



شناسایی داده‌های پرت در آنالیز منطقه‌ای سیلاب به روش آنالیز مولفه‌های اصلی

محمد حسین نوری قیداری

عضو هیات علمی دانشگاه آزاد اسلامی واحد زنجان

noori@azu.ac.ir

خلاصه

داده‌های حدی بالا و پایین که داده‌های پرت نامیده می‌شوند، نتایج تجزیه و تحلیل داده‌ها را تحت تاثیر قرار داده و نتایج بدست آمده را دور از واقعیت می‌کنند. در آنالیز منطقه‌ای سیلاب داده‌های پرت بالا باعث برآورد زیاد سیلاب طرح شده که این موجب بالا رفتن هزینه‌های ساخت سازه‌های آبی می‌گردد. داده‌های حدی پایین نیز به علت برآورد کم سیلاب طرح سبب افزایش هزینه‌های تخریب سازه‌ها می‌شود. در هیدرولوژی روش جامع و مطمئنی برای شناسایی داده‌های پرت وجود ندارد. در این تحقیق از روش جدید آنالیز مولفه‌های اصلی که یک روش آماری چند متغیره است برای شناسایی داده‌های پرت استفاده شده است. با روش آنالیز مولفه‌های اصلی با توجه به ساختار داده‌ها، عوامل ایجاد تغییرات داده‌ها که مولفه‌های اصلی یا متغیرهای نهان نامیده می‌شوند، شناسایی می‌شود. در روش بکار گرفته شده داده‌هایی که مولفه‌های اصلی آنها از بقیه اختلاف زیادی دارد به عنوان داده پرت قلمداد می‌گردد. نتایج بدست آمده از این روش برای حوزه دریاچه ارومیه نشان‌دهنده دقت بالای این روش است.

کلمات کلیدی: داده‌های پرت، آنالیز منطقه‌ای سیلاب، آنالیز مولفه‌های اصلی، حوزه دریاچه

1. مقدمه

داده‌های پرت داده‌هایی هستند که بطور معنی داری از مقدار داده‌های اصلی و واقعی انحراف داشته که می‌تواند ناشی از خطای اندازه‌گیری و جمع‌آوری داده‌ها یا ناشی از عوامل طبیعی باشد. وجود داده‌های باعث بروز مشکلاتی در برآورد سیلاب طرح می‌گردد که چند نمونه آن عبارتند از: 1) وجود داده‌های پرت باعث انتخاب توزیع احتمالاتی غیر واقعی می‌گردد 2) برآورد پارامترهای توزیع‌های احتمالاتی تحت تاثیر داده‌های پرت قرار گرفته و برآورد همراه با خطا خواهد بود 3) داده‌های حدی پایین باعث برآورد کم چندک‌های سیلاب شده که این باعث برآورد کم سیلاب طرح و در نتیجه موجب تخریب سازه در سیلابهای بزرگتر می‌شود 4) داده‌های حدی بالا باعث برآورد زیاد سیلاب طرح شده که این موجب بالا رفتن هزینه‌های ساخت سازه‌های آبی می‌شود.

در هیدرولوژی با وجود مطالعات صورت گرفته بر روی شناسایی داده‌های پرت هنوز روش مناسب و جامع برای شناسایی داده‌های پرت معرفی نشده است. در این تحقیق از آنالیز مولفه‌های اصلی که یک روش آماری چند متغیره است برای شناسایی داده‌های پرت معرفی می‌گردد. در روش آنالیز مولفه‌های اصلی متغیرهای اصلی که وابسته هستند با یک تبدیل خطی به متغیرهای جدید و مستقل تبدیل می‌گردد. این متغیرهای جدید ساختار کوواریانس داده‌ها را توصیف کرده و منبع تغییرات داده‌ها را شناسایی می‌کند. متغیرهای جدید که مولفه‌های اصلی نامیده می‌شوند، قابل رتبه‌بندی هستند. بطوری که اولین مولفه اصلی دارای بیشترین اهمیت بوده و عامل اصلی تغییرات داده‌ها را توصیف می‌کند. و همینطور دومین و سومین مولفه اصلی بلحاظ اهمیت در رتبه دو و سه قرار دارد. این سه مولفه اصلی نخست می‌توانند 50 تا 90 درصد تغییرات داده‌ها را شناسایی کنند. بنابراین می‌توان بجای بررسی داده‌های ایستگاههای یک منطقه، این سه مولفه اصلی نخست را ارزیابی کرد. بطور مثال برای حوضه دریاچه ارومیه 34 ایستگاه هیدرومتری وجود دارد یعنی اگر هر ایستگاه به عنوان یک متغیر منظور شود 34 متغیر خواهیم داشت که از لحاظ تعداد قابل توجه بوده و نتیجه‌گیری از