KTH ROYAL INSTITUTE
OF TECHNOLOGY

Applications of machine learning for safe spillway discharge

Shicheng Li & James Yang
Department of Civil and Architectural Engineering
KTH Royal Institute of Technology

Contents

Introduction

Background: hydropower

- Renewable energy source
- Increased design flood: overtopping, cavitation, etc.
- Discharge determination
- Aeration estimation

Introduction

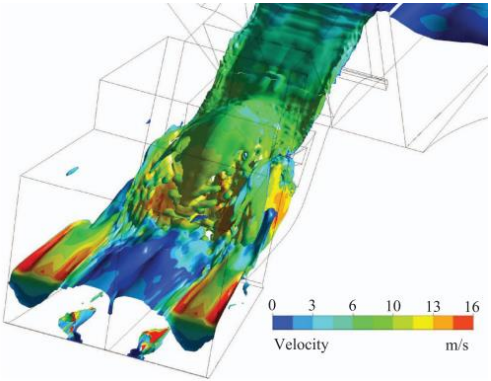
Comparison of the old and new design floods: 20−50% increase

No.	River name	Dam name	Z (m)	Q_o (m^3/s)	Δ (%)
1	Umeälv	Ajaure	45	950	40
2		Stornorrfors	20	3300	35
3		Rusfors	22	1625	27
4	Indalsälven	Midskog	27	2300	35
5		Bergeforsen	29	2300	45
6	Ångermanälven	Stenkullafors	30	1250	40
7		Edensforsen	19	1400	>40
8	Skellefte älv	Gallejaur	55	700	20
9	Klarälven	Höljes	80	1600	>25
10	Ljusnan	Långströmmen	28	1670	50
11		Halvfari	43	650	100
12	Lule älv	Letsi	85	1500	25
13		Porsi	40	2700	15
14		Vittjärv	15	2200	50
15		Boden	21	2800	20

Research methods

Conventional approaches

- Experimental test
- Numerical study (CFD)
- Field observation



Research methods

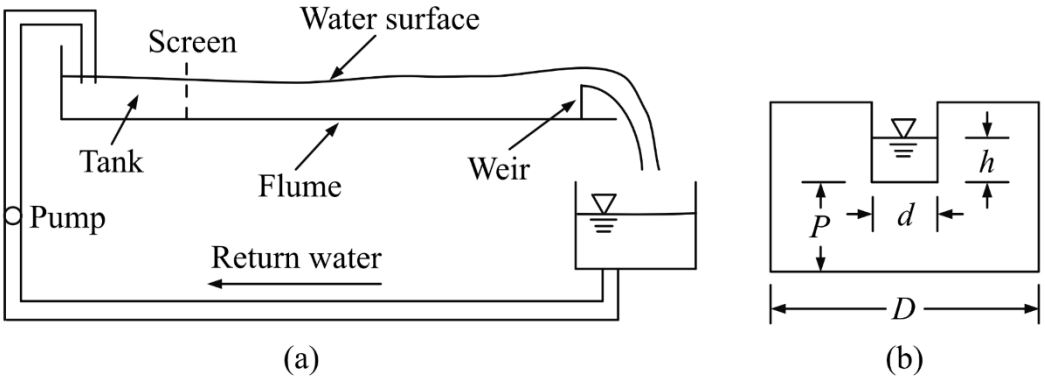
Alternative method

- Machine learning

Advantages

- Map feature data to labels
- Easily identifies trends and patterns
- Handling multi-dimensional and multi-variety data
- Continuous improvement
- Data acquisition

