Exploratory Analysis of Reservoir Water Quality and Catchment Conditions

PAUL YACOBELLIS, DATA SCIENTIST

Today

1. Objective – What Are We Trying to Achieve?
2. Data Analysis Approach
3. Results and Key Insights
4. Data Science Skills
5. Use of AI

Objective

What Are We Trying to Achieve?

a) enhance understanding of drinking reservoir conditions

b) generate actionable insights from data that drive impact and value for Water NSW

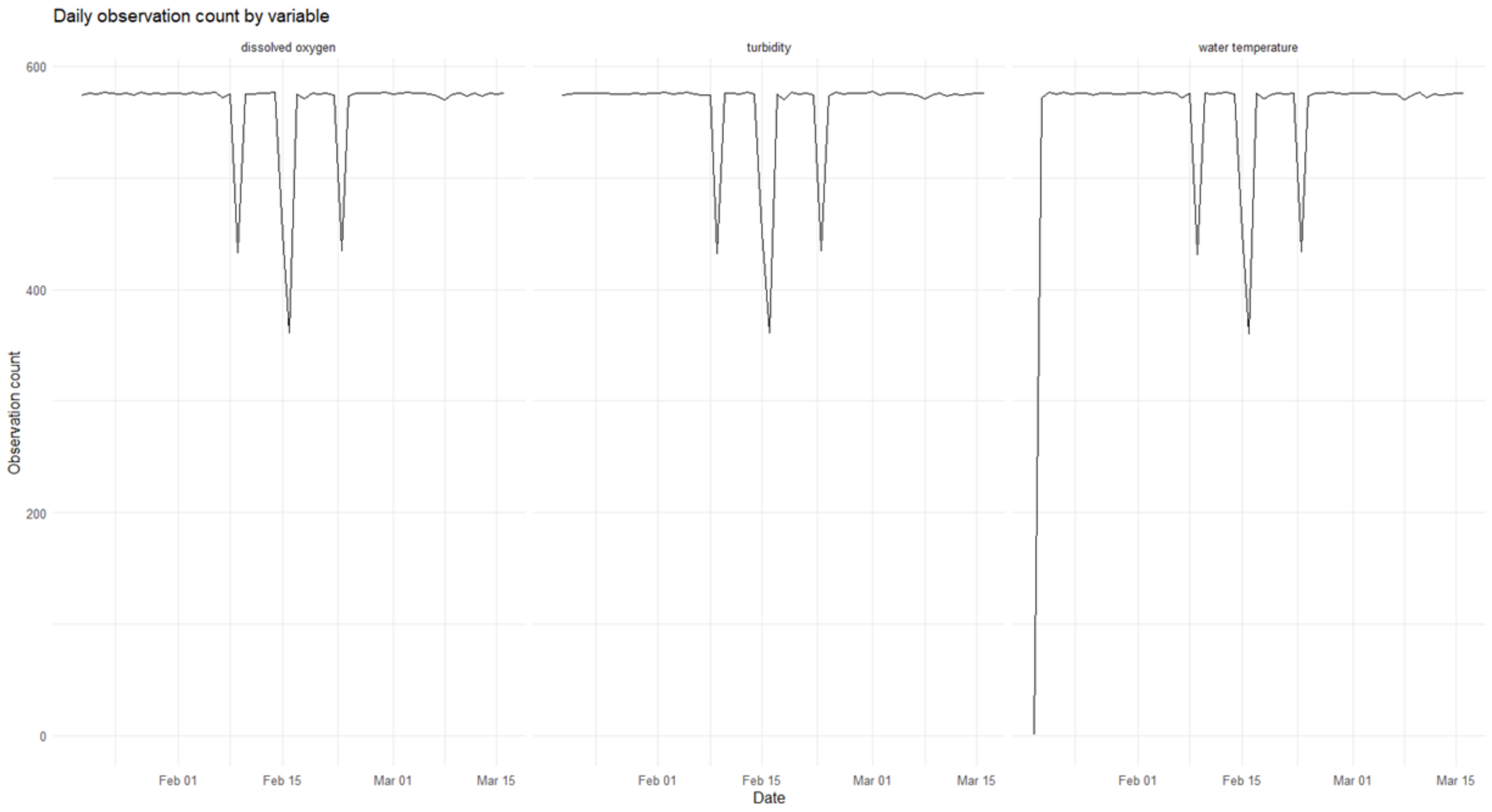
c) to assist water modellers, engineers, and operators to make better decisions

