

# مقایسه روش‌های داده‌محور در ارزیابی اهمیت متغیرها و اندرکنش بین آنها در مدل‌های بارش-رواناب

سید رضی شیخ الاسلامی<sup>۱\*</sup>، سید محمد مهدی معززی<sup>۲</sup>

۱- استادیار، دانشکده مهندسی عمران، دانشگاه صنعتی شریف، تهران، ایران

۲- دانشجوی دکتری، دانشکده مهندسی عمران و محیط زیست، دانشگاه UNSW، سیدنی، استرالیا

پست الکترونیک نویسندگان:

۱- [razi.sheikholeslami@sharif.edu](mailto:razi.sheikholeslami@sharif.edu)

۲- [m.moezzi@unsw.edu.au](mailto:m.moezzi@unsw.edu.au)

## چکیده:

تحلیل حساسیت نقش مهمی در کاهش عدم قطعیت و کمک به بهبود فرایند واسنجی مدل‌ها دارد و به‌عنوان ابزاری کلیدی برای درک رفتار مدل‌ها و شناسایی پارامترهای اثرگذار آن‌ها استفاده می‌شود. مطالعه حاضر به ارزیابی عملکرد سه روش داده‌محور شامل روش جایگشتی، وابستگی جزئی و شاخص آماری H فریدمن برای تحلیل حساسیت مدل‌های بارش-رواناب می‌پردازد. در مقایسه با روش‌های سنتی، روش‌های داده‌محور با اتکا به داده‌های موجود و بدون نیاز به استراتژی‌های نمونه‌برداری ویژه، با تلاش محاسباتی کمتر می‌توانند کارایی بالایی ارائه دهند. در این مقاله، عملکرد این روش‌ها با استفاده از یک مدل ریاضی محک (تابع Sobol G) و یک مدل بارش-رواناب (HBV) ارزیابی شده است. نتایج نشان می‌دهد روش جایگشتی مبتنی بر مدل‌های یادگیری ماشین توانایی بیشتری در برآورد دقیق اهمیت و اندرکنش پارامترها، به‌ویژه در مدل‌های پیچیده و پارامترهای متعدد، دارد. همچنین برای بهبود تفسیرپذیری نتایج، انواع روش‌های گرافیکی شامل نمودار میله‌ای، شبکه‌ای، نقشه‌های حرارتی و نمودارهای وابستگی جزئی مقایسه و تحلیل شده‌اند.

## واژگان کلیدی:

مدلسازی هیدرولوژیکی، تحلیل حساسیت، روش‌های داده‌محور، اندرکنش متغیرها، نمایش بصری

# Comparison of Data-Driven Methods for Assessing Variable Importance and Interactions in Rainfall-Runoff Models

R. Sheikholeslami<sup>1</sup>, S.M.M. Moezzi<sup>2</sup>

1- Assistant Professor, Department of Civil Engineering, Sharif University of Technology, Tehran, Iran

2- Ph.D. Student, School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia

## Abstract:

Global sensitivity analysis is a key component of hydrological modeling, enabling quantification of how changes in input variables influence simulated outcomes and supporting model development, calibration, and decision-making. Conventional sampling-based approaches, such as variance-based global sensitivity analysis, have been widely applied, yet they require substantial computational effort and very large numbers of model runs. These challenges have encouraged the development of data-driven methods that rely on available data, reduce computational burden, and avoid the need for predefined sampling schemes, making them an appealing alternative in many practical modeling situations where computational resources are limited. This study compares three data-driven global sensitivity analysis algorithms: permutation importance, partial dependence, and Friedman's H-statistic. Each algorithm is described in terms of its underlying principles, advantages, and limitations for modeling variable importance and interactions. To evaluate their performance, the Sobol-G function, a widely used benchmark model, and the HBV rainfall-runoff model, a representative conceptual hydrological model, were employed, allowing assessment across both controlled mathematical settings and real-world hydrological conditions. The results indicate that permutation importance, when used with machine learning models such as random forests, typically obtains accurate ranking of influential variables and effective characterization of interaction effects, particularly in complex, nonlinear, and high-dimensional problems. These characteristics are especially important in hydrological applications, where identifying dominant drivers of model behavior is essential for reliable forecasting, uncertainty reduction, and system understanding. In addition to numerical measures, the role of visualization in global sensitivity analysis is highlighted. Visual tools such as bar charts, heatmaps, network diagrams, and partial dependence plots are described and compared, illustrating how they enhance interpretation and communication of findings by providing intuitive summaries of variable effects and interactions. However, caution is advised against over-reliance on visual representations without careful contextual examination, as misleading patterns may appear when graphs are misinterpreted or taken at face value. Overall, the study underscores the significance of data-driven sensitivity analysis as a flexible, efficient, and interpretable approach for improving hydrological modeling and decision support under uncertainty, particularly when computational constraints or model complexity limit the use of traditional sampling-based methods.

**Keywords:** Hydrologic modeling, Sensitivity analysis, Data-driven methods, Variable interaction, Visual representation

## ۱- مقدمه

بدون اینکه نیاز به اجرای مجدد مدل یا استفاده از یک استراتژی نمونه‌برداری ویژه باشد. رویکرد داده‌محور می‌تواند تحلیل حساسیت را از دو جهت بهبود بخشد: نخست، این رویکرد امکان پژوهش در حوزه‌های جدید را فراهم می‌آورد که در آن تحلیل حساسیت می‌تواند برای هر مجموعه داده‌ای اعمال شود، حتی اگر روابط و مکانیزم‌های زیربنایی آن ناشناخته باشند. دوم، این روش امکان انجام تحلیل حساسیت برای مدل‌های محاسباتی پرهزینه را با حداقل هزینه محاسباتی فراهم می‌کند. به عبارت دیگر، در این رویکرد، با در اختیار داشتن نمونه‌ای دلخواه از داده‌های ورودی-خروجی که رفتار سیستم را بیان می‌کند می‌توان معیارهای حساسیت را مستقیماً از این مجموعه داده استخراج کرد؛ بدون نیاز به اجرای مجدد مدل [۷-۳].

باید توجه داشت نیاز روزافزون به تحلیل حساسیت دقیق و قابل اعتماد در مدل‌های بارش-رواناب، اهمیت مقایسه و ارزیابی روش‌های نوین را برجسته می‌کند. همچنین، در ادبیات موجود بررسی نظام‌مند تفاوت‌ها و توانایی‌های روش‌های داده‌محور برای سنجش همزمان اهمیت متغیرها و اندرکنش آن‌ها کمتر مورد توجه قرار گرفته است.

بر این اساس، در مطالعه حاضر روش‌های داده‌محور برای ارزیابی اهمیت و اندرکنش متغیرها در تحلیل حساسیت بررسی شده‌اند. برای محاسبه اهمیت پارامترها از روش‌های جایگشتی، شاخص H فریدمن و نمودارهای وابستگی جزئی تکی استفاده شده‌است. همچنین اندرکنش دوبه‌دو با روش جایگشتی، شاخص H فریدمن و نمودارهای وابستگی جزئی دوتایی برآورد شده و در ادامه اندرکنش گروهی متغیرها نیز با روش جایگشتی ارزیابی گردید. از سوی دیگر، چالش‌های مربوط به انتخاب روش مناسب برای بصری‌سازی نتایج، به‌ویژه در حضور تعداد زیادی متغیر و تعاملات پیچیده، هنوز به‌طور جامع بررسی نشده است. با توجه به اهمیت تفسیرپذیری، روش‌های گرافیکی متداول شامل نمودارهای میله‌ای، نقشه‌های حرارتی، نمودارهای شبکه‌ای و نمودارهای وابستگی جزئی نیز بررسی شده‌اند تا نشان داده شود انتخاب درست ابزار بصری چگونه می‌تواند درک نتایج تحلیل حساسیت را بهبود بخشد.

باید توجه داشت که نمایش بصری نتایج حاصل از تحلیل حساسیت نیز نقش مهمی در درک صحیح آن‌ها دارد. انتخاب

با پیشرفت سریع فناوری‌های محاسباتی، تعداد بیشتری از مدل‌های هیدرولوژیکی مبتنی بر فرآیندهای فیزیکی برای شبیه‌سازی کمیت و کیفیت آب در مقیاس‌های مختلف مکانی و زمانی توسعه یافته‌اند. افزایش پیچیدگی این مدل‌ها به دلیل نیاز به شبیه‌سازی دقیق‌تر پدیده‌های طبیعی منجر به ادغام فرآیندهای فیزیکی متعدد با جزئیات و دقت بیشتری شده‌است [۱]. پیچیدگی بالای این مدل‌ها منجر به افزایش عدم قطعیت‌ها می‌شود که برای درک بهتر رفتار سیستم و تصمیم‌گیری‌های قابل اعتماد، باید کمی‌سازی شوند. در این راستا، تحلیل حساسیت ابزاری مؤثر برای بررسی مدل‌های پیچیده است و به چالش‌هایی نظیر تعیین اهمیت عوامل ورودی برای اولویت‌بندی و جمع‌آوری داده‌ها، شناسایی عوامل بی‌تأثیر، و مقایسه مدل‌ها و فرضیات پاسخ می‌دهد. این رویکرد، همچنین، با ارزیابی اهمیت و تعامل متغیرها، به بهبود دقت مدل‌ها، کاهش عدم قطعیت، ارتقای تصمیم‌گیری و تعمیم بهتر مدل‌ها به شرایط جدید نیز کمک می‌کند [۲].

انجام یک تحلیل حساسیت جامع بر روی مدل‌های هیدرولوژیکی معمولاً نیازمند تعداد زیادی اجرای مدل است که به منابع و هزینه محاسباتی بالایی احتیاج دارد. باید توجه داشت هنگام استفاده از روش‌های سنتی مبتنی بر نمونه‌برداری برای انجام تحلیل حساسیت، دو چالش اصلی بروز می‌کند: مشکل ابعاد بالای فضای متغیرها و ماهیت شدیداً غیرخطی پاسخ در مدل‌ها. به عبارت دیگر، برای بررسی کامل فضای متغیرها و توصیف دقیق غیرخطی بودن پاسخ، روش‌های سنتی تحلیل حساسیت که مبتنی بر نمونه‌برداری هستند، به تعداد زیادی نمونه نیاز خواهند داشت. این موضوع به‌نوبه خود می‌تواند هزینه زمانی و محاسباتی قابل توجهی را به همراه داشته باشد. علاوه بر این، یکی دیگر از نقاط ضعف اصلی روش‌های سنتی تحلیل حساسیت، وابستگی آن‌ها به یک استراتژی خاص برای نمونه‌برداری است. این وابستگی باعث می‌شود که امکان استفاده مجدد از داده‌های موجود حاصل از اجراهای قبلی، در تحلیل حساسیت وجود نداشته باشد [۲].

