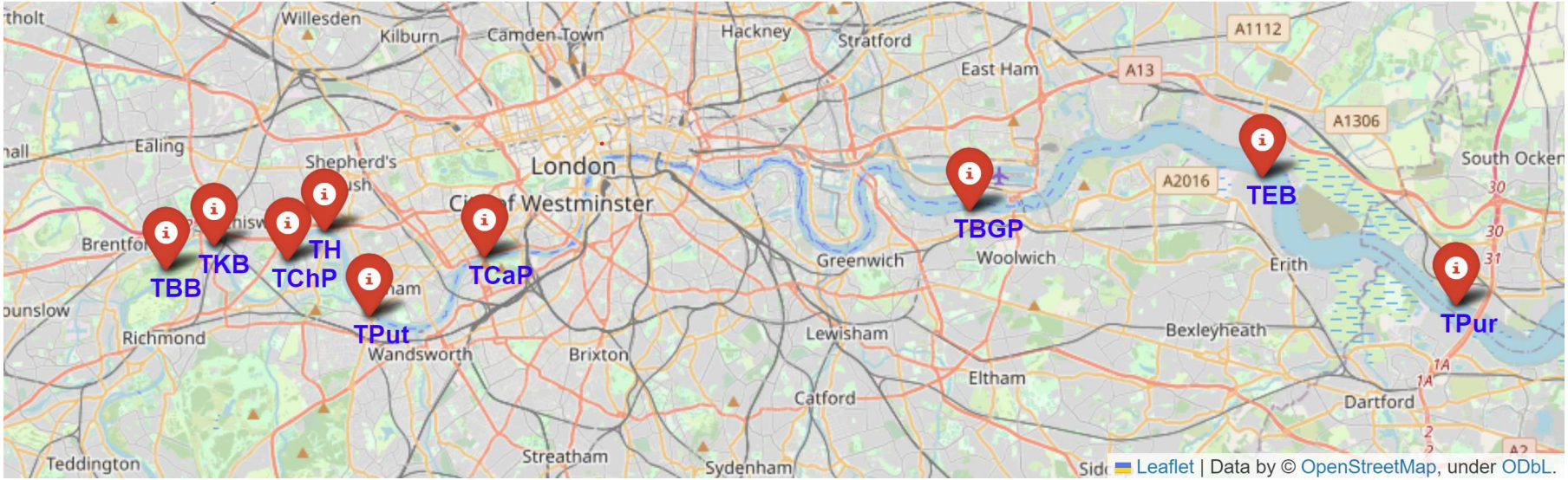
Meteor have developed the Water Quality as a Service (WQaaS) model which gives customers great flexibility and control over the monitoring they require with no upfront capital costs.

The water quality monitoring system has been developed over the past five years, building on Meteor's experience providing highly portable, reliable, robust monitoring systems to the Environment Agency. Utilising the ESNET Modular monitoring system there are a range of monitoring platforms which are suitable for a wide variety of monitoring scenarios.

The geographical visualisations were generated using the Folium Python library with map data sourced from OpenStreetMap.

The data set considered

Thames Brentford Barge (TBB), Thames Kew Barge (TKB), Thames Chiswick Pier (TChP), Thames Hammersmith (TH), Thames Putney (TPut), Thames Cadogan Pier (TCaP), Thames Barrier Gardens Pier (TBGP), Thames Erith Barge (TEB), and Thames Purfleet (TPur).



He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

The geographical visualisations were generated using the Folium Python library with map data sourced from OpenStreetMap.

