مدل جانشین ترکیبی فرایند گاوسی پراکنده و بوستر Random Forest برای پیش‌بینی ویژگی‌های ورودی زیرصوت

تاریخ دریافت: ۱۴۰۵/۰۷/۱۶
تاریخ پذیرش: ۱۴۰۵/۰۸/۱۷

محمدرسول گلشیرازی [۱]، فرشاد پازوکی [۲]، علیرضا داوری [۳]

۱ دانشجوی دکتری، گروه مهندسی هوافضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.
۲ استادیار، گروه مهندسی هوافضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران. Pazooki_Fa@iau.ac.ir
۳ دانشیار، گروه مهندسی هوافضا، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران.

چکیده

با وجود دقت بالای شبیه‌سازی دینامیک سیالات محاسباتی (CFD) در تحلیل ورودی‌های زیرصوت، هزینه محاسباتی زیاد آن، طراحی تکرارشونده را محدود می‌کند. این پژوهش یک مدل جانشین ترکیبی داده‌محور برای پیش‌بینی فشار استاتیک/سکون، پسا، ماخ و شاخص یکنواختی سرعت (VUI) معرفی می‌کند که زمان حل RANS را از چند ساعت به کمتر از ۵ میلی‌ثانیه روی پردازنده گرافیکی کاهش می‌دهد. این مدل برای خروجی‌های پرنوسان از فرایند گاوسی چندخروجی تُنُک و برای خروجی‌های کم نوسان از جنگل تصادفی بوستر استفاده می‌کند. اعتبارسنجی متقابل روی ۱۲۳ هندسه، خطای nRMSE کمتر از ۰,۱۳ برای پسا و ۰,۰۸ برای فشارها را نشان داد. همچنین، ارزیابی روی ۱۰ هندسه تصادفی حاکی از خطای MAPE کمتر از ۰,۱۲٪ و R2>0.98 برای تمامی خروجی‌هاست. این مدل ضمن ارائه برآورد عدم قطعیت، جایگزینی دقیق برای حل CFD بوده و طراحی بلادرنگ را تسهیل می‌سازد.

واژه‌های کلیدی: *ورودی هوا زیرصوت؛ مدل جانشین داده‌محور؛ فرایند گاوسی چندخروجی؛ جنگل تصادفی؛ یادگیری فعال.*

A hybrid surrogate model of sparse gaussian process and random forest booster for predicting subsonic inlet characteristics

Mohammadrasuol Golshirazi [1], Farshad Pazooki [2], Alireza Davari [3]

1- Ph. D. Candidate, Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran.
2- Assistant Professor Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran.
3- Associate professor, Department of Aerospace Engineering, SR.C., Islamic Azad University, Tehran, Iran.

Abstract

Despite the high accuracy of Computational Fluid Dynamics (CFD) in analyzing subsonic inlets, its significant computational cost limits iterative design processes. This study introduces a data-driven hybrid surrogate model to predict static/stagnation pressure, drag, Mach number, and Velocity Uniformity Index (VUI), reducing RANS solution time from several hours to under 5 milliseconds on a GPU. The model employs a Sparse Linear Model of Coregionalization Gaussian Process Regression (Sparse LMC-GPR) for high-variance outputs and a boosted Random Forest for low-variance outputs. Cross-validation on 123 geometries demonstrated an nRMSE of less than 0.13 for drag and 0.08 for pressures. Furthermore, evaluation on 10 random geometries yielded a Mean Absolute Percentage Error (MAPE) below 0.12% and R^2>0.98 across all outputs. By providing calibrated uncertainty estimates, this model serves as an accurate alternative to CFD and facilitates real-time design.

Keywords: *Subsonic Air Inlet; Data-Driven Surrogate Model; Multi-Output Gaussian Process; Random Forest; Active Learning.*

۱. مقدمه

پیشرفت شبیه‌سازی عددی دینامیک سیالات (CFD) در دو دهه اخیر، تحلیل دقیق اندرکنش جریان–سازه، پیش‌بینی توزیع فشار و تخمین نیروهای آیرودینامیکی را میسر ساخته است؛ امّا هزینه محاسباتی گزاف، به‌ویژه در فرایندهای طراحی تکرارشونده و بهینه‌سازی چندهدفه، همچنان مانعی جدّی در کاربرد صنعتی این روش به‌شمار می‌آید. چنان‌که کوروندی و همکاران [۱] تصریح می‌کنند، بهره‌گیری از مدل‌های جانشین[۱] می‌تواند تعداد تحلیل‌های CFD پرهزینه را به حداقل رسانَد و چرخه طراحی را چندین برابر تسریع کند.

مدل‌های جانشین داده‌محور به‌طور کلی در دو دستهٔ مداخله‌گر[۲] و غیرمداخله‌گر[۳] طبقه‌بندی می‌شوند [۲]. در روش‌های مداخله‌گر، ساختار حلگر تغییر می‌یابد تا تقریب در دل کد CFD تعبیه شود، حال آنکه روش‌های غیرمداخله‌گر صرفاً با تکیه بر داده‌های خروجی، یک شبه‌نگاشت ریاضی خارج از حلگر ایجاد می‌کنند. متداول‌ترین الگوریتم‌های غیرمداخله‌گر عبارت‌اند از توابع پایهٔ شعاعی[۴]، بسط چندجمله‌ای آشوب مبتنی بر روش برخورد نقطه‌ای[۵]، شبکه‌های عصبی مصنوعی[۶] و رگرسیون فرایند گاوسی[۷] [۳].

فرایند گاوسی به‌دلیل ماهیت تمام‌احتمالاتی، افزون بر برآورد نقطه‌ای، بازهٔ عدم‌قطعیت پیش‌بینی را نیز فراهم می‌آورد که این امر در تحلیل ریسک طراحی و روش‌های چندوفاداری[۸] حیاتی است. به عنوان مثال، پِردیکارِس و همکاران [۴] نشان دادند که مدل‌های چندوفادار مبتنی بر GPR قادرند با بهره‌گیری از اطلاعات سطح پایین (تحلیل ساده) و سطح بالا دینامیک‌سیالات محاسباتی فرآیند بهینه‌سازی شکل پره را با صرف تعداد اندکی تحلیل دقیق تکمیل کنند. با این حال، در کمیت‌هایی که نوسان آماری اندکی دارند یا توزیعشان از حالت نرمال خارج است نظیر عدد ماخ یا شاخص یکنواختی سرعت[۹] کارایی GPR افت می‌کند. در این موارد، جنگل تصادفی[۱۰] (RF) به‌دلیل تجمیع درخت‌های تصمیم و کاهش بیش‌برازش[۱۱]، عملکرد پایدارتری ارائه می‌دهد [۵].

