تحلیل حساسیت واریانس‌مبنا در مدل‌های تعیینی

مجید جانفدا،[۱] داود شاهسونی [۲]

تاریخ دریافت: ۱۳۹۱/۱۰/۱۹
تاریخ پذیرش: ۱۳۹۵/۱۲/۲۵

چکیده:

امکان مطالعهٔ بسیاری از پدیده‌های علمی و طبیعی در شرایط آزمایشگاهی میسر نیست و لذا آنها را در قالب مدل‌های ریاضی بیان و با کدهای کامپیوتری پیچیده، شبیه‌سازی می‌کنند. اجرای مدل کامپیوتری با ورودی‌های متفاوت را آزمایش کامپیوتری می‌نامند. مباحث آماری، در آزمایش‌های کامپیوتری از جایگاه ویژه‌ای برخوردار است. در این مقاله ضمن تشریح ساختار این مدل‌ها، به معرفی وبیان اهمیت تحلیل حساسیت واریانس‌مبنا می‌پردازیم. تحلیل حساسیت، مجموعهٔ روش‌هایی است که اثربخش بودن پارامترهای ورودی را بر خروجی مدل با شاخص‌های حساسیت تعیین می‌کنند. این شاخص‌ها بر اساس مفاهیم واریانس‌های شرطی بیان می‌شوند. از آن جا که شکل ریاضی این مدل‌ها، به‌صورت صریح مشخص نیست؛ مسئلهٔ برآورد این شاخص‌ها با روش‌های مونته‌کارلومبنا مطرح می‌گردد. از طرف دیگر زمان اجرا، چالش جدی در مدل‌های کامپیوتری محسوب می‌شود. طرح نقاط آزمایش ویژه‌ای مبتنی بر اعداد شبه‌تصادفی، برای کاهش زمان اجرای مدل پیشنهاد شده است. به‌منظور پرداختن به جنبهٔ عملی، از مدل INCA-N که میزان آلایندگی نیتروژن ورودی به آب رودخانه‌ها را شبیه‌سازی می‌کند، استفاده شده است تا بتوان با شاخص‌های حساسیت متغیرهای تأثیرگذار بر این عامل تهدیدکنندهٔ سلامت انسان و محیط زیست را شناسایی کرد.

واژه‌های کلیدی: مدل تعیینی، تحلیل حساسیت واریانس‌مبنا، اعداد شبه‌تصادفی، مدل INCA-N.

۱ مقدمه

امکان مطالعه و بررسی بسیاری از پدیده‌های علمی در شرایط آزمایشگاهی بنا بر هزینه‌های گزاف اجرایی، طولانی بودن زمان اجرا و عملی نبودن انجام آزمایش به‌خصوص در پدیده‌های زیست‌محیطی، مقدور نیست. محققان این پدیده‌ها را با مدل‌های ریاضی بیان می‌کنند. این مدل‌ها شامل معادلات دیفرانسیل معمولی با مشتقات جزئی خطی یا غیرخطی می‌باشند. برنامه یا کد کامپیوتری را که قادر به حل عددی این دستگاه معادلات پیچیده باشد، مدل کامپیوتری و اجرای مدل با ورودی‌های متفاوت را آزمایش کامپیوتری گویند [۵]. خروجی آزمایش‌های کامپیوتری بر خلاف آزمایش‌های فیزیکی، فقط تابعی از متغیرهای ورودی مدل است و متأثر از خطای تصادفی نمی‌باشد. بدین معنا که اجرای مدل با ورودی‌های همانند، خروجی یکسانی را تولید می‌کند. از این رو مدل‌های کامپیوتری، مدل‌های تعیینی[۳] نیز نامیده می‌شوند. در واقع اگر $X_k, \ldots, X_1$ متغیرهای ورودی مستقل با توزیع احتمال $P_i(x_i)$، $i = 1, \ldots, k$ باشند، خروجی مدل کامپیوتری ($f(x)$) تابعی چندمتغیره در فضای k-بعدی به‌صورت $Y = f(x_1, \ldots, x_k)$ خواهد بود که معادلهٔ ریاضی آن به‌طور صریح مشخص نیست، اما امکان محاسبهٔ خروجی مدل برای هر دسته از ورودی‌ها وجود دارد.

فقدان خطای تصادفی در آزمایش‌های کامپیوتری منجر به عدم کارایی مفاهیم کلاسیک طرح آزمایش، نظیر بلوک‌بندی، تکرار و تصادفی‌سازی می‌شود. ساختار و ویژگی‌های مدل

[۱] کارشناس ارشد آمار دانشگاه صنعتی شاهرود، ایران
[۲] گروه آمار دانشگاه صنعتی شاهرود، ایران

[۳] deterministic model
[۴] surrogate model
[۵] sensitivity analysis

کامپیوتری سبب مطرح شدن موضوعاتی از قبیل طرح آزمایش، مدل‌های کامپیوتری، انتخاب مدل‌جانشین،[۴] عدم قطعیت و تحلیل حساسیت[۵] می‌شود که نقش آمار در آن ها حیاتی است [۱۳]. طرح نقاط آزمایش مدل‌های کامپیوتری با توجه به مسئلهٔ زمان اجرای مدل اهمیت فراوانی دارد. فقدان طرح نقاط آزمایش مناسب علاوه بر تحمیل هزینهٔ اجرای مدل، ممکن است موجب رخداد خطای نوع اول (بی‌اهمیت دانستن یک متغیر بااهمیت) و خطای نوع دوم (بااهمیت تلقی کردن یک متغیر بی‌اهمیت) شود. همچنین انتخاب مدل جانشین تابع f به‌منظور کاهش هزینهٔ زمان اجرا با روش‌های پیشرفتهٔ آماری و یادگیری ماشین انجام می‌شود [۵]. تحلیل عدم قطعیت، خود یک مسئلهٔ آماری است. شناسایی عوامل تأثیرگذار بر خروجی مدل توسط تحلیل حساسیت، در عین پیچیدگی مدل، کثرت عوامل ورودی، اثرهای متقابل آن ها و زمان‌بر بودن اجرای مدل از اهمیت ویژه‌ای برخوردار است و موجب ساده‌سازی مدل و تفسیر بهتر نتایج در مسائل مدیریت ریسک و تصمیم‌گیری می‌شود [۱۸]. در این مقاله از میان کاربردهای یاد شده به موضوع تحلیل حساسیت پرداخته‌ایم که لازمهٔ آن مطالعهٔ مقدماتی طرح آزمایش مدل‌های کامپیوتری است.

