

تحلیل حساسیت متغیرهای ورودی در مدل‌های خطی تعمیم‌یافته

محمد خراسانی^۱، فرزاد اسکندری^۲

تاریخ دریافت: ۱۴۰۰/۱۲/۲۵

تاریخ پذیرش: ۱۴۰۱/۰۵/۲۵

چکیده:

در دنیای امروزی با استفاده از فرآیند مدل‌سازی آماری می‌توان پدیده‌های طبیعی را جهت تحلیل و پیش‌بینی اتفاقات مورد بررسی، به کار برد. در بسیاری از روش‌های مدل‌سازی آب‌شناسی از اطلاعات موجود به بهترین وجه استفاده نمی‌شود به این دلیل که مدل‌های آب‌شناسی فرایندهای محیطی را در طیف وسیعی نشان می‌دهند که باعث افزایش پیچیدگی مدل می‌شود. در هنگام پیش‌بینی به‌طور مشخص پارامترها بر عملکرد مدل‌های آماری تأثیر می‌گذارند. وجود عدم قطعیت در بسیاری از مسائل ارزیابی ریسک در پارامترها منجر به عدم قطعیت در پیش‌بینی‌های مدل می‌شود. تحلیل حساسیت کلی ابزاری است که برای نشان دادن عدم قطعیت استفاده می‌شود و در تصمیم‌گیری، ارزیابی ریسک، ساده‌سازی مدل‌ها و مانند آن کاربرد دارد. دو روش تحلیل حساسیت فاصله مینکوفسکی و تحلیل حساسیت منطقه‌ای روش‌هایی هستند که می‌توانند با یک مجموعه نمونه داده‌شده از جفت‌های ورودی - خروجی مدل کار کنند. یک تفاوت قابل توجه بین این دو روش آن است که تحلیل حساسیت فاصله مینکوفسکی، توزیع‌های خروجی را مشروط به مقادیر ورودی (پیشرو) تحلیل می‌کند؛ در حالی که تحلیل حساسیت منطقه‌ای، توزیع‌های ورودی را مشروط به مقادیر خروجی تحلیل می‌کند (معکوس). در این مقاله، به تعیین روابط روش‌های حساسیت کلی (فاصله مینکوفسکی و منطقه‌ای) پرداخته و نشان داده خواهد شد، هنگامی که تحلیل حساسیت منطقه‌ای بر توابع چگالی احتمالی متمرکز شود، به سمت تحلیل حساسیت فاصله مینکوفسکی همگرا می‌شود. همچنین مشابه شاخص‌های حساسیت پیشرو، می‌توان شاخص‌های حساسیت معکوس را به دست آورد. سرانجام به تحلیل حساسیت یک طرح ذخیره‌سازی آب با استفاده از مدل «های‌مد» با ابعاد بالای خروجی‌های مدل پرداخته می‌شود.

واژه‌های کلیدی: تحلیل حساسیت کلی، تحلیل حساسیت فاصله مینکوفسکی، تحلیل حساسیت منطقه‌ای، رده‌بندی خروجی.

۱ مقدمه

شود، آنگاه فرضیه قادر به توضیح فرایند مورد بررسی و رفتار سامانه خواهد بود؛ بنابراین از این دیدگاه، مدل‌سازی را می‌توان فرایند آزمودن فرضیه‌های مختلف در رابطه با چگونگی عملکرد سامانه واقعی دانست.

در بسیاری از روش‌های مدل‌سازی آب‌شناسی، از اطلاعات موجود به بهترین وجه استفاده نمی‌شود؛ زیرا مدل‌های آب‌شناسی که در سامانه‌های مدل‌سازی تعبیه شده‌اند، فرایندهای محیطی را در طیف وسیعی از مقیاس‌های مکانی نشان می‌دهند که باعث افزایش پیچیدگی مدل، کمبود داده‌های مشاهده‌ای مناسب برای محدود کردن حالت‌ها و خروجی‌های مدل می‌شود و از طرف دیگر اطلاعات موجود در داده‌ها برای شناسایی ساختار و پارامترهای مدل محدود هستند؛ در نتیجه

امروزه با توجه به توسعه علم آمار در تجارت و صنایع تولیدی از پیش‌بینی آماری به‌عنوان ابزاری قدرتمند برای بیان رخدادهایی که ممکن است در آینده اتفاق بیفتد، استفاده می‌شود. مدل ابزاری است جهت خلاصه‌سازی و شناخت ما از یک سامانه واقعی که توسط روابط و قوانین قابل‌ارائه می‌باشد.

بیون (۲۰۱۱) مدل را یک فرضیه برای درک چگونگی رفتار یک سامانه می‌داند [۲]؛ بنابراین فرضیه‌ها به‌وسیله‌ی داده‌های اندازه‌گیری و ویژگی‌های فیزیکی سامانه آب‌شناسی بررسی می‌شوند. هنگامی که فرضیه در نظر گرفته شده بر اساس آزمون‌های مطرح شده مورد قبول واقع

^۱ محمد خراسانی (نویسنده مسئول: mokhorasani555@gmail.com)

^۲ فرزاد اسکندری

است که آن را با تابع چگالی احتمال توأم $f_X(x)$ نمایش می‌دهند. تابع چگالی احتمال حاشیه‌ای x_i را نیز با $f_{X_i}(x_i)$ نشان می‌دهند. خروجی‌های مدل m بعدی را می‌توان به صورت $Y = \{Y_1, \dots, Y_m\} = g(X)$ که تابع چگالی احتمال توأم خروجی مدل را با $f_Y(y)$ نمایش داد.

۱.۲ معیار فاصله مینکوفسکی

روش معیار فاصله مینکوفسکی، برای اندازه‌گیری تأثیر ورودی‌های مدل بر اساس تابع چگالی احتمال توأم از توزیع خروجی استفاده می‌شود [۱۵]. از این روش برای مدل‌های دارای چندین خروجی از مرتبه‌ی p استفاده می‌شود. وقتی که متغیر ورودی X در مقدار مشخص x_i ثابت باشد، تابع چگالی احتمال توأم شرطی خروجی مدل را می‌توان به صورت $f_{Y|X_i}(y)$ نشان داد؛ بنابراین میانگین تأثیر در خروجی مدل برابر است

$$E_{X_i}(\sup_{y \in Y} |f_Y(y) - f_{Y|X_i}(y)|) = \int_{X_i} (\sup_{y \in Y} |f_Y(y) - f_{Y|X_i=x}(y)|) f_{X_i}(x) dx \quad (1)$$

هنگامی که $p = 1$ باشد، معادله (۱) با شاخص چگالی مبنا برابر است [۳]. شاخص حساسیت فاصله مینکوفسکی را می‌توان به صورت زیر بیان کرد

$$\delta_i = \frac{1}{p} E_{X_i}(s(X_i)) \quad (2)$$

این شاخص، اثر نرمال‌شده پارامتر ورودی X_i بر روی تابع چگالی احتمال توأم خروجی‌های مدل را نشان می‌دهد. همچنین یک شاخص حساسیت مشابه با فرم بالا وجود دارد که با جایگزینی تابع چگالی احتمال با تابع توزیع تجمعی به صورت زیر به دست می‌آید

$$I_i^{cdf}(p) = E_{X_i} \left(\int_y |F_Y(y) - F_{Y|X_i}(y|x_i)|^p dy \right)^{\frac{1}{p}} = \int_{X_i} \left(\int_y |F_Y(y) - F_{Y|X_i=x}(y|x)|^p dy \right)^{\frac{1}{p}} f_{X_i}(x) dx \quad (3)$$