Data Analysis Approach

- clarified requirements of stakeholders and business
- performed exploratory data analysis to understand data quality and real-world behaviour
- analysed temporal patterns, lag effects
- validated findings with standard statistical techniques
- sensed check interpretations with former colleagues: flood modeller and water economist

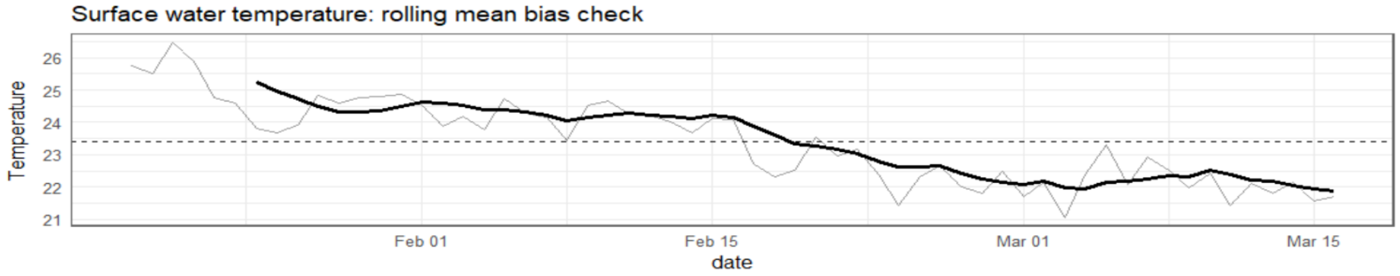
Data Coverage / Bias Considerations

- Daily WQ reading counts
  - Drop in sensor counts on same days
  - Dates line up with peak rain events
  - True WQ values biased downward
  - Missing water temp won't have material impact

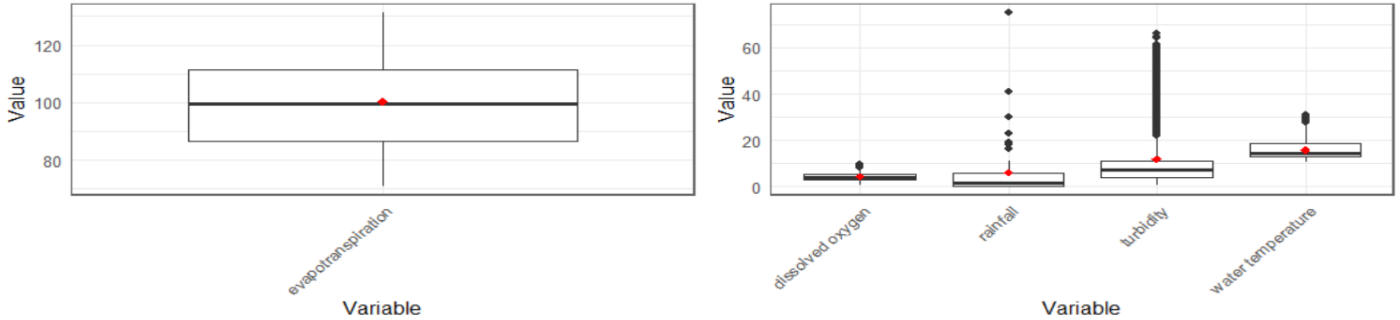


Data Coverage / Bias Considerations

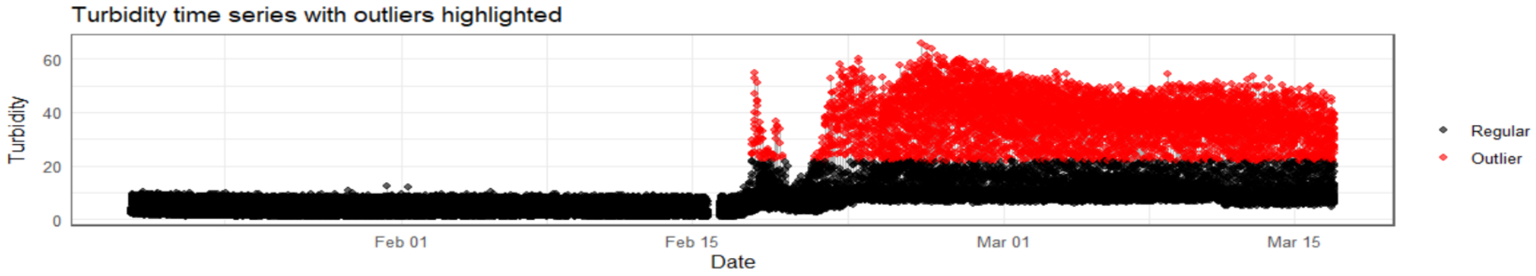
- Bias detection
  - surface temp rolling average