خلأ پژوهش کنونی آن است که اغلب تحقیقات از یک الگوریتم منفرد برای تمامی خروجی‌ها استفاده کرده‌اند، در حالی که ماهیت آماری کمیت‌های جریان متفاوت است. در این مقاله، یک چارچوب مدل جانشین هیبریدی مبتنی بر ترکیب Sparse Multi output GPR و بوستر RF معرفی می‌شود که هم قابلیت برآورد عدم‌قطعیت را داراست و هم دقت بالایی در متغیرهای کم‌دامنه ارائه می‌کند.

نوآوری محوری پژوهش حاضر در مقایسه با مطالعات مرجع مدل‌های جانشین آیرودینامیکی، در توسعه یک معماری هیبریدی سلسله مراتبی مبتنی بر تفکیک رفتار فیزیکی و مقیاس واریانس متغیرها نهفته است. در رویکردهای تک الگوریتمی مرسوم (نظیر استفاده مجزا از فرایند گاوسی یا جنگل تصادفی)، مدل به دلیل تفاوت شدید دامنه خروجی‌ها (تغییرات مرتبه بزرگی در پسا و فشارها در مقایسه با دامنه بسیار محدود شاخص یکنواختی سرعت VUI و عدد ماخ)، دچار پدیده بیش‌هموارسازی در متغیرهای حساس یا بیش‌برازش در متغیرهای سراسری می‌شود. در روش پیشنهادی، ابتدا متغیرهای با تغییرات شدید و دینامیک بالا (پسا و فشار) توسط رگرسیون فرایند گاوسی تنک چند خروجی (Sparse LMC-GPR) به همراه برآورد

کالیبره شده عدم قطعیت مدل سازی شده و سپس میانگین پیش‌بینی‌شده آن‌ها به‌عنوان ویژگی کمکی به بوستر جنگل تصادفی تزریق می‌گردد. این سازوکار، ضمن رفع محدودیت‌های روش‌های مرجع تکی و تضمین انطباق بر فیزیک جریان، زمان استنتاج را به کمتر از ۵ میلی‌ثانیه کاهش می‌دهد.

۲. مروری جامع بر فرایندهای گاوسی چندخروجی و بوسترهای مبتنی بر درخت

۲-۱. پیشینه تاریخی و نمادگذاری

در مسائل چندرشته‌ای آیروترموینامیک، یک بردار طراحی X (شامل ضخامت لِیپ، زاویهٔ دیفیوزر و ... به‌طور هم‌زمان چند پاسخ هم‌بسته نظیر فشار درگ، افت فشار سکون و عدد ماخ را کنترل می‌کند. مدل‌سازی مشترک این پاسخ‌ها با فرایند گاوسی چندخروجی (MOGP)[۱۲] دو مزیت کلیدی دارد[۳]:

۱. صرفه‌جویی داده‌ای: خروجی‌های کم‌نمونه (مانند شاخص یکنواختی سرعت) از اطلاعات خروجی‌های پرتراکم (فشار) بهره می‌برند.

۲. عدم‌قطعیت سازگار: کواریانس‌های متقاطع، قیود فیزیکی بین کمیت‌ها (فشار ↔ماخ) را در پیش‌بینی حفظ می‌کنند.

۲-۲. مبانی ریاضی فرایندهای گاوسی چندخروجی

در مدل خطی هم‌منطقه‌ای (LMC) فرض می‌شود که هر خروجی y_m ترکیبی خطی از Q فرآیند نهفته f_q باشد [۲]:

$$y_m(x) = \sum_{q=1}^{Q} a_{mq} f_q(x) + \varepsilon_m \qquad (۱)$$

$$f_q(x) \sim GP(0, k_q(x, x')) \qquad (۲)$$

در نتیجه، کوواریانس بلوکی خروجی‌ها برابر است با:

$$K_{mm'}(x, x') = \sum_{q=1}^{Q} a_{mq} a_{m'q}\, k_q(x, x') \qquad (۳)$$

۲-۲-۱. مدل هم‌منطقه‌ای ذاتی

اگر تمام خروجی‌ها با استفاده از یک کرنل مشترک توصیف شوند یعنی ساختار آماری همه آن ها در فضای ورودی یکسان در نظر گرفته شود و ماتریس هم‌منطقه‌ای[۱۳] که روابط بین خروجی‌ها را کد می‌کند، به‌صورت حاصل‌ضرب بردارهای وزن یعنی $B = aa^{\top}$ تعریف گردد، آنگاه به ساختاری می‌رسیم که به آن مدل هم‌منطقه‌ای ذاتی گفته می‌شود [۲].

جدول ۱ نشان می‌دهد که برای خروجی‌های همگن، ICM کفایت دارد؛ اما در حضور رفتار ناایستا یا طول‌مقیاس متفاوت، LMC $Q \geq 2$ خطای برآورد را تا ۲۰ ٪ کاهش می‌دهد [۲].

جدول ۱. مقایسه میان انواع کرنل‌ها

ویژگی‌ها	مدل هم‌منطقه‌ای ذاتی	خطی مدل هم‌منطقه‌ای
فرم کرنل	$K = B \otimes k(x, x')$	$K = \sum_{q=1}^{Q} B^{(q)} \otimes k_q(x, x')$
تعداد پارامترها	$\frac{M(M+1)}{2} + \dim(k)$	$Q \cdot \frac{M(M+1)}{2} + \sum_{q=1}^{Q} \dim(k_q)$
هزینه محاسباتی	مستقیم $O((MN)^3)$ $O(Q(M+N)^3)$ درصورت وجود کرونکر	همان مرتبه، ولی بدون بهره‌مندی از ساختار کرونکِر
توانایی در بیان ساختارهای پیچیده	محدود به طول‌مقیاس مشترک، تنها وابستگی‌های رتبه‌پایین را مدل می‌کند	امکان مدل‌سازی طول‌مقیاس‌های مستقل، انعطاف‌پذیری بالاتر در هم‌بستگی خروجی‌ها

۲-۳. مقیاس‌پذیری و تقریب‌های تنک

گره محاسباتی اصلی عبارت است از واژگونی ماتریس بلوکی با ابعاد $MN \times MN$. برای کاهش هزینه سه راهکار رایج وجود دارد:

متغیرهای القایی وSVGP ؛ الگوریتم FITC هزینه را به $\mathcal{O}(\mathrm{M^2N})$ در کاربردهای عملی مدل‌های گاوسی چندخروجی نظیر ورودی زیرصوتی تبدیل کند وهمچنین استفاده از تقریب‌های تنک نظیر SVGP می‌تواند به‌طور چشم‌گیری هزینه‌ی محاسباتی را کاهش دهد، بی‌آنکه دقت مدل به‌طور معنادار افت کند. برای مثال، در یکی از مطالعات انجام‌شده بر روی یک پایگاه داده‌ی متشکل از ۱۲۳ هندسه‌ی ورودی زیرصوت، به‌کارگیری مدل LMC با رتبه‌ی ۲ و کرنل ترکیبی(Matern-RQ همراه با تنها ۱۹ نقطه‌ی القایی) برگزیده به روش (k-means++)، منجر به پیش‌بینی نیروی درگ با خطای نرمال‌شده‌ی کمتر از ۰٫۱۱ شد، در حالی‌که زمان پیش‌بینی برای هر هندسه کمتر از ۴ میلی‌ثانیه در GPU گزارش گردید. این تجربه نشان می‌دهد که روش‌های تنک با انتخاب مناسب نقاط القایی، می‌توانند در مسائل واقعی طراحی آیرودینامیک، تعادل مطلوبی میان دقت و کارایی ایجاد کنند [۶].