Example 1: Discharge determination*



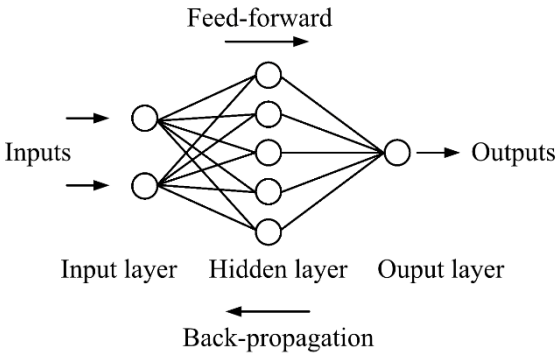
	D (cm)	P (cm)	d (cm)	h (cm)
Test 1 (Aydin et al. 2002)	30.0	4.0−16.0	0.5−7.5	2.1−27.9
Test 2 (Gharahjeh et al. 2015)	32.0	8.0	10.0−32.0	1.0−54.0

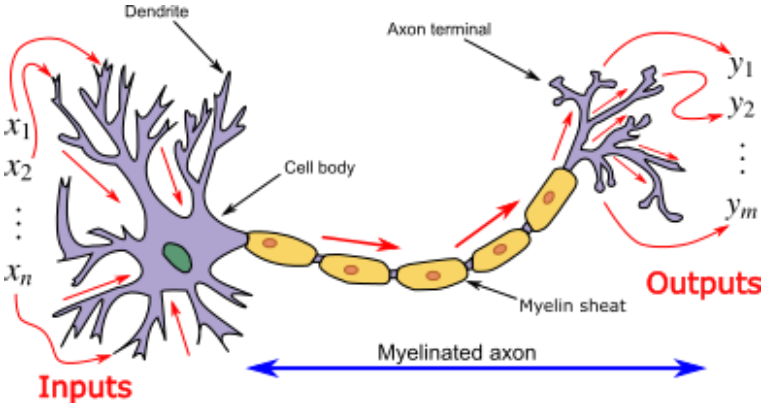
* Li, S., Yang, J., & Ansell, A. (2021). Discharge prediction for rectangular sharp-crested weirs by machine learning techniques. *Flow Measurement and Instrumentation*, 79, 101931.

Example 1: Discharge determination

Artificial neural network

$$o_j = \sum_{i=1}^{n} \beta_i f_a\left(w_i x_j + b_i\right) \quad j = 1,2\ldots,N$$

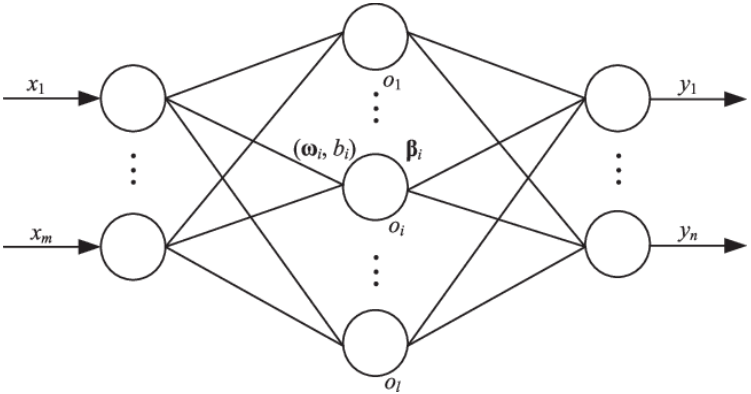




Example 1: Discharge determination

Extreme learning machine

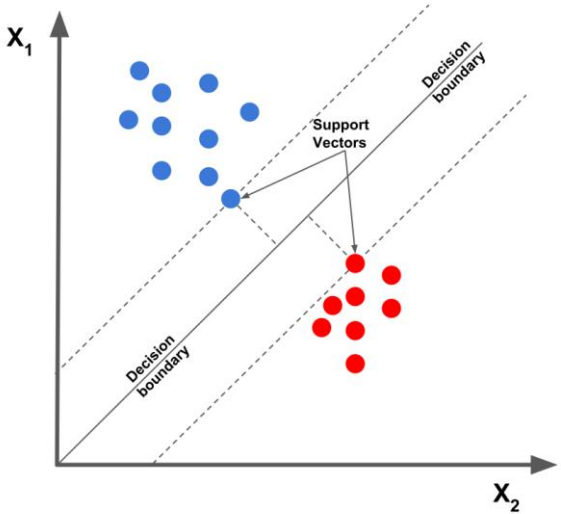
- One hidden layer
- Randomly assigned weights in the input layer and analytically solved in the output layer
- Reduced risks of overfitting