۲.۲ روش تحلیل حساسیت منطقه‌ای

در روش تحلیل حساسیت منطقه‌ای، خروجی مدل را به چند رده جداگانه تقسیم می‌کنند و سپس توزیع شرطی پارامترهای ورودی رده‌ها با یکدیگر مقایسه می‌شوند [۱۳]. اگر توزیع شرطی پارامترهای ورودی X_i نسبت به توزیع قبلی تفاوت معنی‌داری داشته باشد، X_i تأثیر مهمی بر خروجی مدل دارد؛ در غیر این صورت تأثیر

باعث ایجاد عدم قطعیت در ساختار پارامترهای مدل و عدم اطلاعات دقیق در مورد رفتار مدل می‌شود. به طور کلی آنچه مدل‌سازی و هدف پژوهش را بسیار پیچیده می‌کند، عدم قطعیت می‌باشد. عدم قطعیت روش علمی نیست؛ بلکه ماهیت آن است [۹]. برای این منظور از روش حساسیت کلی استفاده می‌شود. تحلیل حساسیت کلی ابزاری است که برای نشان دادن عدم قطعیت استفاده می‌شود و در تصمیم‌گیری، ارزیابی ریسک و ساده‌سازی مدل کاربرد دارد. تحلیل حساسیت شامل مدل‌ها، ورودی و خروجی مدل می‌باشد. در طول دهه‌های گذشته، روش‌های مختلفی از تحلیل حساسیت کلی ارائه شده است. بسیاری از این روش‌ها، برای نشان دادن عدم قطعیت باعث از بین رفتن اطلاعات می‌شود ([۱۱]، [۱۲])؛ لذا می‌توان از شاخص چگالی مبنا (DSA)، به عنوان رویکردی مناسب برای تحلیل حساسیت نام برد [۳]. روش تحلیل حساسیت فاصله مینکوفسکی (MD)، یک روش تعمیم‌یافته از روش تحلیل حساسیت چگالی مبنا از مرتبه‌ی p می‌باشد [۱۵]، هنگامی که مرتبه‌ی $p = 1$ باشد، تحلیل حساسیت چگالی مبنا به دست می‌آید [۳].

واکتر و همکاران (۲۰۰۱) به بیان یک چارچوب تحلیلی برای مدل‌سازی آب‌شناسی پرداختند [۸] که از اطلاعات موجود در داده‌ها برای شناسایی ساختار و پارامترهای مدل استفاده بهینه کند و امکان تحلیل دقیق رفتار مدل را فراهم آورد. در این چارچوب، روش تحلیل حساسیت منطقه‌ای (RSA) استفاده شده است. این روش عدم قطعیت را بر اساس ویژگی‌های خروجی مدل بررسی می‌کند [۱۳]. هدف بررسی روابط در این مقاله، روش‌های حساسیت GSA و هم‌ارزی مجانبی بین آن‌ها است. در هنگام فراموشی جهت علی از ورودی به خروجی مدل، ورودی و خروجی صرفاً مجموعه متغیر تصادفی وابسته با توزیع توأم می‌باشند. به همین امر می‌توان از معیارهای وابستگی بین ورودی و خروجی برای تحلیل حساسیت در پیش‌بینی‌ها استفاده کرد. این مقاله به شرح زیر تنظیم شده است. بخش ۲، مروری کوتاه بر RSA و MD دارد. بخش ۳، ابتدا رابطه بین روش‌های RSA و MD و اثبات هم‌ارزی بودن مجانبی آن‌ها را ارائه می‌دهد. در بخش ۴، به بررسی و مقایسه این دو روش در یک مدل آب‌شناسی و در بخش ۵، به بیان نتیجه‌گیری از مطالب بیان شده در این مقاله پرداخته می‌شود.

۲ مواد و روش‌ها

فرض کنید $Y = g(X)$ یک مدل می‌باشد که در آن $X = \{X_1, X_2, \dots, X_d\}$ ، یک بردار d بعدی از متغیرهای ورودی تصادفی

۳ مقایسه دو رویکرد

۱.۳ ایجاد ارتباط بین تحلیل حساسیت معیار فاصله مینکوفسکی و منطقه‌ای

همان‌طور که بیان شد، MD بر توزیع احتمال خروجی مدل و RSA بر توزیع احتمال ورودی مدل تمرکز دارند. در یک مدل قطعی، توزیع احتمال خروجی مدل به توزیع احتمال ورودی مدل بستگی دارد [۶]؛ بنابراین می‌توان ارتباطی بین ξ_i در رابطه (۲) و $S_{i,M}^{PDF}$ در رابطه (۸) به دست آورد. برای سادگی در اثبات ابتدا $p = ۱$ در نظر گرفته و سپس p را برای درجه‌های بالاتر تعمیم داده خواهد شد.

در ابتدا $S_{i,M}^{PDF}$ را می‌توان به صورت زیر بازنویسی کرد

$$\begin{aligned} S_{i,M}^{PDF} &= \sum_{K=1}^M d_{C_k,i}^{PDF} P_{C_K} \\ &= \sum_{K=1}^M \int_{X_i} |f_{X_i}(x_i) - f_{X_i|C_k}(x_i | C_k)| dx_i \int_{C_k} f_Y(y) dy d_{C_k,i}^{CDF} = \int_{X_i} |F_{X_i}(x_i) - F_{X_i|C_k}(x_i)| dx_i \quad k = 1, \dots, M \\ &= \sum_{K=1}^M \int_{X_i} \left| \int_{C_k} (f_{X_i}(x_i) f_Y(y) - f_{X_i,Y}(x_i, y)) dy \right| dx_i \\ &= \sum_{K=1}^M \int_{X_i} \left| \int_{C_k} (f_Y(y) - f_{Y|X_i}(y | x_i)) dy \right| f_{X_i}(x_i) dx_i \end{aligned} \quad (۹)$$

با توجه به قضیه مقدار میانگین، برای هر $k = 1, \dots, M$ ، $y_k \in C_k$ ، رابطه برقرار است؛ بنابراین داریم

$$\int_{C_k} (f_Y(y) - f_{Y|X_i}(y | x_i)) dy = (f_Y(y_k) - f_{Y|X_i}(y_k | x_i)) V(C_k), \quad k = 1, \dots, M \quad (۱۰)$$

همان‌طور که مشاهده می‌کنید، $V(C_k)$ حجم زیرمجموعه‌ی C_k را نشان می‌دهد؛ پس $S_{i,M}^{PDF}$ را می‌توان به صورت زیر بیان کرد

$$\begin{aligned} S_{i,M}^{PDF} &= \sum_{K=1}^M \int_{X_i} (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| V(C_k)) f_{X_i}(x_i) dx_i \\ &= \int_{X_i} \sum_{K=1}^M (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| V(C_k)) f_{X_i}(x_i) dx_i \end{aligned} \quad (۱۱)$$

می‌توان $\lim_{M \rightarrow \infty} S_{i,M}^{PDF}$ را به صورت زیر به دست آورد

$$\begin{aligned} \lim_{M \rightarrow \infty} S_{i,M}^{PDF} &= \lim_{M \rightarrow \infty} \int_{X_i} \sum_{K=1}^M (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| \\ &\quad \times V(C_k)) f_{X_i}(x_i) dx_i \end{aligned} \quad (۱۲)$$

