

دانشگاه صنعتی شاهرود

دانشکده ریاضی

گروه ریاضی کاربردی

پایان نامه جهت اخذ درجه کارشناسی ارشد آمار ریاضی

بررسی روش‌های واریانس - مبنا در تحلیل حساسیت

مدل‌های تعیینی

دانشجو : مجید جانفدا

استاد راهنما :

دکتر داود شاهسونی

استاد مشاور :

دکتر ناصر رضا ارقامی

آذر ماه ۱۳۹۰

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

سپاس از دو وجود مقدس

آنان که ناتوان شدند تا به توانایی برسم

مویشان سپید گشت تا رو سفید شوم

و عاشقانه سوختند تا کرم باخش وجود و رو منکر را بهم باشند

تقدیم به

پدرم

و

مادرم

بمهر و قدر دانی

حمد و سپاس پروردگار یکتا که لطف و کرم بی کرائش من را نیز در برگرفت تا به وسع توان خویش گامی کوچک در گستره علم و معرفت بردارم و میرگشت تا از خرمن دانش و تجربه بزرگان و نیک اندیشان بهره ببرم.

الکون که به یاری خداوند متعال، این دوره پر خاطره از دوران تحصیل را به پایان رسانده‌ام، هر چند واژه‌ها یارای آن نیست که لطف و محبت و بزرگواری آنانی را که در تمام دوران زندگی‌ام جرحه‌نوش دریای مهر و محبتشان بوده‌ام به تصویر کشم، اما به رسم ادب و احترام بوسه بر دستانشان زده و بر خود واجب می‌دانم، زحمت پدر و مادر مهربانم را که به‌واره راه‌کشی مشکلاتم در تمام مراحل زندگی بوده‌اند ارج نهاده و مراتب شکر قلبی و باطنی را از الطاف و مهربانی‌های آنان ابراز دارم. همچنین لازم می‌دانم که از زحمت فراوان استاد توانمند آقای دکتر شایسته‌فانی که بارها به‌نیای ما و نظرات ارزنده و صبر و حوصله فراوان، نقش مهمی در به‌ثمر رساندن این کار داشته‌اند صمیمانه تشکر و تشکر نمایم، بی‌تردید انجام این پایان‌نامه بدون بهکاری و راهنمایی ایشان امکان‌پذیر نبود. همچنین بر خود لازم می‌دانم از زحمت آقای دکتر ارقامی به‌عنوان مشاور این پایان‌نامه نیز کمال تشکر را دارم. همچنین از زحمت مسئولین آموزش و دفتر دانشکده آقای حسین پور و خانم خاوردی که بهکاری لازم را با حسن داشتند تشکر می‌نمایم. در اینجا همچنین لازم می‌دانم از اعضای خانواده عزیزم و مخصوصاً پدر و مادرم، همچنین از دوستان عزیزم آقایان خوجمی، نوری، امیدیان، لندرانی، گهایگانی و کریمی که مایه دلگرمی من بوده و متحمل زحمت زیادی شدند نیات پاک‌سازی را داشته‌باشم و برای آنها بهترین‌ها را آرزو می‌کنم.

مجید جانعلی

آذر ۱۳۹۰

دانشجو تأیید می نماید که مطالب مندرج در این پایان نامه (رساله) نتیجه تحقیقات خودش می باشد و در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده است.

کلیه حقوق مادی مترتب از نتایج مطالعات ، آزمایشات و نوآوری ناشی از تحقیق موضوع این پایان نامه (رساله) متعلق به دانشگاه صنعتی شاهرود می باشد .

آذر ۱۳۹۰

چکیده

امکان مطالعه بسیاری از پدیده‌های علمی در شرایط آزمایشگاهی وجود ندارد. از این رو این پدیده‌ها، در قالب مدل‌های ریاضی بیان شده و رفتار آنها توسط کدهای کامپیوتری شبیه‌سازی می‌شود. این مدل‌ها شامل دستگاهی از معادلات دیفرانسیل معمولی یا جزئی هستند که حل آن‌ها توسط روش‌های عددی انجام می‌گیرد. برنامه یا کد کامپیوتری که قادر به حل عددی این دستگاه معادلات باشد را **مدل کامپیوتری** و اجرای مکرر این مدل با مقادیر مختلفی از ورودی‌ها را **آزمایش کامپیوتری** می‌گویند. ساختار این مدل‌ها از پیچیده‌گی‌های خاصی برخوردار است که یکی از عوامل آن، تعداد زیاد ورودی‌های است. متغیرهای ورودی می‌تواند به طور تکی یا در تقابل با سایر ورودی‌ها، خروجی مدل را تحت تاثیر خود قرار دهند. از این رو شناخت ورودی‌های تاثیرگذار و همچنین ورودی‌های کم اهمیت یا بی‌اهمیت ضروری است. این مهم در قالب **تحلیل حساسیت** مدل‌های کامپیوتری، مورد مطالعه قرار می‌گیرد. روش‌های مختلفی برای تحلیل حساسیت مدل‌ها پیشنهاد شده است. یکی از این روش‌ها، تحلیل حساسیت **واریانس - مبنا** است. تحلیل حساسیت در این روش با استفاده از اندازه‌هایی به نام‌های **شاخص حساسیت مرتبه اول** و **شاخص حساسیت کل** انجام می‌شود. شاخص‌های مذکور انتگرال‌های چندگانه‌ای هستند که مبتنی بر مفاهیم امیدریاضی شرطی و واریانس شرطی است. مقدار واقعی این انتگرال‌ها به دلیل صریح نبودن رابطه ورودی - خروجی مدل، همواره مجهول است. در نتیجه مسئله برآورد آن‌ها یا تقریب انتگرال‌ها مطرح می‌شود. برای این منظور راهکارهای متعددی توسط محققین پیشنهاد شده است. روش تحلیل حساسیت واریانس مبنا نیز دو رهیافت کلی موسوم به روش مونت کارلو مبنای سالتلی و روش طرح متعادل تصادفی (**RBD**) را برای تقریب شاخص‌ها ارائه می‌دهد که در این پایان‌نامه مورد مطالعه قرار گرفته است. پس از معرفی این دو روش، واسنجی (اعتبار سنجی) آن‌ها با استفاده از تابع

تحلیلی سبل انجام گرفت. نتایج بررسی نشان از دقت کافی و عملکرد مناسب این دو روش در برآورد شاخص‌های حساسیت داشت. به منظور تحلیل حساسیت مدل زیست محیطی *INCA-N* از روش‌های مذکور استفاده شد. مدل کامپیوتری *INCA-N*، جریان نیتروژن ورودی به آب رودخانه را شبیه‌سازی می‌کند. نیتروژن یک آلاینده منابع آب است که اثرات سوئی بر سلامت انسان و جانداران دارد. در این مطالعه، مدل مذکور تابعی از هفت متغیر ورودی (نرخ تبادلات نیتروژنی) و یک متغیر خروجی (متوسط سالانه بار نیتروژن ورودی به رودخانه) در نظر گرفته شد. پس از تحلیل حساسیت نیتروژن ورودی به رودخانه *Tweed* واقع در کشور انگلستان، با استفاده از روش مونت کارلو مبنای سالتلی و روش طرح متعادل تصادفی مشخص شد که چهارمتغیر، میزان جذب نیترات گیاهان، نرخ نیترات‌سازی، آلی‌سازی و معدنی‌سازی مهمترین عوامل موثر بر این آلاینده آب رودخانه *Tweed* است.

فهرست

فصل اول	۱
مقدمه	۲
۱ - ۱ ضرورت و اهداف پایان نامه	۳
۱ - ۲ ساختار پایان نامه	۵
فصل دوم	۷
۲ - ۱ آزمایش کامپیوتری	۹
۲ - ۲ مفاهیم و تعاریف مشترک آزمایش های کامپیوتری و فیزیکی	۱۲
۲ - ۳ نقش آمار در آزمایش کامپیوتری	۱۳
فصل سوم	۱۷
مقدمه	۱۸
۳ - ۱ تحلیل حساسیت	۱۹
۳ - ۲ روش های تحلیل حساسیت	۱۹
۳ - ۳ معیارهای انتخاب روش تحلیل حساسیت	۲۳
۳ - ۴ تاریخچه روش تحلیل حساسیت واریانس - مبنا	۲۵
۳ - ۵ پیکربندی روش واریانس - مبنا	۲۸
۳ - ۶ تجزیه واریانس خروجی مدل بر اساس شاخص های حساسیت	۳۳