Example 1: Discharge determination

Support vector machine

Establish an optimal hyper-plane that separates two classes of samples with the largest margin



Example 1: Discharge determination

Performance criteria

Coefficient of determination (CD)

Correlation coefficient (CC)

Root mean square error (RMSE)

Mean absolute error (MAE)

$$CC = \frac{\sum_{i=1}^{N}\left(O_i - \bar{O}_i\right)\left(S_i - \bar{S}_i\right)}{\sqrt{\sum_{i=1}^{N}\left(O_i - \bar{O}_i\right)^2 \sum_{i=1}^{N}\left(S_i - \bar{S}_i\right)^2}}$$

$$CD = \left(\frac{\sum_{i=1}^{N}\left(O_i - \bar{O}_i\right)\left(S_i - \bar{S}_i\right)}{\sqrt{\sum_{i=1}^{N}\left(O_i - \bar{O}_i\right)^2}\sqrt{\sum_{i=1}^{N}\left(S_i - \bar{S}_i\right)^2}}\right)^2$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(O_i - S_i\right)^2}{N}}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|O_i - S_i\right|$$

Example 1: Discharge determination

Theoretical analysis

$$Q = F\left(h, d, D, P, \rho, g, \mu, \sigma\right)$$

$$\left.\begin{aligned}
\Pi_1 &= F_1\left(\Pi_2, \Pi_3, \Pi_4, \Pi_5, \Pi_6\right) \\
\Pi_1 &= \frac{Q}{\rho^{\alpha 1} g^{\beta 1} h^{\chi 1}} \\
\Pi_2 &= \frac{P}{\rho^{\alpha 2} g^{\beta 2} h^{\chi 2}} \\
\Pi_3 &= \frac{d}{\rho^{\alpha 3} g^{\beta 3} h^{\chi 3}} \\
\Pi_4 &= \frac{D}{\rho^{\alpha 4} g^{\beta 4} h^{\chi 4}} \\
\Pi_5 &= \frac{\mu}{\rho^{\alpha 5} g^{\beta 5} h^{\chi 5}} \\
\Pi_6 &= \frac{\sigma}{\rho^{\alpha 6} g^{\beta 6} h^{\chi 6}}
\end{aligned}\right\}$$

$\Pi_{1*} = \Pi_1/\Pi_3$, $\Pi_{2*} = 1/\Pi_2$, $\Pi_{3*} = \Pi_3/\Pi_2$,
$\Pi_{4*} = \Pi_3/\Pi_4$, $\Pi_{5*} = \Pi_3/\Pi_5$, $\Pi_{6*} = (\Pi_2)^3/\Pi_6$

$$C_d = \frac{Q}{d\sqrt{gh^3}} = F_2\left(\frac{h}{P}, \frac{d}{P}, \frac{d}{D}, \mathrm{R}, \mathrm{W}\right)$$

R = $\rho d(gh)^{0.5}/\mu$ = weir Reynolds number,
W = $\rho g d^2/\sigma$ = Weber number

Example 1: Discharge determination

Input vectors

Input No.	Dimensionless parameters	Input No.	Dimensionless parameters
M1	h/D	M6	h/P, d/P, R
M2	d/P	M7	d/P, d/D, W
M3	h/P, d/D	M8	h/P, d/P, d/D, R
M4	d/P, R	M9	h/P, d/D, R, W
M5	h/P, d/P, d/D	M10	h/P, d/P, d/D, R, W

Example 1: Discharge determination

Optimal input: h/P, d/P, d/D, R (M8)