قابل توجهی ندارد. اگر خروجی‌های مدل در یک مجموعه رده محدود شوند، می‌توان آن را به صورت زیر نمایش داد

$$C = C_1, \dots, C_M \quad (۴)$$

رده‌ها زیرمجموعه‌ی Y هستند که دوبه‌دو مجزا و اجتماع آن‌ها برابر با کل فضای خروجی مدل است. تابع توزیع تجمعی رده‌شرطی با پارامتر ورودی X_i را می‌توان به صورت

$$F_{X_i|C_k}(x_i) = Pr(X_i < x_i | Y \in C_k) \quad (۵)$$

بیان کرد. تفاضل بین تابع توزیع تجمعی قبلی $F_{X_i}(x_i)$ و تابع توزیع تجمعی رده‌شرطی $F_{X_i|C_k}(x_i)$ برای معیار حساسیت خروجی مدل با پارامترهای ورودی X_i را می‌توان از رابطه زیر به دست آورد

$$d_{C_k,i}^{CDF} = \int_{X_i} |F_{X_i}(x_i) - F_{X_i|C_k}(x_i)| dx_i \quad k = 1, \dots, M \quad (۶)$$

از آنجا که خروجی‌های مدل به M رده تقسیم شده‌اند، می‌توان شکل نرمالیده آن را به صورت

$$S_{i,M}^{CDF} = d_{C_k,i}^{CDF} P_{C_k}$$

بیان کرد. P_{C_k} احتمال قرار گرفتن مدل در رده C_k است که آن را می‌توان به صورت $P_{C_k} = \int_{C_k} f_Y(y) dy$ بیان کرد. $S_{i,M}^{CDF}$ را می‌توان به عنوان مقدار مورد انتظار برای $d_{C_k,i}^{CDF}$ در نظر گرفت که یک شاخص حساسیت معکوس برای پارامتر ورودی X_i محسوب می‌شود. هنگامی که در رابطه (۶)، تابع چگالی احتمال جایگزین تابع توزیع تجمعی شود، رابطه زیر به دست می‌آید

$$d_{C_k,i}^{PDF} = \int_{X_i} |f_{X_i}(x_i) - f_{X_i|C_k}(x_i)| dx_i \quad (۷)$$

که شاخص حساسیت آن به صورت زیر می‌باشد

$$S_{i,M}^{PDF} = \sum_{K=1}^M d_{C_k,i}^{PDF} P_{C_K} \quad (۸)$$

شاخص‌های حساسیت منطقه‌ای را می‌توان با یک مجموعه از نمونه‌ها برآورد کرد. کافی است یک انتگرال یک‌بعدی صرف نظر از بعد خروجی از مدل گرفته شود تا برآورد به دست آید. همین امر باعث شده است که RSA برای تحلیل حساسیت چندمتغیره آسان‌تر از DSA باشد.

مدل متغیرهای گسسته باشند، پارامترهای ورودی نیز متغیرهای گسسته می‌باشند که باعث می‌شود تابع چگالی احتمال‌ها به تابع جرم احتمال تبدیل شوند.

حال با توجه به مطالب بیان‌شده بالا، برای به دست آوردن مرتبه p می‌توان به یک تحلیل حساسیت منطقه‌ای با مرتبه‌ی p به صورت زیر رسید

$$E_{X_i} \left(\int_y |f_Y(y) - f_{Y|X_i}(y | x_i)|^p dy \right)^{\frac{1}{p}} = \int_{X_i} \left(\int_y |f_Y(y) - f_{Y|X_i}(y | x_i)|^p dy \right)^{\frac{1}{p}} f_{X_i}(x_i) dx_i \quad (16)$$

بنابر نامساوی مثلث، می‌توان $|f_Y(y) - f_{Y|X_i}(y | x_i)|^p$ را به صورت زیر بیان کرد

$$|f_Y(y) - f_{Y|X_i}(y | x_i)|^p \geq 0 \quad (17)$$

با قرار دادن این نامساوی در رابطه (۱۶)، رابطه‌ی زیر به دست می‌آید

$$\begin{aligned} & \int_{X_i} \left(\int_y |f_Y(y) - f_{Y|X_i}(y | x_i)|^p dy \right)^{\frac{1}{p}} f_{X_i}(x_i) dx_i \\ & \leq \int_{X_i} \left[\int_y |f_Y(y)|^p dy + \int_y |f_{Y|X_i}(y | x_i)|^p dy \right]^{\frac{1}{p}} f_{X_i}(x_i) dx_i \\ & = \int_{X_i} \left[\int_y |f_Y^p(y) f_{X_i}^p(x_i)| dy + \int_y |f_{Y|X_i}^p(y | x_i) f_{X_i}^p(x_i)| dy \right]^{\frac{1}{p}} dx_i \end{aligned} \quad (18)$$

بر اساس قضیه‌ی بی‌زی، می‌توان

$$\begin{aligned} f_{Y|X_i}^p(y) &= \frac{f_{X_i|Y}^p(x_i) f_Y^p(y)}{f_{X_i}^p(x_i)} \\ \Rightarrow \frac{f_{Y|X_i}^p(y)}{f_Y^p(y)} &= \frac{f_{X_i|Y}^p(x_i)}{f_{X_i}^p(x_i)} \end{aligned} \quad (19)$$

را به دست آورد.

در نهایت با جایگذاری رابطه‌ی (۱۹) در رابطه (۱۸) رابطه‌ی زیر به دست می‌آید

$$\begin{aligned} & = \int_y \left[\int_{X_i} |f_{X_i}(x_i)|^p dx_i \right]^{\frac{1}{p}} f_Y(y) dy \\ & + \int_y \left[\int_{X_i} |f_{X_i|Y}^p(x_i)|^p dx_i \right]^{\frac{1}{p}} f_Y(y) dy \end{aligned} \quad (20)$$

با استفاده از نامساوی مثلث داریم

$$\begin{aligned} & \int_y \left[\int_{X_i} |f_{X_i}(x_i)|^p dx_i \right]^{\frac{1}{p}} f_Y(y) dy \\ & + \int_y \left[\int_{X_i} |f_{X_i|Y}^p(x_i)|^p dx_i \right]^{\frac{1}{p}} f_Y(y) dy \\ & \geq \int_y \left(\int_{X_i} |f_{X_i}(x_i) - f_{X_i|Y}(x_i)|^p dx_i \right)^{\frac{1}{p}} f_Y(y) dy \end{aligned}$$

فرض کنید $|f_Y(y) - f_{Y|X_i}(y | x_i)|$ انتگرال‌پذیر نسبت به y برای هر مقدار ممکن x_i باشد؛ بنابراین می‌توان به کمک تعریف انتگرال به رابطه‌ی زیر رسید

$$\lim_{M \rightarrow \infty} \sum_{k=1}^M (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| V(C_k)) = \int_y |f_Y(y) - f_{Y|X_i}(y | x_i)| dy \quad (13)$$

از طریق گرفتن مقادیر مختلف M ، می‌توان توالی توابع $h_M(x_i) = \sum_{k=1}^M (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| V(C_k))$ را به دست آورد. هنگامی که $M \in N^+$ باشد، با استفاده از تعریف انتگرال می‌توان به این نتیجه رسید که $h_M(x)$ به‌طور یکنواخت به $h(x_i)$ همگرا است. با استفاده از رابطه‌ی (۱۳)، می‌توان این رابطه را به صورت زیر به دست آورد

$$\begin{aligned} h_M(x_i) &= \sum_{k=1}^M \left| \int_{C_k} (f_Y(y) - f_{Y|X_i}(y | x_i)) dy \right| \\ &\leq \sum_{k=1}^M \int_{C_k} |f_Y(y) - f_{Y|X_i}(y | x_i)| dy \\ &= \int_y |f_Y(y) - f_{Y|X_i}(y | x_i)| dy \end{aligned} \quad (14)$$