برای حل این مشکلات، روش‌های داده‌محور برای تحلیل حساسیت پیشنهاد شده‌است. چنین روش‌هایی مبتنی بر این ایده هستند که می‌توان اطلاعات مرتبط با حساسیت متغیرها را از یک نمونه محدود و عمومی از داده‌های موجود استخراج کرد؛

روش مناسب برای بصری‌سازی به نوع شاخص‌های حساسیت و اهداف تحلیل بستگی دارد، به‌ویژه زمانی که شاخص‌های مختلف یا مجموعه‌ای از آن‌ها باید ارائه شوند. بنابراین، نمایش گرافیکی مناسب، تفسیر نتایج را ساده‌تر کرده و امکان مقایسه دقیق‌تر و تحلیل جامع‌تری را فراهم می‌کند.

## ۲- روش انجام پژوهش

### ۲-۱- محاسبه اهمیت متغیرها به روش جایگشتی

در روش جایگشتی مراحل محاسبه اهمیت متغیر  $Z$  ام به این صورت است که ابتدا مدل یادگیری ماشین مورد نظر بر روی داده‌های موجود آموزش داده می‌شود و سپس دقت مدل، که در این مطالعه با استفاده از ضریب تعیین ( $R^2$ ) اندازه‌گیری می‌شود؛ در حالتی که متغیرها در وضعیت اولیه خود قرار دارند، محاسبه می‌شود. در مرحله بعد، داده‌های ستون متغیر  $Z$  ام به صورت تصادفی تحت جایگشت قرار می‌گیرند؛ به این معنا که ردیف‌های داده‌های همان ستون به صورت تصادفی جابجا می‌شوند. سپس، دقت مدل در این حالت نیز محاسبه می‌شود. اهمیت متغیر  $Z$  ام از اختلاف دقت اولیه و دقت بعد از جایگشت به دست می‌آید. لذا، هرچه قدر یک متغیر مهم‌تر باشد، تفاوت بیشتری در دقت مدل ایجاد خواهد کرد. شاخص حساسیت در این روش به شکل زیر تعریف می‌شود. [۶]

$$PFI_j = \frac{1}{N} \left( \sum_{i=1}^N (R^2 score_1 - R^2 score_2) \right) \quad (1)$$

در معادله ۱،  $N$  نشان دهنده تعداد جایگشت‌های مختلفی است که روی داده‌ها اعمال شده است. در واقع می‌توان برای افزایش دقت، چندین جایگشت مختلف از داده‌ها را در نظر گرفت و میانگین آنها را به عنوان عدد نهایی گزارش کرد. همچنین  $R^2 score_1$  دقت اولیه مدل و  $R^2 score_2$  دقت مدل در حالتی است که داده‌های ستون متغیر  $Z$  تحت جایگشت قرار گرفته باشند.

### ۲-۱-۱- محاسبه اندرکنش گروهی متغیرها به روش جایگشتی

برای فهم بهتر تأثیر متغیرها، ممکن است بخواهیم بدانیم که یک گروه از متغیرها، با توجه به دسته‌بندی‌های از پیش تعیین شده، چقدر مهم هستند. همچنین، اغلب پیش می‌آید که چندین متغیر تقریباً به یک اندازه مهم باشند یا بین آن‌ها هم‌بستگی

وجود داشته باشد. بنابراین، شناسایی متغیرهای مهم به‌تنهایی کافی نیست و لازم است تأثیر یا اهمیت گروهی آن‌ها نیز بررسی شود. علاوه بر این، برای مدل‌های با تعداد زیادی متغیر، تحلیل تأثیر متغیرها به‌صورت جداگانه و تصویری ممکن است بسیار دشوار باشد.

برای به‌دست آوردن میزان اندرکنش گروهی متغیرهای مختلف نیز از همان روش محاسبه اهمیت جایگشتی می‌توان استفاده کرد، با این تفاوت که این بار به جای تک ستون‌ها، ستون‌های متغیرهای هر گروه به‌صورت هم‌زمان و تصادفی جایگشت داده می‌شوند تا تأثیر و اهمیت گروه‌های از پیش تعیین شده به‌دست آید. این گروه‌بندی‌ها می‌توانند با استفاده از دانش قبلی و با توجه به ارتباطات فیزیکی یا منطقی میان متغیرها انجام شوند.

### ۲-۱-۲- مدل یادگیری ماشین استفاده شده

در این پژوهش، از مدل یادگیری ماشین جنگل تصادفی (Random Forest) به منظور پیاده‌سازی روش‌های مختلف تحلیل حساسیت استفاده شده است. مدل جنگل تصادفی یک مدل رایج در یادگیری ماشین است که نتیجه رگرسیون چندین درخت تصمیم (Decision trees) را ترکیب می‌کند که هر کدام از آنها از زیرمجموعه‌ها و ویژگی‌های تصادفی‌ای از کل داده ساخته شده است. پیش‌بینی نهایی مدل بر اساس میانگین تخمین درخت‌ها، مطابق معادله ۲ به دست می‌آید [۸].

$$\hat{f}_{rf}^B = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (2)$$

در این معادله،  $T_b(x)$  پیش‌بینی هر کدام از درخت‌های تصمیم،  $B$  تعداد کل این درخت‌ها و  $\hat{f}_{rf}^B$  مقدار پیش‌بینی نهایی مدل برای یک داده خاص است. این مدل به دلیل کارایی بالا بسیار شناخته شده است و قادر به کاهش مشکلات مربوط به بیش‌برازش و مدیریت مجموعه داده‌های بزرگ است [۹]. همچنین این الگوریتم برای مقابله با داده‌ها با نوسان زیاد و نشان دادن اهمیت متغیرها نیز مفید است [۱۰-۱۱].

برای پیاده‌سازی الگوریتم جنگل تصادفی ابتدا داده‌های مربوط به متغیرهای ورودی و متغیر خروجی، پیش‌پردازش می‌شوند که طی آن، با استفاده از مقیاس‌کننده MinMax، داده‌های تمام متغیرها بین ۰ و ۱، با توجه به معادله ۳ مقیاس بندی شدند.

$$X_{Sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

$$SS_{total} = \sum_i (y_i - \bar{y})^2 \quad (6)$$

۲-۲- محاسبه اهمیت و اندرکنش متغیرها به روش

### وابستگی جزئی

مقدار وابستگی جزئی (Partial Dependence) میانگین تغییرات در مقدار پیش‌بینی شده مدل را اندازه‌گیری می‌کند. معادله ۷، رابطه ریاضی محاسبه وابستگی جزئی برای متغیرهای  $S$  ( $X_S$ ) را نشان می‌دهد.

$$F_S = \frac{1}{n} \sum_{i=1}^n \hat{f}(X_S, X_C) \quad (7)$$

در این رابطه،  $F_S$  تابع وابستگی جزئی،  $\hat{f}$  مدل یادگیری ماشین و  $n$  تعداد ردیف‌های داده‌های استفاده‌شده در آموزش مدل است. در واقع توابع وابستگی جزئی تکی و دوتایی با استفاده از این رابطه روی هر نقطه از متغیر(های)  $X_S$  و بدون تغییر دادن متغیرهای  $X_C$  حاصل می‌شوند [۱۳]. نمودارهای وابستگی جزئی که از رابطه ۷ برای محاسبه مقادیر هر نقطه استفاده می‌کنند، رابطه بین متغیر (های) مورد نظر و متغیر هدف مدل را نشان می‌دهند.

برای مثال اگر از مدل‌های خطی مانند رگرسیون خطی استفاده شود، رابطه بین تمام متغیرها و متغیر هدف به صورت خطی نمایش داده می‌شود. همانطور که اشاره شد، یکی از مزیت‌های مدل جنگل تصادفی این است که روابط غیر خطی را هم نمایش می‌دهد و می‌تواند مدیریت کند. نمودارهای وابستگی جزئی تکی رابطه بین یک متغیر و متغیر هدف را نشان می‌دهد در صورتی که وابستگی جزئی‌های دوتایی، تاثیرات همزمان دو متغیر بر روی متغیر هدف را نمایش می‌دهند.

۲-۳- محاسبه اندرکنش متغیرها براساس شاخص آماری

### H فریدمن مرتبه دو

یکی دیگر از روش‌های محاسبه اندرکنش متغیرها که مبنای ریاضیاتی نسبتاً پیچیده‌تری دارد، روش H آماری فریدمن است. ایده بدست آوردن آن بر مبنای تابع وابستگی جزئی است که در قسمت قبلی توضیح داده شد. معادله ۸، رابطه پیشنهاد شده برای محاسبه اندرکنش‌های دوتایی با استفاده از تابع وابستگی جزئی را نشان می‌دهد.

$$H_{jk}^2 = \frac{\sum_{i=1}^n [f_{jk}(x_{ij}, x_{ik}) - f_j(x_{ij}) - f_k(x_{ik})]^2}{\sum_{i=1}^n f_{jk}^2(x_{ij}, x_{ik})} \quad (8)$$

که  $X$  نشان دهنده هر کدام از داده‌ها و  $X_{Sc}$  مقدار مقیاس شده آن داده توسط این روش است.  $X_{max}$  و  $X_{min}$  نیز به ترتیب مقدار بیشینه و کمینه داده‌ها را نشان می‌دهند.

در ادامه، کل داده‌ها در ۱۰ مرحله متفاوت، به صورت تصادفی به دو زیرمجموعه برای آموزش و آزمایش مدل تقسیم شدند که سهم داده‌های آزمایشی، ۲۵٪ از کل داده‌ها بود و این داده‌ها برای آموزش دادن مدل استفاده نشدند تا اعتبارسنجی دقیقی از مدل به دست آید. در انتها نیز، پس از به دست آوردن فرآیندهای Hyperparameters) بهینه که بهترین مدل را می‌سازند، با تطبیق مدل به داده‌ها، دقت مدل با استفاده از معیارهای مختلف، سنجیده شد.