۴۰	۳-۷ ویژگی های شاخص های حساسیت
۴۱	۳-۸ اهداف بنیادی یا سبک های تحلیل حساسیت واریانس - مبنا
۴۴	فصل چهارم
۴۵	مقدمه
۴۵	۴-۱ بنیان ریاضی شاخص حساسیت مرتبه اول
۴۷	۴-۲ برآورد شاخص حساسیت مرتبه اول
۵۰	۴-۳ بنیان ریاضی شاخص حساسیت کل
۵۱	۴-۴ برآورد شاخص حساسیت کل
۵۲	۴-۵ دستور کار برای برآورد شاخص مرتبه اول و شاخص کل
۵۵	۴-۶ فاصله اطمینان بوت استرپ برای شاخص های حساسیت
۵۷	۴-۷ اعداد شبه تصادفی
۶۵	فصل پنجم
۶۶	مقدمه
۶۶	۵-۱ روش طرح متعادل تصادفی
۸۰	۵-۲ مقایسه روش های تقریب شاخص های حساسیت
۸۱	فصل ششم
۸۲	مقدمه

۸۲.....	۶- ۱ معرفی تابع تحلیلی سبل
۸۵.....	۶- ۲ ارزیابی تکنیک مونت کارلو- مبنای سالتلی توسط تابع تحلیلی
۸۸.....	۶- ۳ ارزیابی تکنیک طرح متعادل تصادفی با تابع تحلیلی
۹۱.....	۶- ۴ مقایسه روش مونت کارلو - مبنای سالتلی و طرح متعادل تصادفی
۹۹.....	فصل هفتم
۱۰۰.....	مقدمه
۱۰۲.....	۷- ۱ معرفی اجمالی مدل <i>INCA-N</i>
۱۰۲.....	۷- ۲ معرفی ورودی ها و خروجی های مدل
۱۰۴.....	۷- ۳ منطقه مورد مطالعه
۱۰۶.....	۷- ۴ تحلیل حساسیت مدل <i>INCA-N</i> با روش مونت کارلو-مبنای سالتلی
۱۰۶.....	۷- ۵ تحلیل حساسیت مدل <i>INCA-N</i> با استفاده از روش <i>RBD</i>
۱۲۱.....	۷- ۶ نتیجه گیری
۱۲۲.....	نتیجه گیری نهایی
۱۲۵.....	پیشنهادهات
۱۲۶.....	فهرست منابع
۱۳۳.....	ضمیمه
۱۳۴.....	مقدمه

۱۳۵	چرخه نیتروژن
۱۳۹	منابع ورود نیتروژن به آب رودخانه
۱۴۱	مدل سازی چرخه نیتروژن در حوضه آبریز
۱۴۳	معرفی مدل <i>INCA-N</i>
۱۴۴	قابلیت‌های مدل <i>INCA</i>
۱۴۶	ساختار مدل <i>INCA-N</i>

فهرست اشکال

- شکل (۱ - ۱) - ساختار کلی پایان نامه ۶
- شکل (۱ - ۲) - ساختار کلی آزمایشهای کامپیوتری و فیزیکی ۱۰
- شکل (۲ - ۲) - ترکیبات مختلف در چرخه نیتروژن ۱۰
- شکل (۲ - ۳) - فرآیند چرخه نیتروژن در خاک و آبهای زیر زمینی ۱۱
- شکل (۲ - ۴) - ساختار شماتیکی مدل INCA-N ۱۲
- شکل (۲ - ۵) - چگونگی ساخت مدل جانشین و کاربرد آن ۱۵
- شکل (۳ - ۱) - نمودار پراکنش Y در مقابل متغیرهای توضیحی ۳۰
- شکل (۳ - ۲) - تقسیم نمودار پراکنش به باریکتهای با طول Δx_i ۳۱
- شکل (۴ - ۱) - نمودار پراکنش نقاط تصادفی تولید شده از توزیع یکنواخت برای X, Y ۵۸
- شکل (۴ - ۲) - مقایسه روند همگرایی در روشهای سبل، کاذب و یکنواخت ۶۱
- شکل (۴ - ۳) - نمودار پراکنش دنباله سبل در فضای ۲ بعدی (چپ) و فضای ۳۰ بعدی برای متغیرهای ۲۵ و ۲۶ (راست).
..... ۶۳
- شکل (۴ - ۴) - شمای ناهمگنی در فضای دو بعدی ۶۴
- شکل (۵ - ۱) - نمودار توابع تبدیل و توزیع تجربی متناظر با آنها ۷۳
- شکل (۵ - ۲) - نمودار پراکنش نقاط طرح آزمایش در فضای دو بعدی بر اساس تابع تبدیل سالتلی (راست) و تابع تبدیل کدا (چپ) ۷۴
- شکل (۵ - ۳) - تجزیه فوریه تابع $y = 2X_1 + X_1^2 + X_2 + X_3$ در بسامدهای متناظر آن ۷۶
- شکل (۵ - ۴) - الگوریتم روش RBD ۸۰

شکل (۶ - ۱) - نمودار تابع سبیل در حالت یک متغیره (پانل راست) و دومتغیره (پانل چپ، به ازای $A_1=A_2=0$) ۸۳

شکل (۶ - ۲) - روند همگرایی مقادیر برآورد متغیر بسیار مهم X_1 (پانل A)، متغیر مهم X_2 (پانل B)، متغیر نسبتا مهم X_3

(پانل C) و بی اهمیت X_4 (پانل D) ۸۵

شکل (۶ - ۳) - فاصله اطمینان ۹۵٪ مقادیر برآورد شده اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۵۰۰

..... ۸۶

شکل (۶ - ۴) - فاصله اطمینان ۹۵٪ مقادیر برآورد شده اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۱۰۰۰

..... ۸۷

شکل (۶ - ۵) - فاصله اطمینان ۹۵٪ مقادیر برآوردی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۲۰۰۰

..... ۸۷

شکل (۶ - ۶) - روند همگرایی مقادیر برآورد متغیر بسیار مهم X_1 (پانل A)، متغیر مهم X_2 (پانل B)، متغیر نسبتا مهم X_3

(پانل C) و بی اهمیت X_4 (پانل D) ۸۸

شکل (۶ - ۷) - مقادیر واقعی و فاصله اطمینان ۹۵٪ برای متغیر بسیار مهم X_1 (پانل A)، متغیر مهم X_2 (پانل B)، متغیر

نسبتا مهم X_3 (پانل C) و متغیر بی اهمیت X_4 (پانل D) در حجم های نمونه مختلف ۹۰

شکل (۶ - ۸) - مقادیر واقعی و فاصله اطمینان ۹۵٪ برای متغیرهای فاقد اهمیت X_5 (پانل A)، X_6 (پانل B)، X_7 ۹۱

شکل (۶ - ۹) - مقایسه روند همگرایی روش طرح متعادل تصادفی و روش سالتلی به مقادیر تحلیلی در متغیر X_1 ۹۲

شکل (۶ - ۱۰) - مقایسه روند همگرایی روش RBD و روش سالتلی به مقادیر تحلیلی در متغیر X_2 (پانل A)، متغیر X_3

(پانل B)، متغیر X_4 (پانل C) و متغیر X_5 (پانل D) ۹۳

شکل (۶ - ۱۱) - فاصله اطمینان میانگین برآورد اثر اصلی با روش های سالتلی و RBD در حجم نمونه ۵۰۰ ۹۴

شکل (۶ - ۱۳) - فاصله اطمینان میانگین برآورد اثر اصلی با روش های سالتلی و RBD در حجم نمونه ۲۰۰۰ ۹۵

شکل (۶ - ۱۴) - مقایسه مقادیر واقعی با مقدار برآورد شده با روش سالتلی و RBD با متغیرهای با درجه اهمیت مختلف

..... ۹۸

- شکل (۷ - ۱) - چرخه نیتروژن در طبیعت ۱۰۱
- شکل (۷ - ۲) - ساختار شماتیکی مدل INCA-N ۱۰۴
- شکل (۷ - ۳) - حوضه آبریز رودخانه TWEED ۱۰۵
- شکل (۷ - ۴) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) متغیر نیترات زدایی ۱۰۷
- شکل (۷ - ۵) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر ثابت نیتروژن هوا ۱۰۸
- شکل (۷ - ۶) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر جذب نیترات گیاهان ۱۰۸
- شکل (۷ - ۷) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر بیشترین میزان جذب ۱۰۹
- شکل (۷ - ۸) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر معدنی سازی ۱۱۰
- شکل (۷ - ۹) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (چپ) در متغیر آلی سازی ۱۱۰
- شکل (۷ - ۱۰) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (چپ) در متغیر جذب آمونیاک ۱۱۱
- شکل (۷ - ۱۱) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل INCA-N در نمونه به حجم ۴۰۰۰ ۱۱۴
- شکل (۷ - ۱۲) - فاصله اطمینان ۹۵٪ اثر تام متغیرهای مدل INCA-N در نمونه به حجم ۴۰۰۰ ۱۱۴
- شکل (۷ - ۱۳) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۵۰۰ ۱۱۸
- شکل (۷ - ۱۴) - نمودار جعبه‌ای اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۵۰۰ ۱۱۹
- شکل (۷ - ۱۵) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۱۰۰۰ ۱۱۹
- شکل (۷ - ۱۶) - نمودار جعبه‌ای اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۱۰۰۰ ۱۲۰
- شکل (۷ - ۱۷) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۲۰۰۰ ۱۲۰
- شکل (۷ - ۱۸) - نمودار جعبه‌ای اثر اصلی متغیرهای مدل INCA-N با حجم نمونه ۲۰۰۰ ۱۲۱