با توجه به این که $2 \leq E_{X_i}(S(X_i)) \leq 10$ ، می‌توان نتیجه گرفت که توالی تابع $h_M(x_i) \leq 2$ کران‌دار است. همچنین با توجه به قضیه‌ی همگرایی مغلوب، می‌توان حد و انتگرال را پیش از X_i در رابطه (۱۲) به کار برد تا رابطه‌ی زیر را به دست آید

$$\begin{aligned} \lim_{M \rightarrow \infty} S_{i,M}^{PDF} &= \int_{X_i} \lim_{M \rightarrow \infty} \sum_{K=1}^M (|f_Y(y_k) - f_{Y|X_i}(y_k | x_i)| \\ & V(C_k)) f_{X_i}(x_i) dx_i \\ &= \int_{X_i} \int_y |f_Y(y) - f_{Y|X_i}(y | x_i)| dy f_{X_i}(x_i) dx_i \\ &= 2\delta_i \end{aligned} \quad (15)$$

هنگامی که M افزایش می‌یابد تمام رده‌ها دوه‌دو مجزا و اجتماع آن‌ها برابر با کل فضای خروجی مدل می‌شود، افزایش M باعث می‌شود که اندازه هر رده کوچک شود. همان‌طور که تعداد رده‌های M به بی‌نهایت متمایل می‌شود $S_{i,M}^{PDF}$ معادل δ_i می‌شود (بدون ثابت $\frac{1}{2}$) که باعث ایجاد ارتباط بین تحلیل حساسیت فاصله مینکوفسکی و تحلیل حساسیت منطقه‌ای می‌شود. در رابطه (۱۰)، از قضیه مقدار میانگین استفاده شده است که خروجی‌های مدل را متغیرهای تصادفی پیوسته در نظر گرفته است. هنگامی که خروجی‌های

دست می-آید

$$\eta_i^{PDF} = E_Y \left[\int_{X_i} |f_{X_i}(x_i) - f_{X_i|Y}(x_i)| dx_i \right] \quad (24)$$

با باز کردن رابطه-ی بالا، در نهایت رابطه-ی زیر به دست می-آید

$$\begin{aligned} \eta_i^{PDF} &= \int_y \int_{X_i} |f_{X_i}(x_i) - f_{X_i|Y}(x_i)| dx_i f_Y(y) dy \\ &= \int_y \int_{X_i} |f_{X_i}(x_i) f_Y(y) - f_{X_i,Y}(x_i, y)| dx_i dy \\ &= \int_{X_i} \int_y |f_Y(y) - f_{Y|X_i}(y)| dy f_{X_i}(x_i) dx_i \quad (25) \end{aligned}$$

همان طور که مشاهده می-شود، هنگامی که مرتبه $p = 1$ باشد، رابطه-ها-ی (۲) و (۲۴) باهم برابرند. این برابری باعث ایجاد ارتباط کلی-تر بین روش-های تحلیل حساسیت معکوس و پیشرو بر اساس مفهوم منطق مشترک می-شود.

۱.۲.۳ برآورد δ_i در روش معکوس

برآورد شاخص حساسیت پیشرو در رابطه (۲۲) ارائه شده است. با استفاده از RSA ، می-توان یک روش مشابه برای برآورد شاخص حساسیت معکوس ارائه داد. تفاوت روش پیشرو و معکوس در این است که روش پیشرو نمونه-های ورودی را به رده-های مختلف تقسیم و سپس احتمال رده-شرطی خروجی مدل را برآورد می-کند؛ درحالی-که روش معکوس نمونه-های خروجی را به رده-های مختلف تقسیم و سپس احتمال رده-شرطی ورودی مدل را برآورد می-کند. از آنجاکه شاخص حساسیت معکوس تابع چگالی احتمال مبنا در (۲۶) معادل با شاخص حساسیت پیشرو تابع چگالی احتمال مبنا در (۲) می-باشد؛ پس می-توان برای برآورد δ_i از روش معکوس استفاده کرد؛ بنابراین می-توان هم-ارزی بین این دو شاخص حساسیت را تأیید و با استفاده از آن، این دو روش را با یکدیگر مقایسه کرد.

الگوریتم برآورد δ_i در روش معکوس

۱. فرض کنید مجموعه-ای از نمونه-ها $\{x^{(1)}, \dots, x^{(N)}\}$ وجود دارد؛ پس می-توان از طریق $Y = g(X)$ بر اساس تابع چگالی احتمال توأم $f_X(x)$ از پارامترهای ورودی، نمونه-های خروجی $\{y^{(1)}, \dots, y^{(N)}\}$ مربوطه را به دست آورد.

۲. نمونه-های خروجی را به M رده- $\{C_1, \dots, C_M\}$ نامتداخل تقسیم و سپس احتمال P_{C_K} هر رده را برآورد کرد

$$\hat{P}_{C_K} = \frac{N_{C_K}}{N}, k = 1, \dots, M$$

که در آن N_{C_K} تعداد نمونه-های خروجی متعلق به C_K را نشان می-دهد.

$$= E_Y \left(\int_{X_i} |f_{X_i}(x_i) - f_{X_i|Y}(x_i)|^p dx_i \right)^{\frac{1}{p}} f_Y(y) dy$$

این رابطه در واقع هم-ارزی بین تحلیل حساسیت فاصله مینکوفسکی و منطقه-ای را بر اساس مرتبه-ی p را نشان می-دهد.

۲.۳ تحلیل حساسیت پیشرو و معکوس

بورگنوو و همکاران (۲۰۱۶)، یک دلیل منطقی برای تحلیل حساسیت کلی و برآورد داده-ها ارائه داده-اند [۴]. DSA را می-توان مورد خاصی از این تحلیل حساسیت تعمیم-یافته نام برد که در آن \mathbb{P}_Y یک ویژگی احتمالاتی برای خروجی مدل است. مانند واریانس یا تابع چگالی احتمال. همین طور می-توان به طور مشابه، ویژگی احتمالاتی شرطی برای خروجی مدل را با $\mathbb{P}_{Y|X_i}$ نمایش داد؛ بنابراین فاصله بین \mathbb{P}_Y و $\mathbb{P}_{Y|X_i}$ را می-توان به صورت زیر اندازه-گیری کرد

$$\gamma_i(x_i) = \zeta(\mathbb{P}_Y, \mathbb{P}_{Y|X_i=x_i}) \quad (21)$$

همان طور که در رابطه (۲۱) مشاهده می-شود، نمادهای $\zeta(\cdot, \cdot)$ و $\gamma_i(x_i)$ به ترتیب بیانگر عملگر داخلی و آماره-ی داخلی می-باشد. شاخص حساسیت را می-توان به صورت زیر تعریف کرد

$$\xi_i = E[\gamma_i(X_i)] = E[\mathbb{P}_Y, \mathbb{P}_{Y|X_i}] \quad (22)$$

ξ_i شاخص حساسیت کلی X_i بر اساس $\zeta(\cdot, \cdot)$ می-نامند. با در نظر گرفتن انواع مختلف عملگرهای داخلی، شاخص حساسیت در معادله (۲۲)، می-تواند با شاخص-های مختلف حساسیت مانند شاخص حساسیت واریانس مبنا، شاخص حساسیت چگالی مبنا و مانند آن مطابقت داشته باشد.