اصلی‌ترین این فرا پارامترها، تعداد درخت‌های تصمیم مدل است که با افزایش آن عموماً مدل دقیق‌تر شده ولی هزینه زمانی برای آموزش و استفاده از مدل بیشتر می‌شود. علاوه بر آن، فرآیند عمق که نشان دهنده عمق بیشینه درخت‌های تصمیم یک مدل است که افزایش این مقدار، توانایی مدل‌سازی بهتر شرایط پیچیده داده‌ها را می‌دهد اما احتمال بیش‌برازش را نیز افزایش می‌دهد. [۱۲]. مقدار نهایی این فرآیندها که با استفاده از روش GridSearchCV بدست آمد و منجر به ساختن بهترین مدل شده بودند، در جدول (۱) آمده است.

جدول ۱: فرآیندهای بهینه مدل جنگل تصادفی نهایی

مقدار	فرا پارامتر
۵۲	حالت تصادفی
۱۰۰	تعداد درخت‌های تصمیم
۲۵	عمق بیشینه هر درخت

در این مطالعه، معیار اصلی اندازه‌گیری دقت مدل، ضریب تعیین یا همبستگی است که هرچه قدر میزان این معیار به عدد یک نزدیک‌تر باشد، همبستگی بین دو سری داده بیشتر است و نشان‌دهنده بالاتر بودن دقت مدل در پیش‌بینی خواهد بود. در معادلات زیر، اگر  $\gamma$  نشان‌دهنده هر کدام از داده‌های مشاهده‌شده یا داده‌های واقعی باشد،  $f$  نشان‌دهنده داده پیش‌بینی شده،  $n$  تعداد داده‌های استفاده‌شده برای اعتبارسنجی و  $\bar{\gamma}$  میانگین کل داده‌های مشاهده‌شده باشد، رابطه ریاضی محاسبه ضریب تعیین به شکل زیر است [۱۲].

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad (4)$$

$$SS_{residual} = \sum_i (y_i - f_i)^2 \quad (5)$$

## ۲-۴-۳- نقشه‌های حرارتی

نقشه‌های حرارتی یک نمایش گرافیکی از داده‌ها هستند که با استفاده از طیف رنگی، تغییرات مقادیر را نشان می‌دهند. در تحلیل حساسیت، این نمودارها برای نمایش توزیع اهمیت و اندرکنش متغیرها در یک نگاه کلی به کار می‌روند. مزیت اصلی نقشه‌های حرارتی این است که به‌سادگی می‌توان الگوها و روندهای مهم را در میان تعداد زیادی متغیر شناسایی کرد. رنگ‌های مختلف در این نقشه‌ها نشان‌دهنده‌ی شدت اهمیت یا اندرکنش متغیرها هستند که امکان تحلیل سریع و آسان تغییرات را فراهم می‌کنند. این روش به‌خصوص برای نمایش تعاملات در سطوح چندبعدی و پیچیده بسیار مفید است.

## ۲-۴-۴- نمودارهای وابستگی جزئی

نمودارهای وابستگی جزئی روشی قدرتمند برای تحلیل چگونگی تأثیر یک متغیر (وابستگی جزئی تکی) یا دو متغیر (وابستگی جزئی دوتایی) بر خروجی مدل هستند، در حالی که سایر متغیرها ثابت نگه داشته می‌شوند. این روش، به‌ویژه در مدل‌های پیچیده و غیرخطی که روابط میان متغیرها چندلایه و غیرقابل پیش‌بینی است، کمک شایانی به درک تأثیرات منفرد یا دوتایی متغیرها می‌کند. در این نمودارها، محور افقی نشان‌دهنده‌ی مقادیر متغیر ورودی و محور عمودی نشان‌دهنده‌ی تأثیر آن بر متغیر خروجی است. این نمودارها به‌طور ویژه برای بررسی رفتارهای غیرخطی مدل مفید هستند و بینش دقیقی از نقش هر متغیر در خروجی نهایی ارائه می‌دهند.

## ۳- مدل‌های عددی استفاده شده

### ۳-۱- تابع ریاضی Sobol G

تابع Sobol G یک تابع محک معروف برای بررسی عملکرد الگوریتم‌های تحلیل حساسیت است. طبیعت غیرخطی و غیر یکنواخت بودن این تابع به همراه نتایج تحلیلی آنالیز حساسیت که در اختیار می‌گذارد، آن را به یک مدل ریاضی مناسب برای بررسی تحلیل حساسیت تبدیل کرده است. متغیر خروجی این تابع از حاصل ضرب متغیرهای ورودی مختلف در یکدیگر طبق معادله ۹ محاسبه می‌شود.

$$G = \prod_{i=1}^l \frac{(|4x_i - 2| + a_i)}{1 + a_i} \quad (9)$$

در این معادله،  $f_j(x_{ij})$  و  $f_k(x_{ik})$  توابع وابستگی جزئی تکی هر کدام از متغیرهای  $k$  و  $j$  زاند و  $f_{jk}(x_{ij}, x_{ik})$  تابع وابستگی جزئی دو متغیر است. مقدار عدد حاصل‌شده از معادله بالا، به عنوان اندرکنش دوتایی متغیرهای مختلف استفاده می‌شود که  $H$  آماری مرتبه دوم نرمال شده است [۱۳].

در صورتی که دو متغیر هیچ اندرکنشی نداشته باشند، می‌توان تابع وابستگی جزئی آن‌ها را به‌صورت مجموع توابع وابستگی جزئی هر یک از متغیرها نشان داد. به عبارت دیگر، اگر بین دو متغیر تعامل خاصی وجود نداشته باشد، اثر هر متغیر بر پیش‌بینی مدل به‌صورت مستقل از دیگری خواهد بود. شاخص  $H$  میزان انحراف از این فرض را می‌سنجد؛ هرچه مقدار این شاخص به صفر نزدیک‌تر باشد، اندرکنش کمتری بین متغیرها وجود دارد و هرچه به یک نزدیک‌تر باشد، نشان‌دهنده اندرکنش قوی‌تری بین متغیرهاست. در برخی موارد، مقدار شاخص  $H$  می‌تواند بزرگ‌تر از یک نیز باشد که به تعاملات پیچیده‌تر بین متغیرها اشاره دارد.

## ۲-۴-۲- روش‌های بصری نمایش نتایج تحلیل حساسیت

### ۲-۴-۲-۱- نمودارهای میله‌ای

نمودارهای میله‌ای یکی از متداول‌ترین روش‌ها برای نمایش اهمیت متغیرها در تحلیل حساسیت هستند. این نمودارها این امکان را فراهم می‌سازند تا به‌سادگی متغیرها را بر اساس میزان تأثیرشان مرتب کرد. با استفاده از نمودارهای میله‌ای، می‌توان تأثیر هر متغیر بر روی خروجی مدل را به‌صورت دقیق مشاهده و متغیرهای با اهمیت بیشتر را به وضوح از بقیه تفکیک کرد. این نمودارها به‌خصوص برای تحلیل ساده از نتایج و اولویت‌بندی متغیرها مناسب هستند و به تصمیم‌گیری بهتر در مورد اصلاح مدل کمک می‌کنند.

### ۲-۴-۲-۲- نمودارهای شبکه‌ای

نمودارهای شبکه‌ای برای نمایش روابط پیچیده میان متغیرها استفاده می‌شوند. این نمودارها، به‌ویژه زمانی که متغیرهای زیادی با یکدیگر در تعامل هستند، اهمیت دارند. با استفاده از گره‌ها و یال‌ها، می‌توان تعاملات متقابل متغیرها را نمایش داد. هر گره نشان‌دهنده اهمیت یک متغیر است و خطوط ارتباطی بین گره‌ها (یال‌ها)، نشان‌دهنده‌ی اندرکنش بین آن‌هاست. این نوع نمودارها کمک می‌کنند تا ساختار پیچیده شبکه‌ای از متغیرها و اثرات متقابل آن‌ها را به‌خوبی مشاهده و درک کرد.

- زمانی بسیار متفاوتی داشته باشند. تغییرات دمایی در این حوضه گسترده است؛ به طوری که دما در زمستان‌ها تا منفی ۳۰ درجه و در تابستان‌ها تا حدود مثبت ۳۰ درجه می‌رسد. بیش از نیمی از بارش سالانه به صورت برف رخ می‌دهد و ذخیره برف در زمستان بخش عمده ورودی آبی حوضه در فصل ذوب را تشکیل می‌دهد. بنابراین فرآیندهای برف‌گیری، ذوب برف و یخ و همچنین توزیع ارتفاعی دما و بارش نقش تعیین‌کننده‌ای در زمان‌بندی و شدت رواناب در رود Bow دارند.

افزون بر این، ویژگی‌هایی مانند شیب زیاد، وجود لایه‌های نفوذپذیر و حضور نواحی با خاک کم‌عمق بر رفتار هیدرولوژیکی حوضه اثرگذار هستند و باعث می‌شوند حساسیت برخی پارامترهای مدل HBV، از جمله پارامترهای مرتبط با ذخیره برف، تلفات اولیه، و ضریب رواناب سطحی، نسبت به سایر پارامترها بیشتر باشد [۴].

#### ۴- نتایج عددی

##### ۴-۱- نتایج دقت مدل یادگیری ماشین

بعد از تطبیق مدل بر روی داده‌های جدا شده برای آموزش آن، دقت پیش‌بینی مدل با استفاده از داده‌های مخصوص آزمایش، با استفاده از معیار ضریب تعیین ( $R^2$ ) محاسبه شد. همانطور که گفته شد، برای بررسی بهتر دقت مدل، ده بار داده‌های آموزشی و آزمایشی تصادفی مختلفی به مدل داده شد (10-fold cross validation) تا قدرت پیش‌بینی آن در ترکیب‌های مختلف سنجیده شود.

در مطالعه حاضر، مدل یادگیری ماشین، در هر دو حالت با ۱۰۰۰۰ و ۵۰۰۰۰ داده مورد استفاده، دقت بسیار بالایی داشته‌است بطوریکه میانگین  $R^2$  به دست آمده برای مدل برازش شده به داده‌های مدل HBV-SASK، برابر با ۰/۹۸ و انحراف معیار آن ۰/۰۱۵ شد. همچنین، مدل برازش شده به داده‌های تابع Sobol G نیز دقت ۰/۹۰ با انحراف معیار ۰/۰۰۷۳ را داشت. این مقادیر مشخص می‌کند که مدل توانسته به خوبی رابطه بین متغیرهای ورودی و خروجی را تشخیص دهد. این دقت بالای مدل تطبیق داده‌شده، لازمه تحلیل حساسیت و محاسبه دقیق اهمیت و اندرکنش هاست.

