



Estimation of air concentration in chute spillway using metamodel methods

Kiyoumars Roushangar^{1✉} | Reza Saadatjoo² | Hamidreza Abbaszadeh³ | Aydin Panahi⁴

1. Corresponding Author, Department of Civil Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz,

Iran. E-mail: kroshangar@yahoo.com

2. Department of Civil Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran. E-mail:

reza.saadatjoo@gmail.com

3. Department of Civil Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran. E-mail:

ha.abbaszadeh@tabrizu.ac.ir

4. Department of Civil Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran. E-mail:

a.panahi1401@ms.tabrizu.ac.ir

Article Info

Article type: Research Article

Article history:

Received: Jan. 31, 2023

Revised: Feb. 21, 2024

Accepted: Apr. 17, 2024

Published online: June. 2024

Keywords:

Aeration,
Chute Spillway,
Gaussian Process Regression,
Support Vector Machine.

ABSTRACT

One of the ways to prevent creating negative pressure and cavitation in spillways is to introduce air into the flow over the spillways. Understanding the distribution of air concentration variations along the spillway is of significant importance for estimating the aeration level. This study explores the application of GPR and SVM models in predicting air concentration. To achieve this, a dataset of 2268 laboratory experiments obtained from hydraulic models of chute spillways was utilized in the modeling process. Various input models were defined based on different combinations of measured parameters. The results demonstrate the high capability of both methods in estimating the required air concentration over the spillway. In predicting air concentration in the chute spillway under artificial aeration conditions, flow discharge (QW), longitudinal distance ratio from the end of the deflector to the channel width (L/W), and depth ratio (perpendicular to the spillway) to channel width (Y/W) significantly influenced the outcomes. Statistical indices, including R, DC, and RMSE for this case were 0.9214, 0.8451, and 1.008, respectively, in the GPR, and 0.9333, 0.8662, and 0.937 in the SVM. For scenarios without artificial aeration, the model with input parameters QW, L/W, Y/W, and ΔP (pressure difference between atmospheric pressure and the pressure under the jet) achieved the best performance in the GPR method with values of R=0.9222, DC=0.8644, and RMSE=0.914. In the SVM, the same model with values of 0.87, 0.7543, and 0.123 for R, DC, and RMSE, respectively, was selected as the superior model.

Cite this article: Roushangar, K., Saadatjoo, R., Abbaszadeh, H., & Panahi, A. (2024). Estimation of air concentration in chute spillway using metamodel methods, *Iranian Journal of Soil and Water Research*, 55 (4), 601-613.

<https://doi.org/10.22059/ijswr.2024.370643.669640>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijswr.2024.370643.669640>



تخمین غلظت هوا در سرریز شوت با استفاده از روش‌های فرامدل

کیومرث روشنگر^۱ | رضا سعادت‌جو^۲ | حمیدرضا عباس‌زاده^۳ | آیدین پناهی^۴۱. نویسنده مسئول، گروه مهندسی عمران، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران. رایانامه: kroshangar@yahoo.com۲. گروه مهندسی عمران، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران. رایانامه: reza.saadat.joo@gmail.com۳. گروه مهندسی عمران، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران. رایانامه: ha.abbaszadeh@tabrizu.ac.ir۴. گروه مهندسی عمران، دانشکده مهندسی عمران، دانشگاه تبریز، تبریز، ایران. رایانامه: a.panahi1401@ms.tabrizu.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	یکی از راه‌های جلوگیری از ایجاد فشار منفی و کاویتاسیون در سرریزها، هوادهی به جریان عبوری از سرریزها می‌باشد. شناخت نحوه توزیع تغییرات غلظت هوا در طول سرریز جهت تخمین میزان هوادهی از اهمیت زیادی برخوردار است. در پژوهش حاضر کاربرد روش‌های فرامدل رگرسیونی فرآیند گاوسی (GPR) و ماشین بردار پشتیبان (SVM) در پیش‌بینی غلظت هوا مورد بررسی قرار گرفت. بدین منظور مجموعه داده‌های آزمایشگاهی (۲۲۶۸) به‌دست آمده از مدل‌های هیدرولیکی سرریز شوت در فرآیند مدل‌سازی به کار گرفته شد. مدل‌های ورودی متنوعی بر اساس ترکیب مختلفی از پارامترهای اندازه‌گیری شده تعریف گردید. نتایج به‌دست آمده نشان دهنده توانایی بالای هر دو روش در برآورد غلظت هوای مورد نیاز بر روی سرریز است. در برآورد میزان غلظت هوا در سرریز شوت برای حالتی که هوادهی مصنوعی توسط هوادهی انجام می‌گیرد پارامترهای دبی جریان (QW)، نسبت فاصله طولی از انتهای دفلکتور به عرض کانال (L/W) و نسبت عمق (عمود بر سرریز) بر عرض کانال (Y/W) تأثیر زیادی داشتند. نتایج شاخص‌های آماری ضریب همبستگی (R)، ضریب تبیین (DC) و خطای جذر میانگین مربعات برای این حالت در روش GPR به ترتیب ۰/۹۲۱۴، ۰/۸۴۵۱ و ۰/۱۰۰۸ و مقادیر ۰/۹۳۳۳، ۰/۸۶۶۲ و ۰/۰۹۳۷ در روش SVM است. برای حالتی که هوادهی مصنوعی توسط هوادهی انجام نمی‌گیرد، مدل با پارامترهای ورودی Qw، L/W، Y/W و ΔP (اختلاف فشار ما بین فشار اتمسفر و فشار زیر جت) با دارا بودن مقادیر R=۰/۹۲۲۲، DC=۰/۸۶۴۴ و RMSE=۰/۰۹۱۴ در روش GPR و به ترتیب با مقادیر ۰/۸۷، ۰/۷۵۴۳ و ۰/۱۲۳ به‌عنوان بهترین مدل انتخاب گردیدند.
تاریخ دریافت: ۱۴۰۲/۱۰/۱۶	
تاریخ بازنگری: ۱۴۰۲/۱۲/۲	
تاریخ پذیرش: ۱۴۰۳/۱/۲۹	
تاریخ انتشار: تیر ۱۴۰۳	
واژه‌های کلیدی:	
رگرسیون فرآیند گاوسی، سرریز شوت، ماشین بردار پشتیبان، هوادهی.	

استناد: روشنگر، کیومرث؛ سعادت‌جو، رضا؛ عباس‌زاده، حمیدرضا و پناهی، آیدین (۱۴۰۲). تخمین غلظت هوا در سرریز شوت با استفاده از روش‌های فرامدل. مجله تحقیقات

<https://doi.org/10.22059/ijswr.2024.370643.669640> ۶۰۱-۶۱۳، (۴) ۵۵، آب و خاک ایران، ۶۰۱-۶۱۳

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2024.370643.669640>

مقدمه

تأمین آب مورد نیاز بخش‌های مختلف کشاورزی، صنعت و شرب بالاخص در مناطق خشک و نیمه خشک از اهمیت خاصی برخوردار است (Süme et al., 2024). در دهه‌های اخیر و با پیشرفت تکنولوژی، احداث سدهای مخزنی و سازه‌های هیدرولیکی جانبی گسترش قابل توجهی نموده و با روند افزایشی احداث سدهای بلند و نیاز به بالابردن ایمنی سدها، توجه هرچه بیشتر مهندسی هیدرولیک به مقوله طراحی اقتصادی و مطمئن این سازه‌ها را می‌طلبد. یکی از سازه‌های هیدرولیکی که نقش مهمی را در فراهم نمودن ایمنی کافی سد به هنگام ورود سیلاب به داخل مخزن و تخلیه آن به پائین‌دست سد ایفا می‌کند، سرریزها می‌باشند. با توجه به سابقه طولانی استفاده از مدل‌های هیدرولیکی در طراحی سرریزها و دقت قابل توجه نتایج حاصل از آن‌ها، مهندسی را بر آن داشته تا با ساخت این مدل‌ها و انجام آزمایش نسبت به طراحی اقتصادی و ایمنی‌تر سرریزها اقدام نمایند. به دلیل بالا بودن سرعت جریان و افت فشار در سرریزهای بلند، پتانسیل وقوع پدیده کاویتاسیون افزایش یافته و منجر به خسارت شدید به سرریز گردد. بنابراین سازه ویژه‌ای که هواده تنداب نامیده می‌شود باید در محل‌هایی که در آن هواده طبیعی از سطح آزاد جریان برای محافظت کافی نیست تعبیه شود. بنابراین شناخت نحوه توزیع تغییرات غلظت هوا در طول سرریز جهت تخمین میزان هواده و ارائه شیوه نوین و خلاقانه که بتواند در این راستا موثر واقع شود از اهمیت بسیار زیادی برخوردار است.