فهرست جداول

جدول (۳- ۱) - قابلیت های روش های مختلف تحلیل حساسیت	۲۴
جدول (۴- ۱) - روش های برآورد اثر اصلی و اثر کل	۵۴
جدول (۵- ۱) - توابع تبدیل مختلف روش FAST	۷۱
جدول (۵- ۲) - مقایسه روش های مختلف تحلیل حساسیت	۸۰
جدول (۶- ۱) - برد و درجه اهمیت تابع $G_I(X_I)$ به ازای مقادیر مختلف A_I	۸۴
جدول (۶- ۲) - مقادیر A_I ورودی های تابع سبل	۸۴
جدول (۷- ۱) - مشخصات پارامترهای ورودی مدل INCA-N	۱۰۳
جدول (۷- ۲) - کران های بالا و پایین متغیرهای مدل INCA-N در حجم نمونه ۴۰۰۰	۱۱۲
جدول (۷- ۳) - آمار توصیفی برآورد اثر اصلی متغیرهای نیترات زدایی و ثابت نیتروژن هوا	۱۱۵
جدول (۷- ۴) - آمار توصیفی برآورد اثر اصلی متغیرهای جذب نیترات و بیشترین میزان جذب نیترات	۱۱۶
جدول (۷- ۵) - آمار توصیفی برآورد اثر اصلی متغیرهای معدنی سازی و آلی سازی	۱۱۶
جدول (۷- ۶) - آمار توصیفی برآورد اثر اصلی متغیرهای جذب آمونیاک و تمام متغیرهای مدل	۱۱۷
جدول (۷- ۷) - نتایج برآورد اثر اصلی و تام متغیرهای مهم مدل INCA-N	۱۲۲

فصل اول

مقدمه

مقدمه

رفتار بسیاری از پدیده‌های طبیعی و علمی را می‌توان توسط مدل‌های ریاضی شبیه‌سازی نمود. این گونه مدل‌ها شامل دستگاهی با تعداد زیادی از معادلات دیفرانسیل معمولی یا جزئی است که متغیرهای مورد مطالعه را به یکدیگر مرتبط می‌سازند. عملکرد این مدل‌ها را می‌توان همانند فرآیندی فرض کرد که در آن تعدادی از متغیرها، در نقش ورودی و یک یا چند متغیر نیز نقش خروجی را ایفا می‌کنند. غالباً حل این گونه دستگاه معادلات با استفاده از روش‌های عددی انجام می‌گیرد. به برنامه یا کد کامپیوتری که قادر به حل عددی این دستگاه معادلات باشد، **مدل کامپیوتری** و اجرای مکرر مدل با مقادیر مختلفی از ورودی‌ها را **آزمایش کامپیوتری** می‌گویند. ساختار این مدل‌ها از پیچیده‌گی‌های خاصی برخوردار است که یکی از عوامل آن، تعداد زیاد ورودی‌های است. متغیرهای ورودی می‌تواند به طور تکی یا در تقابل با سایر ورودی‌ها، خروجی مدل را تحت تاثیر خود قرار دهند. از این رو شناخت ورودی‌های تاثیرگذار و همچنین ورودی‌های کم اهمیت یا بی‌اهمیت و کنترل آن‌ها باعث ساده‌تر شدن مدل خواهد شد به‌طوری که امکان تفسیر بهتر نتایج در مسایلی از قبیل مدیریت ریسک^۱، تصمیم‌گیری و غیره نیز میسر می‌گردد. این مهم در قالبی تحت عنوان **تحلیل حساسیت**^۲ مدل‌های کامپیوتری، مورد مطالعه قرار می‌گیرد. روش‌های مختلفی برای تحلیل حساسیت مدل‌ها پیشنهاد شده است که در هر یک، میزان اهمیت هر ورودی و اثر متقابل آن با سایر متغیرها، توسط اندازه‌هایی به نام شاخص حساسیت بیان می‌گردد. یکی از این روش‌ها، تحلیل حساسیت **واریانس - مبنا**^۳ است که در آن اندازه‌هایی به نام‌های **شاخص**

^۱ Risk Mangement^۲ Sensitivity Analysis^۳ Variance based

حساسیت مرتبه اول^۱ و شاخص حساسیت کل^۲ معرفی می‌گردد. ساختار این دو اندازه حساسیت مبتنی بر مفاهیم امیدریاضی شرطی و واریانس شرطی است. شاخص‌های مذکور توسط انتگرال‌های چندگانه بیان می‌شوند که مقدار واقعی آنها به دلیل صریح نبودن رابطه ورودی-خروجی مدل، همواره مجهول است و در نتیجه مسئله برآورد آنها یا تقریب انتگرال‌ها مطرح می‌شود. تاکنون راهکارهای متعددی توسط محققین برای تقریب این انتگرال‌ها پیشنهاد شده است. از آنجا که "زمان اجرای مدل‌های کامپیوتری" موضوعی بحرانی و چالش برانگیز است لذا می‌توان راهکارهای پیشنهادی را از این منظر و در قالبی تحت عنوان هزینه محاسبات مورد ارزیابی قرار داد. روش تحلیل حساسیت واریانس مبنا، دو رهیافت کلی برای تقریب شاخص‌ها ارائه می‌دهد که اساس کار این پایان نامه را تشکیل می‌دهد. این دو رهیافت عبارتند از: روش مونت کارلو مبنای سالتلی و روش طرح متعادل تصادفی (*RBD*) که در این پایان نامه آنها را مطالعه نموده و با مد نظر قرار دادن آخرین پیشنهادات محاسباتی، عملکرد این دو روش را در برآورد شاخص‌های حساسیت مورد ارزیابی قرار می‌دهیم.

۱-۱ ضرورت و اهداف پایان‌نامه

ارتباط بین آماردانان و متخصصین سایر علوم که از دستاوردها و روش‌های آماری به عنوان ابزار قوی تجزیه و تحلیل اطلاعات استفاده می‌کنند، امری اجتناب ناپذیر است. این ارتباط دو طرفه باعث می‌شود تا در عمل، آماردانان بتوانند چالش‌های موجود در تحلیل اطلاعات را شناسایی نموده و راه‌حلهایی برای مرتفع کردن آن ارائه نمایند. از سوی دیگر مهندسين نیز می‌توانند با بهره‌گرفتن از دیدگاه‌ها و روش‌های نوین آماری، دقت تجزیه و تحلیل اطلاعات را افزایش دهند. این مسئله انگیزه‌ای شد تا در این پایان نامه

^۱ First order sensitivity analysis

^۲ Total sensitivity analysis

رویکرد کاربردی آمار مورد توجه قرار گیرد تا از این طریق بتوان گوشه‌هایی از جایگاه نقش آمار را شناسایی نمود.

یکی از مقوله‌های کاربردی که علم آمار به‌طور خاص در آن به ایفای نقش می‌پردازد، تحلیل حساسیت مدل‌های شبیه‌ساز پدیده‌های طبیعی است که به نوعی با مدیریت بر سلامت انسان در ارتباط هستند. از جمله این پدیده‌ها می‌توان به فرآیندهایی اشاره نمود که منجر به پدیدار شدن ترکیبات نیتروژنی در منابع آب و تهدید سلامت انسان می‌شود. شایان ذکر است که نیتروژن به‌طور طبیعی چرخه مشخصی دارد که در آن، ترکیبات مختلف نیتروژنی تحت شرایط متفاوت به‌وجود می‌آیند. این ترکیبات از راه‌های مختلف سبب آلودگی منابع آب می‌شوند. نیتروژن نیاز اصلی گیاهان است که به‌طور مستقیم و غیرمستقیم (مانند نیترات) تامین می‌شود. نیترات از طریق تماس منابع آب با فاضلاب‌های شهری، صنعتی و غیره وارد منابع آب شده و به ترکیبات سمی چون نیتريت و نیتروسامین تبدیل می‌شود. بیماری‌هایی مانند متاهموگلوبین، سرطان، ناقص‌الخلقه‌بودن نوزادان، بزرگ شدن تیروئید و غیره از اثرات سوء نیترات بیش از حد در آب آشامیدنی است. با توجه به آثار سوء ترکیبات نیتروژن بر سلامت انسان، دانستن فرآیندها و نحوه ورود این ترکیبات به منابع آب آشامیدنی امری اجتناب‌ناپذیر است. بدین منظور مدل‌های متعددی برای برآورد میزان بار نیتروژن ورودی به رودخانه‌ها بسط و توسعه یافته است. مدل فرآیند-مبنای^۱ $INCA-N^x$ به‌عنوان مورد مطالعاتی این تحقیق، ثمره یکی از این تلاش‌ها است که میزان جریان آب و نیتروژن را در یک حوضه آبریز تعیین می‌کند.