The data set considered

Water quality indicators:

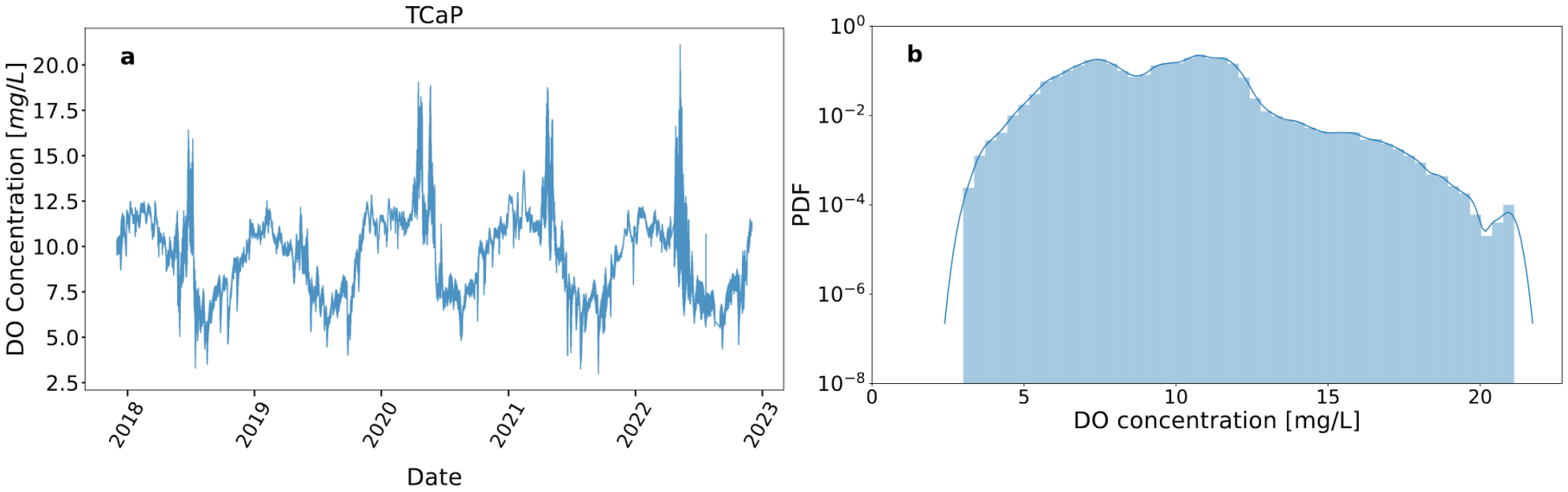
- Dissolved oxygen (DO)
- Temperature
- Electrical conductivity (COND)
- pH
- Turbidity
- Ammonium
- Rainfall

Measured from 01/12/2017 to 01/12/2022. Temporal resolution of 15 minutes, except for rainfall.

Calendar and time features:

- the hour of the day
- the day of the week
- the month of the year
- the time of the half-day
- the time of the year

Water quality statistics



He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

Time series forecasting

The employed forecasting models:

- Last
- Repeat
- Linear
- Sequential Dense Network (Dense)
- Convolutional Neural Network (Conv)
- Long Short-Term Memory (LSTM)
- Informer