در این تابع، هر چقدر که ضریب  $a_i$  مربوط به متغیر  $i$  ام کوچکتر باشد، با توجه به اینکه مقدار این ضریب در مخرج قرار گرفته است، تاثیر آن متغیر روی خروجی بیشتر می‌شود. [۱۴-۱۵]

در مطالعه حاضر جهت بررسی دقت و عملکرد روش‌های داده‌محور در تشخیص اهمیت متغیرها، ابتدا ۱۵ متغیر که ۱۰۰۰۰ داده تصادفی بین ۰ و ۱ برای هر کدام تولید شد را به عنوان متغیرهای ورودی تابع ریاضی Sobol G قرار دادیم و مقدار متغیر خروجی را طبق معادله ۹ محاسبه کردیم. در ادامه، همان متغیرهای ورودی را نیز به عنوان ورودی مدل یادگیری ماشین و خروجی تابع Sobol G را به عنوان متغیر هدف آن در نظر گرفتیم تا مدل رگرسیونی جنگل‌های تصادفی روی داده‌ها برازش شود.

جدول (۲) مقدار پارامتر  $a_i$  را برای متغیرها مختلف استفاده‌شده در مدل نشان می‌دهد. بنابراین طبق لیست ضرایب، اگر  $k < p$ ، آنگاه انتظار داریم تا مدل یادگیری ماشین هم اهمیت متغیر  $x_k$  را بیشتر از اهمیت متغیر  $x_p$  نشان دهد.

جدول ۲: مقادیر ضرایب هر متغیر در تابع Sobol G همراه با گروه بندی

گروه	مقدار ضریب $a_i$
بسیار تاثیر گذار	$\{a_1, a_2, a_3\} = \{0, 1, 2\}$
تاثیر گذار	$\{a_4, a_5, a_6, a_7\} = \{4, 6, 8, 10\}$
کم تاثیر	$\{a_8, a_9, a_{10}, a_{11}\} = \{15, 20, 25, 30\}$
بدون تاثیر	$\{a_{12}, a_{13}, a_{14}, a_{15}\} = \{50, 60, 70, 80\}$

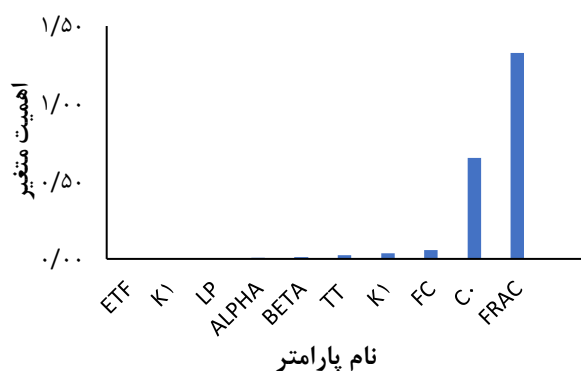
##### ۳-۲- مدل بارش-رواناب HBV-SASK

مدل HBV برگرفته از نام Hydrologiska Byråns Vattenbalansavdelning است که توسط Lindström و همکاران برای اولین بار به منظور کارهای تحقیقاتی و آموزشی توسعه داده شد [۱۶]. این مدل ورودی‌های ده‌گانه‌ای دارد که در انتها می‌توان به وسیله این مدل، دبی و جریان آب خروجی از یک حوضه آبریز را شبیه‌سازی کرد. ۱۰ پارامتر اصلی این مدل، تعاریف آنها و مقادیر ممکن آنها در جدول (۳) آمده است [۱۷].

در این مطالعه، مدل HBV-SASK برای شبیه‌سازی دبی خروجی رودخانه Bow، واقع در شهر Banff در استان Alberta کانادا، استفاده شد. این حوضه آبریز با مساحت تقریبی ۲۲۱۰ کیلومتر مربع در بخش جنوب‌غربی استان و در محدوده کوه‌های راکی قرار دارد. موقعیت کوهستانی حوضه و اختلاف ارتفاع قابل توجه آن سبب می‌شود الگوهای بارش، دما و رواناب مکانی

جدول ۳: نام، حدود و توضیحات پارامترهای اصلی مدل HBV-SASK

نام پارامتر مدل	حد پایین	حد بالا	توضیحات
TT	-۴	۴	حد دما که بر حسب °C برای ذوب یا انجماد که مرز بین باران و برف است.
C0	۰	۱۰	ضریب انجماد پایه، بر حسب mm/°C بر روز
ETF	۰	۱	تصحیح ناهمگنی دما در تبخیر و تعرق بالقوه بر حسب 1/°C
LP	۰	۱	حد برای تبخیر و تعرق بالقوه روزانه (PET) به عنوان ضریب پارامتر FC
FC	۵۰	۵۰۰	ظرفیت خاک، بر حسب میل متر. حداکثر آبی که خاک می تواند نگه دارد.
BETA	۱	۳	پارامتر شکل. توان معادله آزادسازی خاک. پارامتر بدون بعد
FRAC	۰/۱	۰/۹	نسبت مقدار آزاد سازی که وارد مخزن سریع می شود. پارامتر بی بعد
K1	۰/۰۵	۱	ضریب مخزن سریع. تعیین می کند چه نسبتی از ذخیره در هر روز آزاد می شود. پارامتر بی بعد
ALPHA	۱	۳	پارامتر شکل. توان معادله محاسبه خروجی مخزن سریع. پارامتر بدون بعد
K2	۰	۰/۰۵	ضریب مخزن کند. تعیین می کند چه نسبتی از ذخیره در هر روز آزاد می شود. پارامتر بی بعد



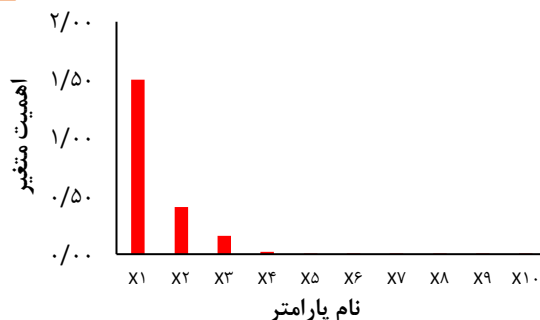
شکل ۲: نمودار میله‌ای اهمیت پارامترهای مختلف مدل HBV-SASK

### ۳-۴ نتایج اندرکنش بین متغیرها

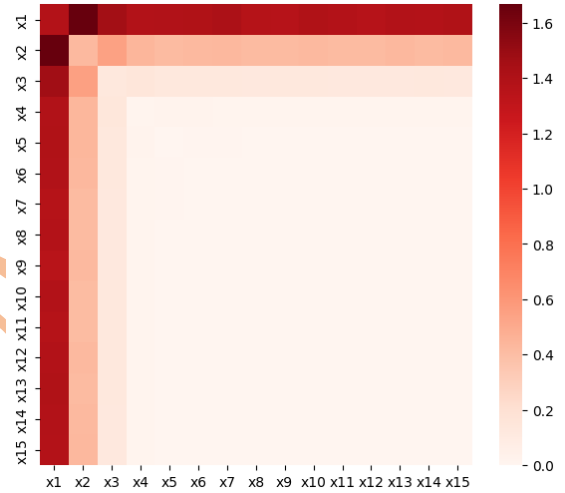
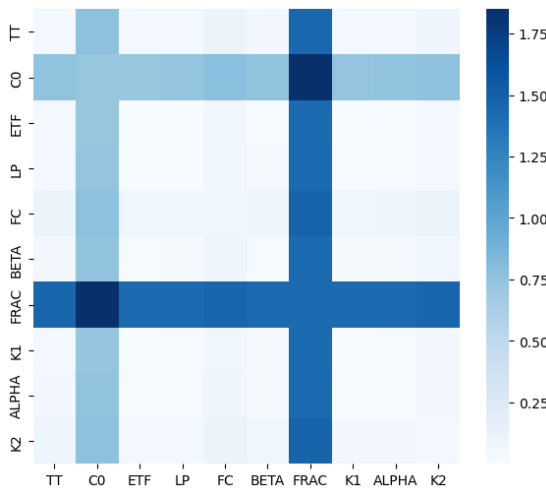
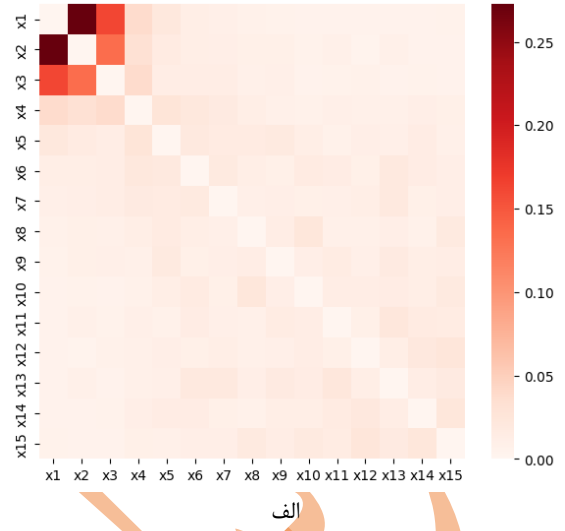
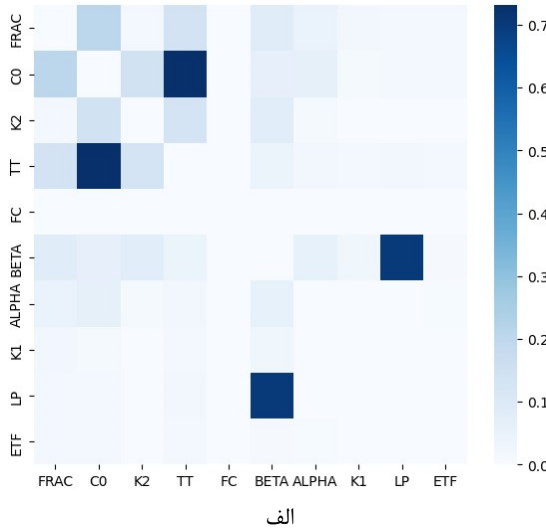
در ادامه، اندرکنش بین پارامترهای مدل عددی از دو روش جایگشتی و H آماری مرتبه دوم به دست آمد. نمودارهای نقشه حرارتی (Heatmap) در شکل‌های (۳ و ۴)، اندرکنش بین متغیرهای مدل Sobol G و پارامترهای مختلف مدل عددی HBV-SASK را با ۱۰۰۰۰ ردیف داده نشان می‌دهند. نتایج مربوط به اندرکنش بین متغیرها به روش جایگشتی روش H فریدمن برای مدل Sobol G که در شکل (۳) نشان داده شده است. همانطور که دیده می‌شود، متغیرهایی که اهمیت تکی بالایی دارند، روی اندرکنش بین متغیرها نیز به مقدار زیادی تاثیر گذاشته‌اند. همچنین، برای مدل HBV-SASK شکل (۴-الف) نتایج اندرکنش بین پارامترها به روش H فریدمن با ۱۰۰۰۰ ردیف داده و شکل (۴-ب) نتایج اندرکنشها به روش جایگشتی را با همان داده‌ها نشان می‌دهد.