	CD	CC	RMSE	MAE
ANN	0.9441	0.9716	0.0093	0.0066
SVM	0.9474	0.9734	0.0090	0.0063
ELM	0.9171	0.9577	0.0112	0.0085

Example 1: Discharge determination

Comparison with empirical models: statistics performance

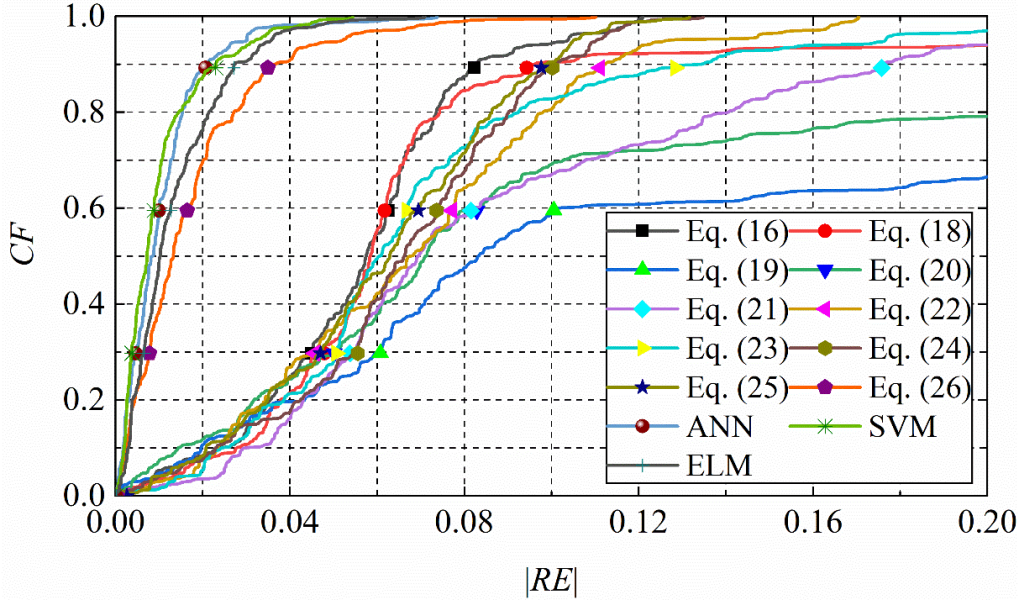
- Only a few empirical models give satisfactory predictions
- Machine learning models are the most accurate: mean error < 1.35%

Method	Max. |RE| (%)	Mean |RE| (%)	Method	Max. |RE| (%)	Mean |RE| (%)
Eq. 16	12.09	5.60	Eq. 24	13.48	6.53
Eq. 18	252.35	10.00	Eq. 25	13.16	6.13
Eq. 19	333.55	36.03	Eq. 26	11.02	1.74
Eq. 20	71.58	13.82	ANN	7.38	1.06
Eq. 21	24.14	8.85	SVM	5.44	0.99
Eq. 22	17.04	6.86	ELM	7.10	1.35
Eq. 23	22.12	7.04	-	-	-

Example 1: Discharge determination

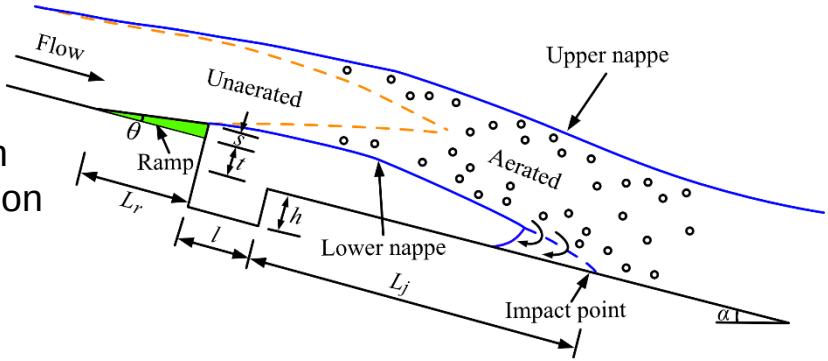
Comparison with empirical models: cumulative frequency

