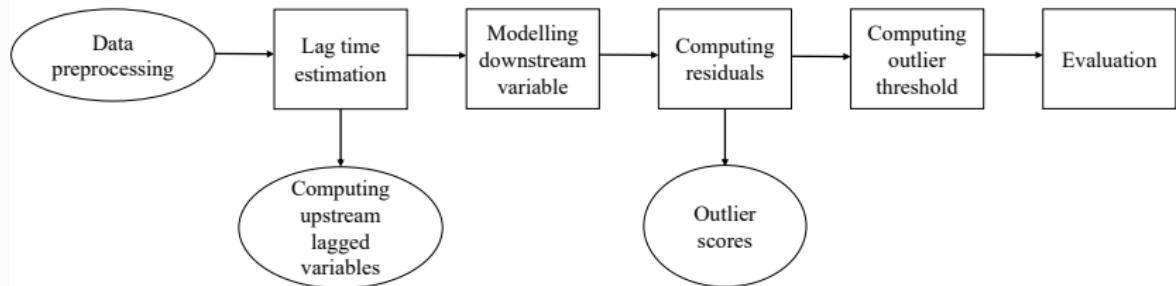


Outlier Detection in Water Quality Variables Utilising the Temporal Correlation

Puwasala Gamakumara

October 18, 2021

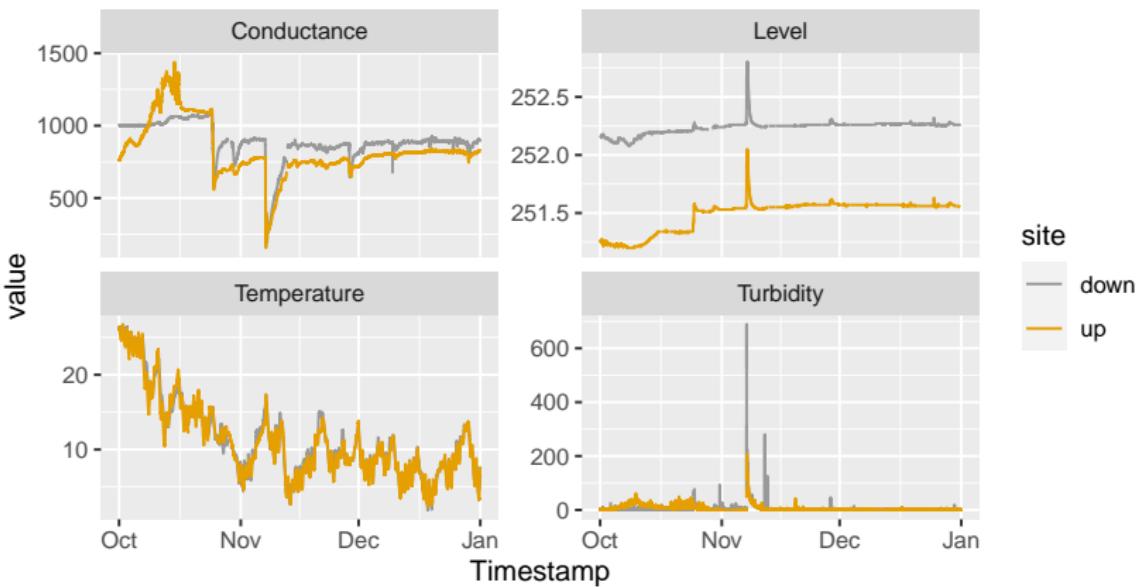
Anomaly detection framework



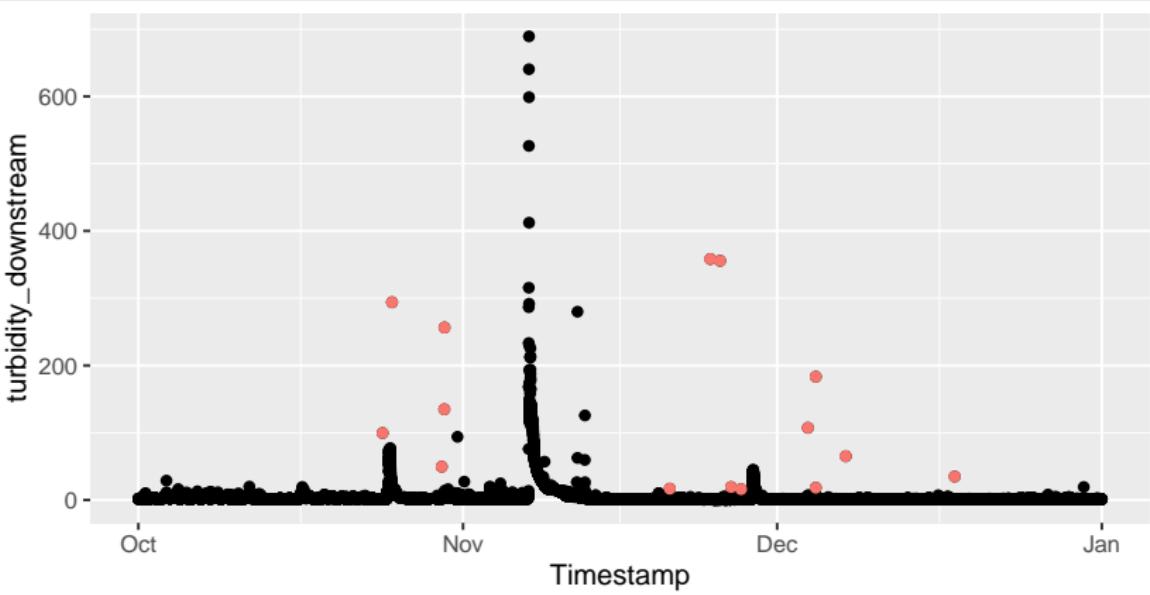
Data

- Pringle Creek - Texas, USA
- Two sensor locations (200m apart)
- **Variables available** - Turbidity, Conductivity, Level and Temperature
- **Time span** - 01-10-2019 to 31-12-2019
- **Frequency** - 5 minute intervals

Time plots



Turbidity downstream



Model - 1

$$\begin{aligned} \log(\text{turbidity_downstream}_t) \sim & s_1(\log(\text{turbidity_upstream}_{t-d_t})) \\ & + s_2(\text{conductance_upstream}_{t-d_t}) + s_3(\text{level_upstream}_{t-d_t}) \\ & + s_4(\text{temperature_upstream}_{t-d_t}) + \epsilon_t \end{aligned}$$

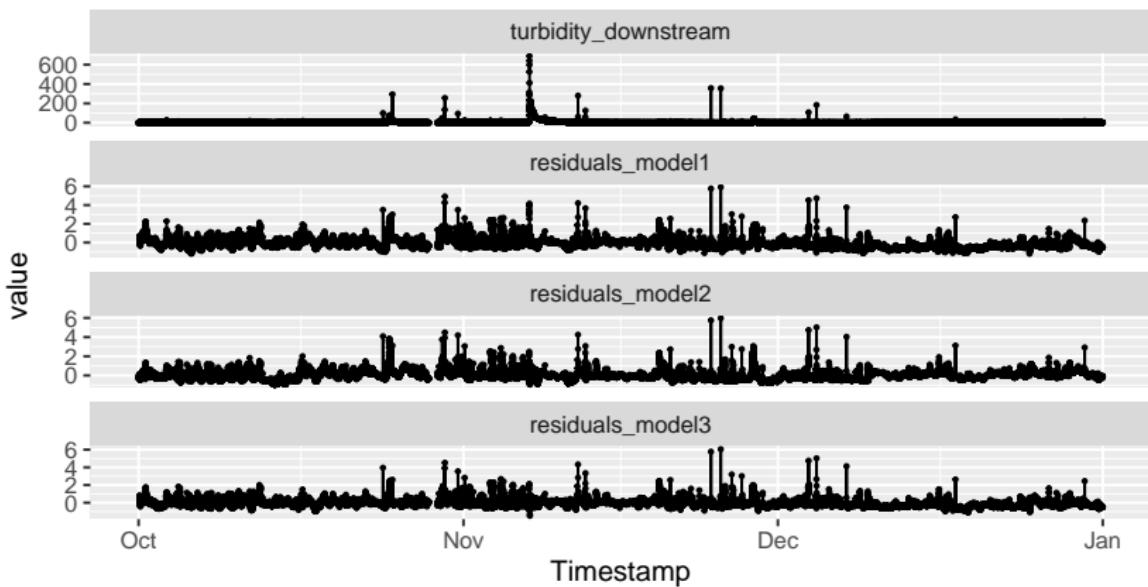
Model - 2