۲ تحلیل حساسیت

تعدد و تنوع متغیرهای ورودی در مدل‌های کامپیوتری سبب پیچیدگی آنها می‌شود. لذا مسئلهٔ ساده‌سازی مدل حائز اهمیت است. این موضوع با شناسایی متغیرهای با اهمیت و کنار گذاشتن متغیرهای بی‌اهمیت و در قالب تحلیل حساسیت بررسی می‌شود. اندازه‌های حساسیت، میزان اهمیت هر یک از متغیرهای ورودی را مشخص می‌کنند. تأثیرگذار بودن یا بی‌تأثیر بودن متغیرها و اثرهای متقابل آن‌ها در مدیریت بحران و ریسک از جایگاه ویژه‌ای برخوردار است. از طرف دیگر، هزینه‌های تحمیلی، نگهداری و کنترل متغیرها تصمیم‌گیری را با چالش جدی مواجه می‌سازد. از این رو برای تصمیم‌گیری بهتر، به‌ویژه در مواقع بحرانی، باید متغیرهای بااهمیت قابل کنترل و غیر قابل کنترل، کم‌اهمیت و بی‌اهمیت شناسایی شوند. از این رو تحلیل حساسیت از دیدگاه مدیریتی موجب می‌شود تا تصمیمات اتخاذ شده از پشتوانهٔ منطقی بیشتری برخودار باشند و نتایج حاصل از این تصمیم‌گیری تا حد امکان بهینه شود. می‌توان از این ابزار برای ساده‌سازی مدل، کشف خطاهای فنی مدل و شناسایی نواحی بحرانی مدل بهره جست. لذا تحلیل حساسیت، رویکردی مهم در مدیریت و پژوهش تلقی می‌شود.

تحلیل حساسیت از دیدگاه آماری به‌صورت «مطالعه چگونگی تخصیص عدم‌قطعیت خروجی یک مدل به منابع عدم قطعیت موجود در ورودی‌های مدل» تعریف می‌شود [۱۶]. به بیان دیگر تحلیل حساسیت، به‌منظور ارائهٔ روشی برای تعیین تأثیر پارامترهای ورودی بر خروجی مدل مطرح می‌شود [۱۹]. تحلیل حساسیت واریانس‌مبنا [۶] از کارایی مناسبی در تحلیل حساسیت مدل‌های خطی و غیرخطی برخوردار است. میزان اهمیت هر یک از متغیرهای ورودی بر اساس شاخص‌های حساسیت مبتنی بر مفاهیم واریانس شرطی تعیین می‌شود [۱]. این روش نخستین بار در دهه ۱۹۷۰ مطرح شد [۸،۴،۳،۲]. در همین زمینه شاخص‌های حساسیت دیگری تحت عنوان اندازهٔ حساسیت یا نسبت همبستگی پیشنهاد شده است [۱۴،۱۲،۱۰،۷،۶].

۱.۲ شاخص حساسیت مرتبهٔ اول

بنا بر رابطهٔ تجزیهٔ واریانس بر حسب عامل شرطی X_i داریم

$$V(Y) = V_{X_i}(E_{X_{-i}}(Y|X_i)) + E_{X_i}(V_{X_{-i}}(Y|X_i)), \quad (۱)$$

که در آن نماد X_{-i} به‌معنای همهٔ متغیرها به‌جز متغیر X_i است. اندازه حساسیت خروجی مدل (Y) نسبت به متغیر X_i مطابق رابطهٔ (۲) بیان می‌شود:

$$V_i = V(Y) - E_{X_i}(V_{X_{-i}}(Y|X_i = x_i)) \quad (۲)$$

دلیل این تعریف آن است که اگر متغیر X_i را در مقدار x_i ثابت نگه داریم، می‌توانیم تأثیر آن را بر خروجی مدل با

$$V(Y|X_i = x_i)$$

بیان کنیم. اما از آن‌جا که مقدار x_i می‌تواند در تکیه‌گاه X_i تغییر کند، تأثیر تثبیت متغیر X_i به‌صورت $E(V(Y|X_i = x_i))$ و کاهش در عدم قطعیت برابر با $V(Y) - E(V(Y|X_i = x_i))$ خواهد بود. این کمیّت به مقیاس Y وابسته بوده که با تقسیم آن

[۶] variance based sensitivity analysis

[۷] frist oredr sensitivity index

بر واریانس Y، عدد مطلق به‌دست‌آمده (رابطهٔ (۳)) را شاخص حساسیت مرتبهٔ اول [۷] یا اثر اصلی متغیر iام می‌نامند.

$$S_i = \frac{V_i}{V(Y)} = \frac{V(E(Y|X_i = x_i))}{V(Y)}, \qquad (۳)$$

که همواره $0 \le S_i \le 1$ است. مسلماً هر قدر اهمیت متغیر X_i در خروجی مدل کمتر باشد، واریانس Y کمتر دستخوش تغییر قرار می‌گیرد و $E(V(Y|X_i = x_i))$ به $V(Y)$ نزدیک‌تر بوده، $V_i = V(Y) - E(V(Y|X_i = x_i))$ دارای مقدار کمتری خواهد بود (و برعکس). می‌توان شاخص حساسیت مرتبهٔ دوم را با ثابت نگه داشتن متغیرهای X_i و X_j مطابق با رابطهٔ زیر اندازه‌گیری کرد.

$$S_{ij} = \frac{V_{ij}}{V(Y)} = \frac{V(E(Y|X_i, X_j))}{V(Y)} - S_i - S_j, \qquad (۴)$$

که در آن $V(E(Y|X_i, X_j))$ اندازهٔ اثر توأم متغیرهای X_i و X_j بر خروجی مدل (Y) است. به همین ترتیب، شاخص‌های مراتب بالاتر تعریف می‌شوند. اهمیت شاخص‌های حساسیت مراتب بالاتر زمانی آشکار می‌شود که نتوان حساسیت خروجی مدل را تنها با شاخص حساسیت مرتبهٔ اول تبیین کرد. ثابت می شود که مجموع همهٔ اثرهای مرتبهٔ اول، دوم و مراتب بالاتر برابر با یک است [۱]؛ یعنی

$$\sum_{i=1}^{k} S_i + \sum_{1 \le i < j \le k} S_{ij} + \cdots + S_{1,2,\ldots,k} = 1$$

این بدان معنا است که اگر مجموع اثرهای اصلی متغیرها $\sum_{i=1}^{k} S_i = 1$ باشد، اثرهای متقابل بین متغیرها وجود نخواهند داشت و مدل مورد بررسی را جمعی [۸] گویند، اگر $\sum_{i=1}^{k} S_i \le 1$ باشد، مدل را غیرجمعی نامند. شایان ذکر است که هزینهٔ محاسباتی اثرهای متقابل ورودی‌های مدل با افزایش تعداد ورودی‌ها، به‌صورت نمایی افزایش می‌یابد و عملاً محاسبهٔ شاخص‌های حساسیت مرتبهٔ دوم، سوم و مراتب بالاتر ناممکن می‌شود و مقرون به‌صرفه نمی‌باشد. در نتیجه بسیاری از محققان به‌دنبال شاخص جایگزینی بودند که علاوه بر کاهش هزینه‌های محاسباتی، حاوی اطلاعات مفیدی در خصوص اثرهای متقابل باشد. سرانجام هما و سالتلی [۶] شاخصی را تحت عنوان شاخص حساسیت کل معرفی کردند.