یکی از تأثیرات جانبی کاویتاسیون در بسیاری از سازه‌ها، سر و صدا و ارتعاش می‌باشد که از ماشین‌های محرکه گرفته تا شیرهای بزرگ در صنعت و سازه‌های هیدرولیکی مانند سرریزها را شامل می‌شود. راحمیر سروصدای ناشی از کاویتاسیون و آزار دهنده بودن آن را مورد بررسی قرار داد (Rahmeyer, 1981). ایشان در نظر گرفتن صدا به‌عنوان فاکتور مهم در طراحی و عملکرد یک سازه هیدرولیکی را باعث ایجاد محدودیت احتمالی دانست. نتایج تحقیق فرنک و میشل نشان داد که دوام داشتن کاویتاسیون باعث کاهش عملکرد سازه هیدرولیکی می‌شود (Franc and Michel, 2004). زبری موضعی دیواره‌ها (مانند دیواره‌های بتنی سرریز سدها) می‌تواند باعث جداشدگی جریان از سطح شده و کاهش فشار را ایجاد کند. نتایج تحقیق فیستر و هاگر بیانگر بیشترین مقدار غلظت هوا در کف سرریز و تا نقطه برخورد جت است (Pfister and Hager, 2010). همچنین نتایج آن‌ها نشان داد که از نقطه برخورد جت، غلظت کاهش یافته و یک گردپایان هواده بزرگ نیز در محل نقطه برخورد جت به کف سرریز حاصل می‌شود. فیستر در سرریز دارای هواده به تعیین ویژگی‌های هیدرولیکی جریان دو فازي پرداخت. تغییرات ویژگی‌های هیدرولیکی جریان دو فازي همچون ضریب هواده، طول جت، گسترش غلظت هوا، با تغییر در پارامترهای زاویه دفلیکتور و زیر فشار داخل کاویتا به‌دست آمد (Pfister, 2011). نتایج تحقیق نشان داد که دفلیکتور با شیب بالاتر دارای عملکرد بهتری در وارد کردن هوا به جریان است. همچنین با افزایش زیر فشار داخل کاویتا، طول جت و ضریب هواده کاهش می‌یابد. همچنین با افزایش زیر فشار، غلظت هوای کف سرریز نیز کاهش می‌یابد. وو و همکاران به بررسی تأثیر پارامترهای هندسی شوت هواده، بروی مقادیر آب برگشتی به داخل کاویتا تشکیل شده در پائین‌دست رمپ هواده پرداختند (Wu et al., 2011). در اعداد فرود پائین و شیب طولی کم شوت، آب برگشتی ناشی از برخورد جت عبوری از روی هواده به کف شوت می‌تواند کاویتا پائین‌دست هواده را پر نماید و خود موجب بروز کاویتاسیون گردد. نتیجه نشان می‌دهد مقدار آب برگشتی به داخل کاویتا، با طول جت پرتابی از روی هواده رابطه دارد. برای هواده‌هایی با ارتفاع بزرگ‌تر یا زاویه رمپ بیشتر، مقدار طول جت پرتابی بزرگ‌تر می‌باشد. چکیب نقطه شروع هواده طبیعی از سطح جریان در سرریز پلکانی را بررسی نمود (Chakib, 2013). نتایج نشان داد که ورود هوا به جریان به دلیل کاستن از اصطکاک دیواره‌ها باعث افزایش سرعت جریان شده و با افزایش آگذری جریان، نقطه شروع هواده از سطح به پائین‌دست حرکت می‌کند. کرمانی و همکاران نقاطی که احتمال آسیب ناشی از کاویتاسیون در سرریز سد وجود داشت را بررسی کردند (Kermani et al., 2013). نتایج بیانگر این است که در مقاطع نزدیک به انتهای سرریز احتمال خسارت به دلیل افزایش سرعت جریان در طول سرریز بیشتر می‌شود. سلماسی و همکاران به بررسی مدل‌های فیزیکی مختلف سرریز پلکانی و تأثیر اختلاط شدید آب بر افزایش غلظت اکسیژن محلول در آب پرداختند (Salmasi et al., 2021). نتایج نشان داد که با افزایش دبی و غوطه‌ور شدن تمام پله‌ها، اثر پله‌های سرریز بر تلاطم و هواده جریان کاهش می‌یابد. در دبی‌های بیش از ۰/۱۵ متر مربع بر ثانیه، با افزایش شیب سرریز، راندمان هواده افزایش می‌یابد. راضا و همکاران به بررسی عددی تأثیر شیب سرریز پلکانی بر حباب هوا و مکان شروع حباب پرداختند (Raza et al., 2021). طول منطقه جریان بدون هواده زمانی افزایش می‌یابد که شیب کانال از شیب تند به شیب ملایم کاهش یابد.

یکی از مهم‌ترین مسائل در بررسی پدیده کاویتاسیون تعیین نحوه توزیع غلظت هوا است. روابط تجربی موجود جهت تخمین غلظت هوا به دلیل اینکه مربوط به مدل‌هایی با هندسه خاص می‌باشند، کاربرد محدودی دارند. ساخت مدل هیدرولیکی سازه‌های آبی هزینه‌بر

است. همچنین در تبدیل داده‌ها از مقیاس آزمایشگاهی به مقیاس اصلی مشکلاتی وجود دارد. لذا استفاده از روش‌هایی که بتوان تخمین خوبی از نحوه توزیع غلظت هوا داشته باشند؛ بسیار موثر است. بدین منظور استفاده از روش‌های بهینه‌سازی چند هدفه همواره از اهمیت ویژه‌ای برخوردار بوده است. در پژوهش حاضر روش‌های ماشین بردار پشتیبان و رگرسیون فرآیند گوسی به عنوان روش‌های فرامدل جهت تخمین پارامترها استفاده گردید و نتایج آن مورد ارزیابی قرار گرفته است. مدل‌های مختلف با پارامترهای ورودی مختلف تعریف شده و پارامتر غلظت هوا تخمین زده شد و پارامترهای تأثیرگذار هر مورد شناسایی شد. این تحقیق تلاشی است در راستای مطالعات قبلی صورت گرفته که با نمایش کاربرد روش ماشین بردار پشتیبان و رگرسیون فرآیند گوسی در تخمین توزیع غلظت هوا و ارزیابی کاربردی بودن این روش‌ها نسبت به داده‌های آزمایشگاهی موجود در این زمینه صورت گرفته است.

مواد و روش‌ها

ماشین بردار پشتیبان (SVM)

ماشین بردار پشتیبان به عنوان یکی از روش‌های یادگیری با نظارت، اولین بار توسط Vapnik (1995) برای طبقه‌بندی و پیش‌بینی مورد استفاده قرار گرفت. ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده کرده و منجر به یک جواب بهینه کلی می‌گردد. در مدل رگرسیون SVM تابعی مرتبط با متغیر وابسته Y که خود تابعی از چند متغیر مستقل X است، برآورد می‌شود. مشابه سایر مسائل رگرسیونی فرض می‌شود رابطه میان متغیرهای مستقل و وابسته با تابع جبری مانند $f(x)$ به علاوه مقداری اغتشاش (خطای مجاز ϵ) مشخص شود (Norouzi et al., 2021; Hassanzadeh and Abbaszadeh, 2023).

$$f(x) = W^T \phi(X) + b \quad \text{رابطه ۱}$$

$$Y = f(x) + \text{noise} \quad \text{رابطه ۲}$$

چنانچه W بردار ضرایب و b ثابت مشخصه‌های تابع رگرسیونی و ϕ نیز تابع کرنل باشد، آنگاه هدف پیدا کردن فرم تابعی برای $f(x)$ است. این مهم با آموزش مدل SVM توسط مجموعه‌ای از نمونه‌ها (مجموعه آموزش) محقق می‌شود. تابع SVM رگرسیونی را می‌توان به فرم زیر بازنویسی کرد.

$$f(x) = \sum_{i=1}^N \bar{a}_i \phi(X_i)^T \phi(X) + b \quad \text{رابطه ۳}$$

پارامتر a_i بیانگر میانگین ضرایب لاگرانژ است. محاسبه $\phi(X)$ در فضای مشخصه آن ممکن است بسیار پیچیده باشد. برای حل این مشکل روند معمول در مدل SVM رگرسیون انتخاب یک تابع کرنل است. انتخاب کرنل برای SVM به حجم داده‌های آموزشی و ابعاد بردار ویژگی بستگی دارد. به عبارت دیگر، باید با توجه به این پارامترها تابع کرنلی را انتخاب نمود که توانایی آموزش برای ورودی‌های مسئله را داشته باشد. در عمل چهار نوع کرنل خطی (Linear)، کرنل چندجمله‌ای (Polynomial)، کرنل سیگموئید (Sigmoid) و کرنل پایه شعاعی (RBF) به کار گرفته می‌شوند (Abbaszadeh et al., 2024).

$$K(X_i, X_j) = (X_i, X_j) \quad \text{رابطه ۴}$$

$$K(X_i, X_j) = (1 + (X_i, X_j))^d \quad \text{رابطه ۵}$$

$$K(X_i, X_j) = \tanh(-a(X_i, X_j) + C) \quad \text{رابطه ۶}$$

$$K(X_i, X_j) = \exp(-\|X - X_i\|^2 / \sigma^2) \quad \text{رابطه ۷}$$

در روابط بالا $K(X_i, X_j)$ تابع کواریانس یا کرنل بوده که در نقاط X_i و X_j محاسبه می‌گردد. a ، C ، d و σ بیانگر توابع کرنل هستند (Roushangar and Homayounfar, 2019). ترم d درجه چندجمله‌ای و C عددی صحیح و مثبت است که عامل تعیین جریمه در هنگام رخ دادن خطای آموزش مدل است.