$$\begin{aligned} \log(\text{turbidity_downstream}_t) \sim & s_2(\text{conductance_dowstream}_{t-1}) \\ & + s_3(\text{level_downstream}_{t-1}) + s_4(\text{temperature_downstream}_{t-1}) \\ & + \epsilon_t \end{aligned}$$

Model - 3

$$\begin{aligned} \log(\text{turbidity_downstream}_t) \sim & s_1(\log(\text{turbidity_upstream}_{t-d_t})) \\ & + s_2(\text{conductance_upstream}_{t-d_t}) + s_3(\text{level_upstream}_{t-d_t}) \\ & + s_4(\text{temperature_upstream}_{t-d_t}) + s_5(\text{conductance_downstream}_{t-1}) \\ & + s_6(\text{level_downstream}_{t-1}) + s_7(\text{temperature_downstream}_{t-1}) \\ & + \epsilon_t \end{aligned}$$

Residuals from the three models



- We use Peak-Over-Threshold method to estimate the outlier threshold

Results

method	TP	TN	FP	FN	OP	accuracy	ER
model3	14	25750	17	0	0.9990	0.9993	7e-04
model1	13	25792	17	1	0.9626	0.9993	7e-04
model2	6	25968	1	9	0.5711	0.9996	4e-04

Classifying the outliers detected from model...3