۲.۲ اثر کل

هما و سالتلی [۶] این شاخص را تحت عنوان شاخص حساسیت کل [۹] یا اثر کل معرفی کردند. این شاخص را با نماد S_T نشان می‌دهیم و شامل همهٔ اثرهایی است که متغیر مربوط در آن دخیل است. به‌عنوان مثال اگر خروجی مدل، تابعی از سه متغیر باشد، اثر کل X_1 به‌صورت $S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$ تعریف می‌شود.

به‌منظور ارائهٔ یک اندازه برای اثر کل متغیر X_i، واریانس خروجی مدل را به نوعی دیگر و با شرط ثابت نگه داشتن همهٔ متغیرهای ورودی به‌جز X_i، یعنی X_{-i}، مطابق با رابطهٔ (۵) تجزیه می‌کنیم.

$$V(Y) = V(E(Y|X_{-i})) + E(V(Y|X_{-i})). \qquad (۵)$$

بنا بر تعریف شاخص حساسیت مرتبهٔ اول می‌توان $V(E(Y|X_{-i}))$ را به‌عنوان اثر مرتبهٔ اول X_{-i} در نظر گرفت. لذا $E(V(Y|X_{-i})) = V(Y) - V(E(Y|X_{-i}))$ دربرگیرندهٔ همهٔ اثرهایی است که متغیر X_i به‌نوعی در آن مشارکت دارد. با تقسیم طرفین معادله بر $V(Y)$ شاخص حساسیت کل مطابق با رابطهٔ (۶) به‌دست می‌آید.

$$S_{T_i} = \frac{E(V(Y|X_{-i}))}{V(Y)} = 1 - \frac{V(E(Y|X_{-i}))}{V(Y)}. \qquad (۶)$$

با استفاده از اثر کل و اثر اصلی می‌توان اثرهای بین متغیر X_i و سایر متغیرهای ورودی را اندازه‌گیری کرد. برای این منظور کافی است مقدار $S_{T_i} - S_i$ محاسبه شود.

۳.۲ محاسبهٔ شاخص‌حساسیت مرتبهٔ اول

اگر X_jها مستقل از هم و هر یک از آن ها دارای توزیع احتمال $P_j(x_j)$ باشد، میانگین و واریانس Y و واریانس‌های شرطی به‌صورت زیر محاسبه می‌شوند.

$$E(Y) = \int_{\Omega^k} YP(\mathbf{x})d\mathbf{x} = \int \ldots \int f(x_1, \ldots, x_k) \times \prod_{i=1}^{k} P_i(x_i)dx_i$$

$$V(Y) = \int \ldots \int f^2(x_1, \ldots, x_k) \prod_{i=1}^{k} P_i(x_i)dx_i - E^2(Y)$$

[۸] additive

[۹] total sensitivity index

$$V(E(Y|X_j)) = \int E^2(Y|X_j = x_j)P_j(x_j)dx_j - E^2(Y)$$
$$= U_j - E^2(Y) \quad (7)$$

که در آن

$$U_j = \int\int \ldots (\int f(x_1, \ldots x_k) \prod_{i=1, i\neq j}^{k} P_i(x_i)dx_i)^2 \times P_j(x_j)dx_j.$$

و $\mathbf{X} = (X_1, \ldots, X_k)$ است. اکنون می‌توان شاخص حساسیت مرتبهٔ اول را به‌صورت رابطهٔ (۸) بازنویسی کرد

$$S_j = \frac{V(E(Y|X_j))}{V(Y)} = \frac{U_j - E^2(Y)}{V(Y)}. \quad (8)$$

ایشیگامی و هما [۸] برای سهولت محاسبهٔ U_j پیشنهاد کردند که‌توان دوم یک انتگرال، به‌صورت حاصل ضرب دو انتگرال با دو آرگومان متفاوت محاسبه شود، یعنی

$$\int E^2(Y|X_j = x_j)P_j(x_j)dx_j =$$
$$\int\int \ldots \int f(x_1, \ldots, x_j, \ldots, x_k) \prod_{i=1, i\neq j}^{k} P_i(x_i)dx_i$$
$$\times \int \ldots \int f(x'_1, \ldots x_j, \ldots, x'_k) \prod_{i=1, i\neq j}^{k} P'_i(x'_i)dx'_i P_j(x_j)dx_j$$
$$= \int \ldots \int f(x_1, \ldots x_k) \prod_{i=1}^{k} P_i(x_i)dx_i$$
$$\times f(x'_1, \ldots x_j, \ldots, x'_k) \prod_{i=1, i\neq j}^{k} P'_i(x'_i)dx'_i. \quad (9)$$

۴.۲ محاسبهٔ شاخص حساسیت کل

نحوهٔ محاسبهٔ $V(E(Y|X_{-j}))$ در شاخص حساسیت کل

$$S_{T_j} = 1 - \frac{V(E(Y|X_{-j}))}{V(Y)}$$

به‌شرح زیر است. در این‌جا نماد X_{-j} به‌معنای همهٔ متغیرها به‌جز متغیر jام است

$$V(E(Y|X_{-j})) = E(E^2(Y|X_{-j})) - E^2(Y)$$
$$= \int \ldots (\int f(x_1, \ldots, x_j, \ldots, x_k)P_j(x_j)dx_j)^2 \times \prod_{i=1, i\neq j}^{k} P_i(x_i)dx_i - E^2(Y). \quad (10)$$

با در نظر گرفتن جملهٔ اول عبارت فوق با نماد U_{-j}، شاخص حساسیت کل به‌صورت زیر بازنویسی می‌شود

$$S_{T_j} = 1 - \frac{V(E(Y|X_{-j}))}{V(Y)} = 1 - \frac{U_{-j} - E^2(Y)}{V(Y)}. \quad (11)$$

در محاسبهٔ شاخص‌های حساسیت، دو چالش اساسی وجود دارد:

الف) بعد فضای انتگرال برابر با $2K - 1$ است.

ب) رابطهٔ ریاضی ورودی - خروجی (f) به‌طور صریح مشخص نیست.