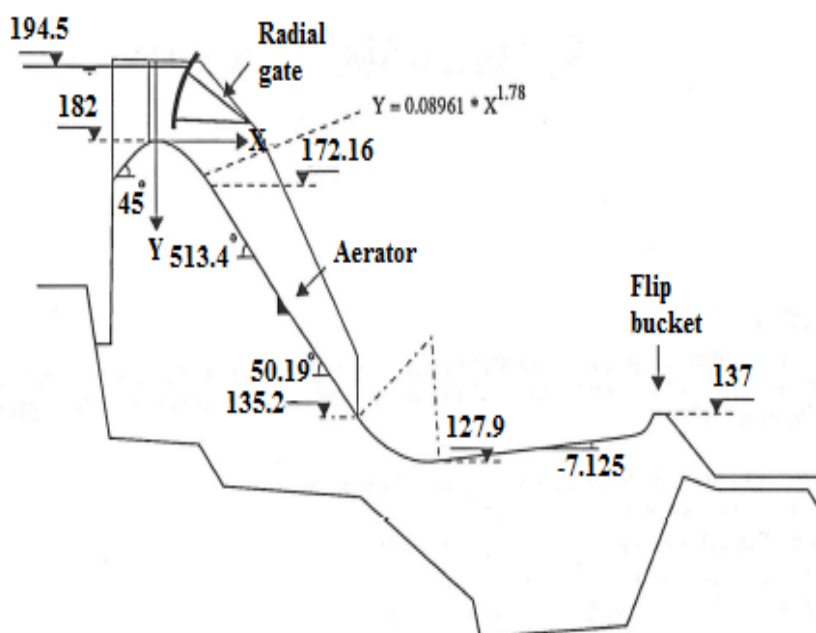
رگرسیون فرآیند گوسی (GPR)

فرآیند گاوسی (Gaussian Process Regression) در تئوری احتمالات، یک فرآیند تصادفی است که به صورت دنباله‌ای از متغیرهای تصادفی شناخته می‌شود. در این فرآیند، متغیرهای تصادفی معمولاً برحسب مثلاً زمان مرتب شده‌اند و هر زیر مجموعه متناهی از این

متغیرها دارای توزیع گاوسی (نرمال) چند متغیره هستند (Rasmussen and Williams., 2006). بنابراین اگر $\{X_i; t \in T\}$ یک دنباله تصادفی باشد آنگاه $X_{t1}, \dots, X_{tk} = (X_{t1}, \dots, X_{tk})$ دارای توزیع نرمال (گاوسی) چند متغیره هستند. به بیان دیگر در یک فرآیند گاوسی، هر ترکیبی خطی از مجموعه متناهی از فرآیند دارای توزیع نرمال یک متغیره است. T ، مجموعه‌ای نامتناهی از اندیس‌های زمان یا مکان است. در این تحقیق به منظور مدل‌سازی داده‌ها از یک برنامه نوشته شده در نرم‌افزار متلب، با فرمت M-File استفاده شده است. در این برنامه داده‌های نرمال شده به صورت فایل xlsX از نرم افزار اکسل فراخوان شده و پس از اجرای برنامه، نتایج به صورت نمودارهای پراکنش قابل ارائه می‌باشد.

سرریز شوت-مدل‌های توزیع غلظت هوا

جهت بررسی عملکرد روش ماشین بردار پشتیبان و رگرسیون فرآیند گوسی در پیش‌بینی توزیع غلظت هوا در سرریزها، از داده‌های آزمایشگاهی (Chanson (1988) که شامل ۲۲۶۸ داده مربوط به مدل سرریز سد کلاید است، استفاده شده است. سد کلاید یک سد بتنی وزنی در نیوزیلند است. ارتفاع این سد ۱۰۰ متر، طول تاج ۴۹۰ متر، عرض تاج ۱۰ متر و عرض در پی ۷۰ متر است. سرریز این سد از نوع سرریز شوت است. در شکل (۱) شماتیک مدل سرریز سد کلاید و در جدول (۱) محدوده مشخصات هیدرولیکی و هندسی مدل ارائه شده است.



شکل ۱. شماتیک مدل هیدرولیکی (Chanson, 1988)

جدول ۱. مشخصات هیدرولیکی و هندسی مدل (Chanson, 1988)

Y_{90} (m)	d (m)	L (m)	ΔP (m)	d_0 (m)	H_{gate} (m)	Q_w (m ³ /s)	آزمایش
۰/۰۵۲۹ - ۰/۰۳۴۲	۰/۰۱۹۸ - ۰/۰۲۱۰	۳ - ۰/۵۹۲	۱۶۶/۳۹	۰/۰۲۲۹	۰/۰۳	۰/۰۵۲۹	۱
۰/۰۴۷۱ - ۰/۰۶۷۲	۰/۰۳۱۱ - ۰/۰۳۲۴	۳ - ۰/۵۹۲	۳۹۶/۴	۰/۰۳۵	۰/۰۵	۰/۰۸۶۳	۲
۰/۰۴۰۴ - ۰/۰۶۶۱	۰/۰۲۸۱ - ۰/۰۳۱۸	۳ - ۰/۵۹۲	۳۱۰/۲۷	۰/۰۳۴۸	۰/۰۵	۰/۰۷۵۹	۳
۰/۰۵۴۸ - ۰/۰۶۸۵	۰/۰۳۳۰ - ۰/۰۳۴۳	۳ - ۰/۵۹۲	۵۶۱/۸۱	۰/۰۳۵۱	۰/۰۵	۰/۰۹۹	۴
۰/۰۳۷۵ - ۰/۰۶۵۰	۰/۰۲۶۵ - ۰/۰۳۱۳	۳ - ۰/۵۹۲	۲۳۳/۴۴	۰/۰۳۴۶	۰/۰۵	۰/۰۶۸۳	۵
۰/۰۳۲۵ - ۰/۰۶۲۶	۰/۰۲۳۷ - ۰/۰۳۰۰	۳ - ۰/۵۹۲	۱۴۱/۴۳	۰/۰۳۴۱	۰/۰۵	۰/۰۵۲۴	۶

به منظور تعیین مدل‌های ورودی ترکیب مختلفی از پارامترهای اندازه‌گیری شده شامل Q_w دبی جریان، L/W نسبت فاصله طولی از انتهای دفلکتور بر عرض کانال، Y/W نسبت عمق (عمود بر سرریز) بر عرض کانال، ΔP اختلاف فشار ما بین فشار اتمسفر و فشار زیر جت برای برآورد توزیع غلظت هوا در طول جت مطابق با جدول (۲) توسعه داده شد.

جدول ۲. مدل‌های تعریف شده برای مدل

سناریو ۲ (هوادهی مصنوعی)		سناریو ۱ (هوادهی طبیعی)	
پارامترهای ورودی	مدل	پارامترهای ورودی	مدل
Q_w	CA(I)	Q_w	C(I)
$Q_w, L/W$	CA(II)	$Q_w, L/W$	C(II)
$Q_w, Y/W$	CA(III)	$Q_w, Y/W$	C(III)
$Q_w, \Delta P$	CA(IV)	$Q_w, \Delta P$	C(IV)
$Q_w, L/W, Y/W$	CA(V)	$Q_w, L/W, Y/W$	C(V)
$Q_w, L/W, \Delta P$	CA(VI)	$Q_w, L/W, \Delta P$	C(VI)
$Q_w, L/W, Y/W, \Delta P$	CA(VII)	$Q_w, L/W, Y/W, \Delta P$	C(VII)
$Q_w, Y/W, \Delta P$	CA(VIII)		
$L/W, Y/W, \Delta P$	CA(IX)		

نرمال‌سازی داده‌ها

یکی از مراحل پیش پردازش داده‌ها، نرمال‌سازی آن‌ها می‌باشد. آموزش نرم‌افزار در صورت اعمال یکسری پیش‌پردازش‌ها بر روی ورودی‌ها و هدف می‌تواند کارایی بالاتری داشته باشد. اصولاً وارد کردن داده‌ها به صورت خام باعث کاهش سرعت و دقت شبکه می‌شود. بنابراین نرمالیزه کردن داده‌ها خصوصاً زمانی که دامنه تغییرات ورودی‌ها زیاد باشد، کمک شایانی به آموزش بهتر و سریع‌تر مدل می‌کند. روش استفاده شده برای پیش‌پردازش داده‌ها مطابق معادله (۸) است (Roushangar and Shahnazi, 2020):

$$X_n = 0.1 + 0.9 \times \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad \text{رابطه ۸}$$

در این رابطه، X_{\min} و X_{\max} به ترتیب حداقل و حداکثر مقدار داده‌های مشاهده شده و X_n داده نرمال شده می‌باشد. برای رسیدن به نتایج بهتر و دقیق‌تر، روند آموزش چند بار تکرار و الگوی ۷۵٪ داده‌ها جهت آموزش و ۲۵٪ داده‌ها جهت آزمون انتخاب گردید.