He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

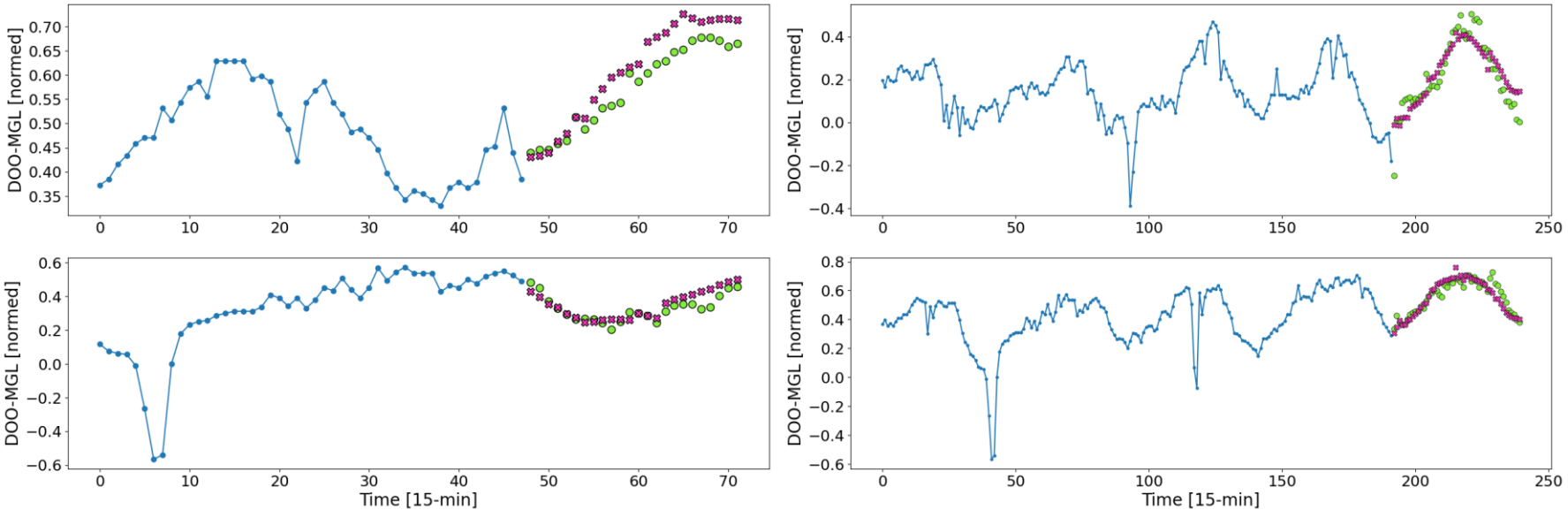
Performance metrics

Input_Pred	Metrics	Last	Repeat	Linear	Dense	Conv	LSTM	Informer
48_1	MAE	**0.076**	0.252	0.083	0.108	0.126	0.080	0.114
	SMAPE [%]	**12.254**	31.632	14.779	19.206	21.357	13.469	16.372
48_12	MAE	0.190	0.178	0.188	0.200	0.214	0.183	**0.160**
	SMAPE [%]	25.248	24.050	24.574	26.310	27.811	24.818	**21.763**
48_24	MAE	0.240	0.178	0.267	0.313	0.329	0.195	**0.176**
	SMAPE [%]	30.375	24.061	31.896	36.330	37.667	25.473	**23.513**
48_48	MAE	0.253	0.178	0.291	0.347	0.339	0.179	**0.175**
	SMAPE [%]	31.703	24.070	34.223	39.092	38.627	25.439	**23.583**
96_1	MAE	**0.076**	0.251	0.076	0.085	0.089	0.080	0.108
	SMAPE [%]	**12.254**	31.451	12.370	15.109	16.097	13.709	15.539
96_12	MAE	0.190	0.178	0.189	0.199	0.211	0.180	**0.146**
	SMAPE [%]	25.273	24.064	24.609	26.306	27.514	23.636	**21.099**
96_24	MAE	0.240	0.178	0.268	0.296	0.313	0.205	**0.153**
	SMAPE [%]	30.391	24.073	36.822	35.525	38.049	33.025	**22.227**
96_48	MAE	0.253	0.178	0.292	0.315	0.319	0.174	**0.157**
	SMAPE [%]	31.727	24.094	34.336	36.817	37.268	24.997	**22.292**
192_1	MAE	**0.076**	0.260	0.076	0.086	0.086	0.082	0.106
	SMAPE [%]	**12.254**	32.544	12.385	15.260	14.794	13.841	15.968
192_12	MAE	0.191	0.179	0.189	0.200	0.220	0.183	**0.135**
	SMAPE [%]	25.337	24.190	24.720	26.332	28.629	24.100	**20.063**
192_24	MAE	0.241	0.179	0.269	0.298	0.313	0.212	**0.143**
	SMAPE [%]	30.468	24.199	32.107	35.674	36.839	27.997	**21.313**
192_48	MAE	0.254	0.179	0.292	0.316	0.319	0.175	**0.150**
	SMAPE [%]	31.811	24.217	34.403	36.896	37.480	24.468	**21.955**

He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

Forecast examples

Example windows for Informer's predictions of t ∈12, 48 with input lengths of 48 and 192 time steps:



He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

Time series forecasting

The encoder generates three distinct representations, the key (K), query (Q), and value (V) for each input.