این چالش‌ها سبب طرح مسئلهٔ تقریب انتگرال‌ها و در نتیجه برآورد شاخص‌های حساسیت می‌شوند.

۳ برآورد شاخص‌های حساسیت

۱.۳ برآورد شاخص حساسیت مرتبهٔ اول

برای برآورد اثر اصلی S_j در رابطهٔ (۸) می‌باید U_j، $E(Y)$ و $V(Y)$ برآورد شوند.

الف) برآورد U_j

ساختار انتگرال U_j در رابطهٔ (۹) به‌گونه‌ای است که برای تقریب انتگرال با روش مونته‌کارلو، این ایده به‌وجود می‌آید که دو ماتریس A و B با N سطر توسط اعداد تصادفی تولید شوند که هر سطر این دو ماتریس، معرف یک بردار kتایی از مقادیر شبیه‌سازی‌شده برای متغیرهای $\mathbf{X} = (X_1, \ldots, X_k)$ به‌صورت زیر باشد [۸]

$$A = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \ldots & x_{Nk} \end{bmatrix},$$

$$B = \begin{bmatrix} x'_{11} & x'_{12} & \ldots & x'_{1k} \\ x'_{21} & x'_{22} & \ldots & x'_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{N1} & x'_{N2} & \ldots & x'_{Nk} \end{bmatrix}.$$

از آن‌جا که در عبارت $f(x_1, \ldots, x_k)f(x'_1, \ldots, x_j, \ldots, x'_k)$ مقدار متغیر jام بردار $(x'_1, \ldots, x_j, \ldots, x'_k)$ از بردار $(x_1, \ldots, x_j, \ldots, x_k)$ انتخاب شده است، باید به‌ازای هر یک از X_iها ماتریس جدیدی بر مبنای ماتریس B تولید شود که فقط ستون jام آن برگرفته از ستون jام ماتریس A باشند. بدین منظور، ماتریس‌های جدید $(C_1, \ldots, C_k)$ را طوری می‌سازیم که

تنها تفاوت ماتریس B و C_j در ستون jام باشد؛ یعنی

$$C_j = \begin{bmatrix} x'_{11} & x'_{12} & \dots & x_{1j} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x_{2j} & \dots & x'_{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x'_{N1} & x'_{N2} & \dots & x_{Nj} & \dots & x'_{Nk} \end{bmatrix}.$$

با روش انتگرال‌گیری مونته‌کارلو، برآورد U_j به‌صورت زیر خواهد بود [۱۵].

$$\begin{aligned} \hat{U}_j =& \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},\dots,x_{rk})\times \\ & f(x'_{r1},\dots,x'_{r(j-1)},x_{rj},x'_{r(j+1)},\dots,x'_{rk}) \\ =& \frac{1}{N-1}\sum_{r=1}^{N} f(A)_r f(C)_r \end{aligned} \qquad (۱۲)$$

که در آن N اندازهٔ نمونه است.

ب) برآورد $E^2(Y)$

به‌منظور برآورد $E^2(Y)$ در رابطهٔ $\hat{S}_j$ می‌توان از میانگین ماتریس A یا B یا تلفیقی از آن دو ماتریس استفاده کرد [۱۴].

$$\hat{E}^2(Y) = (\hat{E}(y))^2 = \left(\frac{1}{N}\sum_{j=1}^{N} y_A^j\right)^2,$$

$$\hat{E}^2(Y) = \frac{1}{N}\sum_{j=1}^{N} y_A^j y_B^j$$

ج) برآورد $V(Y)$

برای برآورد $V(Y)$ می‌توان از ماتریس A به‌صورت زیر استفاده کرد.

$$\hat{V}(Y) = \frac{1}{N}\sum_{j=1}^{N}(y_A^j)^2 - \hat{E}^2(Y).$$

بنا بر آن‌چه گفته شد، برآورد اثر اصلی مستلزم اجرا و محاسبهٔ خروجی مدل به‌ازای هر یک از N سطر ماتریس‌های A، B و $C_1,...,C_k$ بوده که تعداد کل دفعات اجرای مدل برای برآورد اثر اصلی همه متغیرها برابر با $N(K+2)$ است. اگر زمان هر اجرای مدل را t فرض کنیم، زمان کل مورد نیاز برای برآورد اثرات اصلی برابر با $t \times N(K+2)$ خواهد بود، که با توجه به ثابت بودن t و K، نقش N بسیار مهم است.

۲.۳ برآورد شاخص حساسیت کل

از آن‌جا که برآوردگرهای $E(Y)$ و $V(Y)$ در بندهای ب و ج بخش قبل معرفی شدند، برای برآورد کردن شاخص حساسیت کل در رابطه (۱۱) کافیست یک راهکار محاسباتی برای عبارت

$$U_{-j} = \int\dots\int f(x_1,\dots,x_j,\dots,x_k)\prod_{i=1,i\neq j}^{k} P_i^2(x_i)dx_i$$

$$\times f(x_1,\dots,x'_j,\dots,x_k)P_j(x_j)dx_i\ P_j(x'_j)dx'_j$$

پیشنهاد شود. با توجه به ماتریس‌های A و B در عبارت $f(x_1,\dots,x'_j,\dots,x_k)$ لازم است به‌ازای هر یک از X_iها ماتریس جدیدی بر مبنای ماتریس A تولید شود که فقط ستون jام آن بر گرفته از ماتریس B باشد، یعنی

$$D_j = \begin{bmatrix} x_{11} & x_{12} & \dots & x'_{1j} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x'_{2j} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x'_{Nj} & \dots & x_{Nk} \end{bmatrix}.$$

بنا بر این با روش انتگرال‌گیری مونت‌کارلو U_{-j} مطابق زیر برآورد می شود.

$$\begin{aligned} \hat{U}_{-j} =& \frac{1}{N-1}\sum_{r=1}^{N} f(x_{r1},\dots,x_{rk})\times \\ & f(x_{r1},\dots,x_{r(j-1)},x'_{rj},x_{r(j+1)},\dots,x_{rk}) \\ =& \frac{1}{N-1}\sum_{r=1}^{N} f(A)_r f(D)_r \end{aligned} \qquad (۱۳)$$

هزینهٔ محاسبهٔ هر U_{-j} برابر با N و در نهایت هزینهٔ محاسبهٔ شاخص کل همهٔ متغیرهای ورودی، $N \times K$ خواهد بود. موارد بیان‌شده در فصل ۳ را می‌توان در الگوریتم زیر خلاصه کرد.

۴ الگوریتم برآورد اثر اصلی و اثر کل

سالتلی [۶] برای برآورد شاخص‌های حساسیت، راهکار زیر را معرفی کرد.