### ۲-۴ نتایج اهمیت متغیرها به روش جایگشتی

بعد از برازش مدل یادگیری ماشین روی داده‌های تولید شده برای مدل Sobol G، مقدار اهمیت متغیرهای آن به روش جایگشتی محاسبه شد. طبق نتایج به دست آمده در شکل (۱)، روش جایگشتی مبتنی بر جنگل‌های تصادفی توانسته اهمیت متغیرها را متناسب با تاثیری که بر مدل می‌گذارند به درستی مطابق با ضرایب ارائه شده در جدول ۲ محاسبه کند.



شکل ۱: اهمیت متغیرهای مختلف به دست آمده برای تابع Sobol G در آزمون دوم، پارامترهای ورودی مدل HBV-SASK نیز به عنوان متغیرهای ورودی مدل یادگیری ماشین و دقت محاسبه شده (NSE) HBV-SASK به عنوان متغیر هدف مدل یادگیری ماشین تنظیم شدند. شکل (۲)، اهمیت پارامترهای مختلف مدل HBV-SASK را که با استفاده از روش جایگشتی مبتنی بر جنگل‌های تصادفی به دست آمده نشان می‌دهد. طبق این نمودار، پارامترهای FRAC و C0 بیشترین تاثیر و پارامترهای ETF، K1 و LP کمترین تاثیر را روی دقت مدل دارند.



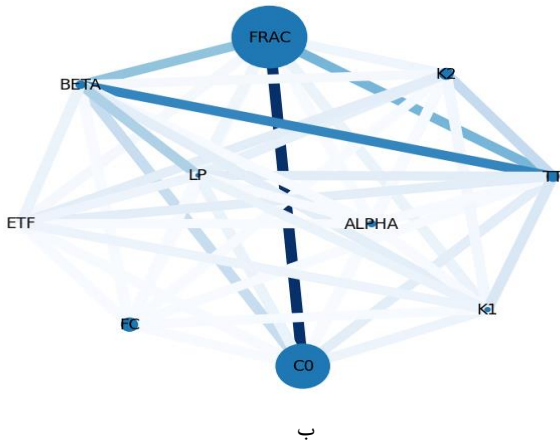
شکل ۴: نتایج اندرکنش بین متغیرها در مدل HBV-SASK با ۱۰۰۰۰ داده (الف) روش H فریدمن. (ب) روش جایگشتی

شکل ۳: نتایج اندرکنش بین متغیرها در تابع Sobol G با ۱۰۰۰۰ داده (الف) روش H فریدمن. (ب) روش جایگشتی

#### ۴-۴- نمودارهای شبکه‌ای

برای سهولت تفسیر و دید بهتر نسبت به اهمیت‌ها و اندرکنش (دو به دوی) بین متغیرها می‌توان از نمودارهای شبکه‌ای مطابق شکل (۶) استفاده کرد. در این شکل‌ها، بزرگی هر گره میزان اهمیت متغیر را نشان می‌دهد و هر چه اندازه آن بزرگتر باشد اهمیت متغیر متناظر آن هم بیشتر است. همچنین رنگ خطوط هر چه تیره‌تر باشد، اندرکنش آن دو متغیر نیز بیشتر بوده‌است.

در آزمون بعدی، هریک از مدل‌ها با ۵۰۰۰۰ نمونه اجرا شد. شکل (۵)، نتایج اندرکنش بین متغیرها با ۵۰۰۰۰ داده را برای هر دو روش در مدل HBV-SASK نشان می‌دهد. طبق شکل (۵)، مشاهده می‌شود روش H فریدمن، با افزایش تعداد داده‌ها نتایج متفاوتی را تولید کرده‌است. درحالی که در نتایج روش جایگشتی تغییری بوجود نیامده‌است. نکته مهم دیگر این است که علاوه بر عدم ثابت‌ماندن نتایج در روش H آماری، این روش هزینه محاسباتی بسیار بیشتری نسبت به روش جایگشتی دارد. لازم به ذکر است، نتایج به‌دست‌آمده از روش جایگشتی، تطابق بسیار بیشتری با نتایج حاصل‌شده از روش وابستگی جزئی که در ادامه بررسی می‌شود دارد.



شکل ۶: نمودارهای شبکه‌ای حاصل از نتایج روش H فریدمن برای الف) تابع Sobol G. ب) مدل HBV-SASK

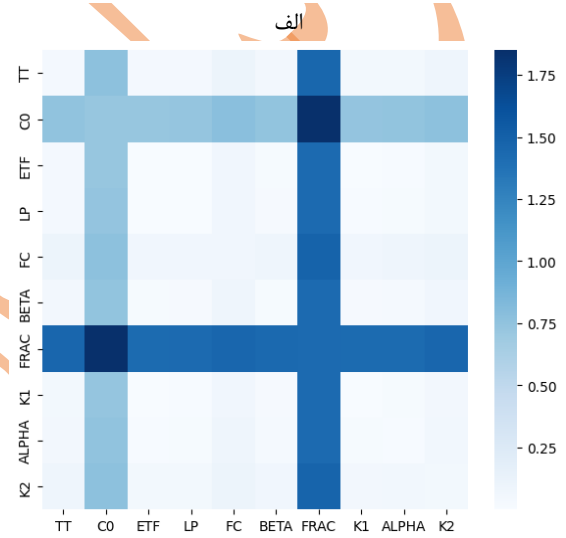
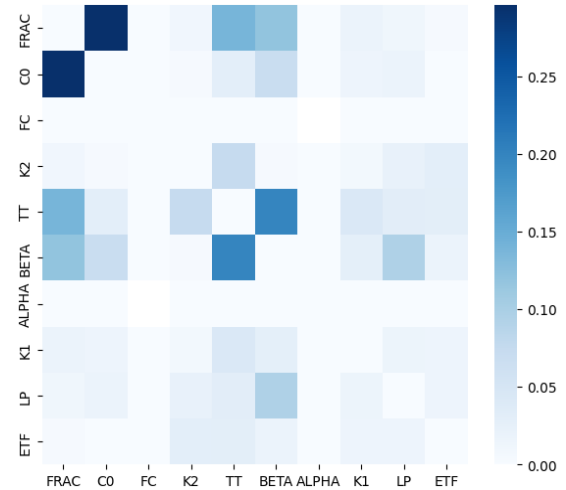
#### ۴-۵- نمودارهای وابستگی جزئی تک

نمودارهای وابستگی جزئی، از محاسبه معادله ۷ به دست می‌آیند که نشان‌دهنده اهمیت متغیرها و اندرکنش بین آنها خواهند بود. شکل (۷)، نمودارهای وابستگی جزئی دو متغیر ( $x_6$  و  $x_1$ ) از تابع Sobol G را نشان می‌دهد. علاوه بر نوع تاثیرگذاری این دو متغیر بر متغیر خروجی، مقدار تاثیرگذاری هر یک با توجه به بازه محور عمودی نیز مشخص می‌شود.

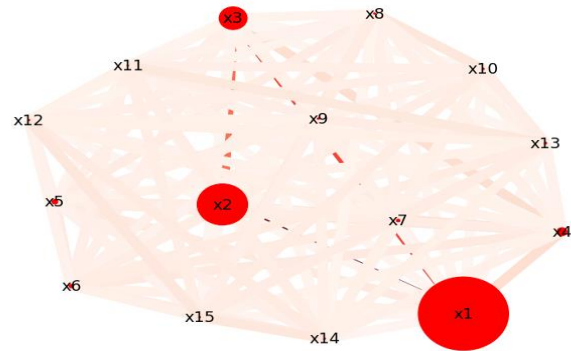
شکل‌های (۸ و ۹) نیز نمودارهای وابستگی جزئی تک‌متغیره برای چهار متغیر FRAC، CO، BETA و ETF از مدل HBV-SASK را نشان می‌دهند. محور عمودی این نمودار، بیانگر تغییرات متغیر هدف، یعنی دقت مدل عددی HBV-SASK است.

بزرگتر بودن بازه تغییرات محور قائم در هر شکل، نشان‌دهنده تاثیر بیشتر بر روی متغیر هدف است. برای مثال، متغیر FRAC که مهم‌ترین متغیر مدل است، بیشترین بازه تغییرات را روی متغیر هدف اعمال کرده و از آن طرف متغیر ETF که کمترین اهمیت را دارد، کمترین بازه تغییرات متغیر هدف را نیز دارد.

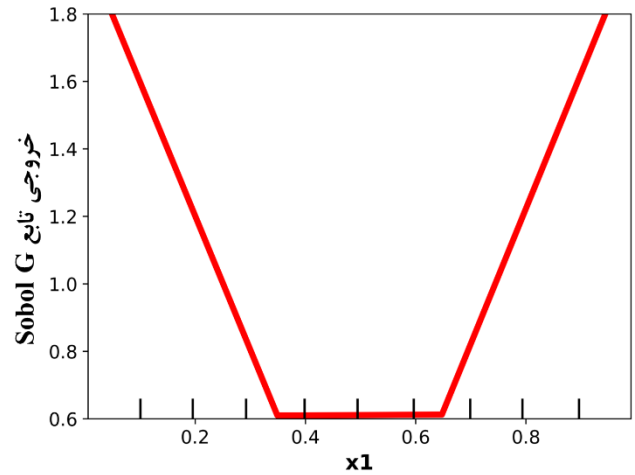
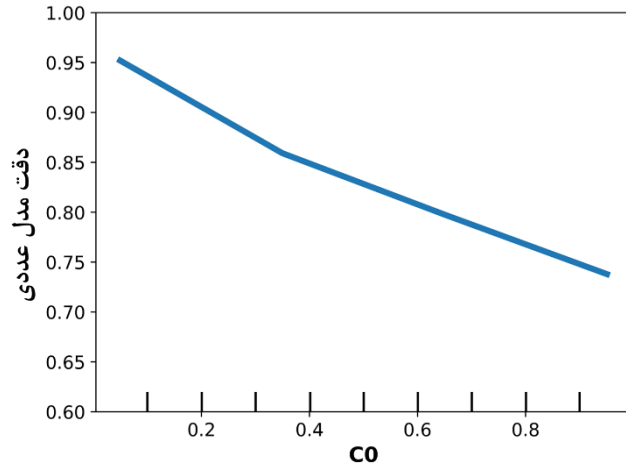
همچنین، روند تغییرات نمودار (صعودی، نزولی، خطی و ...)، نوع تاثیرگذاری پارامتر را نشان می‌دهد. در شکل (۸) روند افزایشی متغیرهای مهم تاثیر منفی بر دقت مدل داشته است.