The ProbSparse Self-attention:

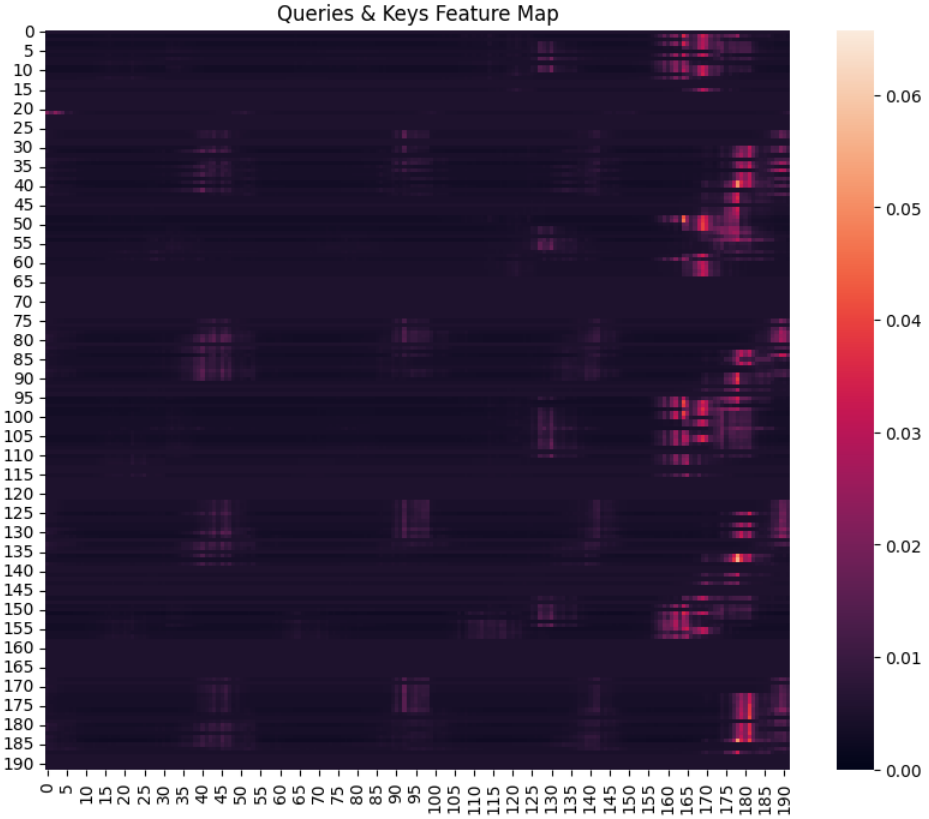
$Attention(Q, K, V) = \text{softmax}(\overline{Q}K^T / \sqrt{d})V$,

where d represents the dimensionality of the Qs and Ks, while $\overline{Q}$ is the sparse Q matrix containing the top-u Qs.

H. Zhou et al. *Proceedings of the AAAI conference on artificial intelligence,* 2021

Interpretability

- Attention to periodicity.
- Attention to the recent time steps.
- Long-range dependencies



He, Boehringer, Schäfer, Heppell, Beck, *Scientific Reports*, 2024

Summary and future outlook

- Demonstrated superior performance in DO long-term forecasts.
- Identified and interpreted the input time windows that drive forecast capabilities.
- Potential sustainable solutions to monitor DO dynamics.

Thank you

Please scan the QR code to view the complete paper

Queen Mary
University of London
Science and Engineering