۱. ماتریسی به ابعاد $N \times 2K$ از اعداد تصادفی تولید و دو ماتریس A و B را که هر کدام شامل نیمی از نمونه‌های تولید شده هستند در نظر می‌گیریم. در این جا K تعداد متغیرهای ورودی مدل و N اندازهٔ نمونه است. پیشنهاد سبل [۲۰]، استفاده از اعداد شبه‌تصادفی[۱۰] برای شبیه‌سازی است.

.

[۱۰] Quasi random number

$$A=\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix},$$

$$B=\begin{bmatrix} x_{1(k+1)} & x_{1(k+2)} & \dots & x_{1(2k)} \\ x_{2(K+1)} & x_{2(K+2)} & \dots & x_{2(2k)} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N(K+1)} & x_{N(K+2)} & \dots & x_{N(2k)} \end{bmatrix}.$$

۲. ماتریس‌های جدید C_i را با ترکیب دو ماتریس A و B می‌سازیم. ماتریس C_i همان ماتریس B است که ستون iام آن از ماتریس A گرفته شده است.

۳. خروجی مدل را برای همهٔ سطرهای ماتریس‌های B، A و $C_1, ..., C_k$ محاسبه می‌کنیم تا بردارهای خروجی Nتایی

$$Y_A = f(A), \quad Y_B = f(B), \quad Y_{C_i} = f(C_i), \quad i = 1, \dots, k$$

به‌دست آیند.

۴. برآورد شاخص حساسیت مرتبهٔ اول و کل هر یک از متغیرهای ورودی مدل، مطابق با رابطه های (۱۵) و (۱۶) بیان می‌شوند.

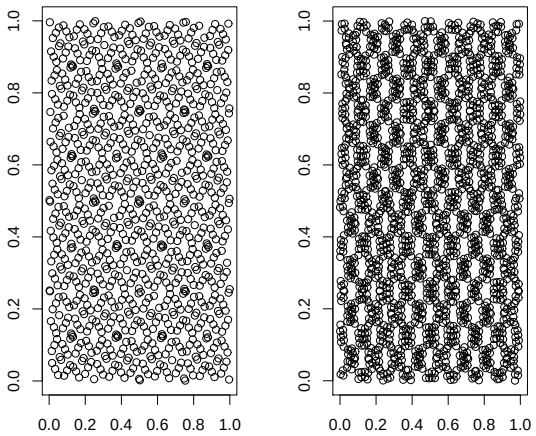
$$\hat{S}_i = \frac{\hat{E}(Y|X_i)}{\hat{V}(Y)} = \frac{\frac{1}{N}\sum_{j=1}^{N} y_A^j . y_{C_i}^j - f_0^2}{\frac{1}{N}(\sum_{j=1}^{N} y_A^j)^2 - f_0^2} \qquad (۱۴)$$

$$\hat{S}_{T_i} = 1 - \frac{\hat{E}(Y|X_{-i})}{\hat{V}(Y)} = 1 - \frac{\frac{1}{N}\sum_{j=1}^{N} y_B^j . y_{C_i}^j - f_0^2}{\frac{1}{N}(\sum_{j=1}^{N} y_A^j)^2 - f_0^2} \qquad (۱۵)$$

که $f_0^2 = \hat{E}^2(Y) = (\frac{1}{N}\sum_{j=1}^{N}(y_A^j))^2$ است. در روابط فوق، رهیافت کاهش هزینهٔ محاسبات لحاظ شده است.

۱.۴ اعداد شبه‌تصادفی

روش مونته‌کارلو رایج‌ترین روش تقریب انتگرال است. اعداد تصادفی تولید شده با این روش، کل فضا را پوشش نمی‌دهند. با افزایش تعداد متغیرهای ورودی، این فضای خالی به ابرفضاهایی تبدیل می‌شود که اطلاعات رفتاری تابع در آن‌ها موجود نیست و سبب خطای قابل ملاحظه‌ای در برآورد انتگرال‌ها می‌شود. از آن‌جا که مدت زمان اجرای مدل‌های کامپیوتری نیز به‌عنوان هزینه تلقی می‌شود، باید الگوریتمی طراحی شود که با کمترین تعداد نقاط، بیشترین اطلاعات رفتاری تابع را در اختیار قرار دهد و خطای تقریب به کمترین مقدار خود کاهش یابد. در کنار مسئله خطای تقریب انتگرال، موضوع سرعت همگرایی برآوردها نیز بسیار حائز اهمیت است. بسیاری از پژوهشگران برای حل این مسائل، از اعداد تصادفی کاذب[11] استفاده می‌کنند [۲۱]. دنبالهٔ اعداد شبه‌تصادفی، نسخهٔ پیشرفتهٔ اعداد شبه‌تصادفی است. این دنبالهٔ اعداد به‌منظور برخورداری از سطح بالای یکنواختی نقاط آزمایش در فضاهای چندبعدی طراحی شده است که اعضای آن از نظر آماری مستقل از هم نیستند [۹]. نرخ همگرایی توابع با استفاده از این دنباله در مقایسه با روش یکنواخت و کاذب بسیار بیشتر است [۱۱]. دنبالهٔ اعداد شبه‌تصادفی فضای متغیرهای ورودی را کاوش می‌کنند و امکان پوشش یکنواخت فضای متغیرها را فراهم می‌سازد [۱۶]. دنباله‌های سبل و فاوره[12] و نیدریتر[13] نمونه‌هایی از این دنباله‌ها هستند. دنبالهٔ سبل یک طرح آزمایش مناسب و کارامد در این زمینه محسوب می‌شود [۲۱]. شکل ۱ نمودار پراکنش دو متغیر از دنبالهٔ سبل را در فضاهای دو و بعدی با اندازهای نمونهٔ و نشان می‌دهد.



شکل ۱. نمودار پراکنش دنبالهٔ شبه تصادفی در فضای دوبعدی (چپ) و فضای ۳۰ بعدی (راست)

[11] Pseudo random number

[12] Faure

[13] Niederriter

۵ مدل نیتروژن ورودی به آب رودخانه

نیتروژن یک آلایندهٔ منابع آب است که آثار سوئی بر سلامت انسان و سایر جانداران دارد که مدل‌سازی آن از اهمیت بسزایی برخوردار است. مدل کامپیوتری INCA-N[۱۴] یک مدل فرآیندمبنا و نیمه‌توزیعی است که جریان نیتروژن ورودی به آب رودخانه را شبیه‌سازی می‌کند [۲۲،۲۳،۲۴]. پارامترهای ورودی مدل در سه دسته تقسیم‌بندی شده و مشخصات آنها در جدول ۱ ذکر شده است. خروجی این مدل شامل سری‌های زمانی برآورد روزانهٔ میزان جریان آب و غلظت آمونیاک ونیترات در ایستگاه‌های مختلفی است که در طول میسر اصلی رودخانه قرار دارند. به‌منظور تک‌مقداری کردن خروجی مدل، این سه سری زمانی توسط رابطهٔ زیر با هم ترکیب شده، میزان بار نیتروژن غیرآلی روزانه را نتیجه می‌دهد.