^۱ Process-based

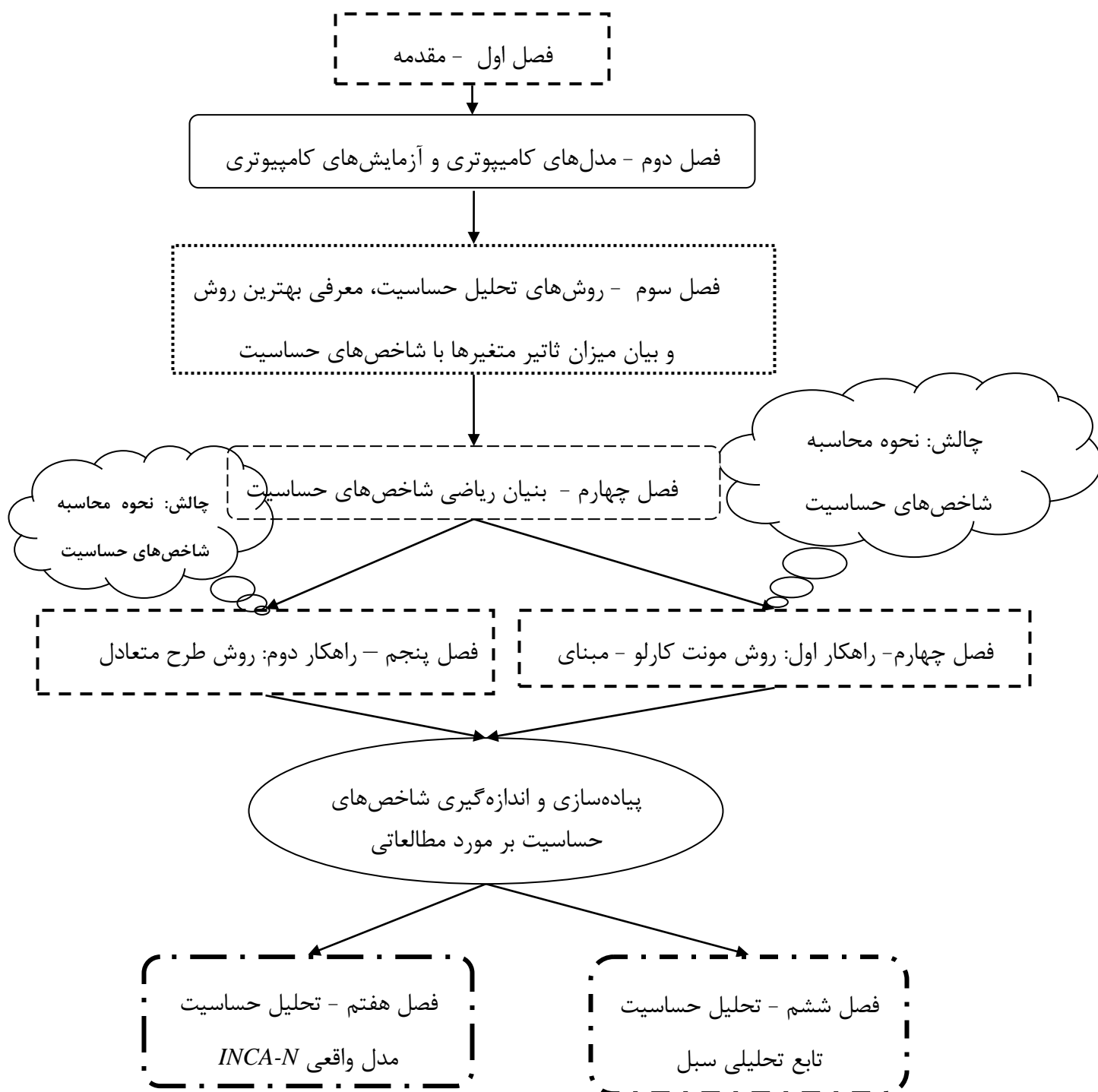
^۲ Integrated Nitrogen of Catchment

در این تحقیق قصد بر آن است تا جنبه‌هایی کاربردی از مفاهیم عمیق آماری را توسط تحلیل حساسیت مدل‌های تعیینی مورد بررسی قرار دهیم. در این راستا و با انجام تحلیل حساسیت مدل *INCA-N* این امکان نیز فراهم می‌شود که عوامل تاثیرگذار بر آلاینده‌گی منابع آب شناسایی شوند تا سیاست‌گذاران منابع آب بتوانند با کنترل آن‌ها از آسیب‌های تاثیر این عوامل در تهدید سلامت جامعه کاسته و خسارت‌های مالی و زیست محیطی آن را تعدیل نمایند.

۱ - ۲ ساختار پایان‌نامه

بخش‌های مختلف این پایان‌نامه مطابق ذیل تنظیم شده است:

فصل دوم شامل معرفی مدل‌ها و آزمایش‌های کامپیوتری و نیز بیان نقش آمار در این مدل‌ها است. در فصل سوم ابتدا با ذکر روش‌های مختلف تحلیل حساسیت، به معرفی کامل روش واریانس-مبنا و شاخص‌های حساسیت، که اساس کار این پایان‌نامه را تشکیل می‌دهند، پرداخته‌ایم. شاخص‌های حساسیت واریانس-مبنا توسط روش‌های مختلفی برآورد می‌شوند که در فصل چهارم، روش مونت کارلو - مبنا، سالتلی و در فصل پنجم روش طرح متعادل تصادفی مورد بحث و بررسی قرار گرفته است. این روش‌ها، ابتدا با یک تابع تحلیلی و سرانجام توسط مدل تعیینی *INCA-N*، به ترتیب در فصل‌های ۶ و ۷ مورد ارزیابی واقع شده‌اند. شکل زیر ساختار کلی این تحقیق را نشان می‌دهد.



شکل (۱-۱) - ساختار کلی پایان نامه

فصل دوم

مدل‌ها و آزمایش‌های کامپیوتری

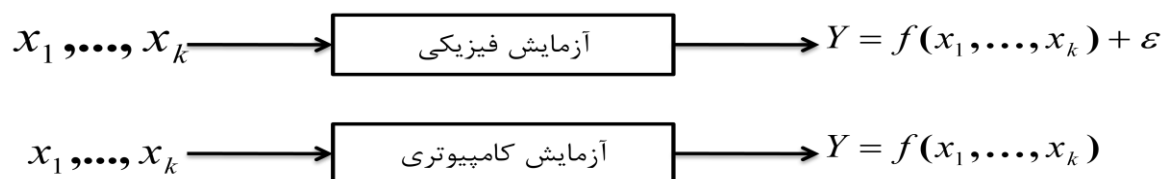
تغییرات آب و هوایی، جزر و مد دریا، سیلاب‌های مهیب، آتشفشان‌ها، زلزله، رانش زمین و . . . هر یک پدیده‌های مختلفی هستند که محیط اطراف انسان را به وجود آورده‌اند. برخی از این پدیده‌های پیچیده و منظم با زندگی انسان در ارتباط است و می‌تواند شرایط کنونی انسان را به کلی دگرگون سازند. به‌منظور استفاده بهینه از محیط پیرامون و پیشگیری از تبعات و آثار مخرب برخی از این پدیده‌ها بایستی قوانین حاکم بر آن‌ها را شناخت و رابطه خود را با محیط، بر اساس آن‌ها تنظیم نمود. اگر چه جهان هستی و پدیده‌های موجود در آن در قیاس با توانایی‌های ذهنی بشر دارای پیچیدگی‌های فراوانی است، اما می‌توان در جهت کشف و تحلیل بسیاری از این پدیده‌های طبیعی گام برداشت و در این راستا از ابزارهای گوناگونی استفاده نمود. یکی از طرق شناخت پدیده‌های طبیعی یا رخداد‌های علمی، طراحی و انجام آزمایش‌های فیزیکی بر اساس فرضیات ارائه شده است. آزمایش فیزیکی به دسته‌ای از آزمایش‌ها اطلاق می‌شود که در محیط‌های آزمایشگاهی، کارخانه‌ها، زمین‌های کشاورزی و . . . با اهدافی نظیر بهبود فرآیند تولید و کیفیت محصولات تولیدی، کاهش تغییرپذیری و افزایش پایایی، سرعت بخشیدن به روند پیشرفت و توسعه فرآیند و کاهش هزینه‌های مازاد انجام می‌گردد. یکی از مشخصه‌های آزمایش‌های فیزیکی بدست آمدن نتایج متفاوت تحت شرایط آزمایشگاهی یکسان است. دلیل حادث شدن این رویداد، حضور خطای تصادفی در آزمایش‌های فیزیکی است که باعث پیچیدگی در تفسیر نتایج و گاهی اوقات انحراف و سردرگمی آزمایشگر می‌شود. وجود این خطاها، مبدا ورود آمار به بحث آزمایش‌های فیزیکی است به‌طوری که در روش‌های کلاسیک طرح و تجزیه آزمایش از اعمالی همچون بلوک‌بندی، تصادفی کردن و تکرار به منظور افزایش اعتبار مقایسه تیمارها و کاهش تاثیر خطاهای تصادفی استفاده می‌کنند (ساکس^۱ و همکاران، ۱۹۸۹).