- Box Plot analysis
  - skewness, outlier detection

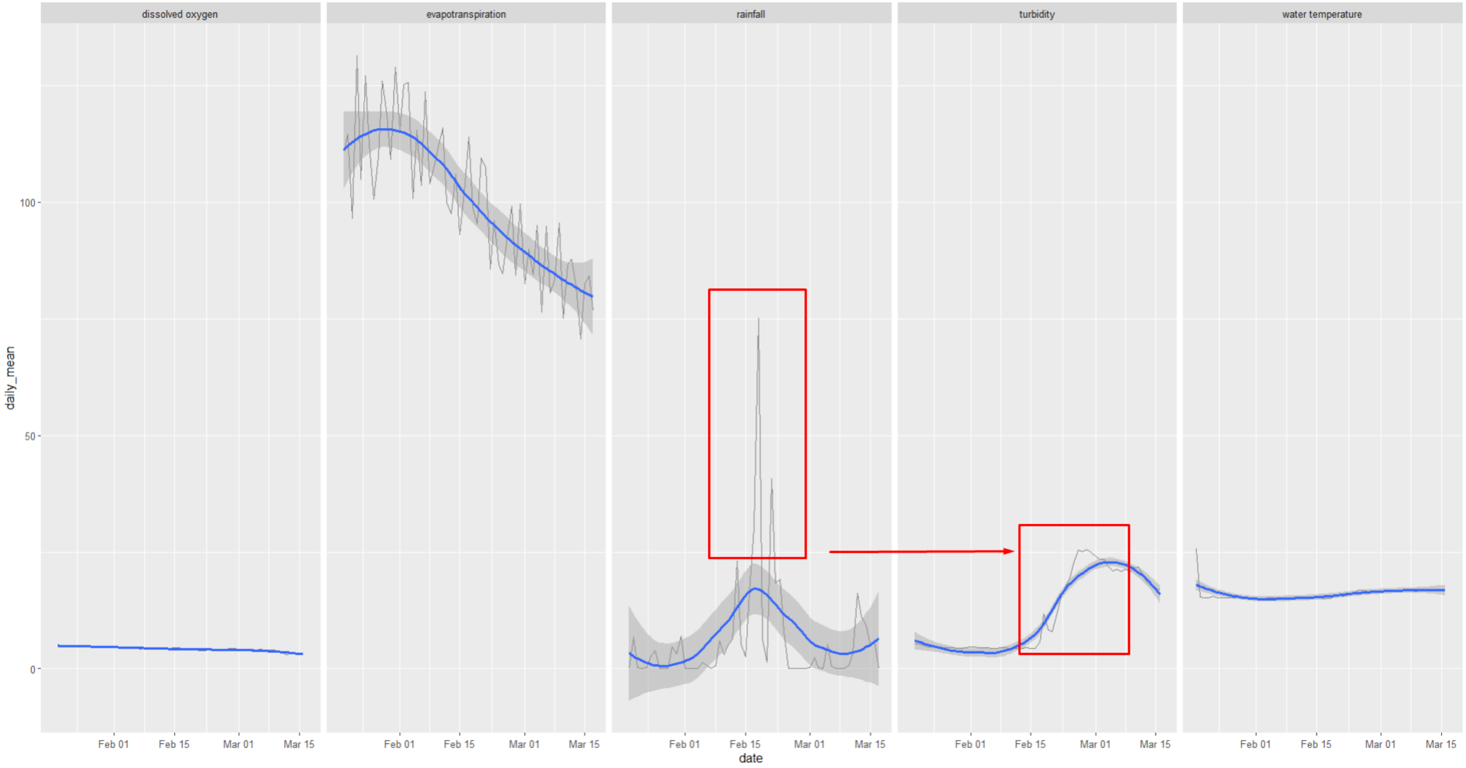


- Outlier analysis
  - Turbidity time series



Key Patterns Observed