تفاوت اصلی بین RSA و منطق رایج در تحلیل حساسیت بر اساس معادله (۲۲)، این است که RSA بر ویژگی-های احتمالی ورودی مدل متمرکز است؛ درحالی-که منطق رایج بر ویژگی-های احتمالی خروجی مدل متمرکز است؛ بنابراین، RSA را می-توان به عنوان تحلیل حساسیت معکوس (ورودی-گرا) و منطق توأم را می-توان تحلیل حساسیت پیشرو (خروجی-گرا) دانست؛ پس هنگام تحلیل حساسیت می-توان از ویژگی-های احتمالاتی هر یک از ورودی- یا خروجی-های مدل استفاده کرد. یک چارچوب کلی برای شاخص حساسیت معکوس (ورودی-گرا) مشابه با رابطه (۲۲) را می-توان به صورت زیر بیان کرد

$$\eta_i = E [E[\zeta(\mathbb{P}_{X_i}, \mathbb{P}_{X_i|Y})]] \quad (23)$$

اگر از PDF به-عنوان ویژگی احتمالاتی و معیار فاصله (L^1 -نرم) در رابطه (۷) نیز به-عنوان عملگر داخلی استفاده شود، رابطه-ی زیر به

برآورد PDF رده-شرطی در کنار هیستوگرام می‌توان از برآورد چگالی هسته‌ای نیز استفاده کرد [۱۳].

در مرحله ۴، برای برآورد انتگرال بر اساس هیستوگرام می‌توان از قاعده دوزنقه‌ای مرکب نیز استفاده کرد. با توجه به مقادیر مرزی برای همه‌ی رده‌ها $x_1^p < \dots < x_i^p < \dots < x_{i+1}^p < \dots < x_{I+1}^p$ ، معیار فاصله‌ای در رابطه‌ی (۷) را می‌توان به صورت زیر بیان کرد

$$\hat{d}_{C_K}^{PDF} = \sum_{i=1}^{p-1} (x_i^{I+1} - x_i^I) \frac{Q_{I+1} + Q_I}{2}$$

که در آن

$$Q_I = |f_{X_i}(x_i^I) - f_{X_i|C_K}(x_i^I | C_K)|$$

همان‌طور که بیان شد برای روش معکوس از یک مجموعه نمونه ورودی-خروجی (تک حلقه) و یک انتگرال یک-بعدی برای برآورد استفاده می‌شود. هزینه‌ی محاسباتی، مستقل از ابعاد پارامترهای ورودی می‌باشد. مشابه با روش پیشرو، $\hat{\delta}_i$ یک برآوردگر سازگار برای δ_i است. روش پیشرو به یک مجموعه نمونه ورودی-خروجی (تک حلقه) نیاز دارد. ایده اصلی این روش در نظر گرفتن PDF شرطی خروجی مدل، هنگامی که پارامترهای ورودی به رده‌های مختلف تعلق دارد است؛ بنابراین لازم است پارامتر ورودی X_i را نسبت به رده‌های مختلف $\{C_1, C_2, \dots, C_M\}$ جدا و سپس تابع چگالی احتمال رده-شرطی $f_{Y|C_K}(y)$ خروجی مدل را برآورد کنید. شاخص حساسیت δ_i را می‌توان به صورت زیر تعریف کرد

$$\delta_i^p = \frac{1}{2} \sum_{K=1}^M P_{iC_K} \int_y |f_Y(y) - f_{Y|C_K}(y | C_K)| dy \quad (26)$$

$P_{iC_K} = \int_{C_K} f_{X_i}(x_i) dx_i$ احتمال قرار گرفتن X_i در رده‌ی C_K را نشان می‌دهد. در مسئله‌هایی با چندین خروجی، استفاده از روش پیشرو دشوار می‌باشد؛ زیرا می‌بایست تابع چگالی احتمال را برآورد کرده و انتگرال چندبعدی از رابطه (۲۶) گرفته شود. درحالی که روش معکوس، نمونه‌های خروجی را نسبت به رده‌های مختلف جدا می‌کند؛ بنابراین در این حالت استفاده از روش خوشه‌بندی $K - means$ پیشنهاد می‌شود.

۳. نمونه‌های ورودی را بر اساس رده‌های خروجی به M رده جدا کنید.

$$B_k = \{x^j | y^j \in C_k\}, \quad k = 1, \dots, M$$

سپس تابع چگالی احتمال حاشیه‌ای رده-شرطی $f_{X_i|C_K}(x_i | C_K), i = 1, \dots, d$ را بر اساس نمونه‌های B_K برآورد کنید. برآورد آن را با $\hat{f}_{X_i|C_K}(x_i | C_K) (i = 1, \dots, d)$ نمایش می‌دهند.

۴. معیار فاصله $d_{C_K}^{PDF} = \int_{X_i} |f_{X_i}(x_i) - f_{X_i|C_K}(x_i | C_K)| dx_i$ را با استفاده از انتگرال-گیری عددی برآورد کنید. برآورد آن را با $\hat{d}_{C_K}^{PDF}$ نشان می‌دهند.

۵. مقدار تقریبی δ_i به صورت زیر نمایش می‌دهند

$$\hat{\delta}_i = \frac{1}{2} \sum_{k=1}^M \hat{d}_{C_K}^{PDF} \hat{P}_{C_K}$$

با اعمال تغییراتی در رابطه (۲۳)، می‌توان از روش فوق برای برآورد هر نوع شاخص حساسیت معکوسی استفاده کرد. برای مثال هنگامی که احتمال رده-شرطی $\mathbb{P}_{X_i|C_i}$ جایگزین PDF حاشیه‌ای رده-شرطی $f_{X_i|C_K}(x_i | C_K) (i = 1, \dots, d)$ در مرحله‌ی ۳ شود، باعث می‌شود در مرحله‌ی ۴ برای برآورد از عملگرهای داخلی $\zeta(\mathbb{P}_{X_i}, \mathbb{P}_{X_i|C_i})$ استفاده شود.

در مرحله‌ی ۲ لازم است نمونه‌های خروجی را به رده‌های مختلف جدا کرد. برای خروجی یک-بعدی، نمونه‌ها را می‌توان مطابق با رویکردی که بورگنوو و همکاران ارائه داده‌اند رده‌بندی کرد [۴]؛ درحالی که برای خروجی چندبعدی می‌توان نمونه‌های خروجی را به وسیله‌ی روش خوشه‌بندی $K - means$ به رده‌های مختلف جدا کرد.