(غلظت نیترات + غلظت آمونیاک) × دبی آب = بار نیتروژن

مقدار بار نیتروژن، یک سری زمانی هفت‌ساله است. خروجی نهایی، با میانگین‌گیری از متوسط‌های سالانه به‌دست آمده که آن را "متوسط سالانه بار ورودی نیتروژن غیرآلی ورودی به آب رودخانه" می‌نامند. منطقهٔ مورد مطالعه در این مقاله، حوضهٔ آبریز رودخانهٔ Tweed در اسکاتلند و انگلستان است. اطلاعات دما، رطوبت و بارش در ایستگاه‌های واقع در مسیر اصلی رودخانه به‌مدت ۷ سال اندازه‌گیری و داده‌های مربوط ثبت شده است. در این مطالعه، مدل INCA-N سامانه‌ای از پارامترهای نرخ تبادلات نیتروژنی $(X_{۹} - X_{۱۵})$ به‌عنوان ورودی مدل در نظر گرفته شده است. سایر پارامترهای ورودی در مقدار میانگین دامنه تغییراتشان ثابت هستند. در واقع مدل INCA-N را تابعی از هفت متغیر در نظر گرفته‌ایم. برای تحلیل حساسیت با روش سالتلی، اندازهٔ نمونهٔ اولیه را برابر با در نظر گرفته‌ایم که هزینه محاسبات آن (تعداد اجراهای مدل) برابر با $N \times (۲ + K) = ۱۰۰۰۰ \times ۹$ است. نقاط آزمایش روش سالتلی با استفاده از دنبالهٔ شبه‌تصادفی توسط نرم‌افزار R تولید شده است. ویژگی اعداد شبه‌تصادفی این است که اعداد به‌صورت غیرتصادفی (تعیینی) تولید می‌شوند و فضای متغیرهای ورودی را به‌طور یکنواخت پوشش می‌دهند. تعیینی بودن دنبالهٔ اعداد شبه‌تصادفی امکان برآورد شاخص‌های حساسیت را در اندازهای نمونهٔ کمتر از اندازهٔ نمونهٔ اولیه و بدون اجرای مجدد مدل مقدور می‌سازد. از آن‌جا که اندازهٔ نمونهٔ اولیه را در نظر گرفته‌ایم،روند همگرایی مقادیر برآورد شدهٔ شاخص‌های حساسیت در اندازهای نمونهٔ کوچکتر نیز قابل رؤیت خواهد بود. به‌عنوان نمونه شکل های ۲ و ۳ به‌ترتیب روند همگرایی اثر اصلی و اثر کل متغیرهای بیشترین میزان جذب $(X_{۱۱})$ و ثابت نیتروژن هوا $(X_{۱۰})$ را نشان می‌دهند. برآورد اثر اصلی و اثر کل این دو متغیر در اندازهٔ نمونهٔ پایدار می‌شود. مقدار همگرایی اثرکل متغیر $X_{۱۱}$ در حدود ۰/۵ و اثر اصلی آن تقریباً برابر با ۰/۴۵ می باشد، این مقادیر حاکیت از مؤثر بودن این متغیر در واریانس خروجی مدل دارد. برآورد اثر اصلی و کل متغیر ثابت نیتروژن هوا به مقدار صفر گرایش دارد. این بدان مفهوم است که $X_{۱۰}$ متغیر بی‌تأثیر در واریانس خروجی مدل است. سایر متغیرهای مورد بررسی از نظر پایداری رفتار مشابه داشتند.

۱.۵ نتایج تحلیل حساسیت با اندازهٔ نمونه بهینه

از آن‌جایی که برآورد شاخص‌های حساسیت اکثر متغیرهای ورودی مدل در اندازهٔ نمونه ۴۰۰۰ پایدار شد، اندازهٔ نمونه مناسب را ۴۰۰۰ درنظر گرفته و اثرات اصلی و کل را با جزئیات بیشتر مورد بررسی قرار می‌دهیم. جهت بررسی عدم قطعیت براوردها، این فرآیند را ۵۰ بار تکرار نموده که نتایج حاصل در جدول (۲) و فاصله اطمینان بوت استرپی در شکل‌های (۴) و (۵) گنجانده شده است. با توجه به میانگین مقادیر برآورد شده اثر اصلی و اثر کل در جدول (۲) مهمترین پارامترهای تاثیرگذار بر متوسط بارنیتروژن سالانه به‌ترتیب عبارتند از میزان جذب توسط گیاهان $(X_{۱۱})$، نرخ نیترات‌زدایی $(X_{۹})$، آلی‌سازی $(X_{۱۴})$ و نرخ معدنی‌سازی $(X_{۱۳})$. میزان اثر سایر متغیرها بسیار ناچیز است و می‌توان از آن‌ها صرفه‌نظر نمود. جدول (۳) اثرات متقابل بین متغیرها و سایر متغیرهای ورودی مدل نشان می‌دهد. فواصل اطمینان اثر اصلی و اثر کل (شکل‌های ۴ و۵) متغیرهای مدل به درک بهتر این مسئله کمک می‌کنند.

۶ بحث و نتیجه گیری

هدف مقاله حاضر معرفی و بکارگیری روش تحلیل حساسیت واریانس‌مبنای سالتلی با شاخص‌های تحلیل حساسیت بود. این شاخص‌ها بر اساس مفاهیم به‌ظاهر ساده واریانس شرطی،

