Multivariate Polynomial Regression Modeling of Boundary E coli Concentrations using Boundary Flows and Baseflow Indices

JAGUPILLA, SARATH CHANDRA K., GURUMURTHY, PRANEETH.*, RAMASWAMY, VENKAT SUNDAR., VACCARI, DAVID A.

STEVENS INSTITUTE OF TECHNOLOGY
Contents

- Introduction
- Objective
- Methodology
  - Baseflow Index (BFI)
  - Models
- Models and Statistics
- Conclusions
Introduction

- Objective: Modelling *E coli* concentrations using Multivariate Polynomial Regression (MPR) with boundary flows and base flow index

- Location: Lower Passaic River, at Paterson, NJ (stretch of 8.65 miles)

- Sampling Events
  - Between July 2009 and October 2011
  - Events – 9 sampling events: 3 dry and 6 wet

- Modeling
  - Baseflow Index: HydroOffice 2015 BFI 3.0 module
  - Software: TaylorFit 1.99 for MPR modelling
Study Area
Objective

- To model the *E. coli* concentrations at
  - Totowa (upstream boundary)
  - Pennington (Tributary)
  - Westside (Tributary)
  - Goffle (Tributary)
  - Elberon (Stormwater Outfall)

- Predictors
  - Boundary flows (at Little Falls and Dundee Dam)
  - Baseflow indices (at Little Falls and Dundee Dam)

- Model Performance (relative to symbolic regression models built earlier)
  - Goodness of fit
  - Complexity
Methodology (Data and BFI)

- **Data**
  - Boundary flows from USGS at Little Falls and Dundee Dam
  - *E coli* concentrations (log transformed) collected by SIT

- **Baseflow Index as predictor**
  - Sliding interval method
  - 94 hour interval
Methodology (MPR Models and Criteria)

- **MPR Models:**
  - \( Y = \sum_{i=1}^{n} a_i Q^{b_i} R^{c_i} S^{d_i}, \quad 1 \leq i \leq n \)
  - \( n \) is the number of terms in the model,
  - \( Q, R, \) and \( S \) are predictors; \( a_i \) is the coefficient
  - \( b_i, c_i, d_i \) are exponents; \( Y \) is the dependent variable

- **Model Building**
  - Multiplicands tested: 1, 2, and 3
  - Exponents tested: -2, -1, 1, and 2

- **Model Selection**
  - \( R^2 \) – goodness of fit
  - Model simplicity
Notations for Variables in Models

- Baseflow index at Dundee Dam – $B_{dd}$
- Baseflow index at Little Falls – $B_{lf}$
- Flow at Dundee Dam – $Q_{dd}$
- Flow at Little Falls – $Q_{lf}$
- Net flow – $Q_{dd-lf}$
- Log transformed concentration of E Coli – log EC
Totowa Models

Totowa model using symbolic regression

\[
\log EC = 3.686 - 1.891e^{\frac{x^2}{2}}
\]

Where \( x = \frac{39.47 \times \frac{dQ_{lf}}{dt} - 473.8 \times \tan h(y)}{Q_{lf} - \frac{dQ_{lf}}{dt} - 1250 - 4.355 \frac{dQ_{dd}}{dt}} \), where \( y = \frac{dQ_{lf}}{dt} - 673.1 \times \left( \frac{dQ_{dd}}{dt} / Q_{lf} \right) - 0.365 \)

G.O.F = 0.67

Totowa model model using MPR

\[
\log EC = 2.54 + 0.37 \times \left( \frac{Q_{lf}}{Q_{dd} \times B_{dd}} \right) - 0.84 \times (Q_{dd} - Q_{lf}) \\
\times B_{lf}^2 \times B_{dd}^2 + 0.15 \times (Q_{dd} \times B_{dd} \times B_{lf}^2) \\
- 1.97 \times \frac{(B_{lf} \times Q_{lf}^2)}{Q_{dd}^2} - (1.08 \times 10^{-10}) \\
\times (Q_{dd} - Q_{lf} \times Q_{lf}^2 \times Q_{dd}^2)
\]

G.O.F = 0.87
Observed vs Predicted – Totowa MPR Model (Upstream Boundary)

$R^2 = 0.8705$

$N = 46$
Westside Models

Westside model using symbolic regression

\[ \log EC = 1.771 + 1.388 \ln \left( 12.72 + \frac{dQ_{lf}}{dt} \right) \]
\[ - 1.388 e^{-\left( \frac{dQ_{lf}}{dt} + \frac{dQ_{dd}}{dt} \right)^2} \]