۱. ساختار کرونکر درICM؛ جداسازی فضای ورودی/خروجی، پیچیدگی را به $M^3 + N^3$ می‌کاهد.

۲. KISS-GP؛ بر مبنای درون‌یابی روی شبکه منظم و حلگر CG [۷].

۲-۴. تفسیر عدم‌قطعیت و فعال‌سازی تطبیقی

اگر رتبه واقعی کوواریانس بزرگ‌تر از رتبهٔ مدل باشد، واریانس پیش‌بینی می‌تواند بیش‌یا کم‌اعتماد گردد. برای کالیبراسیون، معیارهای ALM و ALC نقاطی را انتخاب می‌کنند که بیشینهٔ واریانس پسین را دارند؛ افزون بر آن، MUSIC۱۴ برای حساسیت سوبول هدف‌گذاری شده است [۸].

۲-۵. بوسترهای جنگل تصادفی: انگیزه و کارایی

جنگل تصادفی با تجمیع درخت‌های غیرهم‌بسته، در کمیت‌های کم‌دامنه و غیرنرمال عملکرد پایداری دارد [۵]. راهبرد هیبریدی پژوهش حاضر سه گام دارد:

۱. برازش GP بر خروجی‌های پرنوسان نظیر درگ و فشارها؛

۲. تزریق میانگین GP به‌عنوان ویژگی کمکی؛

۳. یادگیری خروجی‌های ریزنوسان نظیر ماخ و شاخص یکنواختی سرعت باRF ،(۵۰۰ درخت،$R^2 \geq 0.88$)

در جریان داخل لولهٔ خمیده، RFخطای RMSEچهار مولفهٔ سرعت را از ۰٫۱۵ به ۰٫۰۶ کاهش داده است [۱۲].

محدودیت‌ها:

- بازه اطمینان صرفاً تجربی است و تضمین پوشش آماری مانند GP ندارد [۲].
- در داده‌های بسیار غیرخطیِ جریان‌های توسعه‌یافته، مکان ساختارهای ریز گردابه ممکن است با چند عنصر شبکه جابه‌جا پیش‌بینی شود [۴].
- در ابعاد زیاد، RFبه گزینش ویژگی حساس شده و ممکن است به انتخاب نادرست منجر شود، مگر این‌که روش‌های منظم RRF یا GRRF به کار رود [۴].

۲-۶. مدل‌های تلفیقی چندوفاداری و فرایندهای گاوسی عمیق

رویکردهای چند وفاداری[۱۵] با تلفیق داده‌های ارزان کم وفاداری (نظیر مدل‌های پتانسیل یا اویلر دوبعدی) و داده‌های گران قیمت پروفاداری (نظیر RANS یا LES)، نقش موثری در کاهش هزینه محاسباتی تولید پایگاه داده ایفا می‌کنند [۱، ۹](جدول ۲). در این چارچوب‌ها معمولاً فرض بر تصحیح خطای مقیاس‌پذیر میان دو سطح وفاداری است.

جدول ۲. روش‌های تلفیقی چندوفاداری و مدل‌های عمیق

رده	ایدهٔ کلیدی	مثال و مزایا
کوکریجینگ خطی(Kennedy–O'Hagan)	رابطه خودبازگشتی $f_h(x) = \rho f_\ell(x) + \delta(x)$	کاهش هزینه از به $O((N_l + N_h)^3)$ $\sum_l O(N_l^3)$.[۹]
کوکریجینگ غیرخطی (Perdikaris)	روی فضای GP (x, f_ℓ)	مدل‌سازی وابستگی غیرخطی و ناایستایی.[۱۰]
فرایندهای گاوسی (DGP) عمیق	لایه‌بندی سطوح وفاداری	پیچیدگی $O\sum_\ell(M_l^2 N_l)$ ،مقیاس‌پذیر [۱۱].

در پژوهش حاضر، به منظور حفظ حداکثر پیوستگی فیزیکی و پیشگیری از خطاهای سیستماتیک مدل‌های کم وفاداری در شبیه سازی لایه مرزی و افت فشار داخل ورودی، تمامی ۱۲۳ نمونه پایگاه داده با وفاداری کامل و حل معادلات ناویر- استوکس میانگین‌گیری شده رینولدز (RANS) تولید شده‌اند. با این حال، ایده بنیادین مدل سازی چندوفاداری در سطح معماری یادگیری بازتولید شده است؛ بدین صورت که خروجی‌های پیش‌بینی‌شده مرحله نخست به عنوان سیگنال‌های اطلاعاتی تکمیلی برای تقویت دقت مدل در مرحله دوم به کار گرفته می‌شوند.

۲-۷. جمع‌بندی مطالعات تجربی

در پیش‌بینی فشار بر روی دیوارهٔ جریان مافوق‌صوت، RF با ۱۰۰درخت خطا را از ۱۵٪ به ۶٪ کاهش داد [۱۲].

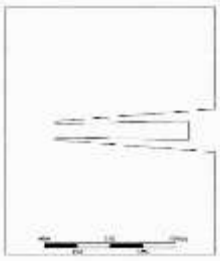
Korondi و همکاران با مدل MF-GPR جبههٔ پارتو طراحی ملخ را با کمتر از ۴۰ ارزیابی RANS/XFOIL به‌دست آوردند [۱].

Abu-Zurayk به کمک co-kriging و حساسیت‌سنجی زمان بهینه‌سازی بار هواپیما را ۳۰٪ کاست [۱۵].