[۱۴] Integrated Nitrogen of Catchment

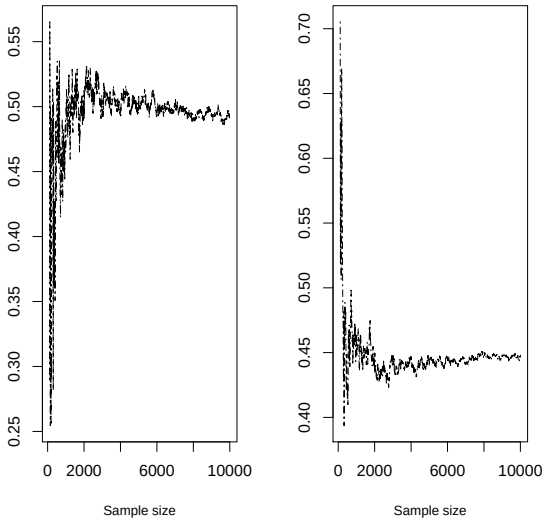
به‌گونه‌ای زیبا توانسته است سرمنشا ایجاد سرفصل بزرگی در تحلیل حساسیت مدل‌های تعیینی باشند. ساختار شاخص‌های حساسیت مرتبه اول و کل به‌گونه‌ای است که علیرغم صریح نبودن فرم ریاضی مدل قادر به شناسایی میزان اهمیت متغیرها و وجود اثرات متقابل بین آن‌ها است. زمان اجرای مدل از جمله پارامترهای کلیدی در زمینه تحلیل حساسیت است که موجب معرفی طرح آزمایش مدل‌های کامپیوتری شده است. در این مقاله به‌منظور برآورد شاخص‌های حساسیت و کاهش تعداد اجراهای مورد نیاز مدل، از رهیافت سالتلی و اعداد شبه‌تصادفی استفاده شد. عدم قطعیت برآوردگرهای با فواصل اطمینان بوت‌استرپی محاسبه شد. نتایج مورد مطالعاتی، مدل INCA-N در رودخانه Tweed، نشان داد چهار متغیر میزان جذب توسط گیاهان، نرخ نیترات‌زدایی، آلی‌سازی و نرخ معدنی‌سازی به‌ترتیب مهمترین عوامل موثر بر نیتروژن ورودی به آب رودخانه محسوب می‌شوند. بنا بر این می‌توان با کنترل این عوامل که نقش تعیین‌کننده بر نیتروژن آب دارند، از آسیب‌ها و پیامدهای ناشی از تاثیر این آلاینده آب کاست. از آن‌جا که مجموع اثرات اصلی مدل تقریبا برابر با یک است می‌توان نتیجه گرفت مدل INCA-N بر حسب پارامترهای نرخ تبادلات نیتروژنی، یک مدل تقریبا جمعی است که ناچیز بودن اثرات متقابل بین متغیرهای وردی نیز حاکی از این حقیقت است.

جدول۱. مشخصات پارامترهای ورودی مدل INCA-N

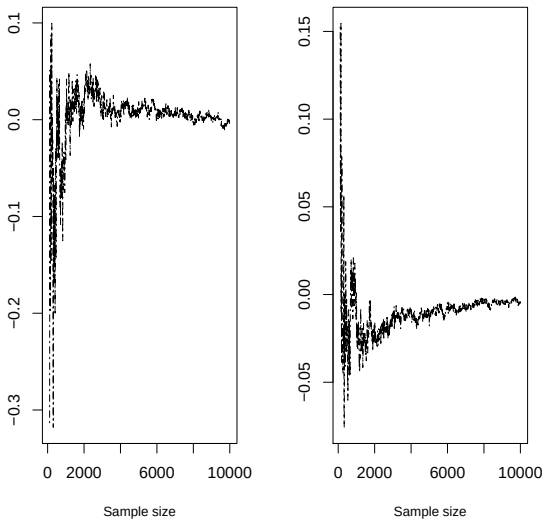
گروه	نام متغیر	واحد	دامنه تغییرات
	جریان سطحی :X_1	m^3/s	[۰, ۰/۰۱]
شرایط	نیترات سطحی :X_2	mg N/l	[۰, ۱۰]
آغارین	آمونیوم سطحی :X_3	mg N/l	[۰, ۲]
	اندازهٔ زه‌کش سطحی :X_4	m^3	[۱۰ * ۵, ۲/۱۰ * ۷]
	جریان زیرسطحی :X_5	m^3/s	[۰, ۰/۰۱]
	نیترات زیر سطحی :X_6	mg N/l	[۰, ۱۰]
	آمونیوم زیرسطحی :X_7	mg N/l	[۰, ۲]
	اندازهٔ زه‌کش زیرسطحی :X_8	m^3	[۱۰, ۲۰]
	نیترات‌زدایی :X_9	mol/day	[۰, ۰/۰۱]
نرخ تبادلات	ثابت نیتروژن هوا:.X_{10}	kg N/ha/day	[۰,/۰۱]
	نیترات جذبی گیاهان :X_{11}	mol/day	[۰, ۰/۰۵]
نیتروژنی	بیشینه نیترات جذبی :X_{12}	kg N/ha/year	[۸۰, ۱۴۰]
	معدنی‌سازی :X_{13}	mol/day	[۰, ۱]
	آلی‌سازی :X_{14}	mol/day	[۰, ۰/۱]
	آمونیوم جذبی گیاهان :X_{15}	mol/day	[۰, ۰/۰۵]
	کمبود رطوبت خاک :X_{16}	Mm	[۱۰۰, ۱۷۰]
پارامترهای	زمان ماندگاری خاک در آب :X_{17}	day	[۰/۰۵, ۵]
هیدرولوژیکی	زمان ماندگاری آب در زیرزمین :X_{18}	day	[۱۰ * ۵, ۲/۱۰ * ۷]
	بیشترین اندازهٔ نگهداری آب در خاک :X_{19}	M	[۰/۰۱, ۱]

جدول۲. کران‌های بالا و پایین متغیرهای مدل INCA-N در اندازهٔ نمونهٔ ۴۰۰۰

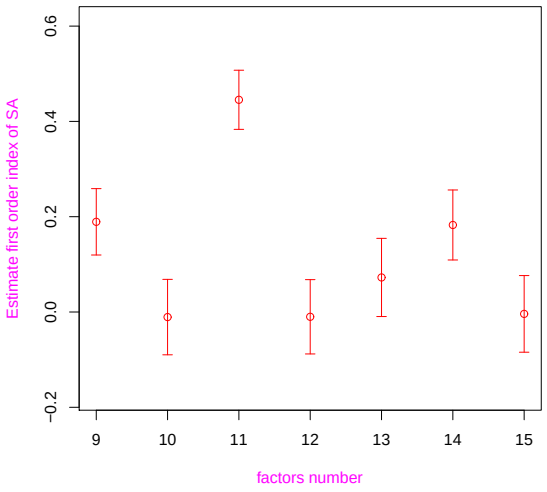
		اثر اصلی			اثر کل		
متغیر	رتبه	*Lower*	*Center*	*Upper*	*Lower*	*Center*	*Upper*
$X_۹$	۲	۰/۰۷	۰/۱۹	۰/۳۱	−۰/۰۴	۰/۲۵	۰/۵۱
$X_{۱۰}$	۷	−۰/۱۵	−۰/۰۱	۰/۱۳	−۰/۳	۰	۰/۳۱۱
$X_{۱۱}$	۱	۰/۳۲	۰/۴۴	۰/۵۷	۰/۲۵	۰/۵	۰/۷۴
$X_{۱۲}$	۶	−۰/۱۴	−۰/۰۱	۰/۱۳	−۰/۲۸	۰/۰۳	۰/۲۲۱
$X_{۱۳}$	۴	−۰/۰۶	۰/۰۷	۰/۲۱	−۰/۱۹	۰/۱۰	۰/۴۰
$X_{۱۴}$	۳	۰/۰۷	۰/۱۸	۰/۳۱	−۰/۰۵	۰/۲۲	۰/۵۱
$X_{۱۵}$	۵	−۰/۱۴	۰	۰/۱۳	−۰/۲۹	۰	۰/۲۲