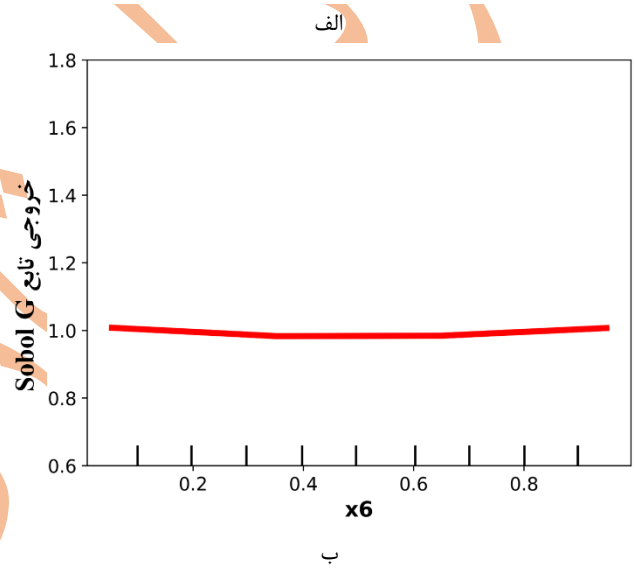
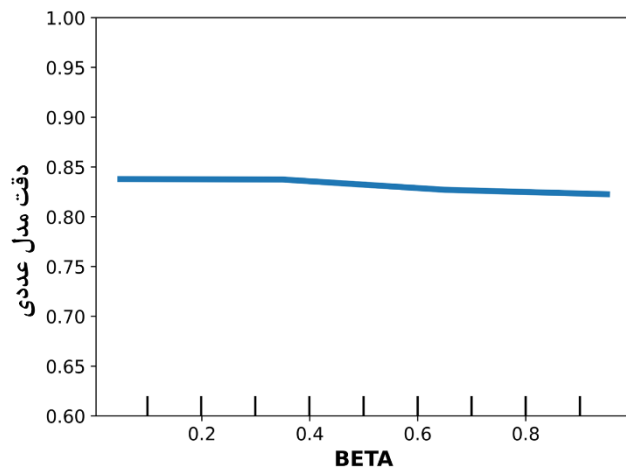
شکل ۵: نتایج اندرکنش متغیرهای مدل HBV-SASK با ۵۰۰۰۰ داده. الف) روش H فریدمن. ب) روش جایگشتی



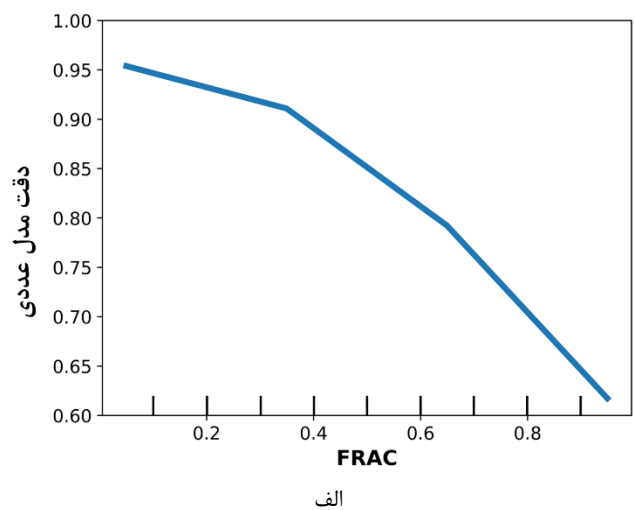
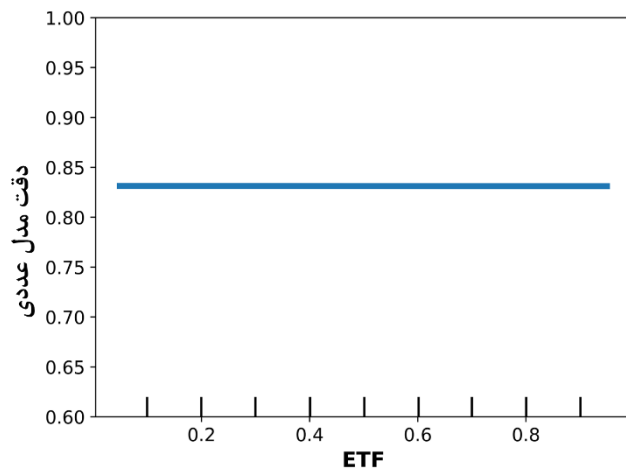
الف



شکل ۸: نمودارهای وابستگی جزئی در مدل HBV-SASK برای پارامترهای بسیار تاثیرگذار الف (FRAC ب) C0



شکل ۷: نمودار وابستگی جزئی در تابع Sobol G برای الف متغیر بسیار تاثیرگذار (x1). ب) متغیر کم اهمیت (x6)

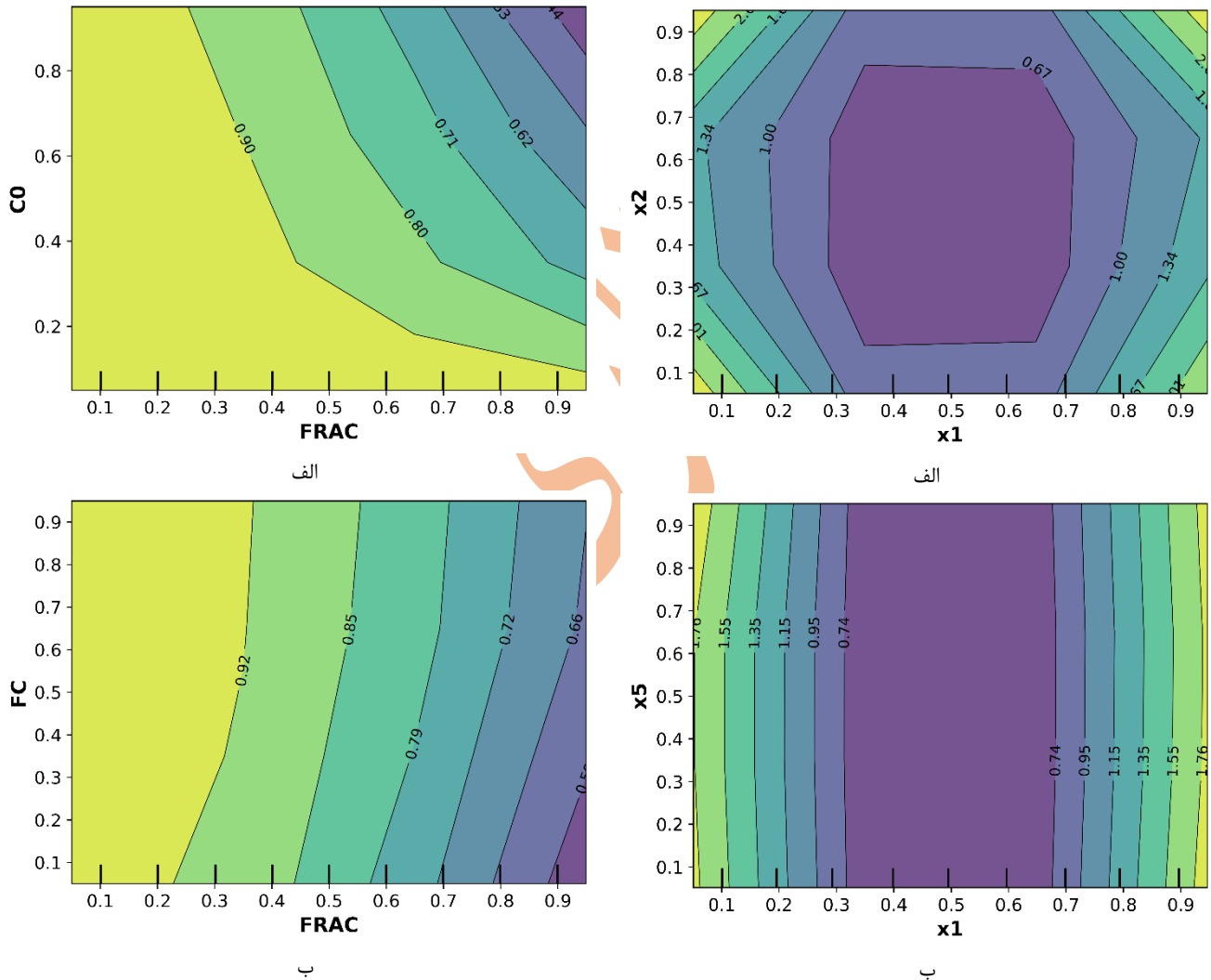


شکل ۹: نمودارهای وابستگی جزئی در مدل HBV-SASK برای پارامترهای کم اهمیت الف (BETA ب) ETF

#### ۴-۶- نمودارهای وابستگی جزئی دوتایی

همانطور که گفته شد، این نمودارها میزان اندرکنش دو به دو متغیرها را در بازه تغییراتشان نشان می‌دهند. هرچقدر اعداد روی خطوط، بازه تغییرات بیشتری در اثر تغییر هم‌زمان متغیرها داشته باشند، مقدار اندرکنش آن دو متغیر بیشتر است. همچنین، جهت خطوط نشان می‌دهد که اثر اندرکنش بین دو متغیر بیشتر در جهت تغییرات کدام متغیر است. در شکل (۱۰)، نمودار وابستگی جزئی دوتایی متغیرهای اول و دوم، و اول و پنجم برای تابع Sobol G رسم شده است.