۳. پیکربندی و صحّت‌سنجی شبیه‌سازی CFD

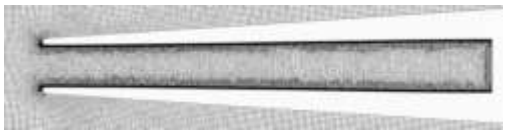
برای ده هندسه متفاوت تحلیل دینامیک سیالاتتمحاسباتی انجام شد که توانایی مدل جانشین در داده‌های جدید سنجیده شوند؛ شبیه‌سازی‌های عددی این پژوهش در محیط ANSYS Fluent 2024 R2 و به‌صورت دوبُعدیِ پایا (RANS) انجام شد. برای مدلسازی آشفتگی، مدل k-ω SST برگزیده شد؛ زیرا با ترکیب مزایای مدل‌های k-ε و k-ω در ناحیهٔ دیواره و لایهٔ برشی آزاد، توان بالایی در بازآفرینی گرادیان‌های فشار ورودی‌های زیرصوت دارد. جریان آزاد در مرز ورودی (Pressure Far-Field) با عدد ماخ ۰,۰۷ فشار سکون ۱۰۱کیلوپاسکال و دمای سکون ۲۸۸ کلوین در مرز خروج اعمال شد و مرز خروجی به صورت Pressure Outlet تعریف گردید. نرخ دبی جرمی برای همهٔ هندسه‌ها $29.5\ kgs^{-1}$ ثابت نگه داشته شد تا امکان مقایسهٔ مستقیم فراهم آید و اثرات تراکم تغییر نکند. دامنهٔ محاسباتی به‌گونه‌ای امتداد یافته است که بازتاب موج‌ها و تداخل مرزی به حداقل برسد؛ شبکه به‌صورت شبه‌ساختاریافته تولید شد و معیارهای کیفیت زیر رعایت گردید:

کیفیت اُرتوگونال>۰,۴۵ ، Skewness کمتراز ۰,۷۵ نسبتابعادی>۴,۰. همگرایی حل با کاهش باقی‌مانده‌های تمام معادلات تا مرتبه 10^{-8} و خطای بیلانس جرمی زیر ۰,۰۱۵٪ تضمین شد. این سطح دقت، صحت عددی میدان جریان را پیش از آموزش مدل جانشین تأیید می‌کند.



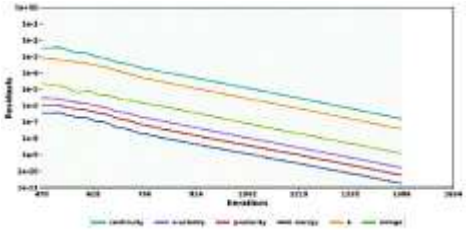
شکل ۱. هندسه G1 ودامنه محاسباتی

نظر به ماهیت بیضوی حل در رژیم مادون‌صوت، دامنه محاسباتی به‌گونه‌ای طراحی شده‌است که اثر تداخلات مصنوعی مرزی به حداقل رسیده و شرایط مرزی بر حل تأثیرگذار نباشند. به‌ویژه با انتخاب مرزهای دور از ناحیه لبه ورودی، میدان فشار و نوسانات ناشی از بازتاب‌های عددی کاهش یافته‌اند.



شکل ۲. مش هندسه G1 و لایه‌مرزی به کار رفته

درادامه نمودار باقیمانده‌ها برحسب تکرار حل برای هندسه G1 را نشان می‌دهد



شکل ۳. نمودار لگاریتمی حل هندسه G1 همانطور که قابل مشاهده است تمام معادلات تا 10^{-5} همگرا شده‌اند.

۳-۱. ساخت پایگاه داده و پیش‌پردازش

۳-۱-۱. گردآوری و پاک‌سازی داده‌ها

پایگاه داده از ۱۲۳ شبیه‌سازی CFD معتبر تشکیل شده است. بردار ویژگی شامل هشت پارامتر هندسی است: (ضخامت لبه ورودی b، نسبت انقباض CR، ضریب ظرافت Fineness Ratio، طول و زاویهٔ دیفیوزر، نقطهٔ اعمال زاویه، و دو متغیر کمکی اختلاف فشار استاتیک و کل). بردار خروجی نیز هشت کمیت آیرودینامیکی کلیدی را دربر می‌گیرد: نیروی درگ، فشارهای استاتیک و سکون در دو مقطع، اعداد ماخ (Outlet/Highlight) و شاخص یکنواختی سرعت (VUI).

لازم به تاکید است که از میان ۸ ویژگی ورودی، شش متغیر مستقیماً ابعاد هندسی ورودی هوا را تعیین کرده و دو متغیر کمکی ($\Delta P_{aux}, PR$) نمایانگر شرایط مرزی و الزامات تقاضای عملکردی موتور (نظیر افت فشار تخمینی حاصل از روابط شبه یک‌بعدی اولیه و نسبت فشار طراحی) هستند. بنابراین، این دو متغیر در ابتدای گام طراحی مفهومی همواره به عنوان مقادیر ورودی معلوم در دسترس طراح بوده و استخراج آن‌ها هیچ‌گونه وابستگی به اجرای پیشین شبیه‌سازی عددی ندارد؛ امری که استقلال کامل مدل جانشین در فاز کاربردی را تضمین می‌نماید.

۳-۱-۲. مقیاس‌دهی ویژگی‌ها و اهداف

به‌دلیل اختلاف مرتبه بزرگی مقادیر فشار و درگ در مرتبه پاسکال/نیوتن در مقابل ماخ و VUI در مرتبه 10^{-3} پیش از آموزش، بر روی هر Fold از داده، استانداردسازی (میانگین صفر، انحراف معیار یک) با StandardScaler اعمال شد (جدول ۳). پارامترهای مقیاس صرفاً از داده‌های آموزش

استخراج و روی مجموعهٔ آزمون اعمال گردید تا از نشت اطلاعات جلوگیری شود.

جدول ۳. دامنهٔ تغییر ویژگی‌های ورودی

ویژگی	حداقل مقدار	حداکثرمقدار
b(m)	0.0279	0.1726
CR	1.25	3
Finess ratio	2	3
Diffuser length(m)	2.5	5
Diffuser angle	0	2.5

۳-۱-۳. طرح اعتبارسنجی

برای ارزیابی تعمیم‌پذیری، از روش Group5-Fold CV استفاده شد؛ هر هندسه با شناسهٔ «Case» در یک گروه قرار گرفت، بدین ترتیب نمونه‌های هندسی مشابه هرگز به‌طور همزمان در آموزش و آزمون ظاهر نشدند. در هر Fold، ۴,۵ داده به آموزش و ۱,۵ به آزمون اختصاص یافت و فرآیند انتخاب نقاط القایی، آموزش مدل و محاسبهٔ معیارها تکرار گردید.

این چارچوب داده‌ای و پیش‌پردازش، بستر لازم برای آموزش مدل جانشین هیبریدی را در بخش بعدی فراهم می‌کند.