شاخص‌های ارزیابی

به منظور ارزیابی کارایی روش‌های استفاده شده در تخمین توزیع غلظت هوا، از پارامترهای آماری R ضریب همبستگی، DC ضریب تبیین، RMSE خطای جذر میانگین مربعات و KGE شاخص کلینگ گوپتا استفاده شد (Abbaszadeh et al., 2023a,b; Daneshfaraz et al., 2023a). به طور کلی از بین مدل‌ها، مدلی به عنوان مدل برتر انتخاب شد که شاخص‌های خطای جذر میانگین مربعات نزدیک به صفر، ضریب تبیین، ضریب همبستگی و شاخص کلینگ گوپتا آن نزدیک‌ترین مقدار به یک باشد (Daneshfaraz et al., 2022a, 2023b).

$$R = \frac{\sum_{i=1}^N (C_m - \bar{C}_m) \times (C_p - \bar{C}_p)}{\sqrt{\sum_{i=1}^N (C_m - \bar{C}_m)^2 \times (C_p - \bar{C}_p)^2}} \quad \text{رابطه ۹}$$

$$DC = 1 - \frac{\sum_{i=1}^N (C_m - C_p)^2}{\sum_{i=1}^N (C_m - \bar{C}_p)^2} \quad \text{رابطه ۱۰}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (C_m - C_p)^2}{N}} \quad \text{رابطه ۱۱}$$

$$KGE = 1 - \sqrt{(R - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$$

$$\beta = \frac{C_{d\,cal}}{C_{d\,obs}}, \gamma = \frac{CV_{cal}}{CV_{obs}} = \frac{\sigma_{cal}/\bar{C}_{d\,cal}}{\sigma_{obs}/\bar{C}_{d\,obs}} \quad \text{رابطه ۱۲}$$

$0.7 < KGE < 1$ Very good

- $0.6 < KGE < 0.7$ Good
- $0.5 < KGE \leq 0.6$ Satisfactory
- $0.4 < KGE \leq 0.5$ Acceptable
- $KGE \leq 0.4$ Unsatisfactory

در روابط بالا، C_m توزیع غلظت هوای اندازه‌گیری شده، \bar{C}_m متوسط توزیع غلظت هوای اندازه‌گیری شده، C_p توزیع غلظت هوای پیش‌بینی شده، \bar{C}_p متوسط توزیع غلظت هوای پیش‌بینی شده و N بیانگر تعداد داده‌ها است. در رابطه (۱۲)، β نسبت میانگین داده‌های محاسباتی به میانگین داده‌های مشاهداتی و γ بیانگر نسبت انحراف استاندارد مقادیر محاسباتی به انحراف استاندارد مقادیر مشاهداتی می‌باشند (Daneshfaraz et al. 2022b).

نتایج و بحث

در استفاده از روش ماشین بردار پشتیبان جهت انتخاب تابع کرنل مناسب، مدل C(VII) با کرنل‌های مختلف مورد ارزیابی قرار گرفت. مطابق جدول (۳) نتایج به‌دست آمده نشان می‌دهد که تابع کرنل (RBF) در پیش‌بینی توزیع غلظت هوا نتایج مطلوبی در مقایسه با سایر کرنل‌ها دارد.

جدول ۳. پارامترهای آماری روش SVM با توابع کرنل متفاوت-مدل C(VII)

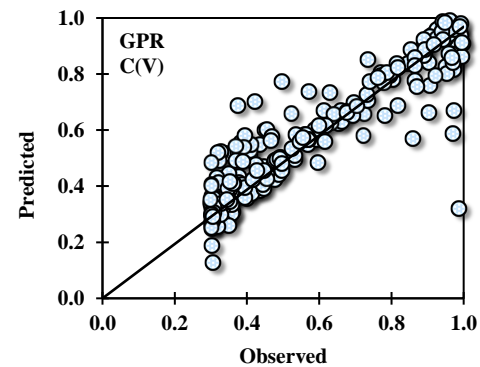
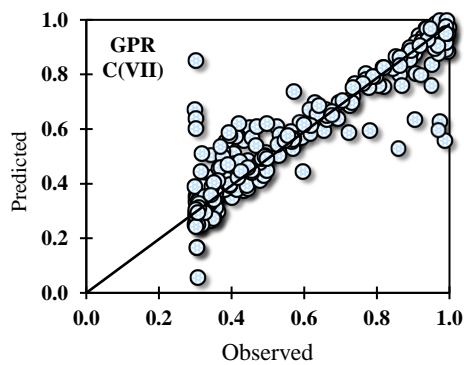
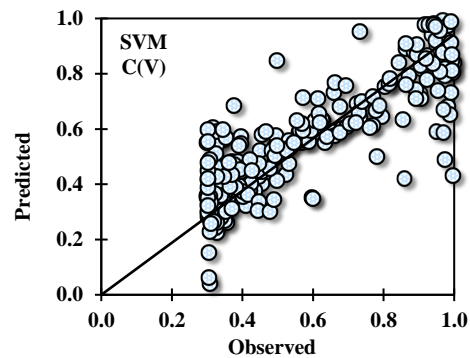
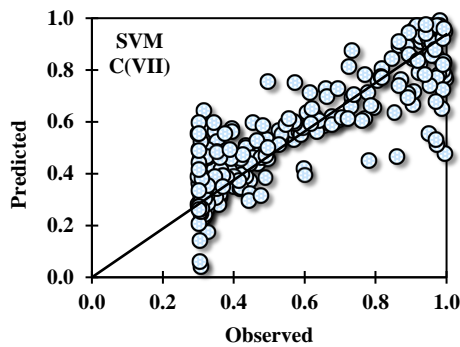
صحت‌سنجی				آزمون				تابع کرنل
KGE	RMSE	DC	R	KGE	RMSE	DC	R	
۰/۴۸۰	۰/۲۲۷۲	۰/۴۱۲۶	۰/۴۸۹۴	۰/۵۴۲	۰/۲۲۶۹	۰/۵۴۵۴	۰/۵۳۵۵	Linear
۰/۴۵۴	۰/۲۲۳۳	۰/۴۵۰۸	۰/۴۷۱۸	۰/۴۸۵	۰/۲۳۰۹	۰/۵۱۹۲	۰/۴۷۲۴	Polynomial
۰/۸۷۸	۰/۱۲۳۰	۰/۷۵۴۳	۰/۸۷۰۰	۰/۹۷۲	۰/۰۵۸۹	۰/۹۴۹۱	۰/۹۷۴۳	RBF
۰/۴۹۸	۰/۲۴۲۴	۰/۵۵۰۸	۰/۵۰۴۵	۰/۵۷۲	۰/۲۴۵۰	۰/۵۸۶۹	۰/۵۸۶۹	Sigmoid

نتایج مدل‌های تعریف شده تحت سناریو ۱

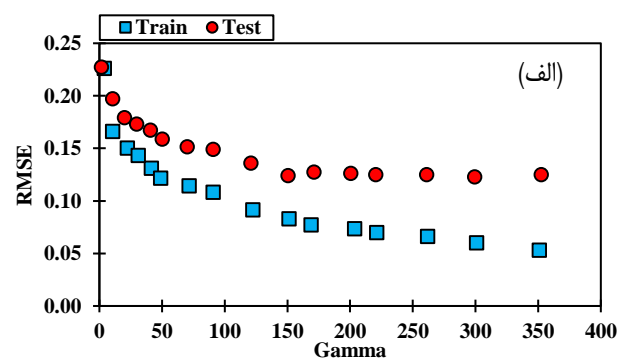
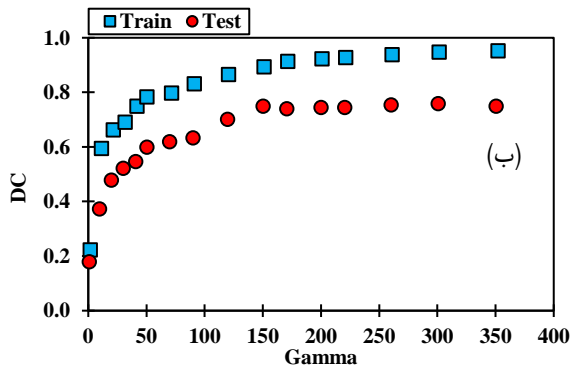
نتایج حاصل از دو روش ماشین بردار پشتیبان و روش رگرسیون فرآیند گوسی برای حالتی که هوادهی مصنوعی توسط هواده انجام نمی‌گیرد (۱۱۱۳ داده)، جهت پیش‌بینی توزیع غلظت هوا در طول سرریز در جدول (۴) ارائه شده است. با توجه به نتایج نشان داده شده در جدول (۴) مشاهده گردید مدل C(VII) با چهار پارامتر ورودی، مقادیر R و DC بیشتر و RMSE کم‌تری را نسبت به سایر مدل‌ها ارائه داده و به‌عنوان مدل برتر در فرآیند مدل‌سازی انتخاب گردید. در این مدل مقایسه مقادیر معیارهای ارزیابی نشان می‌دهد که نتایج حاصل از روش GPR برتری محسوسی نسبت به SVM دارد. بررسی نتایج نشان می‌دهد در این سناریو مدل‌هایی با یک و دو پارامتر ورودی به هیچ عنوان از دقت کافی به‌منظور برآورد توزیع غلظت هوا برخوردار نمی‌باشند. همچنین مقایسه دو مدل C(V) و C(VI) با سه پارامتر ورودی نشان می‌دهد جایگزینی نسبت عمق بر عرض کانال (Y/W) با پارامتر ΔP موجب بهبود چشمگیر دقت مدل‌سازی می‌گردد که همین عامل نشان دهنده تأثیر بالای پارامتر Y/W در پیش‌بینی توزیع غلظت هوا در طول سرریز است. نمودار پراکندگی داده‌ها برای دو مدل برتر در شکل (۲) نشان داده شده است. همچنین شکل (۳) نشان دهنده تغییرات DC و RMSE در برابر مقادیر مختلف پارامتر γ برای مدل برتر می‌باشد. مطابق شکل (۳)، مقدار بهینه پارامتر γ در ماشین بردار پشتیبان برابر با ۳۰۰ است. افزایش مقدار این پارامتر موجب کاهش تدریجی دقت ماشین بردار پشتیبان می‌گردد. برای بهینه‌سازی مشخصه‌های روش SVM رگرسیونی یعنی مقادیر (ε) و (C) در تابع کرنل مورد استفاده در این تحقیق که تابع (RBF) می‌باشد باید مشخصه (γ) بهینه شود. بنابراین در حالت کلی برای پیش‌بینی توزیع غلظت هوا در سرریز توسط روش SVM رگرسیونی، لازم است که مقادیر بهینه سه مشخصه مذکور به‌دست آیند. در پژوهش حاضر برای رسیدن به این منظور دو مشخصه (ε) و (C) توسط الگوریتم بهینه‌سازی جست‌وجوی شبکه و مشخصه (γ) نیز به‌صورت سعی و خطا بهینه شده است. برای این منظور محدوده مشخصه‌های (ε) و (C) به‌ازای مقدار ثابت (γ) تعیین شده است. سپس توسط الگوریتم جست‌وجوی شبکه مقدار دقیق دو مشخصه (ε) و (C) مشخص شدند، روند مذکور برای دیگر مقادیر (γ) نیز تکرار شد و بدین طریق مدل‌های متفاوتی با تغییر در مقدار (γ) حاصل شدند و درنهایت مقادیری از (ε، C) و (γ) که بهترین مقدار شاخص‌های آماری را داشته‌اند، انتخاب شده‌اند.