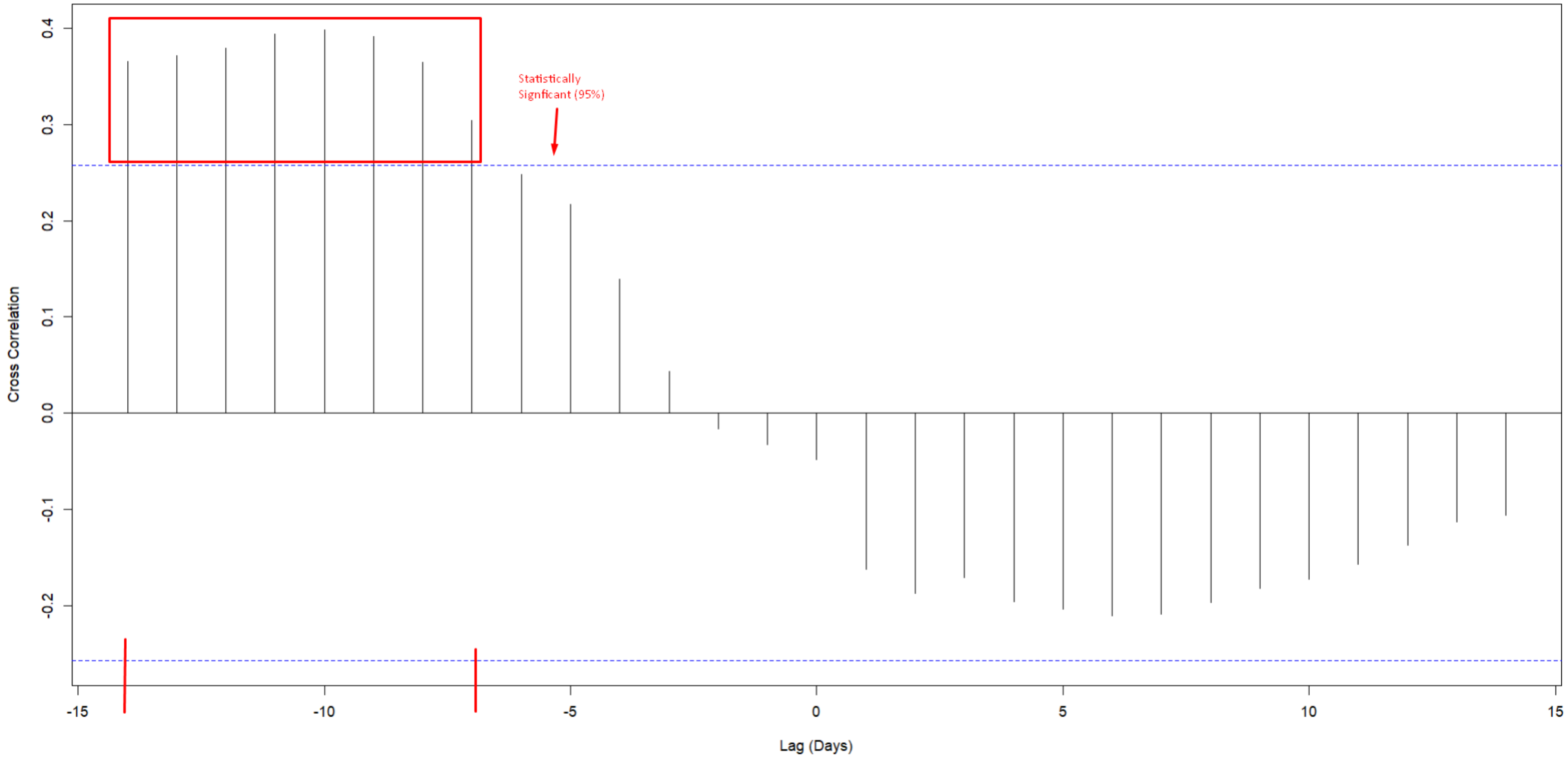
- DO and Water Temp constant
  - But different story at depth
- Evapotranspiration
  - normal seasonal decline
- Lag effect from rainfall -> turbidity