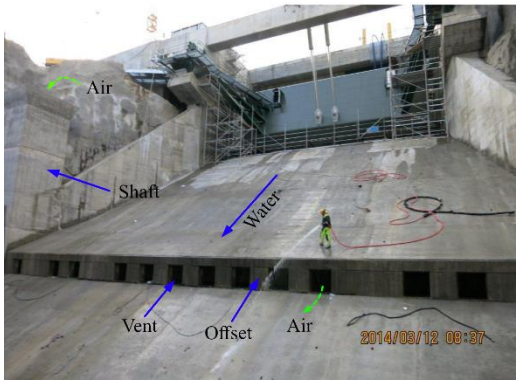
- AI models: 99% of predicted data show a |RE| < 5%
- AI models outperform all the empirical correlations



Example 2: Aeration estimation*

1–2% increase in air concentration leads to 5–7% reduction in cavitation



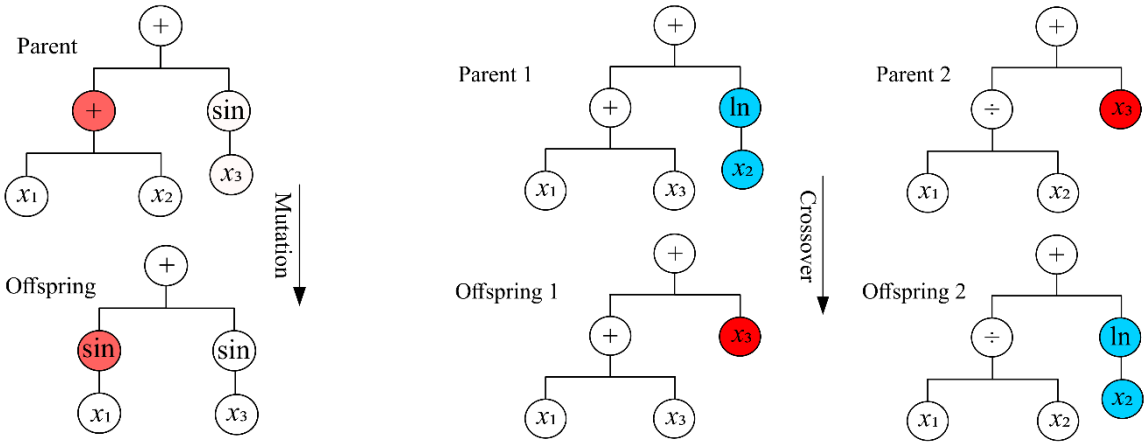


* Li, S., Yang, J., & Liu, W. (2021). Estimation of aerator air demand by an embedded multi-gene genetic programming. *Journal of Hydroinformatics*, 23 (5): 1000–1013 .