جدول ۴. نتایج حاصل از روش‌های SVM و GPR برای برآورد غلظت هوا تحت سناریو ۱

مدل	روش	آزمون				صحت‌سنجی			
		KGE	RMSE	DC	R	KGE	RMSE	DC	R
C(I)	SVM	-/۲۳۱۸	-/۲۵۵۵	-/۲۳۰۰	-/۲۲۵	-/۱۳۹۰	-/۱۴۳۳	-/۳۲۷۸	-/۱۲۰
	GPR	-/۱۱۲۷	-/۰۱۲۷	-/۲۵۹۶	-/۱۰۸	-/۰۸۶۲	-/۰۰۰۹	-/۲۴۸۴	-/۰۷۵
C(II)	SVM	-/۰۶۰۵	-/۰۹۹۱	-/۲۳۳۹	-/۰۵۰	-/۰۴۸۱	-/۱۷۰۳	-/۲۶۸۵	-/۰۴۵
	GPR	-/۱۴۲۴	-/۱۴۰۸	-/۲۵۵۹	-/۱۴۵	-/۱۹۳۵	-/۱۸۰۳	-/۲۵۸۰	-/۱۸۴
C(III)	SVM	-/۵۷۰۱	-/۳۰۲۲	-/۲۱۸۲	-/۵۷۲	-/۵۰۲۷	-/۱۹۸۵	-/۲۲۲۲	-/۵۰۰
	GPR	-/۶۸۴۴	-/۴۶۸۳	-/۱۹۰۵	-/۶۵۵	-/۵۹۳۱	-/۳۲۶۶	-/۲۰۳۷	-/۵۸۴
C(IV)	SVM	-/۲۳۶۰	-/۲۰۷۹	-/۳۶۰۹	-/۳۳۴	-/۱۲۹۴	-/۱۵۵۰	-/۳۵۰۶	-/۱۲۱
	GPR	-/۱۱۲۷	-/۱۱۲۷	-/۲۵۹۶	-/۱۱۲	-/۰۸۶۲	-/۰۸۴۵	-/۲۴۸۴	-/۰۸۵
C(V)	SVM	-/۹۶۹۱	-/۹۳۸۷	-/۰۶۴۷	-/۹۵۴	-/۸۶۳۷	-/۷۴۴۶	-/۱۲۵۵	-/۸۵۷
	GPR	-/۹۷۴۳	-/۹۴۹۰	-/۰۵۹۰	-/۹۷۲	-/۸۵۹۴	-/۷۳۶۶	-/۱۲۹۸	-/۸۴۰
C(VI)	SVM	-/۰۸۴۳	-/۰۷۶۳	-/۳۷۱۰	-/۰۷۹	-/۰۸۶۰	-/۱۱۱۰	-/۲۶۱۷	-/۰۸۱
	GPR	-/۲۰۱۸	-/۰۴۰۶	-/۲۵۵۹	-/۱۹۸	-/۲۰۴۵	-/۱۸۱۳	-/۲۵۸۱	-/۰۷۹
C(VII)	SVM	-/۹۷۴۳	-/۹۴۹۱	-/۰۵۸۹	-/۹۷۲	-/۸۷۰۰	-/۷۵۴۳	-/۱۲۳۰	-/۸۷۸
	GPR	-/۹۷۸۹	-/۹۵۸۲	-/۰۵۳۴	-/۹۷۸	-/۹۲۹۹	-/۸۶۴۴	-/۰۹۱۴	-/۹۰۸



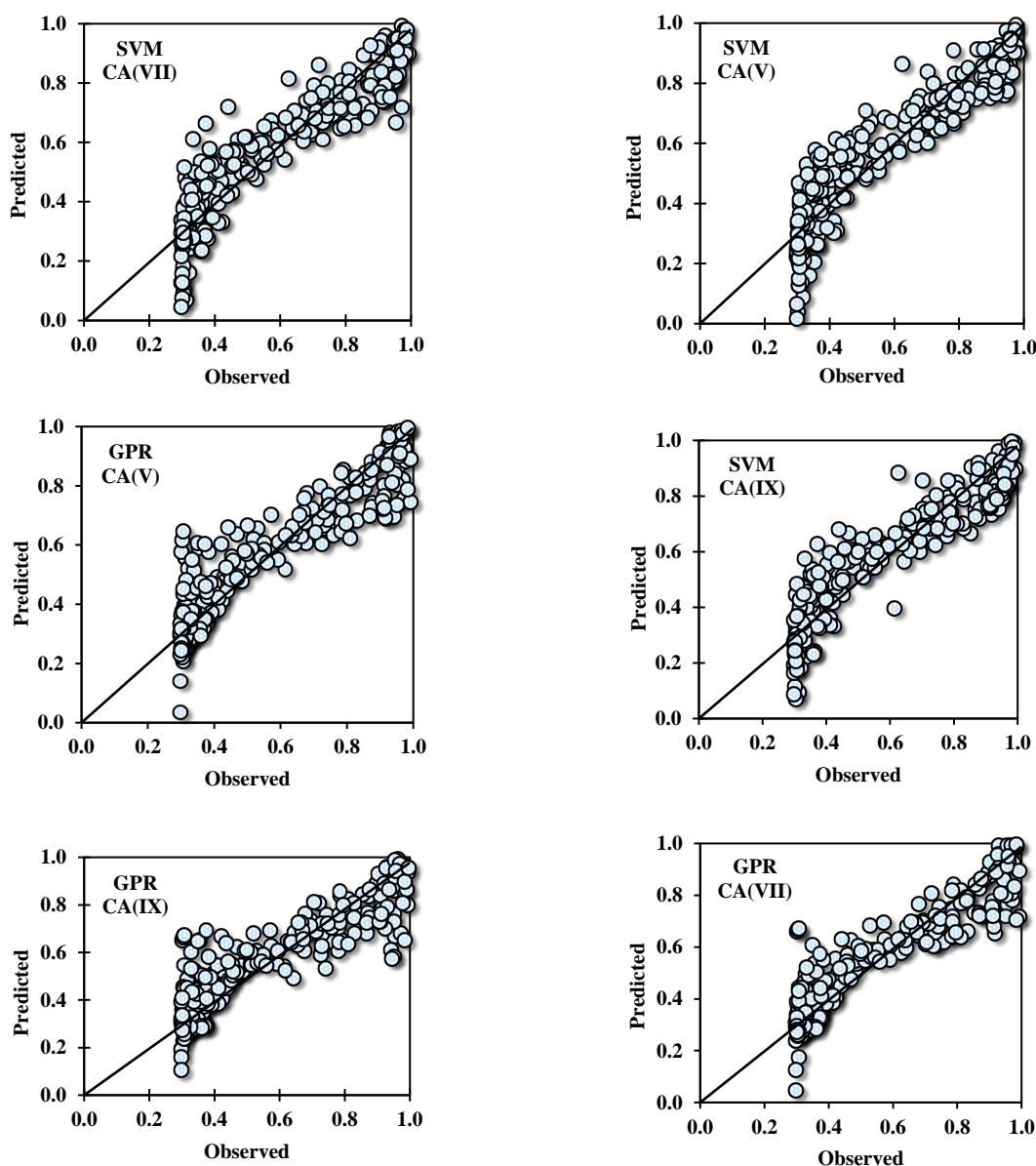
شکل ۲. نمودار پراکندگی بین داده‌های مشاهداتی و پیش‌بینی شده تحت سناریو ۱ (مجموعه داده‌های صحت‌سنجی)