نشان می‌دهد. طبق شکل (۱۱-الف) مشاهده می‌شود که FRAC (تأثیر گذارترین پارامتر) بیشترین اندرکنش را با متغیر مهم بعدی یعنی C0 دارد. خطوط کانتور در نمودارهای وابستگی جزئی FRAC با BETA و ETF تقریباً عمود به محور FRAC است که نشان می‌دهد تغییرات FRAC تأثیر اصلی را روی دقت مدل داشته و سایر پارامترها تأثیر چندانی ندارند. لازم به ذکر است، تطابق زیادی بین نتایج اندرکنش‌های محاسبه شده با روش جایگشتی و نتایج نمودارهای وابستگی جزئی می‌توان دید.

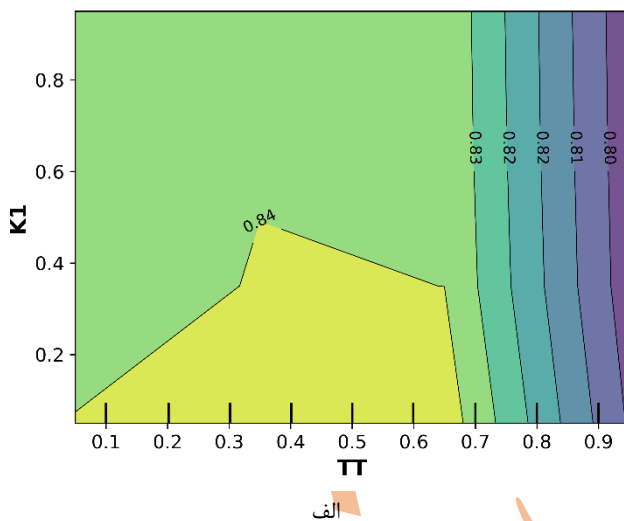


شکل ۱۰: نمودار وابستگی جزئی دوتایی برای تابع Sobol G (الف) اندکنش  $x_1$  با  $x_2$  (ب) اندکنش  $x_1$  با  $x_5$

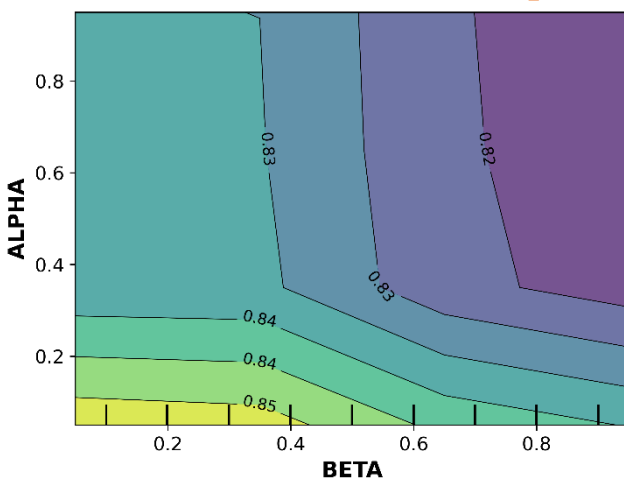
برای مدل HBV-SASK نیز شکل (۱۱)، نمودار وابستگی جزئی دوتایی پارامتر FRAC با پارامترهای C0، FC، BETA و ETF را

ب) برای پارامترهای FRAC و FC که مقدار بازه تغییرات ۰/۳۵ است، کم است. اما طبق شکل (۱۱-الف)، که نتیجه اندرکنش بین متغیرها به روش H آماری را نشان می‌دهد، اندرکنش بین BETA و TT بیشتر از اندرکنش بین FRAC و FC است. بنابراین نتایج روش H فریدمن، بر خلاف روش جایگشتی، با نمودارهای وابستگی جزئی تطابق ندارد.

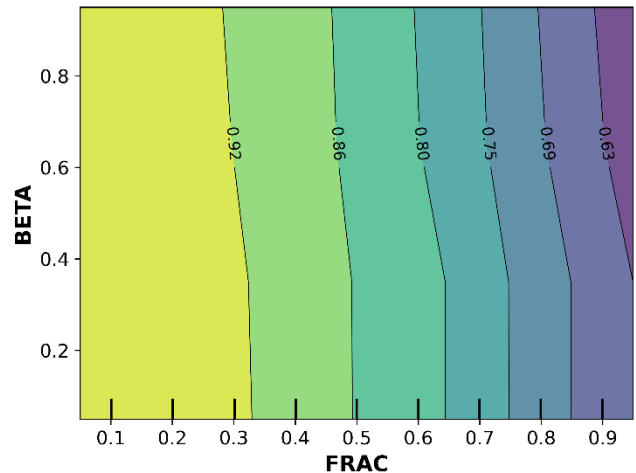
علاوه بر این، از نمودارهای وابستگی جزئی، می‌توان برای تحلیل حساسیت‌های موضعی نیز استفاده کرد. برای مثال، نوع تاثیرگذاری پارامترها در شکل های (۱۲-الف و ث)، در ناحیه‌های مختلف، متفاوت است.



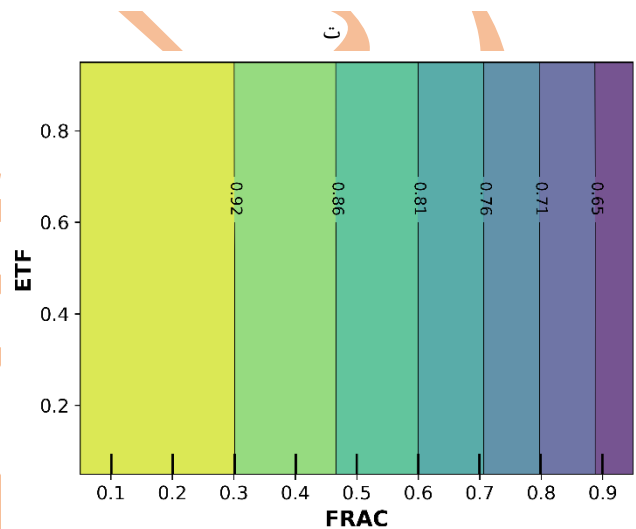
الف



ب



FRAC



FRAC

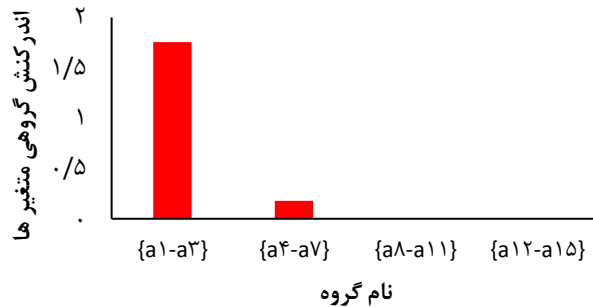
ث

شکل ۱۱: نمودار وابستگی جزئی دوتایی برای پارامتر FRAC در مدل HBV-SASK با پارامترهای الف (C0 ب) FC ت) BETA ت) ETF

شکل (۱۲)، نمودارهای وابستگی جزئی دوتایی برای سایر پارامترهای مدل HBV-SASK را نشان می‌دهد. در شکل (۱۲-ت)، بازه تغییرات دقت مدل تقریباً صفر است. همان‌طور که روش جایگشتی هم اهمیت بسیار کمی برای پارامترهای ETF و LP نشان داده‌بود، نتایج نمودارهای وابستگی تکی و دوتایی آن‌ها هم همین موضوع را تایید می‌کند. همچنین در شکل (۱۲-ت)، جهت خطوط عمود به محور پارامتر LP است که نشان می‌دهد پارامتر ETF بسیار کم اهمیت‌تر است.

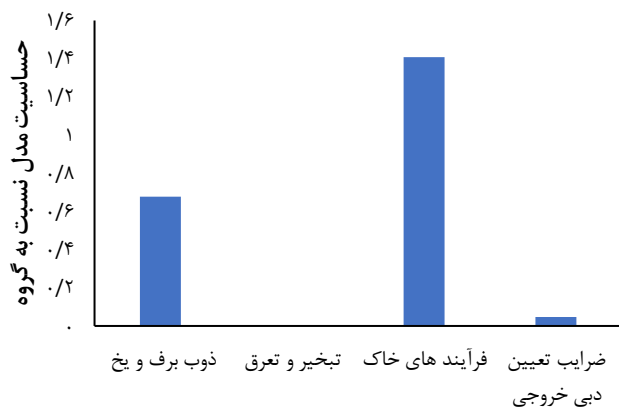
علاوه بر این، طبق شکل (۱۲-ث) که نمودار وابستگی جزئی دو پارامتر BETA و TT است، بازه تغییرات متغیر هدف، تقریباً ۰/۰۵ است که در مقایسه با نمودارهای وابستگی جزئی دوتایی نظیر شکل (۱۱-

شکل (۱۳)، نتیجه اندرکنش گروهی متغیرهای مدل Sobol G را به روش جایگشتی نشان می‌دهد.



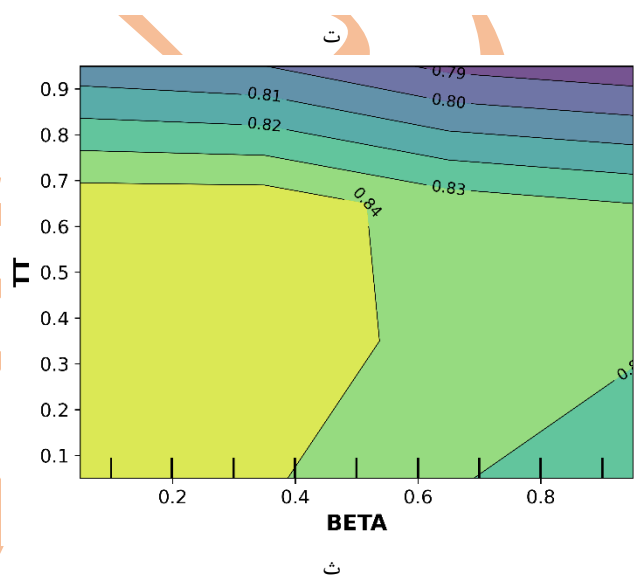
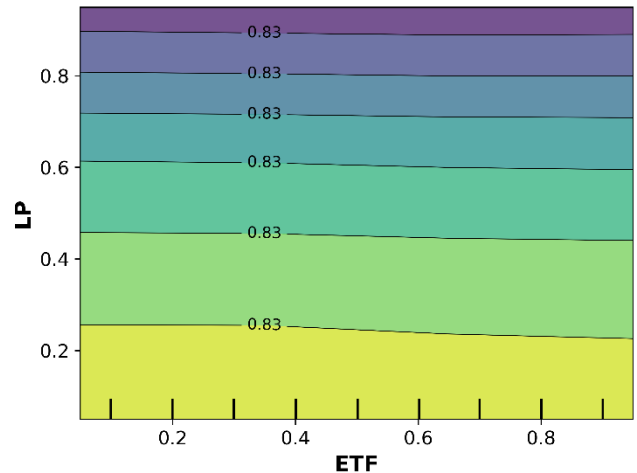
شکل ۱۳: اندرکنش گروهی متغیرهای تابع Sobol G با روش جایگشتی

نتیجه تحلیل حساسیت اندرکنش گروهی پارامترهای مدل HBV-SASK به روش جایگشتی در شکل (۱۴) آمده است. طبق این شکل، پارامترهای مربوط به فرآیندهای خاک بیشترین اهمیت را دارند. ادامه نیز پارامترهای ذوب برف تاثیر زیادی داشته و سایر گروه‌ها تاثیر چندانی روی دقت مدل ندارند.