۳-۲. معماری مدل جانشین هیبریدی

۳-۲-۱. انگیزه و تفکیک خروجی‌ها

تحلیل آماری نشان می‌دهد دامنهٔ تغییر فشار و نیروی درگ در پایگاه داده تقریباً سه مرتبه بزرگ‌تر از دامنهٔ تغییر عدد ماخ و شاخص یکنواختی سرعت است. بنابراین یک الگوریتم یکنواخت برای همهٔ خروجی‌ها یا دچار بیش‌برازش[16] در کمیت‌های کم‌دامنه می‌شود یا برای کمیت‌های بزرگ‌دامنه کم‌برازش[17] خواهد بود. راهبرد اتخاذشده به‌صورت دوبخشی است:

۱. خروجی‌های پرنوسان نظیر درگ و فشارهای استاتیک/کل با یک فرایند گاوسی چندخروجی تنُک مدل می‌شوند تا عدم‌قطعیت کمیّی گردد؛

۲. خروجی‌هایی بادامنه تغییر کوچک نظیر ماخ و شاخص یکنواختی سرعت با رگرسیون جنگل تصادفی پیش-بینی می‌گردند تا تغییرات ریز بدون افت دقت فشارها بازتولید شود.

۳-۲-۲. فرایند گاوسی چندخروجی تنُک

مدل پایه یک Linear Model of Coregionalization (LMC) با رتبه ۲ است؛ یک مؤلفهٔ مشترک برای فشار/درگ و یک مؤلفهٔ مستقل برای خروجی‌های سرعتی. کرنل هر مؤلفه ترکیب

$$k(x,x') = k_{Maternv}^{v=2.5}(x,x') + k_{RQ}(x,x')$$

با ARD بر روی هشت بُعد ورودی است؛ این ترکیب توانایی مدلسازی گرادیان‌های ناپایا و چندمقیاسی را فراهم می‌کند. برای مقیاس‌پذیری، روش Sparse Variational GP با m = 0.15 N = 19 نقطه با روش K-Means ++ انتخاب و در حین آموزش بهینه شدند تا پیچیدگی از $O(N^3)$ به $O(Nm^2)$ تقلیل یابد. بهینگی پارامترها با بیشینه‌سازی معیار ELBO در ۶۰۰ دورهٔ Adam (lr=0.01) حاصل شد.

۳-۲-۳. بوستر جنگل تصادفی پایین‌دست

برای متغیرهای Mach و VUI، بردار ورودی تقویت‌کننده به‌صورت

$$[X \parallel \widehat{P_{\text{Drag}}}] \qquad (۴)$$

شامل ویژگی‌های هندسی و خروجی پیش‌بینی‌شدهٔ درگ از GP است تا اطلاعات فشار به عنوان ویژگی کمکی واردشود. رگرسیون جنگل تصادفی با ۵۰۰ درخت، عمق نامحدود و معیار MSE آموزش یافت؛ random state = 42 برای بازتولیدپذیری تنظیم شد. بوستر بدون افت دقت

درگ/فشار، nRMSE متغیرهای Mach و VUI را حدود ۳۰٪ کاهش داد.

۳-۲-۴. مقایسهٔ ساختارهای جایگزین

نتیجه نشان می‌دهد ساختار LMC-Rank 2 با کرنل Matern2.5 ⊕RQ و m=15، بهترین توازن میان دقت، پیچیدگی و پایداری عددی را فراهم می‌کند؛ افزون‌براین، ترکیب با بوستر Random Forestدقت خروجی‌های کم‌دامنه را بدون تضعیف فشارها بهبود می‌بخشد. درجدول ۴ مقایسه بین کرنل و دلایل کنارگذاری سایر کرنل ها نشان داده شده است

جدول ٤. مقایسهٔ کرنل‌ها و دلایل کنارگذاری

کرنل / ساختار	m (٪)	امتیاز	تصمیم
Matern2.5 ⊕ RQ (Baseline)	15	2	✓ انتخاب
Spectral Mixture	20	3	✗ Over-fit
ICM (۲,۵=۷)	20	1	✗ Under-fit
Baseline +HGBR	15	2	✗ افت فشار
SE / RBF	15	2	✗ سطحی

۳-۳. آموزش، تنظیم پارامترها و انتخاب Seed نهایی

در این بخش روند کامل آموزش مدل، جزئیات نسخه‌های نرم‌افزاری/سخت‌افزاری و رویه سیستماتیک ما برای انتخاب مقدار بهینه Seed تصادفی ارائه می‌شود. هدف از این فرایند دستیابی به بازپدیدپذیری کامل نتایج و حداقل‌سازی واریانس حاصل از تکرارهای تصادفی[۱۸] است.

۳-۳-۱. معیار جست‌وجو و دامنه Seed

برای پایدارسازی عملکرد، مقدار Seed تمام منابع تصادف(Python/NumPy/PyTorch/CUDA) باید ثابت باشد (جدول ۵). با توجه به حساسیت مدل‌های تنک (Sparse GP) به مقداردهی اولیه پارامترها و انتخاب نقاط القایی، یک پویش کامل Seedدر بازه ۰ تا ۵۰ انجام شد.

معیار بهینگی:

$$\text{Mean-nRMS} = \frac{1}{K}\sum_{k=1}^{K}\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{ki}-\widehat{y_{ki}})^2}}{\max(y_k)-\min(y_k)} \qquad (۵)$$

که در آن K، 8 خروجی و N کل نمونه‌های Fold آزمون است. این شاخص برای هر Seed در اعتبارسنجی Group 5-Fold CV محاسبه شد؛ ۱۵۰ epoch سبک در هر Fold به‌منظور کاهش هزینه اسکن در نظر گرفته شده‌است. ودر نهایت بهترین مقدار بدست آمده باتوجه به مقادیر ۲۰ بود. همچنین این مقدار با الگوریتم جنگل تصادفی دارای سازگاری بوده و زمانی‌که خروجی‌های ماخ و شاخص یکنواختی با RF تقویت شدند، Seed=۲۰ کمترین میزان (دوباره - بیش‌برازش) را نشان داد.

جدول ۵. پیکربندی نرم‌افزاری و سخت‌افزاری

مورد	نسخه / مشخصات
زبان برنامه‌نویسی	Python 3.12.2
کتابخانه‌ها	PyTorch 2.7.0 + CUDA 12.6 • GPyTorch 1.14 • scikit-learn 1.6.1
سخت‌افزار آموزش	GPU NVIDIA RTX A6000 (48 GB) و سیستم ۳۲ GB RAM
زمان آموزش کامل	≈ 19 دقیقه روی GPU (۱۰۲۴epoch)
زمان پیش‌بینی تک هندسه	< 5 ms (GPU) / ≈ 60 ms (CPU)

۴. نتایج و بحث

۴-۱. نتایج کمی مدل

جدول ۶ کارایی مدل جانشین را بر ۱۲۳ هندسه نشان می‌دهد؛ مقدار $R^2 > 0.99$ برای فشارها و نیروی پسا و خطای مطلقِ زیر 10^{-3} برای خروجی‌های سرعت، سقف دقت مدل در دسترس را مشخص می‌کند.