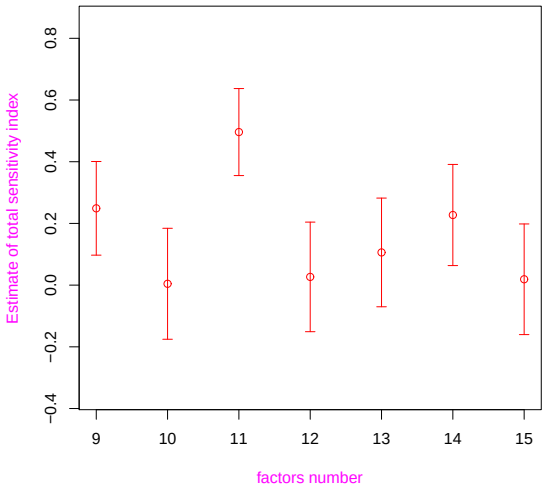
شکل۲. اثر اصلی (سمت راست) و اثر کل (سمت چپ) متغیر $X_{۱۱}$



شکل۳. اثر اصلی (سمت راست) و اثر کل (سمت چپ) متغیر $X_{۱۰}$



شکل ۴.فاصله اطمینان ۰/۹۵ اثراصلی متغیرهای مدل



شکل ۵.فاصله اطمینان ۰/۹۵ اثرکل متغیرهای مدل

مراجع

[1] Archer, G. E. B. saltelli, A. and Sobol I.M. (1997), Sensitivity measures, Anova-like techniques and the use of bootstrap, *Journal of Statistical Computation and Simulation*, **58**, 99-120.

[2] Cuiker, R.I Schaiby, J. H. and Schular, A. G. (1975). Study of the sensitivity of coupled measures, Anova-like techniques and the raction system to uncertainties in rate coefficinet III analysis of approximation, *Journal of Chemical Physics*, **63**, 1140-1149.

[3] Cuiker, R.I. Fourtin, K. E. and Schular, A. G. (1973). Study of the sensitivity of coupled measures, Anova-like techniques and the raction system to uncertainties in rate coefficinet. I theory, Journal of Computational Physics, **26**, 1-42.

[4] Cuiker, R.I Levine, H. and Schular, A. G. (1978). Nonlinear sensitivity analysis of multiparameter model systems, Journal of Computational Physics, **26**, 1-42.

[5] Fang, K.T. Runze, L. and Sudjianto, A. (2006). *Design and modeling for computer expriments*, 1 edition, Chapman and Hall/CRC.

[6] Homma, T. and Saltelli, A. (1996). Important measures in global sensitivity analysis of nonlinear model systems, Relia. Engineering and system safety, **52**, 1-17.

[7] Iman, R. and Hora, S. (1990). A robust measure of uncertainty importance for use in fault tree analysis, *Risk analysis*, **10(3)**, 401-403.

[8] Ishighami, T. and Homma, T. (1996). An Important Qualification Technique Uncrtainty Analysis for Computer Model,in: Proceeding of the Isuma 90. *Frist International Sympotion on Uncertainty Modeling and Analysis*, University of maryland, 3-5 December.

[9] Karaivanova, A., Dimov, I., and Ivanovska, S. (2001). A quasi-Monte Carlo method for integration with improved convergence, *Large-Scale Scientific Computing*, 158-165.

[10] Krzykacz-Hausman, B. (1990). Gesellschaft fuer reaktor sicherheit (GRS) Mbh. Technical report GSR, 1700, garching.

[11] Levy, G. (2002). A introducation to quasi-random number integration with improved convergence . NAG ltd, Oxford, Uk.

[12] McKay, M. (1996). Variance based methods for assessing uncertainty importance in nureg-1150 analysis, Technical report LA-UN-96-2695, Los almos labratories.

[13] Sacks, J. Welch, W. J. Mitchell, T. J. and Wynn, H. P. (1989). Design and analysis of computer expriments, *Statistical Science*, **4**, 409-435.

[14] Saltelii, A. Andres, T.H. and Homma, T. (1993). Some new techniques in sensitivity analysis of model output, *Computational Statistics and Data Analysis*, **15**, 211-238.

[15] Saltelii, A. (2002). Making best use of model valuations to compute sensitivity analysis indices, *Computer physics communication*, **145**, 280-297.

[16] Saltelii, A. Annoni, P. Azzini, I. Compolongo, F. Ratto, M. and Tarantola, S. (2010). Varince based sensitivity analysis of model output. Design and estimate for the total sensitivity index, *Computer physics communaction*, **181**, 259-270.

[17] Saltelii, A. Ratto, M. Andres, T. Compolongo, F. Cariboni, J. and Gatelli, D. (2008).*Golobal sensitivity analysis the primer*, 1 edition, john wiley and sons ltd.

[18] Saltelii, A. Tarantola, S. Compolongo, F. and Ratto, M. (2004). *Sensitivity analysis in practice*, 1 edition, John Wiley and Sons ltd.

[19] Santner, T.J. Williams, B. J. and Notz, W. I. (2003). *The design and analysis of compueter expriments*, 1 edition, Springer, New York ltd.

[20] Sobol, I.M. (1998). On quasi-monte carlo integrations, *Math. and computation expriment*, **47**, 103-112.

[21] Sobol, I.M. (1993). Sensitivity analysis for non-linear mathmatical model, *Mathematical Modeling and Computational Experiment*, **1**, 407-414.

[22] Wade, A. J. Durand, P. Beaujouan, V. Wessel, W. W. Raat, K. J. Whitehead, P.G. Buterfield, D. Rankinen k. leptiso, A. (2002). A nitrogen model for European catchment: INCA, new model structure and equation, *Hydrology and earth system science*, **6**, 559-582.

[23] Whitehead, P.G. Wilson, E. J. Buterfield, D. (1998a). A semi-distributed nitrogen model for multiple source assessment in catchment (INCA): model structure and procces equation, *Science of the total enviroment*, **210/211**, 547-558.

[24] Whitehead, P.G. (1990). Modeling nitrate from agriculture to public water supplies, *Philosophical Trans-actions of the Royal Society*, **329**, 403-410.