شکل ۳. نمودار تغییرات DC و RMSE در برابر پارامتر γ تحت سناریو ۱

نتایج مدل‌های تعریف شده تحت سناریو ۲

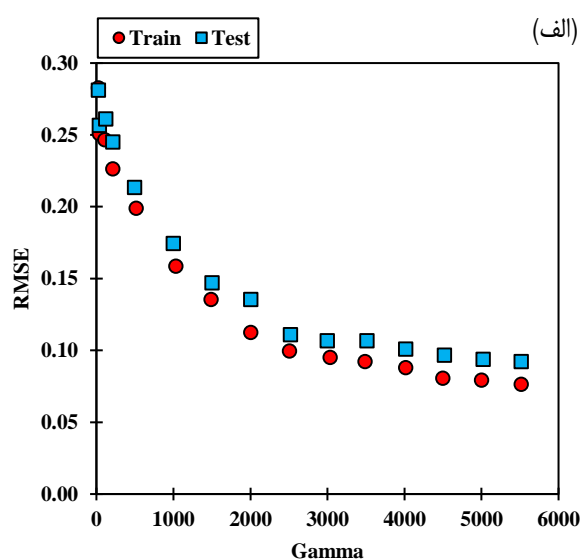
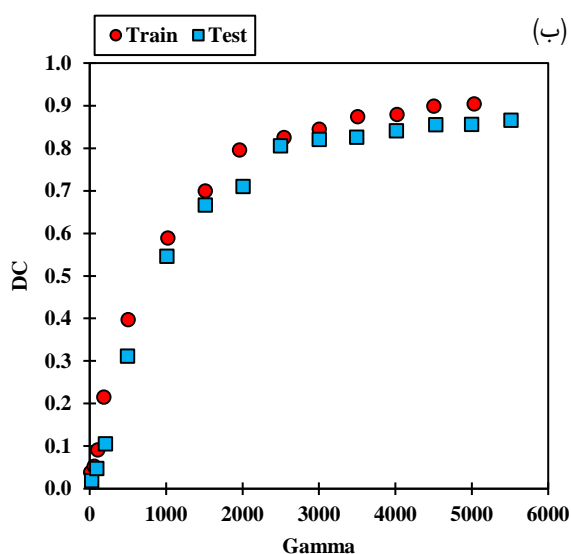
مطابق جدول (۵)، نتایج حاصل از دو روش ماشین بردار پشتیبان و روش رگرسیون فرآیند گوسی برای حالتی که هوادهی مصنوعی توسط هواده انجام می‌گیرد (۱۱۵۵ داده)، ارائه شده است. مطابق جدول (۵)، مدل CA(V) مقادیر R و DC بیشتر و RMSE کمتری نسبت به سایر مدل‌ها ارائه می‌دهد. بنابراین مدل مذکور به‌عنوان مدل برتر در تخمین غلظت هوا در طول سرریز انتخاب می‌گردد. در این سناریو روش‌های مبتنی بر کرنل مورد استفاده در این تحقیق با سه پارامتر ورودی دارای نتایج بهینه بوده و افزایش تعداد پارامترهای ورودی تأثیری در بهبود نتایج ندارد. همانند سناریو اول، در این سناریو نیز مدل‌هایی با یک و دو پارامتر ورودی با دارا بودن مقادیر منفی همبستگی خطی از دقت لازم برخوردار نمی‌باشند. همچنین بررسی نتایج به‌دست آمده در سناریو دوم نشان می‌دهد هر دو روش SVM و GPR دارای عملکرد نسبتاً مشابهی در تخمین غلظت هوای مورد نیاز هستند. نمودار پراکندگی داده‌ها برای سه مدل برتر در شکل (۴) نشان داده شده است. همچنین شکل (۵) نمودار تغییرات DC و RMSE را در برابر مقادیر مختلف پارامتر γ نشان می‌دهد. در روش ماشین بردار پشتیبان در اغلب موارد مقادیر کوچک پارامتر γ موجب خطای بیش آموزش شده و افزایش مقادیر این پارامتر نیز پیچیدگی مدل را افزایش می‌دهد. مقدار بهینه پارامتر γ در ماشین بردار پشتیبان برابر با ۵۰۰۰ بوده است که نشان دهنده پیچیدگی بالای مدل‌سازی در سناریو دوم نسبت به سناریو اول است (شکل ۵).



شکل ۴. نمودار پراکندگی بین داده‌های مشاهداتی و پیش‌بینی شده تحت سناریو ۲ (مجموعه داده‌های صحت‌سنجی)

جدول ۵. نتایج حاصل از روش‌های SVM و GPR برای برآورد غلظت هوا تحت سناریو ۲

مدل	روش	آزمون				صحت‌سنجی			
		R	DC	RMSE	KGE	R	DC	RMSE	KGE
CA(I)	SVM	۰/۰۱۳۷	۰/۰۲۹۴	۰/۳۵۸۹	۰/۰۰۸	۰/۱۰۲۰	۰/۰۴۳۳	۰/۳۴۷۸	۰/۱۱۴
	GPR	۰/۱۲۹۴	۰/۱۴۶۸	۰/۲۵۶۲	۰/۱۲۵	۰/۰۸۸۸	۰/۱۴۲۲	۰/۲۵۶۴	۰/۰۸۷
CA(II)	SVM	۰/۰۶۶۴	۰/۰۶۷۸	۰/۲۶۷۰	۰/۰۵۴	۰/۲۱۸۹	۰/۱۶۱۳	۰/۲۷۶۰	۰/۲۰۵
	GPR	۰/۲۲۴۶	۰/۰۵۰۴	۰/۲۵۱۸	۰/۲۱۴	۰/۰۷۶۷	۰/۰۸۳۰	۰/۲۶۶۶	۰/۰۷۲
CA(III)	SVM	۰/۲۷۶۵	۰/۰۵۵۳	۰/۲۶۵۴	۰/۲۷۲	۰/۳۶۰۹	۰/۳۸۰۴	۰/۲۵۷۵	۰/۳۵۰
	GPR	۰/۵۷۳۹	۰/۳۲۹۴	۰/۲۱۱۶	۰/۵۷۲	۰/۵۸۲۱	۰/۳۳۴۲	۰/۲۰۹۰	۰/۵۸۰
CA(IV)	SVM	۰/۰۳۷۳	۰/۰۶۶۲	۰/۳۴۳۴	۰/۰۲۱	۰/۱۲۱۴	۰/۱۴۹۰	۰/۳۴۸۳	۰/۱۱۸
	GPR	۰/۱۲۹۴	۰/۰۱۶۸	۰/۲۵۶۲	۰/۱۲۵	۰/۵۸۸۸	۰/۰۵۸۸۴	۰/۲۵۶۴	۰/۵۷۷
CA(V)	SVM	۰/۹۵۰۵	۰/۹۰۳۳	۰/۰۸۰۳	۰/۹۴۸	۰/۹۳۳۲	۰/۸۶۶۲	۰/۰۹۳۷	۰/۹۲۱
	GPR	۰/۹۵۱۷	۰/۹۰۵۵	۰/۰۷۹۴	۰/۹۴۵	۰/۹۲۱۴	۰/۸۴۵۱	۰/۱۰۰۸	۰/۹۱۴
CA(VI)	SVM	۰/۱۷۷۶۰	۰/۱۸۸۶	۰/۲۷۵۷	۰/۱۷۴	۰/۲۵۶۸	۰/۲۸۳۱	۰/۲۹۰۲	۰/۲۴۱
	GPR	۰/۲۲۸۵	۰/۰۵۲۲	۰/۲۵۱۵	۰/۲۱۸	۰/۱۰۱۰	۰/۰۹۸۰	۰/۲۶۸۴	۰/۱۰۴
CA(VII)	SVM	۰/۹۵۲۶	۰/۹۰۶۸	۰/۰۷۸۹	۰/۹۴۸	۰/۹۲۴۸	۰/۸۵۴۳	۰/۰۹۷۸	۰/۹۱۰
	GPR	۰/۹۵۶۸	۰/۹۱۵۱	۰/۰۷۵۳	۰/۹۴۷	۰/۹۲۲۸	۰/۸۴۹۳	۰/۰۹۹۴	۰/۹۱۷
CA(VIII)	SVM	۰/۳۱۴۰	۰/۳۵۷۱	۰/۱۶۰۶	۰/۳۰۵	۰/۳۵۴۷	۰/۳۸۰۳	۰/۱۴۷۵	۰/۳۴۰
	GPR	۰/۵۷۵۷	۰/۳۳۱۴	۰/۲۱۱۳	۰/۵۷۲	۰/۵۸۲۸	۰/۳۳۴۸	۰/۲۰۸۹	۰/۵۷۲
CA(IX)	SVM	۰/۹۴۷۷	۰/۸۹۷۸	۰/۰۸۲۶	۰/۹۴۱	۰/۹۲۱۹	۰/۸۴۹۱	۰/۰۹۹۵	۰/۹۰۷
	GPR	۰/۹۲۰۱	۰/۸۴۵۷	۰/۱۰۱۵	۰/۹۲۰	۰/۸۹۷۳	۰/۸۰۴۵	۰/۱۱۳۳	۰/۸۷۰