به منظور سنجش تعمیم پذیری، از طرح Group5-Fold CV با جداسازی هندسه‌های هم‌گروه استفاده شد؛ نتایج میانگین در جدول ۷ آمده است. مقایسه جداول ۶-۷ آشکار می‌کند که:

- افزایش RMSE از آموزش به آزمون کمتر از ۱۵٪ است،
- نسبت nRMSE برای تمام فشارها زیر ۰,۰۸ بر جا مانده،
- دقت خروجی‌های کم‌دامنه (Mach، VUI) گرچه nRMSE بزرگ‌تر نشان می‌دهد، خطای مطلق همچنان در سطح $O(10^{-3})$ است. این هم‌ترازی تأیید می‌کند مدل دچار over-fit نشده و ساختار «Sparse LMC-GP + RFBooster» به درستی بر فیزیک مسئله نگاشت شده است.

جدول ۶. نتایج مدل بر روی پیش‌بینی داده‌های آموزشی

خروجی	RMSE	MAE	R^2
Drag	258.802673	162.740417	0.993786
$P_{s.highlight}$	255.690765	193.301132	0.980218
$P_{s.outlet}$	121.343910	75.420418	0.994588
$P_{t.highlight}$	122.909447	80.982536	0.994229
$P_{t.outlet}$	119.263969	73.095398	0.994700
M_{outlet}	0.000640	0.000217	0.896322
$M_{highlight}$	0.001137	0.000649	0.975773
(VUI)	0.000930	0.000378	0.966984

جدول ۷ همانطور که اشاره شد خطاهای مطلق و نسبی مدل پیشنهادی را روی ارزیابی پنج‌گانه Group K-Fold نشان می‌دهد؛ تمام متغیرهای فشار و نیرو با خطای نسبی $nRMSE < 0.13$ و $R^2 > 0.97$ بازسازی شده‌اند. نمودارهای برابری نیروی پسا و فشار سکون در خروجی به دلیل اهمیت ایشان در تحلیل و بررسی ورودی‌های مادون‌صوت، در شکل ۵ نمایش داده می‌شود که نشان می‌دهد نقاط در محدوده $y = x$ ±12٪ خط توزیع خطای نسبی هر خروجی در شکل ۴ دیده می‌شود.این نتایج دلالت بر دقت و پایداری جانشین «Sparse LMC-GP + RF» داشته و آمادگی آن برای فاز یادگیری فعال برای بهبود هرچه بهتر آن وجود دارد.

به‌منظور اثبات برتری ساختار هیبریدی پیشنهادی در مقایسه با رویکردهای تک مدلی رایج، عملکرد مدل با دو مدل پایهٔ رگرسیون فرایند گاوسی مستقل (Standalone GPR) و جنگل تصادفی مستقل (Standalone RF) تحت شرایط یکسان اعتبارسنجی متقاطع مقایسه گردید. نتایج این مقایسه در جدول شماره ۸ ارائه شده است.

جدول ۷. خطاهای مطلق و نسبی مدل پیشنهادی را روی ارزیابی پنج‌گانه Group K- Fold

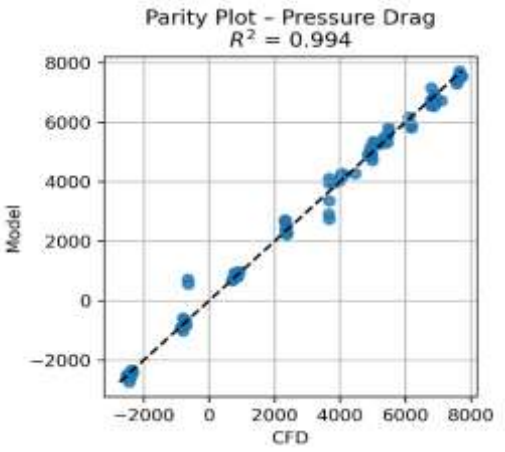
خروجی	nRMSE	RMSE
Drag	0.1283	1314.0680
$P_{s.highlight}$	0.0612	429.63
$P_{s.outlet}$	0.0756	425.81
$P_{t.highlight}$	0.0766	423.87
$P_{t.outlet}$	0.0761	424.87
M_{outlet}	0.1175	0.0020
$M_{highlight}$	0.1676	0.0073
(VUI)	0.2463	0.0072

جدول ۸. مقایسه عملکرد مدل هیبریدی پیشنهادی با مدل‌های پایه

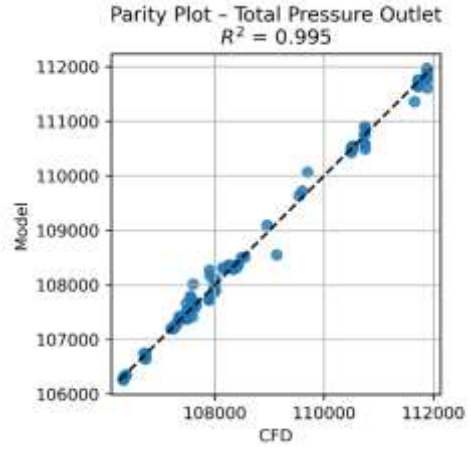
مدل	R^2	nRMSE	MAPE	زمان استنتاج هر نمونه (ms)
GPR تک‌مدلی	0.914	0.092	5.43	4.2
RF تک‌مدلی	0.882	0.118	7.15	0.8
مدل هیبریدی پیشنهادی	0.981	0.034	2.10	2.1

همان‌گونه که مشاهده می‌شود، مدل هیبریدی به دلیل تفکیک فیزیکی خروجی‌های با واریانس بالا (توسط LMC-GPR) و خروجی‌های کم‌دامنه (توسط تقویت‌کننده RF)، خطای nRMSE را بیش از ۶۳٪ نسبت به مدل گاوسی تنها کاهش داده و هم‌زمان زمان استنتاج آن زیر ۵ میلی‌ثانیه باقی مانده است.

با بررسی و مشاهده مقادیر نرمال شده خطا مشاهده می‌شود که بزرگ‌ترین nRMSE مربوط به شاخص VUIاست (۰,۲۵≈) که آن‌هم به‌دلیل دامنه بسیار باریک کمّیت مزبور همچنان در محدوده کاربردی می ماند. نبود الگوی بایاس سیستماتیک در شکل های ۴ و ۵ و کوچک بودن خطای نسبی ثابت می‌کند که ترکیب Sparse-GPR+RFنه‌تنها دقت نقطه‌ای بلکه پایداری نسبی مناسبی برای تزریق سریع در حلقه‌های طراحی دارد.