Example 2: Aeration estimation

Multi-gene genetic programming

- Evolutionary based method
- Linear combination of low-depth GP trees

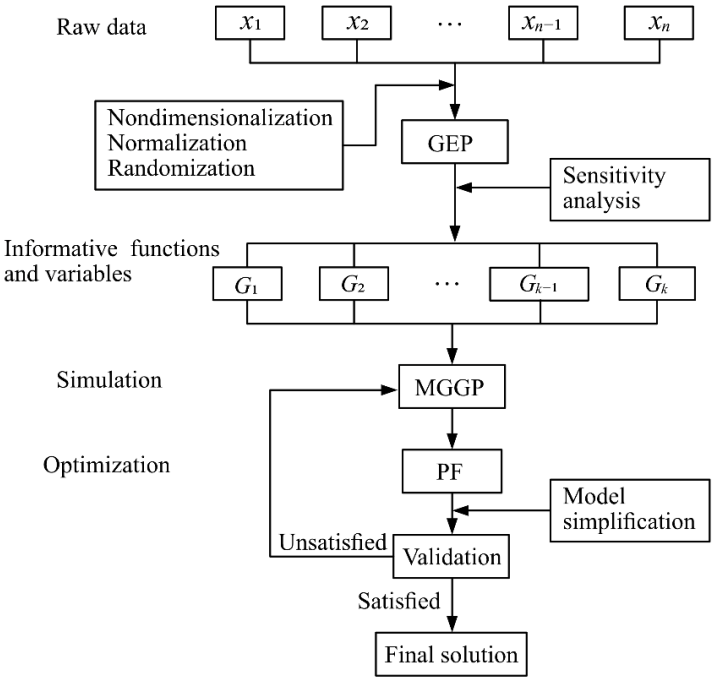


$$T = c_1 G_1 + c_2 G_2 + \cdots + c_n G_n + b$$

Example 2: Aeration estimation

Proposed embedded multi-gene genetic programming

- Feature extraction
- Solution optimization



Example 2: Aeration estimation

Theoretical analysis

$$Q_a = F_1\left(Q_w, W, k_s, s, h, l, \alpha, \theta, \rho_a, \rho_w, d, V, \Delta p, \mu, \sigma, g\right)$$

$$\left.\begin{aligned}
&\Pi_1 = \frac{Q_a}{g^{1/2}d^{5/2}}, \Pi_2 = \frac{W}{d}, \Pi_3 = \frac{k_s}{d} \\
&\Pi_4 = \frac{s}{d}, \Pi_5 = \frac{h}{d} \\
&\Pi_6 = \frac{l}{d}, \Pi_7 = \alpha, \Pi_8 = \theta \\
&\Pi_9 = \frac{\rho_a}{\mu g^{-1/2}d^{-3/2}}, \Pi_{10} = \frac{\rho_w}{\mu g^{-1/2}d^{-3/2}} \\
&\Pi_{11} = \frac{V}{g^{1/2}d^{1/2}}, \Pi_{12} = \frac{\Delta p}{\mu g^{1/2}d^{-1/2}} \\
&\Pi_{13} = \frac{Q_w}{g^{1/2}d^{5/2}}, \Pi_{14} = \frac{\sigma}{\mu g^{1/2}d^{1/2}}
\end{aligned}\right\}$$

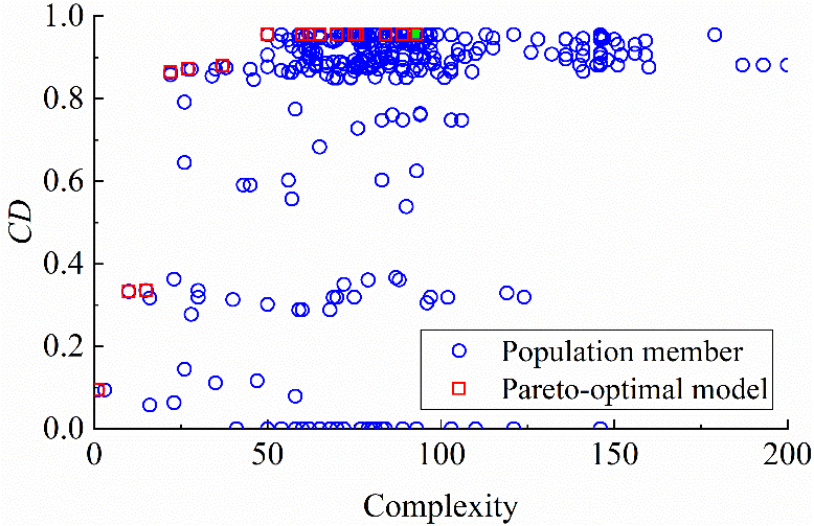
$$\left.\begin{aligned}
&\Pi_{1,13} = \frac{\Pi_1}{\Pi_{14}} = \frac{Q_a}{Q_w} = \beta \\
&\Pi_{2-8} = f(\Pi_2, \Pi_3, \cdots, \Pi_8) = \varphi \\
&\Pi_{11} = \frac{V}{g^{1/2}d^{1/2}} = \mathrm{F} \\
&\Pi_{10,11} = \Pi_{10}\,\Pi_{11} = \frac{\rho_w d V}{\mu} = \mathrm{R} \\
&\Pi_{10,11,14} = \frac{\Pi_{10}\,\Pi_{11}}{\Pi_{14}} = \frac{\rho_w^{1/2} d^{1/2} V}{\sigma^{1/2}} = \mathrm{W} \\
&\Pi_{9,11,12} = \frac{\Pi_9^{1/2}\,\Pi_{11}}{\Pi_{12}^{1/2}} = \frac{\rho_a^{1/2} V}{(\Delta p)^{1/2}} = \mathrm{E} \\
&\Pi_{12,10} = \frac{\Pi_{12}}{\Pi_{10}} = \frac{\Delta p}{\rho_w g d} = P
\end{aligned}\right\}$$