^۱ Sacks

۲- ۱ آزمایش کامپیوتری

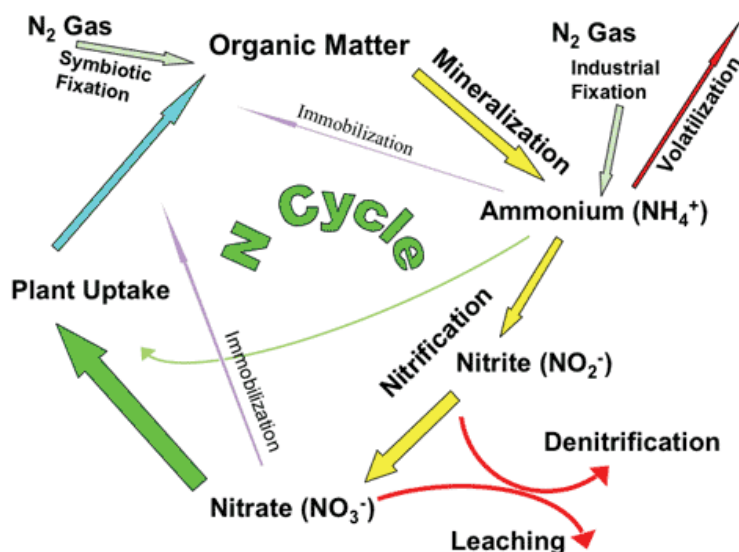
مطالعه بسیاری از پدیده‌های علمی در شرایط آزمایشگاهی بنا بر دلایلی همچون داشتن هزینه‌های گزاف انجام آزمایش، طولانی بودن زمان اجرا و عملی نبودن انجام آزمایش به خصوص در پدیده‌های زیست محیطی، مقدور نیست. از این رو محققین، به شبیه‌سازی این پدیده‌ها توسط مدل‌های ریاضی روی آورده‌اند. مدل‌های ریاضی غالباً شامل یک سری معادلات دیفرانسیل معمولی و یا معادلاتی با مشتق‌های جزئی از نوع خطی و غیرخطی هستند که متغیرهای موجود در آن پدیده را به نوعی با یکدیگر مرتبط می‌سازند. به برنامه یا کد کامپیوتری که قادر به حل عددی این دستگاه معادلات پیچیده باشد، **مدل کامپیوتری** و اجرای این مدل با ورودی‌های متفاوت را **آزمایش کامپیوتری** گویند. ورودی مدل‌های کامپیوتری (فاکتورهای ورودی) چند متغیره بوده و خروجی این مدل‌ها نیز ممکن است یک یا چند متغیره باشد. در بعضی موارد، همانند مورد مطالعاتی این پایان نامه، خروجی‌های مدل وابسته به زمان و به صورت دسته‌ای از سری‌های زمانی است که می‌توان یک یا چند متغیر پاسخ را از آن استخراج نمود. از دیدگاه ریاضی، خروجی یک مدل کامپیوتری را می‌توان تابعی چند متغیره در فضای با بعد بالا به فرم $Y = f(x_1, \dots, x_k)$ در نظر گرفت که x_1, \dots, x_k ورودی‌ها، Y خروجی و معادله ریاضی این تابع (f) به طور صریح مشخص نیست اما قابلیت محاسبه خروجی را برای هر دسته از ورودی‌ها دارد. مدل‌های کامپیوتری مورد بحث در این پایان نامه **مدل‌های تعیینی**^۱ نیز نامیده می‌شوند بدین معنا که اجرای مدل با ورودی‌های همانند، خروجی یکسانی را تولید می‌کند. لذا خروجی آزمایش‌های کامپیوتری، برخلاف آزمایش‌های فیزیکی، تابعی تنها از ورودی‌های مدل بوده و متاثر از خطای تصادفی نیست. شکل (۲-۱) ساختار کلی آزمایش‌های کامپیوتری و فیزیکی را نشان می‌دهد که در آن ε مولفه خطای تصادفی است.

^۱ Deterministic Models



شکل (۲-۱) - ساختار کلی آزمایش‌های کامپیوتری و فیزیکی

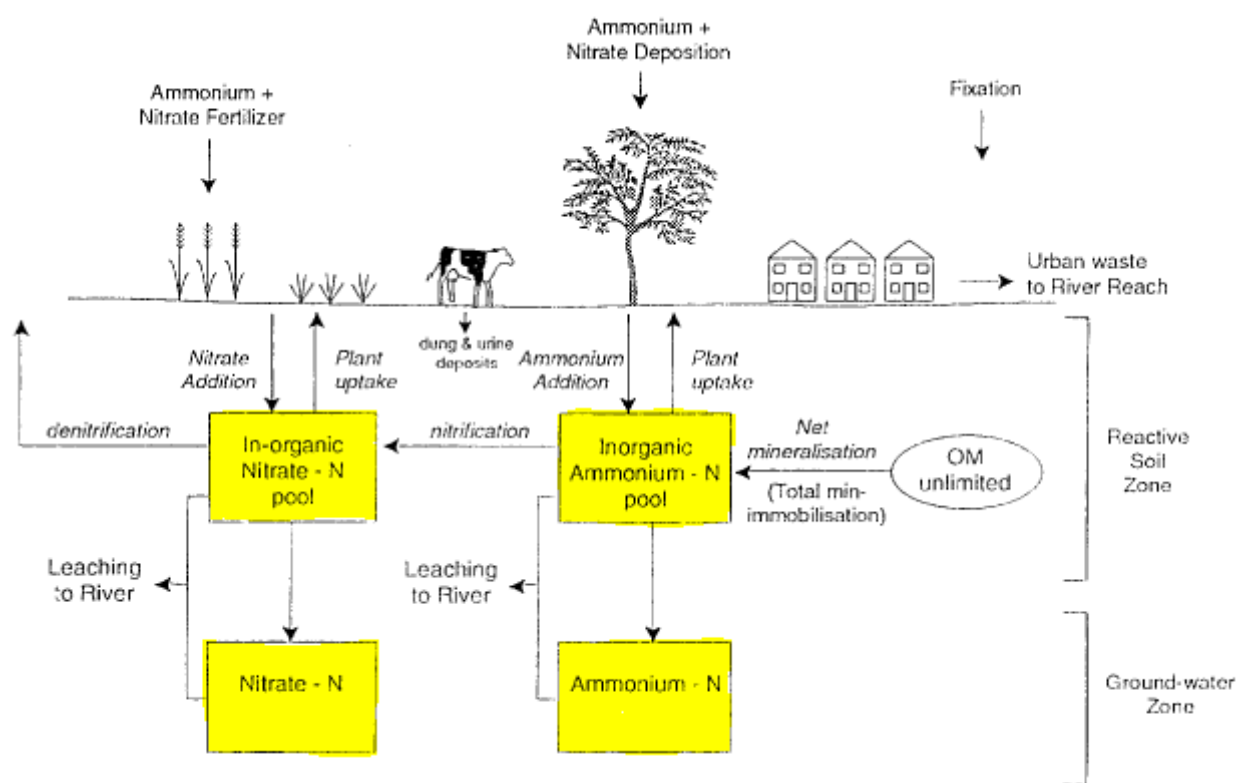
توصیف بسیار مختصر مدل تعیینی *INCA-N* که در این پایان‌نامه به‌عنوان مورد مطالعاتی استفاده شده است می‌تواند به‌درک بهتر ساختار و پیچیدگی این‌گونه مدل‌ها کمک کند. می‌دانیم که عنصر نیتروژن به‌طور طبیعی چرخه مشخصی دارد که در آن، ترکیبات مختلف نیتروژن‌دار تحت شرایط متفاوت به‌وجود می‌آیند. شکل (۲-۲) فرآیند تبادل نیتروژن را در طبیعت نشان می‌دهد.