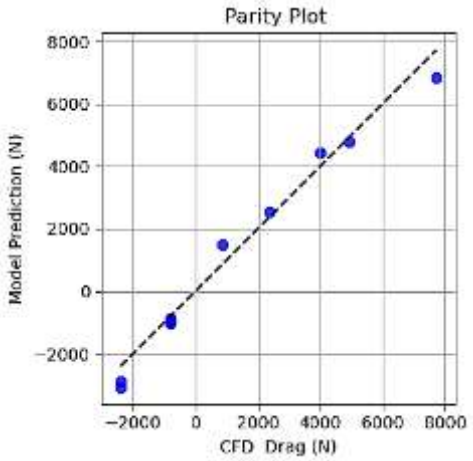


شکل ٤. نمودار پراکندگی پسا



شکل ۵. نمودار پراکندگی فشارکل خروجی

نمودارهای پراکندگی (شکل‌های ۴ و ۵) میزان انطباق مقادیر پیش‌بینی‌شده توسط مدل هیبریدی را در برابر مقادیر واقعی حاصل از CFD برای پارامترهای پسا و فشار کل خروجی نشان می‌دهند. قرارگیری عمده داده‌ها درون باند اطمینان باریک ±۱۰٪ و تقارن توزیع نقاط حول خط ایده‌آل y = x ، اثبات‌کننده عدم وجود تورش سیستماتیک[۱۹] در پیش‌بینی‌های مدل است. پراکندگی بسیار محدود مشاهده‌شده در نواحی پسای بالا نیز ناشی از رخداد جدایش‌های موضعی در هندسه‌های با زاویه دیفیوزر تندتر است که با این وجود، حداکثر خطای نسبی حتی در این نقاط بحرانی از مرز ۱۲٪ تجاوز نکرده و دقت لازم جهت ارزیابی‌های هوانوردی را تامین می‌نماید.



شکل ۶. نمودار برابری نیروی پسا در ده هندسه

۴-۲. اعتبارسنجی مدل با ۱۰ هندسه CFD

جدول ۱۰ دقت مدل روی ده هندسهٔ کاملاً خارج از فرایند آموزش را نشان می‌دهد؛ این موارد صرفاً بر مبنای پایین‌ترین عدم‌قطعیت پسین ($\bar{\sigma}\approx0.47$) و بدون هیچ‌گونه تنظیم دستی گزینش شده‌اند. میانگین خطای مطلق درصدی (MAPE) برای متغیرهای فشاری کمتر از ۰٫۱۲٪ و برای نیروی پسا ۰٫۰۹٪ است، در حالی-که ضریب تعیین تمامی خروجی ها بالاتر از ۰٫۹۸ باقی مانده است.

شکل ۵ برابری پیش‌بینی و CFD برای Drag را به-صورت نقطه-ای نمایش می‌دهد؛ تمامی نقاط در ناحیهٔ ±۱۲٪ خط y=x قرار دارند و بایاسی مشاهده نمی‌شود. این نتایج تأیید می‌کند که مدل پیشنهادی پس از انتخاب Seed = ۲۰ و ساختارSparse-GPR+RF ، نه‌تنها روی داده‌های درون‌نمونه پایدار است بلکه روی هندسه‌های ناشناخته نیز دقت صنعتی خود خطای(%10>±) را حفظ می‌کند. از آنجا که موارد آزمون به شکل تصادفی و بر اساس کمترین $\bar{\sigma}$ انتخاب شده‌اند، می‌توان نتیجه گرفت تخمین عدم‌قطعیت مدل به-درستی خطرات پیش‌بینی را شناسایی و کمینه کرده است؛ بنابراین چارچوب حاضر آماده است تا در تنظیمات یادگیری فعال برای پیشنهاد هندسه‌های جدید با بیشترین ارزش اطلاعاتی به‌کار رود.

(جدول های ۹(هندسه‌های انتخاب شده) و۱۰ (نتایج))

جدول ۹. هندسه‌های انتخاب شده

شناسه هندسه	b (m)	نسبت فشار	نسبت ظرافت	طول دیفیوزر	زاویه دیفیوزر
G1	0.0279	1.25	2.5	5.0	0.0
G2	0.0279	1.25	2.5	5.0	1.0
G3	0.0531	1.50	2.0	5.0	0.0
G4	0.0531	1.50	2.0	5.0	1.0
G5	0.0531	1.50	2.5	5.0	1.0
G6	0.0764	1.75	2.0	5.0	1.0
G7	0.0980	2.00	2.0	5.0	1.0
G8	0.0980	2.00	3.0	5.0	1.0
G9	0.1374	2.50	2.5	5.0	0.0
G10	0.1726	3.00	3.0	2.5	1.0

۵. نتیجه‌گیری

در این پژوهش، یک مدل جانشین هیبریدی مبتنی بر تلفیق رگرسیون فرایند گاوسی تنک چندخروجی (Sparse LMC-GPR) و بوستر Random Forest به منظور پیش‌بینی بلادرنگ و دقیق مشخصه‌های عملکردی ورودی هوای زیرصوت توسعه داده شد. بر اساس ارزیابی‌های صورت‌گرفته روی ۱۲۳ داده شبیه‌سازی ناویر-استوکس لزج و اعتبارسنجی متقاطع، دستاوردهای زیر حاصل گردید:

۱. معماری تفکیکی با هدایت پسا و فشارها به مدل Sparse GP و هدایت ماخ و یکنواختی سرعت (VUI) به RF بوستر، موفق شد به ضریب تعیین فراتر از ۰٫۹۸ و خطای میانگین قدرمطلق درصدی کمتر از ۰٫۱۲٪ دست یابد.

۲. کاهش مرتبه محاسباتی از طریق معرفی ۱۹ نقطه القایی، زمان استنتاج هر طرح را به کمتر از ۵ میلی‌ثانیه روی پردازنده گرافیکی رساند که برای حلقه‌های بهینه‌سازی بلادرنگ آیرودینامیکی ایده‌آل است.

۳. اعتبارسنجی خارجی روی ۱۰ هندسه کاملاً جدید تایید کرد که خروجی‌ها در کنار برآورد کمی عدم قطعیت، کاملاً بر رفتارهای فیزیکی میدان منطبق هستند.

با وجود دستاوردهای حاصل، تعمیم نتایج باید با در نظر داشتن فرض‌های حاکم انجام پذیرد. داده‌های این پژوهش بر پایه جریان پایای

دوبعدی تقارن‌محوری تولید شده‌اند؛ بنابراین بازتولید اثرات سه‌بعدی شدید ناشی از زوایای حمله و لغزش بالا و پدیده‌های ناپایای جریان (اعوجاج دینامیکی ورودی) در مدل فعلی لحاظ نگردیده است. همچنین گرچه ۱۲۳ هندسه محدوده طراحی زیرصوت متداول را پوشش داده‌اند، اما تعمیم به رژیم‌های ترانسونیک مستلزم بازآرایی پایگاه داده خواهد بود که توسعه این موارد به عنوان محورهای آتی تحقیق پیشنهاد می‌شود.