$$\beta = F_2\left(\mathrm{F}, P, \varphi\right)$$

Example 2: Aeration estimation

Model optimization

- Identification of Pareto-optimal models



$$\beta = 21.4P\left(\tan\left(P\right) - P^2\right) - 15.0P + 27.6\mathrm{F}^{1/16} + \frac{39.21P^{1/8}\tan\left(P\right)}{\mathrm{F} + 1.17\tan\left(P\right)} - 30.0$$

Example 2: Aeration estimation

Model optimization: selection of an accurate but simple solution

Model No.	Δ(CD) (%)	Δ(Complexity) (%)	Model No.	Δ(CD) (%)	Δ(Complexity) (%)
1 (benchmark)	-	-	9	0.08	35.5
2	0	4.3	10	0.08	46.2
3	0.03	7.9	11	7.98	60.2
4	0.05	18.3	12	8.81	71.0
5	0.05	19.4	13	9.57	76.3
6	0.06	24.7	14	64.91	83.9
7	0.07	30.1	15	65.14	89.2
8	0.08	33.3	16	90.16	98.9

$$\beta = 8.21P + \frac{22.3P}{\mathrm{F} - P} + 4.51\mathrm{F}^{1/8} + 7.07\mathrm{F}^{1/8}\tan\left(P + 5.47\right) - 14.2$$

Example 2: Aeration estimation

Comparison of different methods

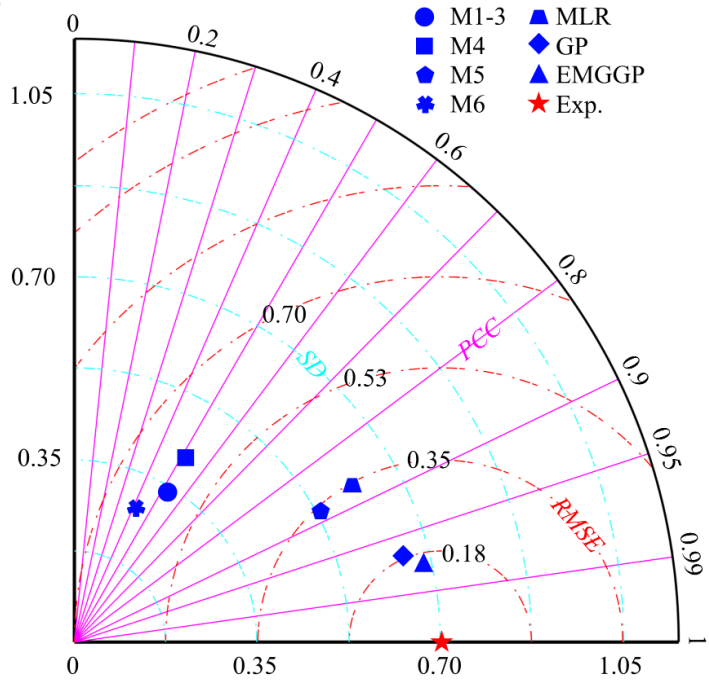
Method	Recalibrated			
	CD	NSC	RMSE (m^3/s)	MAE (m^3/s)
M1–3	0.281	0.279	0.596	0.447
M4	0.261	0.253	0.607	0.439
M5	0.776	0.776	0.288	0.235
M6	0.163	0.163	0.642	0.491
MLR	0.759	0.759	0.345	0.271