در مرحله‌ی ۳، از نمودار هیستوگرام می‌توان برای برآورد تابع چگالی احتمال رده-شرطی $f_{X_i|C_K}(x_i | C_K), (i = 1, \dots, d)$ بر اساس نمونه‌ای که در B_K وجود دارد استفاده کرد. ابتدا دامنه مقادیر پارامتر ورودی را به یک سری رده جدا کنید، سپس نمونه‌ها را شمارش کرده تا فراوانی به دست آید. از طریق نرمال‌ساز فراوانی می‌توان برای هر رده مقادیر تابع چگالی احتمال را برآورد کرد. برای این منظور فریدمن قوانینی را به صورت زیر بیان کرد [۷]:

$$h = \frac{IQR(X_i)}{N^{\frac{1}{3}}} N_{bin} = \left\lceil \frac{max(X_i) - min(X_i)}{h} \right\rceil$$

$IQR(X_i)$ دامنه بین چارکی X_i و h پهنای هر رده را نشان می‌دهد. [۷] تابع حد بالا و N_{bin} هم تعداد رده‌ها را نشان می‌دهد. برای

۴ مورد آزمایشی: مدل آب شناسی

۱۰۴ شرح مدل

در این بخش به بررسی دو روش تحلیل پیشرو و معکوس در یک مدل آب شناسی «های مد»^۳ پرداخته می شود. در مطالعات آب شناسی، اغلب از مدل ها برای کمک به مدیریت منابع آب، پیش بینی آب شناسی، پیش بینی تغییرات آب و هوا و غیره استفاده می شود. رودخانه کارون از مرتفع ترین قله سلسله جبال زاگرس سرچشمه می گیرد و به خلیج فارس می ریزد. شاخه کارون متشکل از سه شعبه عمده کارون علیا، بازفت و خرسان مساحتی در حدود ۲۴ هزار کیلومتر مربع را در محل پل شال و زهکشی می کند. داده های مورد نیاز برای انجام این تحقیق، شامل بارش و تبخیر می باشد که داده های استفاده شده از سایت هواشناسی <https://en.tutiempo.net/climate/05-2014/ws-407980.html>

در یک سال آبی جمع آوری شده است. در این طرح از مدل سازی، برای پیش بینی آب و هوا و از روش های حساسیت کلی به عنوان ابزاری برای توسعه مدل های زیست محیطی استفاده می شود. در این خصوص مطالعات زیادی انجام شده است که در این مقاله مورد بررسی قرار می گیرد. در این طرح با استفاده از پودمان های هواشناسی به جمع آوری داده ها پرداخته که منجر به تشکیل مدل «های مد» می شود. مدل «های مد» یک مدل آب شناسی برای شبیه سازی رواناب بارندگی است که از یک پودمان برف، پودمان رطوبت خاک و پودمان مسیرگزینی جریان تشکیل شده است [۸]. پودمان برف از طریق روش درجه ی روز، میزان ذوب شدن برف را محاسبه می کند [۱]. هنگامی که متوسط دمای فصل پاییز کمتر از متوسط دما برای برف برسد، ذخیره سازی برف اتفاق می افتد. هنگامی که متوسط دمای روزانه بالاتر از آستانه دما باشد ذوب برف اتفاق می افتد. پودمان رطوبت خاک، یک تابع توزیع معین برای ذخیره سازی حداکثر میزان رطوبت خاک (C_{max}) برای عناصر حوضه آبریز می باشد. پارامتر نمایشی β درجه ی خاصی از تغییر پذیری مخزن را نشان می دهد. رطوبت خاک در کسری بالقوه تبخیر می شود که با استفاده از «هامون»^۴ می توان آن را برآورد کرد. از باقی مانده بارندگی و ذوب برف برای پر کردن ذخایر رطوبت خاک استفاده می شود. مقدار اضافی بارندگی به پودمان مسیرگزینی جریان ارسال می شود که این پودمان میزان بارندگی اضافی را با ضریب α افزاز کرده و از طریق مخازن خطی مفهومی موازی جهت

^۴Hamon

شبیه سازی پاسخ های جریان سریع و آهسته سامانه را هدایت می کند. این ضریب سریع مخزن (R_q) و آهسته مخزن (R_s) است که جریان خروجی از هر مخزن را به جریان رودخانه کنترل می کنند؛ بنابراین جریان شبیه سازی شده، مجموع خروجی های هر یک از این مخازن است.

۲۰۴ تحلیل و بررسی مدل

مدلی ساده به صورت دو پارامتری برای مدل هیدرولوژیکی «های مد» را می توان به صورت زیر به بیان کرد [۱۴]

$$F(c) = 1 - \left(1 - \frac{c(t)}{C_{max}}\right)^\beta \quad 0 \leq c(t) \leq C_{max}$$

$F(c)$: تابع «های مد»

C_{max} : حداکثر ظرفیت مخزن

β : ظرفیت رطوبت خاک که این پارامتر رابطه مستقیم با α دارد

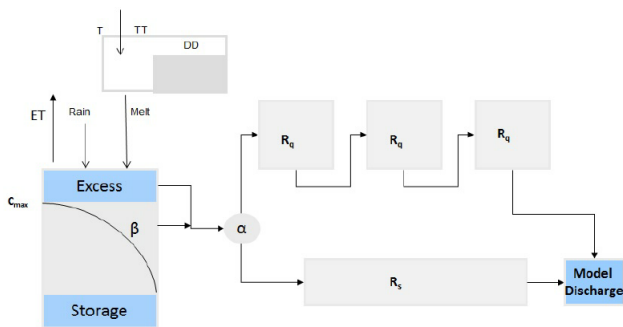
$c(t)$: ظرفیت مخزن در زمان t

α : مقدار افزاز آب اضافی به مخازن سریع و آهسته

R_s : مخزن آهسته

R_q : مخازن سریع

با توجه به شکل ۱ می توان به این نتیجه رسید پارامتر α بخشی از پارامتر β می باشد و هنگامی که رواناب بیش از ظرفیت C_{max} باشد باعث می شود مقادیر اضافی بارندگی را که ناشی از ذوب برف و بارندگی ایجاد می شود را در مخازن سریع و آهسته منتقل و هدایت کند.



شکل ۱. مدل «های مد» بعدی

همچنین می توان برای کالیبره کردن مدل نیز از رابطه ی زیر استفاده

کرد

$$RMSE = \left(\frac{1}{N} \sum_{t=1}^N (q_t^{sim} - q_t^{obs})^2 \right)^{\frac{1}{2}}$$

همان طور که مشاهده می شود q_t^{sim} و q_t^{obs} به ترتیب مقدار جریان

شبیه سازی شده و مقدار مشاهده شده در زمان t را نشان می دهند و N

تعداد مقادیر جریان می باشد. این طرح نشان می دهد که پذیرش ماهیت

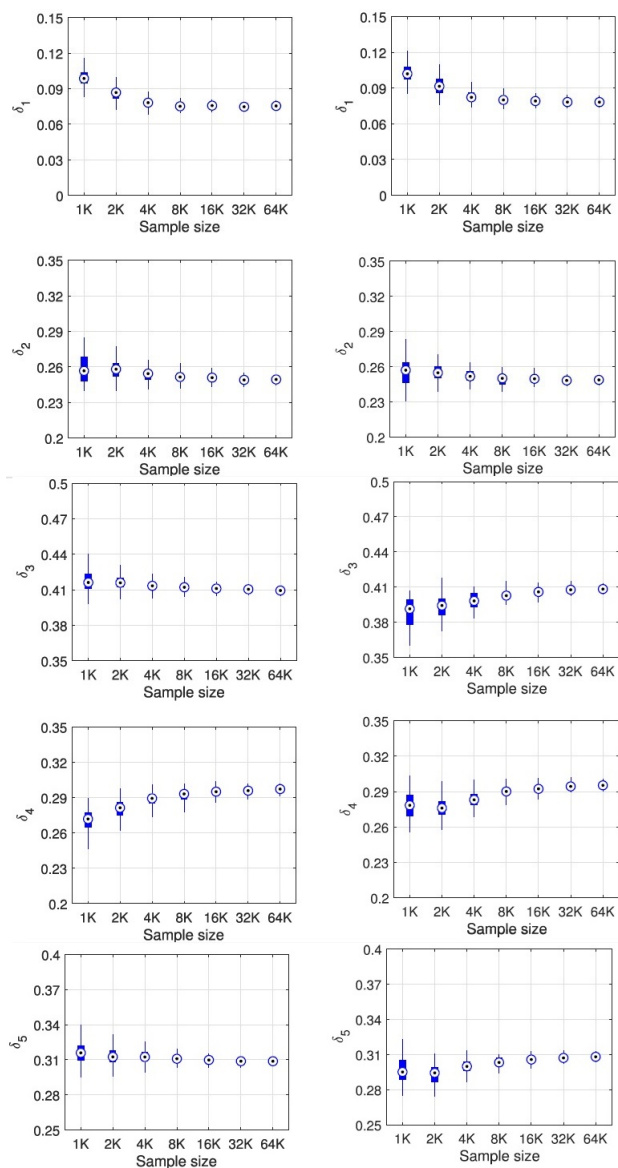
(ب) از یک توزیع یکنواخت استفاده شوند.