Key Patterns Observed

- Cross Correlation
  - Positive lagged relationship 7-14 days after rainfall

Rainfall – Turbidity Cross Correlation

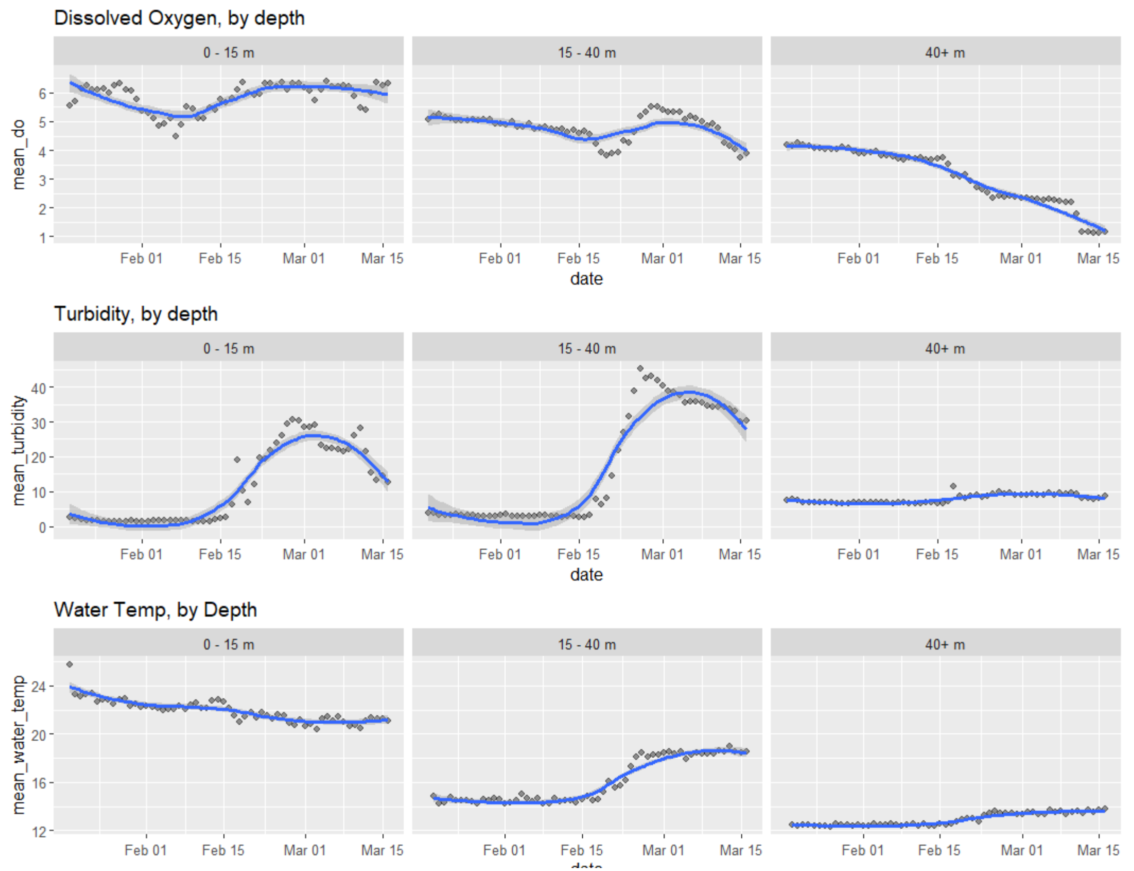


Key Patterns Observed

- DO
  - stable at surface (lower risk of Algal Blooms)
  - DO declines at depth; warrant further investigation with hydrologist

- Turbidity
  - spikes consistent with rainfall lag story

- Water Temp
  - spike at middle depth – requires further investigation



Results , Key Insights, Recommendations

1. rainfall → turbidity is delayed
   - creates opportunity to mitigate risk

2. spikes in turbidity are event-driven
   - driven by rainfall, not randomness

3. WQ at depth warrants further investigation
   - water temperature and DO

Future Data Science Applications

1. more data
   - other reservoir same time
   - same reservoir, different time
   - longer time horizon

2. ARIMAX forecast of turbidity (use rainfall as input to model)
   - only if more data available

3. spatial modelling – examine upstream catchment profiles
   - land use analysis to isolate higher nutrient loading source

4. self-service application
   - lightweight Streamlit app for repeated exploration / analysis for nontechnical users

Use of AI

1. used to generate some boiler plate code
   - decision and judgement human led
   - manually revised for better precision
   - unexplainable outputs discarded

2. middleman of sense checking:
   - results -> hypothesis -> AI sense check -> flood modeller sense check

3. conservative, incremental approach