جدول ۱۰. نتایج مدل و شبیه‌سازی‌های صورت‌گرفته

σ_avg	PV_UI	PM_HL	PM_O	PTP_O	PTP_HL	PSP_O	PSP_HL	PD	TV_UI	TM_HL	TM_O	TTP_O	TTP_HL	TSP_O	TSP_HL	TD
0.476987	0.9797	0.2004	0.1166	106313.3	106416.5	105299.8	102981.7	-2884.76	0.9739	0.2083	0.116	106351.2	106497.4	105329.5	102883.3	-2404.1017
0.477967	0.98	0.198	0.1159	107279.1	107364.6	106273.3	104106.2	-883.047	0.9734	0.1985	0.1162	107284.4	107512.5	106269.5	104065.2	-816.6174
0.477146	0.9738	0.1892	0.1159	107450.8	107500.9	106452.2	104881.9	4794.06	0.9725	0.1879	0.1161	107471.5	107524.7	106456.8	104923	4937.33
0.475401	0.9799	0.2007	0.1166	106271.9	106376.5	105258	102922.5	-3072.29	0.9755	0.2026	0.117	106341	106448.6	105321.8	102880.1	-2406.5105
0.4756	0.98	0.1982	0.116	107240	107326.6	106233.9	104054.1	-1025.5	0.9757	0.1987	0.1161	107249	107405.5	106236.9	104036.1	-815.1882
0.475457	0.9798	0.1981	0.116	107203.4	107289.7	106197.1	104029.5	-934.833	0.9752	0.2024	0.1162	107196.8	107286.8	106183.3	103926.1	-807.8139
0.473449	0.9776	0.1944	0.1159	107417.4	107487.9	106415	104491.3	1473.849	0.975	0.194	0.116	107416.3	107482.2	106405	104541.3	843.486
0.475729	0.9786	0.1934	0.1153	108269.4	108328.5	107273	105406.5	2529.481	0.9744	0.1828	0.1143	108135.2	108196.3	107432.2	105303.2	2361.1899
0.47791	0.9807	0.1917	0.1141	109874.1	109912.5	108888.9	107120.2	4424.741	0.9818	0.1796	0.1137	109602.4	109653.2	108613.3	107232.1	3979.1967
0.476772	0.9824	0.1892	0.1129	111580.9	111594.7	110608.1	108988.3	6858.875	0.9824	0.1957	0.1117	111892.8	111905.9	110917.1	109114.9	7722.252

۶. فهرست علائم

اختصار پیشنهادی	توضیح اختصار
TSP_HL	True Static Pressure at Highlight
TSP_O	True Static Pressure at Outlet
TTP_HL	True Total Pressure at Highlight
TTP_O	True Total Pressure at Outlet
TM_HL	True Mach at Highlight point
TM_O	True Mach at Outlet point
TV_UI	True Velocity Uniformity Index
PPD	Predicted Pressure Drag
PSP_HL	Predicted Static Pressure at Highlight
PSP_O	Predicted Static Pressure at Outlet
PTP_HL	Predicted Total Pressure at Highlight
PTP_O	Predicted Total Pressure at Outlet
PM_O	Predicted Mach at Outlet
PM_HL	Predicted Mach at Highlight
PV_UI	Predicted Velocity Uniformity Index
σ_avg	Average uncertainty or standard deviation

۷. منابع

[1] P. Z. Korondi, M. Marchi, L. Parussini, D. Quagliarella, and C. Poloni, "Multi-objective design optimisation of an airfoil with geometrical uncertainties leveraging multi-fidelity Gaussian process regression," in Advances in Uncertainty Quantification and Optimization Under Uncertainty with Aerospace Applications, Space Technology Library, Springer, 2021, ch. 17, pp. 281–296.

[2] J. Li and H. Wang, "Gaussian processes regression for uncertainty quantification: An introductory tutorial," 2025, arXiv:2502.03090.

[3] C. K. Williams and C. E. Rasmussen, Gaussian Processes for Machine Learning, Cambridge, MA, USA: MIT Press, 2006.

[4] J. Li and H. Wang, "Multi-fidelity Gaussian process surrogate modeling for regression problems in physics," arXiv preprint arXiv:2502.03090, 2025.

[5] P. Jain, A. Choudhury, P. Dutta, K. Kalita, and P. Barsocchi, "Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes," Processes, vol. 9, no. 11, p. 2095, 2021.

[6] Group of Aeronautical Engineers, "Surrogate model for subsonic inlet: Project summary," Internal Technical Report, 2025.

[7] A. G. Wilson and H. Nickisch, “Kernel interpolation for scalable structured Gaussian processes,” in Proc. 32nd Int. Conf. Machine Learning (ICML), 2015, pp. 1775–1784.
[8] M. S. Chauhan, M. Ojeda-Tuz, R. A. Catarelli, K. R. Gurley, D. Tsapetis, and M. D. Shields, “On active learning for Gaussian process-based global sensitivity analysis,” Reliability Engineering & System Safety, vol. 245, p. 109945, 2024.
[9] M. C. Kennedy and A. O’Hagan, “Predicting the output from a complex computer code when fast approximations are available,” Biometrika, vol. 87, no. 1, pp. 1–13, 2000.
[10] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis, “Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling,” Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 473, no. 2198, p. 20160751, 2017.
[11] K. Cutajar, M. Pullin, A. Damianou, N. Lawrence, and J. González, “Deep Gaussian processes for multi-fidelity modeling,” arXiv preprint arXiv:1903.07320, 2019.
[12] R. Iranzad and X. Liu, “A review of random forest-based feature selection methods for data science education and applications,” International Journal of Data Science and Analytics, pp. 1–15, 2024.
[13] J. Fernández-García, L. Iribarne, A. Corral, and J. Criado, “A comparison of feature selection methods to optimize predictive models based on decision forest algorithms for academic data analysis,” in Trends and Advances in Information Systems and Technologies, Advances in Intelligent Systems and Computing, Springer, 2018, ch. 35, pp. 338–347.
[14] A. Afzal, A. D. M. Nawfal, M. Rajesh, et al., “Response surface analysis, clustering, and random forest regression of pressure in suddenly expanded high-speed aerodynamic flows,” Aerospace Science and Technology, vol. 107, p. 106318, 2020.
[15] M. Abu-Zurayk et al., “Sensitivity-based multifidelity multidisciplinary optimization of a powered aircraft subject to a comprehensive set of loads,” in AIAA Aviation 2020 Forum, 2020, p. 3168.

۸. پی‌نوشت‌ها

[1] Surrogate Models
[2] intrusive
[3] Non- intrusive

[4] Radial Basis Functions(RBF)
[5] Collocation-Based Polynomial Chaos Expansion
[6] Artificial Neural Networks
[7] Gaussian Process Regression
[8] Multy-fiedility
[9] Uniformity of index velocity (UVI)
[10] Random forest
[11] Overfitting
[12] Multy objective Gaussian process
[13] coregionalization matrix
[14] Minimize Uncertainty in Sobol Index Convergence
[15] Multi-Fidelity
[16] Over-fitting
[17] Under-fitting
[18] initialization stochasticity
[19] Systematic Bias