Method	Training				Testing			
	CD	NSC	RMSE (m^3/s)	MAE (m^3/s)	CD	NSC	RMSE (m^3/s)	MAE (m^3/s)
MGGP	0.955	0.953	0.155	0.119	0.947	0.946	0.152	0.118
EMGGP	0.955	0.950	0.161	0.121	0.945	0.936	0.166	0.124
GP	0.933	0.931	0.188	0.155	0.931	0.923	0.178	0.144

Example 2: Aeration estimation

Comparison of different methods using Taylor diagram

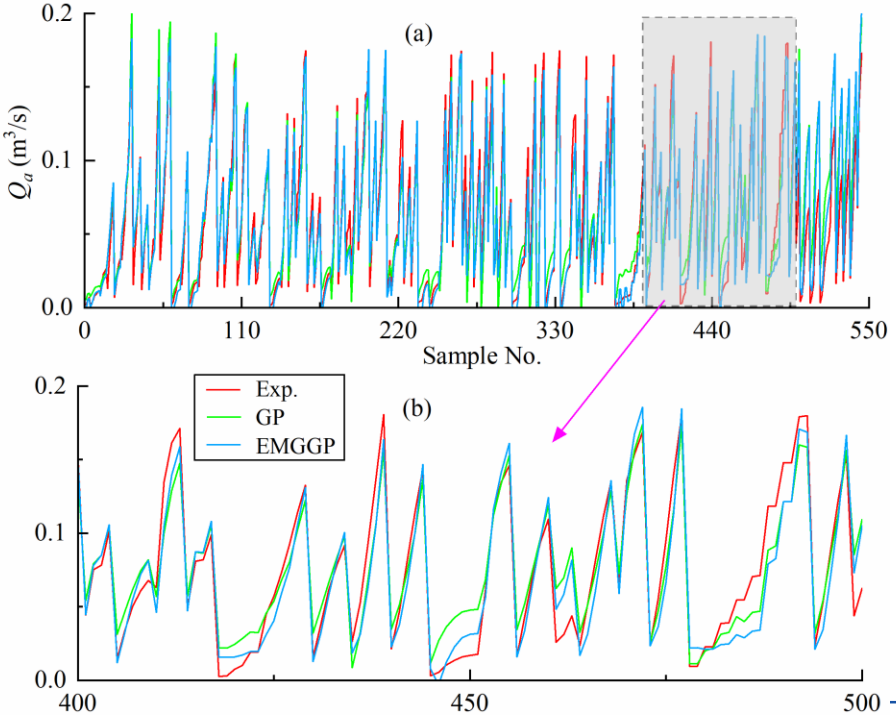
- Simultaneously capture multiple efficiency metrics
- The closer to observation, the better
- EMGGP is the most accurate one



Example 2: Aeration estimation

EMGGP results

- In good agreement with experiments



Summary and future work

Summary

- Multiple machine learning models are developed to determine flow discharge and aeration efficiency
- Machine learning methods show superior performance over experimental approaches

Future work

- Reservoir water level forecasts, time series modelling
- Deep learning, e.g. long short-term memory (LSTM) and nonlinear autoregressive with external input (NARX)

Li, S. & Yang, J. A hybrid gene expression programming model for discharge prediction. *Water Management,* 2021 (*in press*)

Li, S., Yang, J., & Liu, W. Estimation of aerator air demand by an embedded multi-gene genetic programming. *Journal of Hydroinformatics*, 2021, 23(5), 1000−1013.

Li, S., Yang, J., & Ansell, A. Discharge prediction for rectangular sharp-crested weirs by machine learning techniques. *Flow Measurement and Instrumentation*, 2021, 79, 101931.