G.O.F = 0.61

Westside model model using MPR

\[ \log EC = 2.819 - 38.28 \times \left( \frac{B_{lf} \times Q_{dd-lf}}{Q_{lf}} \right)^2 + 0.0216 \]
\[ \times (Q_{dd-lf} \times (B_{dd} \times Q_{lf})^2) - 0.018 \times (B_{lf} \times B_{dd} \times Q_{dd})^2 + 78.24 \times (Q_{dd-lf} \times \left( \frac{B_{lf}}{Q_{lf}} \right)^2 ) - 5.86 \]
\[ \times (B_{dd}^2 \times \frac{Q_{dd-lf}}{Q_{dd}}) \]

G.O.F = 0.84
Goodness of fit- Westside MPR Model

\[ R^2 = 0.8437 \]
\[ N = 28 \]
Goffle Models

Goffle model using symbolic regression

\[ \log EC = 0.917 \times \ln(Q_{lf}) - 0.712 \]
\[ - \left( \frac{-7.915 - 0.652 \frac{dQ_{lf}}{dt}}{Q_{dd} - 132.5} \right)^2 \]
\[ - 1.944e \]

G.O.F = 0.83

Goffle model model using MPR

\[ \log EC = 4.75 - 2.257 \times (Q_{dd} \times \frac{B_{dd}}{Q_{lf}}) \]

G.O.F = 0.71
Goodness of fit- Goffle MPR Model

\[ R^2 = 0.7091 \]

\[ N = 33 \]
Pennington Models

Pennington model using symbolic regression

\[
\log EC = 3.855 + \frac{49.9}{Q_{dd}} - 0.0002257Q_{dd} \\
+ 1.023 \left( e^{-\frac{(10.24 - 6413Q_{dd})^2}{2Q_{dd}}} - e^{-\frac{(Q_{dd} - 7.072)^2}{2}} \right) \\
- e^{-\frac{(33.67 - 0.06709Q_{dd})^2}{2}} - e^{-\frac{(0.1959Q_{lf} - 56.47)^2}{2}}
\]

G.O.F = 0.81

Pennington model model using MPR

\[
\log EC = 11.97 - 10.22 \times \left( \frac{B_{lf} \times Q_{dd}-lf}{Q_{dd}} \right) - 4.44 \\
\times \left( \frac{B_{lf} \times Q_{dd}-lf \times B_{dd}}{Q_{dd}} \right) + 0.021 \\
\times \left( \frac{B_{lf} \times Q_{dd}-lf}{Q_{lf}^2} \right) - 0.478 \\
\times \left( \frac{B_{lf} \times Q_{dd}-lf}{Q_{lf}} \right) + 6.01 \\
\times \left( \frac{Q_{dd}-lf \times B_{lf} \times Q_{dd}-lf}{Q_{dd}} \right)^2 + 1.403
\]

G.O.F = 0.63
Goodness of fit - Pennington MPR Model

\[ R^2 = 0.6315 \]
\[ N = 29 \]
Stormwater Models

Stormwater model using symbolic regression

\[
\log EC = 3.78 + \frac{0.0133}{\tanh(0.053Q_{dd})} \frac{dQ_{lf}}{dt} - 0.4753 \frac{dQ_{dd}}{dt}
\]

G.O.F = 0.88

Stormwater model model using MPR

\[
\log EC = 1.866 + 0.968 \times \left( B_{dd} \times B_{lf} \times Q_{dd} - l_f \right)^2 + \frac{0.524}{Q_{lf} \times (B_{dd} \times Q_{dd} - l_f)^2} + 1.401 \times \frac{Q_{lf}}{(Q_{dd} \times B_{dd})^2}
\]

G.O.F = 0.94
Goodness of fit - Stormwater MPR Model (Elberon)

$R^2 = 0.9468$

$N = 16$
Conclusions and Further Work

- Baseflow Index is a significant predictor of Log EC in the lower Passaic River
- Work demonstrates the importance of identifying suitable predictors in modeling
- Further work – Sensitivity analysis of the models to delineate effects of individual predictors on *E coli* concentrations
Acknowledgements

- Passaic Valley Sewerage Commissioners
- United States Geological Survey
- New Jersey Department of Environmental Protection
- Department of Civil, Environmental and Ocean Engineering at Stevens Institute of Technology, NJ
Thank You

Questions and suggestions ?