(۲) برای تعیین تعداد رده‌ها برای روش‌های پیشرو و معکوس از $N \sim N^{\ddagger}$ استفاده می‌شود که در آن N اندازه کل نمونه است [۴].

(۳) در روش معکوس، نمونه‌های خروجی با استفاده از روش $K - means$ به رده‌های مختلف تقسیم می‌شوند. همه این رده‌ها به صورت دوبه‌دو مجزا و اجتماع آن‌ها برابر با کل فضای خروجی مدل است.

(۴) در نهایت می‌توان شاخص‌های حساسیت را با توجه به اندازه نمونه و خروجی مدل برآورد کرد.

شکل‌های ۲ و ۳ برآورد شاخص‌های حساسیت را با توجه به افزایش اندازه نمونه‌ها برای خروجی‌های مدل با اندازه‌های به ترتیب ۶ و ۳۶ نشان می‌دهند.



شکل ۲. نمودار جعبه‌ای برای خروجی‌های ۶ بعدی

چند هدفه واسنجی و ادغام آن در فرایند مدل‌سازی، میزان اطلاعات بازیابی شده از باقی‌مانده‌های مدل را افزایش می‌دهند تا:

۱- تعداد پارامتر لازم برای تطابق با تمام جنبه‌های سری زمانی خروجی مشاهده‌شده را فراهم آورد (به‌طور جداگانه).

۲- قابلیت شناسایی پارامترهای مدل را افزایش دهد.

۳- مناسب بودن ساختار مدل برای نشان دادن سامانه طبیعی را ارزیابی کند (یعنی شناسایی نارسایی‌های ساختاری مدل).

برای بررسی جامع بین دو روش شاخص حساسیت (شاخص حساسیت معکوس و شاخص پیشرو) در مدل، ابتدا باید پارامترهای ورودی و خروجی مدل را تعیین کرد. در این مطالعه از اطلاعات بارش روزانه، تبخیر و تعرق به‌عنوان داده استفاده می‌شود. در این طرح پنج پارامتری که روی مدل «های‌مد» تأثیر می‌گذارد عبارت‌اند از: β , α , C_{max} , R_s و R_q که می‌توان آن‌ها را به‌عنوان پارامترهای ورودی در نظر گرفت (جدول ۱)؛ بنابراین یک بردار ۵ بعدی از پارامترهای ورودی تصادفی به‌صورت زیر به وجود می‌آید

$$X = (X_1, \dots, X_5) \quad i = 1, \dots, 5$$

جدول ۱. پارامترهای مدل «های‌مد»

پارامتر	حداقل	حداکثر
$C_{max}(X_1)$	۰	۲۹۷۰
$\beta(X_2)$	۰	۲
$\alpha(X_3)$	۰	۱
$R_s(X_4)$	۰	۰/۲
$R_q(X_5)$	۰/۲	۱

مقادیر و پارامترهای به‌دست‌آمده در جدول ۱ از وب‌سایت رسمی رود کارون به‌دست‌آمده است، همان‌طور که بیان شد، طرح ذخیره‌سازی آب در یک دوره ۳۱-ساله موردبررسی قرار می‌گیرد که مراحل بررسی آن به‌صورت زیر انجام می‌شود.

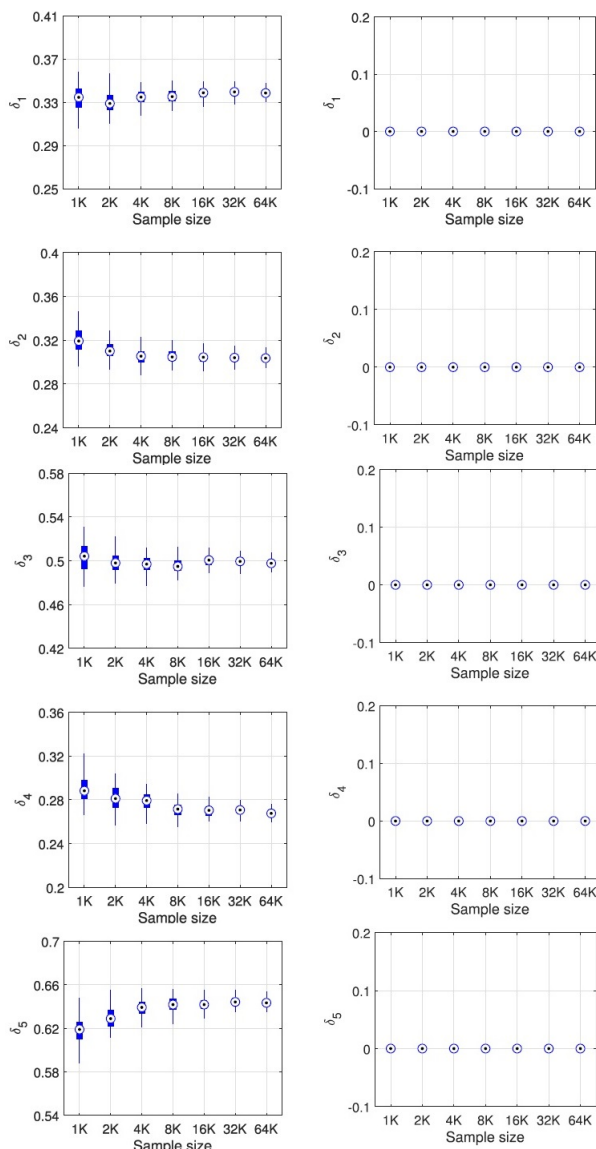
با استفاده از جریان‌های شبیه‌سازی و با توجه به تعداد روزهایی که می‌خواهیم بررسی انجام شود، خروجی مدل از آن به دست می‌آید؛ بنابراین با توجه به بررسی ۶۰ و ۱۰ روزه، خروجی مدل ۶ و ۳۶ بعدی است

$$Y = (Y_1, \dots, Y_6) \quad m = 1, \dots, 6$$

$$Y = (Y_1, \dots, Y_{36}) \quad m = 1, \dots, 36$$

فرض‌های آزمون به ترتیب عبارت است.

(الف) داده‌ها از یک مجموعه نمونه تصادفی انتخاب شوند.