 شکل ۵. نمودار تغییرات DC و RMSE در برابر پارامتر γ تحت سناریو ۲

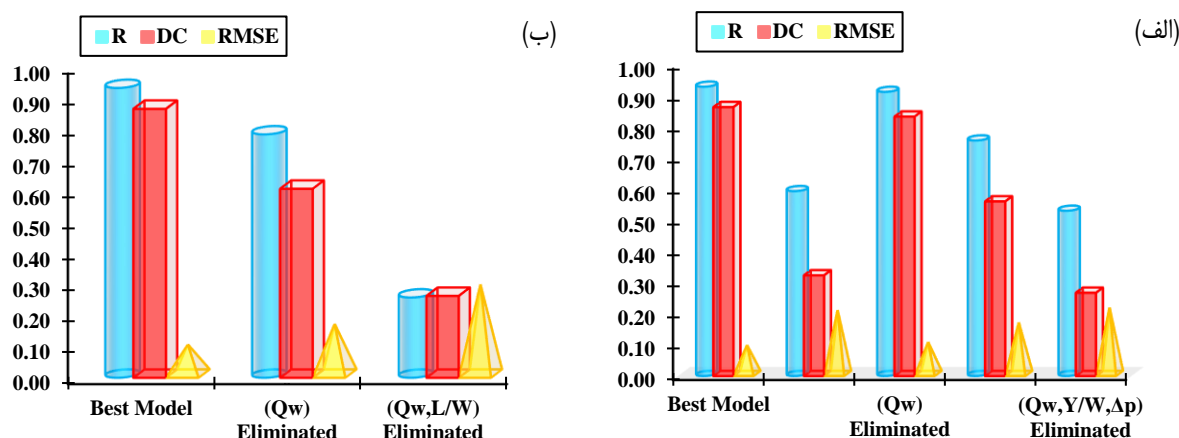
آنالیز حساسیت مدل برتر

به منظور بررسی تأثیر پارامترهای مختلف در تخمین توزیع غلظت هوا از آنالیز حساسیت استفاده شده است. به این صورت که با حذف پارامترهای مدل برتر و اجرای مجدد مدل با استفاده از ماشین بردار پشتیبان و رگرسیون فرآیند گوسی و همچنین محاسبه معیار ارزیابی مدل (R , DC , $RMSE$ and KGE)، میزان تأثیر پارامترهای حذف شده مورد بررسی گرفت. مقادیر آزمون تحلیل حساسیت برای مدل برتر در سناریو ۱ و ۲ در جدول (۶) نشان داده شده است. با توجه به نتایج آنالیز حساسیت مشاهده می‌گردد پارامترهای L/W و Y/W در برآورد توزیع غلظت هوا در طول سرریز دارای اهمیت بسیاری می‌باشد. همچنین حذف پارامتر Q_w تأثیر کمی در نتایج نهایی دارد. مطابق سناریو ۲ در جدول (۶) مشاهده می‌گردد پارامتر L/W در برآورد غلظت هوا در طول سرریز مهم است. همچنین مشابه با سناریو ۱، حذف پارامتر Q_w تأثیر کمی در نتایج نهایی دارد. شکل (۶) تأثیر حذف هر کدام از پارامترها را در مدل‌سازی‌ها را نشان می‌دهد.

جدول ۶. نتایج حاصل از روش آنالیز حساسیت برای برآورد توزیع غلظت هوا مدل برتر تحت سناریو ۱ و ۲

صحت‌سنجی				آزمون				روش	پارامتر حذف شده	مدل برتر سناریو ۱
KGE	RMSE	DC	R	KGE	RMSE	DC	R			
-۰/۴۲۱	۰/۲۴۰۰	-۰/۰۶۵۱	۰/۴۳۲۶	-۰/۵۳۸	-۰/۲۲۰۰	-۰/۲۹۰۷	-۰/۵۵۸۹	SVM	L/W	C(VII)
-۰/۲۱۴	۰/۲۰۴۱	-۰/۳۲۳۷	۰/۵۹۴۳	-۰/۶۷۲	-۰/۱۹۰۰	-۰/۴۷۰۹	-۰/۶۸۶۲	GPR		
-۰/۷۲۱	-۰/۱۲۸۳	-۰/۷۳۲۹	-۰/۸۵۶۶	-۰/۹۷۰	-۰/۰۶۱۲	-۰/۹۴۵۱	-۰/۹۷۲۵	SVM	Q _w	
-۰/۸۱۴	-۰/۱۰۱۲	-۰/۸۳۳۹	-۰/۹۱۳۳	-۰/۹۷۱	-۰/۰۶۳۱	-۰/۹۴۱۷	-۰/۹۷۰۶	GPR		
-۰/۶۷۲	-۰/۱۹۳۲	-۰/۳۹۴۰	-۰/۶۷۹۳	-۰/۸۱۴	-۰/۱۵۴۸	-۰/۶۴۹۰	-۰/۸۱۴۴	SVM	Q _w , ΔP	
-۰/۷۴۰	-۰/۱۶۴۴	-۰/۵۶۴۳	-۰/۷۵۷۱	-۰/۸۴۱	-۰/۱۳۷۶	-۰/۷۲۲۷	-۰/۸۵۰۲	GPR		
-۰/۳۵۰	-۰/۲۶۶۳	-۰/۱۵۰۵	-۰/۳۷۸۰	-۰/۴۵۲	-۰/۲۴۶۶	-۰/۱۰۹۳	-۰/۴۵۹۸	SVM	Q _w , ΔP, Y/W	
-۰/۲۵۴	-۰/۲۱۲۴	-۰/۲۶۷۸	-۰/۵۳۱۲	-۰/۵۸۰	-۰/۲۱۰۱	-۰/۳۵۳۱	-۰/۵۹۴۲	GPR		

صحت‌سنجی				آزمون				روش	پارامتر حذف شده	مدل برتر سناریو ۲
KGE	RMSE	DC	R	KGE	RMSE	DC	R			
-۰/۷۷۱	۰/۱۶۰۲	-۰/۰۸۸	۰/۷۸۳۸	-۰/۷۹۱	-۰/۱۵۸۵	-۰/۶۲۳۴	-۰/۷۹۴۱	SVM	Q _w	CA(VII)
-۰/۷۷۱	۰/۱۶۲۷	-۰/۵۹۶۴	۰/۷۷۲۵	-۰/۷۷۵	-۰/۱۶۰۷	-۰/۶۱۳۲	-۰/۷۸۳۳	GPR		
-۰/۳۴۵	-۰/۲۸۷۹	-۰/۲۶۳۶	-۰/۲۵۸۵	-۰/۲۷۵	-۰/۲۹۳۹	-۰/۲۹۳۶	-۰/۲۸۸۵	SVM	Q _w , L/W	
-۰/۵۲۵	-۰/۲۱۶۴	-۰/۲۸۶۵	-۰/۵۳۶۵	-۰/۴۴۱	-۰/۲۳۱۱	-۰/۱۹۹۶	-۰/۴۴۶۸	GPR		



شکل ۶. مقادیر خطاهای آزمون آنالیز حساسیت مدل برتر (الف) سناریو ۱ (ب) سناریو ۲

نتیجه‌گیری

در پژوهش حاضر از داده‌های آزمایشگاهی به منظور بررسی قابلیت روش‌های فرامدل در برآورد غلظت هوای مورد نیاز در سرریز سد کلاید استفاده گردید. بدین منظور ۲۲۶۸ داده آزمایشگاهی به دست آمده از مدل هیدرولیکی سد کلاید در فرآیند مدل‌سازی به کار گرفته شد. مدل‌های متعددی بر اساس ترکیب مختلفی از پارامترهای اندازه‌گیری شده توسعه داده شده و با بررسی نتایج به دست آمده، میزان تأثیر هر یک از پارامترها در تخمین غلظت هوادهی مشخص گردید. بررسی نتایج به دست آمده نشان داد هر دو روش ماشین بردار پشتیبان و رگرسیون فرآیند گوسی از دقت بالایی در حل مسئله مورد نظر برخوردار می‌باشند. در استفاده از روش رگرسیون بردار پشتیبان از بین توابع کرنل مختلف، تابع کرنل RBF بهترین نتایج را در تخمین غلظت هوای مورد نیاز در سرریزها به دنبال دارد. در برآورد میزان غلظت هوا در سرریز برای حالتی که هوادهی مصنوعی توسط هواده انجام نمی‌گیرد (۱۱۱۳ داده)، مدل با چهار پارامتر ورودی شامل $Q_w, L/W, Y/W$ و ΔP با دارا بودن مقادیر $DC=0/9229, R=0/8644$ و $RMSE=0/0914$ در رگرسیون فرآیند گوسی و $DC=0/7543, R=0/87$ و $RMSE=0/123$ در ماشین بردار پشتیبان به عنوان مدل برتر انتخاب گردید. در برآورد میزان غلظت هوا برای حالتی که هوادهی مصنوعی توسط هواده انجام می‌گیرد (۱۱۵۵ داده)، مدل با سه پارامتر ورودی شامل $Q_w, L/W, Y/W$ با دارا بودن مقادیر $DC=0/8451, R=0/9214$ و $RMSE=0/1008$ در رگرسیون فرآیند گوسی و $DC=0/8662, R=0/9333$ و $RMSE=0/0937$ در ماشین بردار پشتیبان به عنوان مدل