شکل ۱۴: اندرکنش گروه‌های فیزیکی بر مبنای روش جایگشتی

در مدل HBV-SASK، گروه اول شامل پارامترهای CO و TT است که پارامترهای مربوط به ذوب برف هستند و مقدار اندرکنش گروهی آن‌ها ۰/۶۸ است. گروه دوم، پارامترهای ETF و LP را در بر می‌گیرد که شامل پارامترهای مربوط به تبخیر و تعرق است و مقدار اندرکنش گروهی تقریباً صفر دارند. گروه سوم، شامل سه پارامتر مربوط به فرآیندهای ذخیره و رهاسازی آب در خاک است که پارامترهای BETA، FRAC و FC را در بر می‌گیرد و مهم‌ترین گروه از نظر میزان اندرکنش (۱/۴۱) هستند. گروه چهارم که ضرایب



شکل ۱۲: نمودارهای وابستگی جزئی دوتایی الف) TT و k1 و BETA و ب) ALPHA (ت) و LP و BETA (ث) و TT و BETA

#### ۴-۷- نتایج اندرکنش گروهی

در این مطالعه، پارامترهای مدل هیدرولوژیکی HBV-SASK براساس فرآیندهای فیزیکی مربوط به آن‌ها به چهار دسته تقسیم شدند: پارامترهای مربوط به (۱) ذوب برف، (۲) فرآیندهای خاک، (۳) تبخیر و تعرق و (۴) دبی خروجی. همچنین در تابع Sobol G، متغیرها براساس میزان اهمیت‌شان (در جدول (۲)) به گروه‌های (۱) بسیار تأثیرگذار، (۲) تأثیرگذار، (۳) کم‌تأثیر و (۴) بی‌تأثیر دسته‌بندی شدند. علاوه بر این که محاسبه اندرکنش بین متغیرها به روش جایگشتی، قابل اعتمادتر است و هزینه زمانی کمتری نسبت به روش H آماری دارد، این قابلیت را می‌دهد تا اندرکنش‌های گروهی برای متغیرها نیز محاسبه شود.

مربوط به روندیابی دبی خروجی هستند شامل ALPHA، K1 و K2 است و حساسیتی به اندازه ۰/۰۵ را دارند.

## ۵- نتیجه گیری

همان طور که گفته شد، تحلیل حساسیت و ارزیابی اهمیت اندرکنش متغیرها در مدل سازی هیدرولوژیکی، بسیار مهم است. در این مطالعه، از روش های مختلفی برای ارزیابی اهمیت تکی و گروهی متغیرها استفاده شد. نتایج این پژوهش نشان داد که ترکیب روش های تحلیلی و تصویری در تحلیل حساسیت مدل های هیدرولوژیکی می تواند درک جامع تری از نقش، اهمیت و نحوه اثرگذاری متغیرها فراهم کند.

بررسی دو مدل مرجع ریاضی Sobol G و بارش-رواناب-HBV SASK نشان داد که روش جایگشتی نسبت به روش H-فریدمن از پایداری محاسباتی بیشتر، حساسیت کمتر به همبستگی بین متغیرها و سازگاری بالاتری با نمودارهای وابستگی جزئی برخوردار است. این یافته با نتایج گزارش شده در ادبیات پیشین که به ناپایداری روش H در شرایط وجود همبستگی و نیاز آن به حجم نمونه بالا اشاره دارند، همسو است. همچنین، بر اساس مطالعات گذشته، روش های جایگشتی به دلیل عدم وابستگی به ساختار مدل و امکان استخراج اندرکنش ها، رویکردی مناسب برای مدل های غیرخطی و پیچیده معرفی شده اند که نتایج این پژوهش نیز آن را تأیید می کند [۵].

در بررسی مدل HBV-SASK، پارامترهای FRAC و C0 به عنوان مؤثرترین پارامترها بر عملکرد مدل شناسایی شدند؛ موضوعی که با نتایج پژوهش های پیشین درباره نقش غالب پارامترهای مربوط به توزیع مکانی رواناب و ویژگی های اولیه مخزن سازگار است [۱۷]. همچنین تحلیل اندرکنش ها نشان داد که گروه پارامترهای مرتبط با ذوب برف و فرآیندهای خاک تأثیر قابل توجهی در رفتار مدل دارند؛ یافته ای که با گزارش های پیشین درباره حساسیت بالای منطقه مورد مطالعه نسبت به پارامترهای برف-یخ و رطوبت خاک همخوان است [۴].

در مجموع، نتایج این پژوهش تأکید می کند که استفاده هم زمان از روش های تحلیلی (مانند روش جایگشتی) و روش های تفسیری (مانند نمودارهای وابستگی جزئی) می تواند علاوه بر تعیین میزان اهمیت نسبی متغیرها، چگونگی اثرگذاری آنها، میزان غیرخطی بودن رابطه، و اندرکنش های چندسطحی را نیز آشکار کند. این

رویکرد به ویژه برای مدل های هیدرولوژیکی با رفتار پیچیده و وابستگی شدید بین پارامترها توصیه می شود. در آینده، به کارگیری این چارچوب در مدل های هیدرولوژیکی مبتنی بر فیزیک یا مدل های حوضه های مختلف می تواند به توسعه ابزارهای دقیق تر در تحلیل حساسیت و بهبود قابلیت اعتماد مدل های هیدرولوژیکی کمک کند.

## منابع

1. Faticchi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B. and Jones, N., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, 537, pp.45-60. <https://doi.org/10.1016/j.jhydrol.2016.03.026>
2. Razavi, S., Sheikholeslami, R., Gupta, H.V. and Haghnegahdar, A., 2019. VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environmental modelling & software*, 112, pp.95-107. <https://doi.org/10.1016/j.envsoft.2018.10.005>
3. Sheikholeslami, R. and Razavi, S., 2020. A fresh look at variography: measuring dependence and possible sensitivities across geophysical systems from any given data. *Geophysical Research Letters*, 47(20), p.e2020GL089829. <https://doi.org/10.1029/2020GL089829>
4. Sheikholeslami, R., Gharari, S., Papalexioiu, S.M. and Clark, M.P., 2021. VISCOUS: A variance-based sensitivity analysis using copulas for efficient identification of dominant hydrological processes. *Water Resources Research*, 57(7), p.e2020WR028435. <https://doi.org/10.1029/2020WR028435>
5. Sheikholeslami, R., Khanjani, M. and Razavi, S., 2023, December. Estimating Variance-Based Sensitivity Indices Using Random Forests: A Convergence Analysis. In *AGU Fall Meeting Abstracts* (Vol. 2023, pp. H23H-04). 2023AGUFM.H23H..04S
6. Kaneko, H., 2022. Cross-validated permutation feature importance considering correlation between features. *Analytical Science Advances*, 3(9-10), pp.278-287. <https://doi.org/10.1002/ansa.202200018>
7. Liu, H., Clark, M.P., Gharari, S., Sheikholeslami, R., Freer, J., Knoben, W.J., Marsh, C.B. and Papalexioiu, S.M., 2024. An improved copula-

16. Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology*, 201(1-4), pp.272-288. [https://doi.org/10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3)
17. Gupta, H.V. and Razavi, S., 2018. Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, 54(11), pp.8692-8717. <https://doi.org/10.1029/2018WR022668>
8. Jog, A., Carass, A., Roy, S., Pham, D. L., & Prince, J. L. (2017). Random forest regression for magnetic resonance image synthesis. *Medical image analysis*, 35, 475-488. <https://doi.org/10.1016/j.media.2016.08.009>
9. Gong, H., Sun, Y., Shu, X., & Huang, B. (2018). Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials*, 189, 890-897. <https://doi.org/10.1016/j.conbuildmat.2018.09.017>
10. Reis, I., Baron, D. and Shahaf, S., 2018. Probabilistic random forest: A machine learning algorithm for noisy data sets. *The Astronomical Journal*, 157(1), p.16. DOI 10.3847/1538-3881/aaf101
11. Matsuki, K., Kuperman, V. and Van Dyke, J.A., 2016. The Random Forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, 20(1), pp.20-33. <https://doi.org/10.1080/10888438.2015.1107073>
12. Moezzi, S.M.M., Mohammadi, M., Mohammadi, M., Saloglu, D. and Sheikholeslami, R., 2024. Machine learning insights into PM2. 5 changes during COVID-19 lockdown: LSTM and RF analysis in Mashhad. *Environmental Monitoring and Assessment*, 196(5), p.453. <https://doi.org/10.1007/s10661-024-12567-5>
13. Inglis, A., Parnell, A. and Hurley, C.B., 2022. Visualizing variable importance and variable interaction effects in machine learning models. *Journal of Computational and Graphical Statistics*, 31(3), pp.766-778. <https://doi.org/10.1080/10618600.2021.2007935>
14. Glen, G. and Isaacs, K., 2012. Estimating Sobol sensitivity indices using correlations. *Environmental Modelling & Software*, 37, pp.157-166. <https://doi.org/10.1016/j.envsoft.2012.03.014>
15. Sheikholeslami, R., Razavi, S., Gupta, H.V., Becker, W. and Haghnegahdar, A., 2019. Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. *Environmental modelling & software*, 111, pp.282-299. <https://doi.org/10.1016/j.envsoft.2018.09.002>