شکل (۲-۲) - ترکیبات مختلف در چرخه نیتروژن

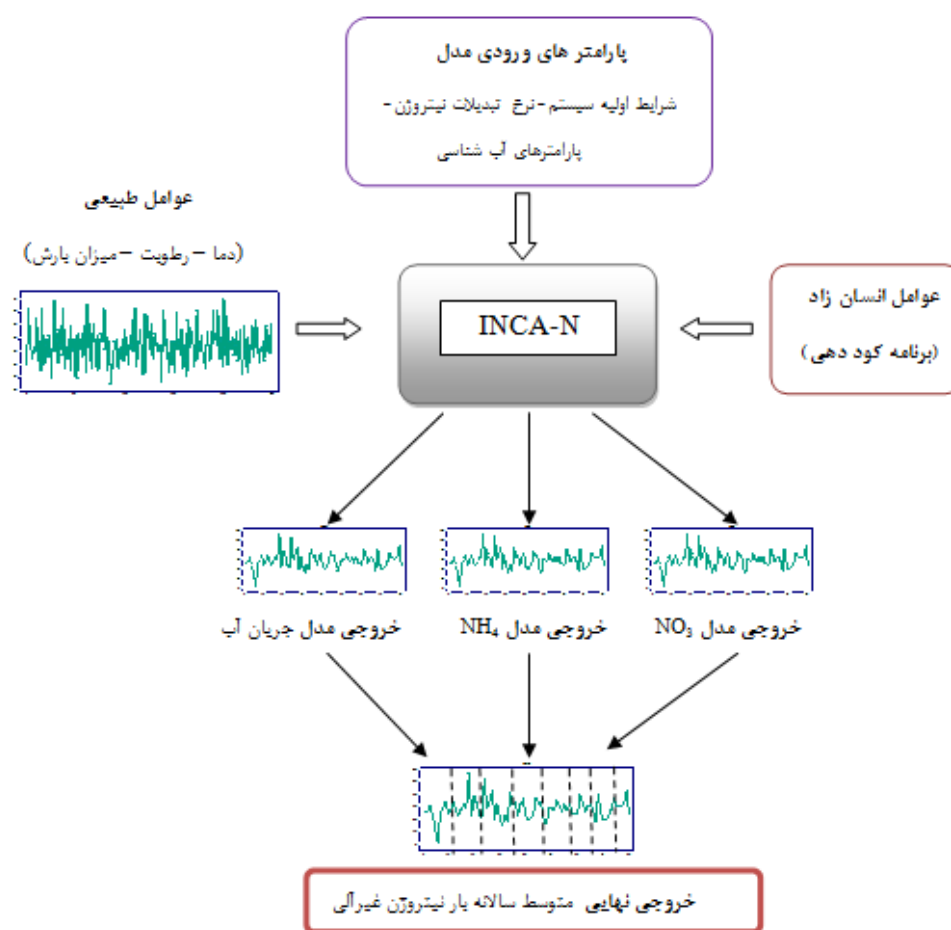
نیتروژن نیاز حیاتی موجودات زنده است. حیوانات و گیاهان نمی‌توانند مستقیماً از نیتروژن موجود در جو استفاده کنند. هنگامی که گیاهان و حیوانات می‌میرند، باقیمانده اجساد آن‌ها بوسیله موجودات ذره بینی موجود در خاک تجزیه و ترکیبات نیتروژن به‌دست می‌آید. این ترکیبات توسط گیاهان مصرف می‌شوند. جانوران با خوردن این گیاهان، ترکیبات نیتروژن را وارد بدن خود می‌کنند. باکتری‌های تجزیه‌کننده‌ی

موجود در خاک، مقداری از ترکیبات نیتروژن‌دار آن را به نیتروژن گازی تبدیل می‌کند. به این ترتیب تقریباً همان اندازه نیتروژنی که از هوا گرفته شده و مصرف شده است، مجدداً به اتمسفر برمی‌گردد. شمای تبادلات نیتروژن در خاک و آب‌های زیر زمینی در شکل (۲-۳) آمده است.



شکل (۲ - ۳) - فرآیند چرخه نیتروژن در خاک و آب‌های زیر زمینی

مدل *INCA-N*، عوامل این چرخه را توسط معادلات دیفرانسیلی به یکدیگر مرتبط ساخته و با استفاده از سری‌های زمانی میزان بارش، دما و رطوبت روزانه در حوضه آبریز رودخانه، میزان نیتروژن ورودی به آن را شبیه سازی می‌کند. ساختار شماتیکی این مدل در شکل (۲-۴) نشان داده شده است.



شکل (۲-۴) - ساختار شماتیکی مدل INCA-N

همان‌طور که در شکل (۲-۴) مشهود است، پارامترهای ورودی مدل را می‌توان در سه بخش گروه بندی کرد. لیست این پارامترهای موثر بر میزان نیتروژن ورودی به رودخانه، در فصل هفتم ذکر شده است.

۲-۲ مفاهیم و تعاریف مشترک آزمایش‌های کامپیوتری و فیزیکی

- عامل^۱: متغیرهای قابل کنترل در آزمایش را عامل و یا ورودی نامند.
- دامنه آزمایش: مقادیر قابل اخذ عامل ورودی می‌باشد که فضای متغیر ورودی نیز گفته می‌شود.

^۱ Factor

- اجرا : انجام آزمایش با سطوح مختلفی از متغیرهای ورودی را اجرای آزمایش نامند.
- پاسخ : نتیجه یک آزمایش، خروجی یا پاسخ نامیده می‌شود.

۲ - ۳ نقش آمار در آزمایش کامپیوتری

فقدان خطای تصادفی در خروجی مدل آزمایش‌های کامپیوتری منجر به عدم کارایی اندازه‌های برگرفته از مجموع مربعات خطا می‌شود. بنابراین مفاهیم کلاسیک طرح آزمایش نظیر بلوک بندی، تکرار و تصادفی-سازی موضوعیت نخواهند داشت. لذا این سوال مطرح می‌شود که نقش آمار در مباحث مرتبط با آزمایش-های کامپیوتری چیست. در پاسخ به این سوال می‌توان به وجود و اهمیت نقش آمار، تحت عناوین زیر اشاره کرد.

طرح آزمایش مدل‌های کامپیوتری

در مطالعه پدیده‌های علمی با مدل‌های کامپیوتری، طولانی بودن زمان اجرای این مدل‌ها از جمله چالش‌های موجود به شمار می‌آید. لذا محققین همواره در صدد ارائه رهیافتی برای تحقق اهداف مطالعات با کمترین تعداد اجرای مدل هستند. یافتن مجموعه مناسبی از نقاط ورودی $S_n = \{X_1, X_2, \dots, X_n\}$ برای اجرای مدل کامپیوتری و به دست آوردن مقادیر $f(X_1), f(X_2), \dots, f(X_n)$ موضوعی است که تحت عنوان طرح آزمایش مدل‌های کامپیوتری مطرح می‌شود. لازم به یادآوری است که X_i بردار مشاهده i -ام از متغیرهای ورودی (x_1, \dots, x_k) است و $i = 1, \dots, n$ می‌باشد. در این زمینه تالیفات متعددی انجام گرفته

است که از جمله آنها می‌توان به گردآوری‌های سانتنر^۱ و همکاران (۲۰۰۳)، ساکس و همکاران (۱۹۸۹) و شاهسونی و گریموال^۲ (۲۰۰۹) اشاره نمود.

انتخاب مدل جانشین^۳ یا متا مدل^۴

با استفاده از یک طرح آزمایش مناسب و با داشتن تعدادی از مقادیر ورودی و خروجی $(X_i, f(X_i) \ i = 1, \dots, n)$ این امکان فراهم می‌شود که تابع f توسط روش‌های پیشرفته آماری برآورد گردد. سپس به جای اجرای پر هزینه مدل کامپیوتری f ، از مدل آماری \hat{f} با هزینه زمانی بسیار کمتر استفاده نمود. موضوع مدل‌سازی آماری که گستره وسیعی از روش‌های رگرسیونی پارامتری و ناپارامتری و روش‌های یادگیری ماشین را در بر دارد، سبب حفظ نقش آمار در مدل‌های کامپیوتری شده به‌طوری‌که نیاز به توسعه این روش‌ها نیز می‌تواند دلیلی بر افزایش اهمیت آمار در زمینه‌های کاربردی باشد. در مباحث مربوط به مدل‌های کامپیوتری و تحلیل حساسیت، \hat{f} را مدل جانشین و یا متامدل می‌گویند (فنگ^۵ و همکاران، ۲۰۰۶). ساختار مدل‌های جانشین در شکل (۲-۵) نشان داده شده است.