شکل ۳. نمودار جعبه‌ای برای خروجی‌های ۳۶ بعدی

نمودارهای سمت چپ در شکل‌های ۲ و ۳ بیانگر نتایج به‌دست‌آمده با استفاده از روش تحلیل معکوس و نمودارهای سمت راست، در شکل‌های ۲ و ۳ بیانگر نتایج به‌دست‌آمده با استفاده از روش تحلیل پیشرو است [۴]. نمودارهای جعبه‌ای با ۳۰۰ بار اجرای مستقل از روش‌های تحلیل بیان‌شده به‌دست‌آمده‌اند. همان‌طور که مشاهده می‌شود، بالا و پایین این جعبه‌ها صدک‌های ۲۵ و ۷۵ مربوط به نتایج می‌باشد.

شکل ۲، بیانگر این است که این دو روش حساسیت با افزایش اندازه نمونه برآورد مشابهی را برای مدل‌های با تعداد خروجی ۶ نشان می‌دهند. همچنین جدول ۲، نتایج متوسط نهایی مربوط به این دو روش با اندازه نمونه ۲۰۰۰۰ را نشان می‌دهد. با توجه به نتایج به‌دست‌آمده در جدول ۲، می‌توان به این نتیجه دست‌یافت که هنگامی که مدل تعداد کمی خروجی داشته باشند این دو روش نتایج یکسانی را ارائه می‌دهند.

جدول ۲. برآوردهای متوسط نهایی $\hat{\delta}_i$ (پارامتری)

روش	$\hat{\delta}_5$	$\hat{\delta}_4$	$\hat{\delta}_3$	$\hat{\delta}_2$	$\hat{\delta}_1$
معکوس	۰/۳۰۳	۰/۲۹۲	۰/۴۰۴	۰/۲۴۴	۰/۰۷۱
پیشرو	۰/۳۰۳	۰/۲۹۰	۰/۴۰۵	۰/۲۴۵	۰/۰۷۳

۵ بحث و نتیجه‌گیری

با توجه به مطالب بیان‌شده، می‌توان به این نتیجه رسید: هنگامی که تعداد خروجی‌های مدل زیاد باشد، برای به دست آوردن نتایج می‌بایست از روش معکوس برای تحلیل مدل استفاده شود؛ به این دلیل که در این روش از تابع چگالی احتمال‌های یک‌بعدی استفاده می‌شود؛ درحالی‌که در روش پیشرو برای برآورد نیاز به تابع چگالی احتمال‌های چندبعدی دارد، همین امر باعث سادگی و بهتر بودن این روش نسبت به سایر روش‌های مشابه مانند پیشرو، کپول تاک و مانند آن است. هم‌چنین لازم به ذکر است هنگامی که تعداد خروجی‌های مدل کوچک باشد، لزومی ندارد از کدام روش استفاده شود؛ به این دلیل که نتایج یکسانی را نشان می‌دهند.

در شکل ۳، نتایج مربوط به ۳۶ خروجی را نشان می‌دهد. روش معکوس همچنان با افزایش اندازه نمونه همگرا است؛ درحالی‌که در روش پیشرو برای همه‌ی پارامترها برآورد را نزدیک صفر نشان می‌دهد؛ لذا برای برآورد روش معکوس می‌بایست از تابع چگالی احتمال‌های یک‌بعدی استفاده شود، درحالی‌که روش پیشرو نیاز به برآورد تابع چگالی احتمال‌های ۳۶ بعدی دارد؛ بنابراین برآورد دقیق چنین تابع چگالی احتمال‌هایی با ابعاد بالا، کار آسانی نیست و تفاضل میان تابع چگالی احتمال‌های شرطی و غیرشرطی را نمی‌توان به راحتی به دست آورد. از این رو برآوردهای روش پیشرو همواره نزدیک به صفر است.

مراجع

- [1] Bergstroem, S. (1975), The development of a snow routine for the HBV-2 model, *Hydrology Research*, **6**, 73.
- [2] Beven, K. J. (2011), *Rainfall-runoff modelling: the primer*, John Wiley & Sons, West Sussex.
- [3] Borgonovo, E. (2007), A new uncertainty importance measure, *Reliability Engineering & System Safety*, **92**, 771–784.
- [4] Borgonovo, E., Hazen, G. B., and Plischke, E. (2016), A common rationale for global sensitivity measures and their estimation, *Risk Analysis*, **36**, 1871–1895.
- [5] Botev, Z. I., Grotowski, J. F., Kroese, D. P., et al. (2010), Kernel density estimation via diffusion, *The annals of Statistics*, **38**, 2916–2957.
- [6] Da Veiga, S. (2015), Global sensitivity analysis with dependence measures, *Journal of Statistical Computation and Simulation*, **85**, 1283–1305.
- [7] Freedman, D. and Diaconis, P. (1981), On the histogram as a density estimator: L² theory, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **57**, 453–476.
- [8] Kollat, J., Reed, P., and Wagener, T. (2012), When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resources Research*, **48**(3).
- [9] Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994), Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, **263**, 641–646.
- [10] Plischke, E., Borgonovo, E., and Smith, C. L. (2013), Global sensitivity measures from given data, *European Journal of Operational Research*, **226**, 536–550.
- [11] Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010), Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer physics communications*, **181**, 259–270.
- [12] Sobol, I. M. (1993), Sensitivity analysis for non-linear mathematical models, *Mathematical modelling and computational experiment*, **1**, 407–414.
- [13] Spear, R. C., Grieb, T. M., and Shang, N. (1994), Parameter uncertainty and interaction in complex environmental models, *Water Resources Research*, **30**, 3159–3169.
- [14] Wagener, T., Boyle, D. P., Lees, M. J., Wheatler, H. S., Gupta, H. V., and Sorooshian, S. (2001), A framework for development and application of hydrological models, *Hydrology and Earth System Sciences*, **5**, 13–26.
- [15] Zhai, Q., Yang, J., Xie, M., Zhao, Y. (2014). Generalized moment-independent importance measures based on Minkowski distance. *European Journal of Operational Research*, **239**(2), 449-455.

Sensitivity analysis of input variables in generalized linear models

Mohammad Khorasani¹ Dr. Farzad Eskandari²

Abstract:

In today's world, using the statistical modeling process, natural phenomena can be used to analyze and predict the events under study. Many hydrological modeling methods do not make the best use of available information because hydrological models show a wide range of environmental processes that complex the model. In particular, when predicting, parameters affect the performance of statistical models. In many risk assessment issues, the presence of uncertainty in the parameters leads to uncertainty in predicting the model. Global sensitivity analysis is a tool used to show uncertainty and is used in decision making, risk assessment, model simplification and so on. Minkowski distance sensitivity analysis and regional sensitivity analysis are two broad methods that can work with a given sample set of model input-output pair. One significant difference between them is that minkowski distance sensitivity analysis analyzes output distributions conditional on input values (forward), while regional sensitivity analysis analyzes input distributions conditional on output values (reverse). In this dissertation, we study the relationship between these two approaches and show that regional sensitivity analysis (reverse), when focusing on probability density functions of input, converges towards minkowski distance sensitivity analysis (forward) as the number of classes for conditioning model outputs in the reverse method increases. Similar to the existing general form of forward sensitivity indices, we derive a general form of the reverse sensitivity indices and provide the corresponding reverse given-data method. Finally, the sensitivity analysis of a water storage design with high dimensions of the model outputs is performed.

Keywords: Global sensitivity analysis, Minkowski distance sensitivity analysis, regional sensitivity analysis, Classification of output.

¹ Allameh Tabataba'i University

² Allameh Tabataba'i University