برتر انتخاب گردید. با توجه به نتایج به دست آمده، مشخص گردید که روش‌های SVM و GPR می‌توانند به‌عنوان روش‌های قابل اطمینان با نتایج قابل قبول و مطلوب برای تخمین غلظت هوای مورد نیاز مورد استفاده قرار بگیرند.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

REFERENCES

- Abbaszadeh, H., Norouzi, R., Sume, V., Kuriqi, A., Daneshfaraz, R., & Abraham, J. (2023a). Sill role effect on the flow characteristics (experimental and regression model analytical). *Fluids*, 8(8), 235.
- Abbaszadeh, H., Daneshfaraz, R., & Norouzi, R. (2023b). Experimental investigation of hydraulic jump parameters in sill application mode with various synthesis. *Journal of Hydraulic Structures*, 9(1), 18-42.
- Abbaszadeh, H., Daneshfaraz, R., Sume, V., & Abraham, J. (2024). Experimental investigation and application of soft computing models for predicting flow energy loss in arc-shaped constrictions. *AQUA—Water Infrastructure, Ecosystems and Society*, 73(3), 637-661.
- Chakib, B. (2013). Numerical computation of inception point location for flat-sloped stepped spillway. *International Journal of Hydraulic Engineering*, 2(3), 4752.
- Chanson, H. (1988). Study of air entrainment and aeration devices on spillway model. A thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in Civil Engineering at the University of Canterbury, New Zealand.
- Daneshfaraz, R., Norouzi, R., Abbaszadeh, H., Kuriqi, A., & Di Francesco, S. (2022a). Influence of sill on the hydraulic regime in sluice gates: an experimental and numerical analysis. *Fluids*, 7(7), 244.
- Daneshfaraz, R., Norouzi, R., Abbaszadeh, H., & Azamathulla, H. M. (2022b). Theoretical and experimental analysis of applicability of sill with different widths on the gate discharge coefficients. *Water Supply*, 22(10), 7767-7781.
- Daneshfaraz, R., Norouzi, R., Ebadzadeh, P., Di Francesco, S., & Abraham, J. P. (2023a). Experimental study of geometric shape and size of sill effects on the hydraulic performance of sluice gates. *Water*, 15(2), 314.
- Daneshfaraz, R., Norouzi, R., Ebadzadeh, P., & Kuriqi, A. (2023b). Influence of sill integration in labyrinth sluice gate hydraulic performance. *Innovative Infrastructure Solutions*, 8(4), 118
- Franc, J. P., & Michel, J. M. (2006). *Fundamentals of cavitation* (Vol. 76). Springer science & Business media.
- Hassanzadeh, Y., & Abbaszadeh, H. (2023). Investigating discharge coefficient of slide gate-sill combination using expert soft computing models. *Journal of Hydraulic Structures*, 9(1), 63-80.
- Kermani, E. F., Barani, G. A., & Hessaroeyeh, M. G. (2013). Investigation of cavitation damage levels on spillways. *World Applied Sciences Journal*, 21(1), 73-78.
- Norouzi, R., Sihag, P., Daneshfaraz, R., Abraham, J., & Hasannia, V. (2021). Predicting relative energy dissipation for vertical drops equipped with a horizontal screen using soft computing techniques. *Water Supply*, 21(8), 4493-4513.
- Pfister, M. (2011). Chute aerators: Steep deflectors and cavity subpressure. *Journal of Hydraulic Engineering*, 137(10), 1208-1215.
- Pfister, M., & Hager, W. H. (2010). Chute aerators: Air transport characteristics. *Journal of Hydraulic Engineering*, 136(6), 352-359.
- Rahmeyer, W. J. (1981). Cavitation damage to hydraulic structures. *American Water Works Association*, 73(5), 270-274.
- Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning* (Vol. 1, p. 159). Cambridge, MA: MIT press.
- Raza, A., Wan, W., & Mehmood, K. (2021). Stepped spillway slope effect on air entrainment and inception point location. *Water*, 13(10), 1428.
- Roushangar, K., & Homayounfar, F. (2019). Prediction characteristics of free and submerged hydraulic jumps on horizontal and sloping beds using SVM method. *KSCE Journal of Civil Engineering*, 23(11), 4696-4709.
- Roushangar, K., & Shahnazi, S. (2020). Prediction of sediment transport rates in gravel-bed rivers using Gaussian process regression. *Journal of Hydroinformatics*, 22(2), 249-262.
- Salmasi, F., Abraham, J., & Salmasi, A. (2021). Effect of stepped spillways on increasing dissolved oxygen in water, an experimental study. *Journal of Environmental Management*, 299, 113600.
- Süme, V., Daneshfaraz, R., Kerim, A., Abbaszadeh, H., & Abraham, J. (2024). Investigation of clean energy production in drinking water networks. *Water Resources Management*, 38(6), 2189-2208.
- Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Wu, J., Ma, F., & Dai, H. C. (2011). Influence of filling water on air concentration. *Journal of Hydrodynamics*, 23(5), 601-606.

Estimation of air concentration in chute spillway using metamodel methods

EXTENDED ABSTRACT

Introduction

Water supply for various agricultural, industrial, and particularly drinking purposes holds significant importance, especially in arid and semi-arid regions. In recent decades, with technological advancements, the construction of reservoirs and hydraulic structures has expanded considerably. Given the increasing trend in building tall dams and the need to enhance dam safety, hydraulic engineers are increasingly focusing on the economic and reliable design of these structures. One hydraulic structure playing a crucial role in ensuring sufficient dam safety during flood inflow into the reservoir and discharge downstream is spillways. Due to the long-standing use of hydraulic models in spillway design and the considerable accuracy of their results, engineers are compelled to construct these models and conduct experiments to achieve more economical and safer spillway designs. Due to the high flow velocity and pressure drop in tall spillways, the potential for cavitation phenomena has increased, leading to severe damage to the spillway. Therefore, a specialized structure called an aerator, also known as an air vent, must be incorporated in locations where natural aeration from the free surface of the flow is insufficient for proper protection. Understanding the distribution of air concentration variations along the spillway is crucial for estimating aeration levels. Introducing innovative and creative methods that can effectively address this issue holds immense importance in this context.

Materials and Methods

Here, Support Vector Machines and Gaussian Process Regression methods were employed as modeling approaches for parameter estimation, and their results were subjected to evaluation. Various models were defined with different input parameters, aiming to estimate air concentration. The concentration of air pollutants was predicted, and the influential parameters for each case were identified. This research endeavors to contribute to previous studies by demonstrating the application of Support Vector Machines and Gaussian Process Regression in estimating air concentration distribution and assessing the practicality of these methods in comparison to laboratory experimental data in this domain. For assessing the performance of the Support Vector Machine and Gaussian Process Regression methods in predicting air concentration distribution in spillways, experimental data from Chanson (1988) comprising 2268 data points have been utilized.

Results and Discussions

To investigate the impact of various parameters on the estimation of air concentration distribution, sensitivity analysis has been employed. This involved systematically excluding superior model parameters and re-running the model using Support Vector Machines and Gaussian Process Regression. Evaluation metrics (R, DC, RMSE and KGE) were calculated, and the influence of the removed parameters was examined. According to the sensitivity analysis results, parameters Y/W and L/W play a crucial role in estimating air concentration distribution along the spillway. Moreover, the exclusion of the Q_w parameter has a minor impact on the results. It is observed that the L/W parameter is significant in predicting air concentration along the spillway. Similar to scenario 1, the removal of the Q_w parameter has a negligible effect on the final results.

Conclusion

Here, a wide range of experimental data was utilized to examine the modeling capabilities in predicting the required air concentration in the spillway of the Clyde Dam. For this purpose, 2268 experimental data obtained from the hydraulic model of the Clyde Dam were employed in the modeling process. Multiple models were developed based on different combinations of measured parameters, and by analyzing the obtained results, the impact of each parameter on estimating air concentration was determined. The results indicated that both Support Vector Machine (SVM) and Gaussian Process Regression (GPR) methods exhibit high accuracy in solving the specified problem. In the application of Support Vector Regression, among various kernel functions, the Radial Basis Function (RBF) kernel yielded the best results in estimating the required air concentration in spillways. The estimation of air concentration in the spillway, particularly for scenarios where artificial aeration is not performed (1113 data), resulted in a model with four input parameters, including Q_w , L/W , Y/W , and ΔP . This model, with values of $R=0.9229$, $DC=0.8644$, and $RMSE=0.0914$ in Gaussian Process Regression, and $R=0.87$, $DC=0.7543$, and $RMSE=0.123$ in Support Vector Machine, was selected as the superior model. For the estimation of air concentration in scenarios where artificial aeration is conducted (1155 data points), a model with three input parameters, Q_w , L/W , and Y/W , demonstrated superior performance. This model, with values of $R=0.9214$, $DC=0.8451$, and $RMSE=0.1008$ in Gaussian Process Regression, and $R=0.9333$, $DC=0.8662$, and $RMSE=0.0937$ in Support Vector Machine, was chosen as the top-performing model. According to the obtained results, it is evident that SVM and GPR methods can serve as reliable approaches with acceptable and desirable results for estimating the required air concentration.

Keywords: Aeration, Chute Spillway, Gaussian Process Regression, Support Vector Machine.