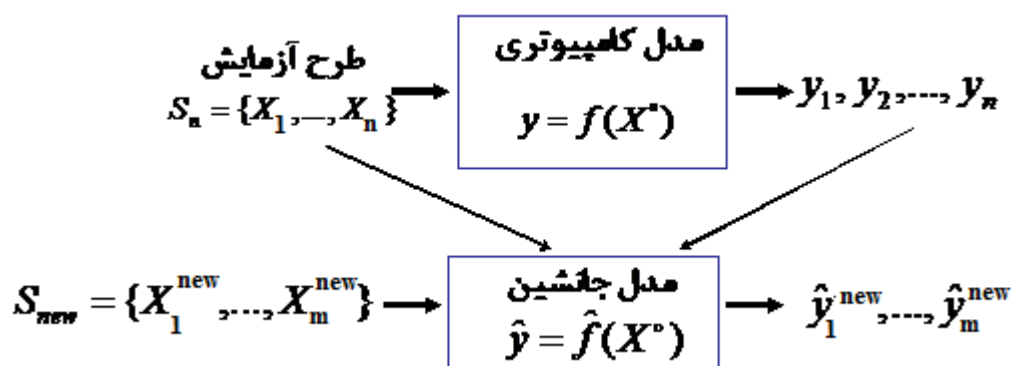
^۱ Santner

^۲ Grimvall

^۳ Surrogate Model

^۴ meta model

^۵ Fang



شکل (۲ - ۵) - چگونگی ساخت مدل جانشین و کاربرد آن

عدم قطعیت

عدم قطعیت مرتبط با انجام پیش‌بینی توسط مدل‌ها و پارامترهای برآوردشده، خود یک مسئله آماری است.

تحلیل حساسیت

نقش آمار در موضوع تحلیل حساسیت مدل‌های کامپیوتری، که هدف آن شناسایی فاکتورهای تاثیرگذار بر خروجی مدل با استفاده از شاخص‌های حساسیت است، را می‌توان در قالب موارد فوق نیز بررسی نمود. انتخاب طرح آزمایش جزو اساسی‌ترین اصول در برآورد شاخص‌های حساسیت است. زیرا مقادیر فاکتورهای ورودی باید به گونه‌ای انتخاب شوند که پوشش‌دهنده کل فضای ورودی مدل باشد. فقدان یک طرح آزمایش مناسب علاوه بر اعمال هزینه زمانی اجرا ممکن است موجب رخداد خطای نوع اول و دوم^۱ شود. از آن‌جا که روش‌های برآورد شاخص‌های حساسیت، نیاز به تعدد اجرای مدل دارند لذا استفاده از مدل‌های جانشین نیز در تحلیل حساسیت حائز اهمیت است.

^۱ خطای نوع اول، بی‌اهمیت دانستن یک ورودی با اهمیت - خطای نوع دوم، با اهمیت دانستن یک ورودی بی اهمیت است.

در این پایان‌نامه، رویکرد آمار در تحلیل حساسیت توسط روش واریانس - مبنای مبتنی بر مفاهیم امیدریاضی و واریانس شرطی مورد مطالعه قرار گرفته است. این مفاهیم دارای تعابیر بسیار ظریف و زیبایی هستند که شاخص‌های حساسیت به بیان آن‌ها می‌پردازد (بخش ۳-۵-۱).

فصل سوم

تحلیل حساسیت مدل‌های کامپیوتری

مقدمه

تعدد پارامترهای ورودی در مدل‌های کامپیوتری علاوه بر پیچیدگی مدل سبب می‌شود که شناسایی متغیرهای تاثیرگذار بر خروجی مدل ضرورت پیدا کند. اهمیت این مسئله در مدل‌هایی که با سلامت انسان، محیط زیست و مسائلی از این دست سروکار دارند، دو چندان است. این مهم یکی از اهداف آزمایش‌های کامپیوتری است که در قالب تحلیل حساسیت مدل بیان می‌شود. در تحلیل حساسیت، میزان اهمیت متغیرهای ورودی با استفاده از اندازه‌های حساسیت بیان می‌شود. میزان اهمیت هر یک از ورودی‌های مدل موجب تغییر در نگرش نسبت به آن ورودی و مطرح شدن مدیریت ریسک می‌شود. در مدیریت ریسک مسائلی مانند تاثیر گذار بودن یا بی تاثیر بودن عامل و نیز تقابل هر یک از عامل‌ها با سایر عوامل حائز اهمیت است. از طرف دیگر هزینه‌های تحمیلی، هزینه‌های نگهداری و کنترل عوامل نیز اهمی نیرومند است که تصمیم‌گیری را با چالش مواجه می‌سازند. در نتیجه به منظور تصمیم‌گیری بهتر بایستی عوامل با اهمیت و قابل کنترل، عوامل با اهمیت و غیر قابل کنترل و عوامل کم اهمیت و بی اهمیت شناسایی شوند. این عمل موجب می‌شود تا تصمیمات اتخاذ شده از پشتوانه محکم و منطقی برخوردار باشند و نتایج حاصل از این تصمیم‌گیری نیز تا حد امکان بهینه گردند

تحلیل حساسیت بر اساس روش‌های مختلفی صورت می‌پذیرد. در ابتدای این فصل به معرفی اجمالی روش‌های مختلف تحلیل حساسیت می‌پردازیم و در ادامه، روش تحلیل حساسیت واریانس- مبنا که اساس کار در این پایان‌نامه است را به‌طور مفصل مورد بحث قرار خواهیم داد. در این روش میزان اهمیت هر یک از متغیرهای ورودی با استفاده از اندازه‌های حساسیت مبتنی بر امیدریاضی شرطی و واریانس شرطی تعیین می‌شود. این اندازه‌های حساسیت، **شاخص حساسیت مرتبه اول و شاخص حساسیت کل** نامیده می‌شوند.

۳- ۱ تحلیل حساسیت

مطالعه چگونگی تسهیم عدم قطعیت خروجی یک مدل به منابع عدم قطعیت موجود در ورودی‌های مدل را **تحلیل حساسیت** می‌نامند (سالتلی^۱ و همکاران، ۲۰۰۸). به بیان دیگر تحلیل حساسیت، به منظور ارائه روشی جهت تعیین تاثیر پارامترهای ورودی بر خروجی مطرح می‌شود (سالتلی، ۲۰۰۴). تحلیل گران مدل نیز از تحلیل حساسیت برای کشف خطاهای فنی در مدل، شناسایی نواحی بحرانی متغیرهای ورودی، مشخص کردن اولویت بندی متغیرهای مورد بررسی و ساده سازی مدل بهره می‌گیرند.

۳- ۲ روش‌های تحلیل حساسیت

روش‌های تحلیل حساسیت از تنوع بسیاری برخوردارند که در این جا ابتدا چهار روش زیر را به اختصار معرفی نموده و سپس روش به کار رفته در این پایان نامه، یعنی روش **واریانس - مبنا^۲** را در بخش‌های بعدی به طور جامع معرفی خواهیم کرد. این روش‌ها عبارتند از

۱. مشتق - مبنا^۳

۲. نمونه‌گیری - مبنا^۴

۳. فیلترسازی مونت کارلو - مبنا^۵

۴. غربالگری^۱

^۱ Saltelli

^۲ Variance-based Methods

^۳ Derivative-based

^۴ Sampling-based

^۵ MonteCarlo filtering- based

۳ - ۲ - ۱ روش مشتق - مبنا

در این روش، معیار اهمیت نسبی هر متغیر ورودی مانند X_i براساس مشتق جزئی آن یعنی $\frac{\partial f(x_1, \dots, x_k)}{\partial x_i}$ بیان می‌شود. به عنوان مثال در مدل $f(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon$ که

در آن $(\beta_0, \dots, \beta_k)$ ضرائب و (x_1, \dots, x_k) متغیرهای مستقل می‌باشند، داریم

$$\frac{\partial f(x_1, \dots, x_k)}{\partial x_i} = \beta_i \quad i = 1, 2, \dots, k \quad (۱-۳)$$

یعنی با ثابت در نظر گرفتن سایر متغیرها، ضریب β_i را می‌توان به عنوان میزان اهمیت نسبی متغیر X_i در پاسخ Y در نظر گرفت که از نظر مبانی رگرسیونی نیز این مطلب تایید می‌شود. رابطه فوق در مدل‌های کامپیوتری منجر به ایجاد دستگاه معادلاتی می‌شود که حل آن با روش‌های عددی امکان پذیر است. این روش زمان زیادی از اجرای برنامه کامپیوتری را به خود اختصاص می‌دهد (سالتلی و همکاران، ۲۰۰۸).

۳ - ۲ - ۲ روش نمونه گیری - مبنا

روش نمونه گیری - مبنا برای تحلیل حساسیت سیستم‌های پیچیده استفاده می‌شود. این روش که توسط هلتن^۲ و همکاران، (۲۰۰۶) پیشنهاد گردید، در پنج مرحله زیر انجام می‌شود. مراحل ۱ تا ۳ را "تولید" و مراحل بعدی را "بررسی" می‌نامند.

۱. یافتن تابع توزیع مناسب برای هر یک از متغیرهای ورودی به طوری که توصیف کننده مناسبی برای خصوصیات و ویژگی‌های متغیر ورودی باشد.

^۱ Screening

^۲ Helton

۲. تولید نمونه برای متغیرهای $X=[x_1, \dots, x_k]$ بر اساس توابع توزیعی آنها

۳. اجرا مدل برای نمونه‌های به دست آمده در مرحله ۲

۴. نمایش عدم قطعیت نتایج خروجی مدل همانند تقریب تابع توزیع و...

۵. تحلیل حساسیت خروجی مدل یعنی بررسی در نگاشت $X_i \rightarrow Y$ به ازای $i=1, \dots, k$

در واقع مرحله بررسی پس از اجرای مدل و به منظور ایجاد یک نگاشت از متغیرهای ورودی $X=[x_1, \dots, x_k]$ (تعداد متغیرها) به خروجی Y طراحی شده است. لازم به ذکر است که تحلیل حساسیت در مرحله پنجم با استفاده از تکنیک‌هایی مانند نمودار پراکنش Y در مقابل X_i ، انجام روش‌های رگرسیون پارامتری و گام به گام، ضریب همبستگی جزئی، رگرسیون ناپارامتری، تجزیه واریانس خروجی، آزمون‌های کلموگروف - اسمیرنوف و غیره انجام می‌شود.

۳ - ۲ - ۳ روش فیلترسازی مونت کارلو - مبنا

تحلیل حساسیت معمولاً بر اساس کلیه مقادیر خروجی در سرتاسر فضای متغیرهای ورودی انجام می‌شود. اما در برخی از موارد تنها بخش (مقادیر) خاصی از خروجی مدل مورد توجه است. مقادیر خاص می‌توانند توسط یک مقدار آستانه‌ای مشخص، مقادیر فرین و یا ناحیه‌ای از خروجی تعیین شوند. این مقادیر را مقادیر (نواحی) مطلوب می‌گوییم و آن‌را با نماد B نشان می‌دهیم. همچنین سایر مقادیر خروجی مدل را مقادیر نامطلوب نامیده و آن‌را با \bar{B} نمایش می‌دهیم. ناحیه B متناظر با افرازی از فضای متغیرهای ورودی است. بنابراین با هر بار اجرای مدل، مقدار خروجی در ناحیه مطلوب یا ناحیه نامطلوب قرار می‌گیرد. تقسیم‌بندی فضای خروجی را فیلترسازی مونت کارلو می‌نامند (سالتلی و همکاران، ۲۰۰۸). هدف در این روش، شناسایی آن دسته از متغیرهای ورودی است که بیشترین نقش را در تولید ناحیه مطلوب ایفا

می‌کنند. برای این منظور، با اعمال مرحله "تولید" در روش نمونه‌گیری- مبنا و با N بار اجرای مدل، تحلیل حساسیت هر یک از متغیرهای ورودی مدل (X_i) مطابق الگوریتم زیر انجام می‌شود.

۱. آن دسته از مقادیر X_i که در ایجاد ناحیه مطلوب B نقش داشته‌اند را جدا نموده و آن را با A_B نشان می‌دهیم. فرض کنید تعداد این نقاط برابر n باشد.

۲. سایر مقادیر X_i که در ایجاد ناحیه نامطلوب \underline{B} نقش دارند را نیز به‌عنوان مجموعه دیگری به نام $A_{\underline{B}}$ در نظر می‌گیریم. تعداد این نقاط برابر با $m=N-n$ است.

۳. فرض کنید $f_n(X_i/B)$ و $f_m(X_i/\underline{B})$ به ترتیب توزیع‌های نامعلوم مجموعه‌های A_B و $A_{\underline{B}}$ باشند.

۴. تحلیل حساسیت نسبت به متغیر X_i ، توسط آزمون استقلال دو نمونه‌ای اسمیرنف به‌صورت زیر انجام می‌شود.

$$H_0 : f_n(X_i | B) = f_m(X_i | \underline{B})$$

$$H_1 : f_n(X_i | B) \neq f_m(X_i | \underline{B})$$

اگر متغیر X_i در تولید ناحیه مطلوب نقشی ایفا نکند انتظار می‌رود که توابع چگالی مذکور مشابه یکدیگر باشند و فرضیه صفر رد نشود. برعکس اگر X_i در ایجاد ناحیه مطلوب تاثیرگذار باشد آن‌گاه بایستی فرضیه صفر رد شود. این روش را تحلیل حساسیت منطقه‌ای^۱ (RSA) نیز می‌نامند. به‌طور کلی روش RSA در پی شناسایی ورودی‌هایی از مدل است که نواحی مطلوب و نامطلوب خروجی را شکل می‌دهند (هورنبرگر^۲ و اسپیر، ۱۹۸۱).

^۱ Regionalized Sensitivity Analysis

^۲ Horenberg and Spear

۳ - ۲ - ۴ روش غربالگری

در برخی از موارد، مدل مورد بررسی شامل ده‌ها و یا صدها پارامتر ورودی است. با توجه به گستردگی بعد فضای متغیرهای ورودی لازم است تا اطلاعات جامعی از رفتار مدل داشته باشیم. این مسئله مستلزم انتخاب یک طرح آزمایش مناسب است که تعداد اجرای مدل را تا حد امکان کاهش دهد. موریس^۱ (۱۹۹۱) برای حل این مسئله استفاده از طرح آزمایش‌های گزینشی را پیشنهاد داد. این طرح آزمایش بیشتر برای مدل‌هایی کاربرد دارد که تعداد متغیرهای تاثیرگذار نسبت به متغیرهای بی‌تاثیر کم باشند. موریس مفهوم مهم و پرکاربرد روش‌های گزینشی را تحت عنوان اثر مقدماتی معرفی نمود و دو اندازه حساسیت را برای هر متغیر ورودی مدل پیشنهاد کرد. کمپلانگو^۲ و سالتلی (۲۰۰۶) نیز بعداً این روش را بهبود بخشیدند. لازم به ذکر است که طرح‌های آزمایش مدل‌های تعیینی که موضوعی بسیار جامع است توسط سانتنر و همکاران (۲۰۰۳) گردآوری شده است. روش‌های گزینشی در مسائل مرتبط با مدیریت ریسک و انتخاب گروهی متغیرهای تاثیرگذار بر خروجی مدل کاربرد دارند.

۳ - ۳ معیارهای انتخاب روش تحلیل حساسیت

تاکنون روش‌های مختلف تحلیل حساسیت به‌طور اجمالی معرفی شد. حال این سوال مطرح می‌شود که معیارهای انتخاب روش تحلیل حساسیت مناسب و ایده‌آل چیست؟ سالتلی (۲۰۰۴) در پاسخ به سوال فوق ملاک‌های روش ایده‌آل را مطابق زیر معرفی نمود.

- معمولاً تولید نقاط آزمایش براساس فرم توزیع متغیرهای ورودی انجام می‌پذیرد. روش تحلیل حساسیت ایده‌آل نباید به مقیاس متغیرهای ورودی و نوع توزیع آن‌ها بستگی داشته باشد.

^۱ Morris

^۲ Campolongo

- امکان محاسبه متوسط چند بُعدی یا $E(Y | X_i)$ وجود داشته باشد.
 - اندازه‌های حساسیت نباید به فرم تابعی مدل وابسته باشند. در واقع تحلیل حساسیت در هر روشی باید بدون توجه به خطی یا غیر خطی بودن مدل، انجام گیرد و قادر به شناسایی اثر متقابل بین متغیرها، به خصوص در مدل‌های غیر خطی باشد.
 - روشی ایده آل است که قابلیت بیان رفتار متغیرهای ورودی را به صورت گروهی داشته باشد. در غیر این صورت با کثرت و تعدد اندازه‌های حساسیت مواجه شده و تفسیر نتایج دشوار خواهد بود.
- بنابر مطالب فوق می‌توان هر یک از روش‌های معرفی شده بخش (۳-۲) را مطابق جدول (۳-۱) ارزیابی کنیم.

جدول (۳-۱) - قابلیت‌های روش‌های مختلف تحلیل حساسیت

ویژگی روش تحلیل حساسیت	امکان مقابله با اثر شکل و مقیاس	امکان محاسبه میانگین چندبُعدی	امکان مستقل از فرم تابعی مدل	امکان بررسی گروهی متغیرها
مشتق - مبنا	x	x	x	✓
نمونه‌گیری - مبنا	✓	✓	x	x
روش‌های گزینشی	x	✓	✓	✓
فیلترسازی مونت کارلو	✓	✓	✓	x

نمادهای ✓ و x در جدول فوق به ترتیب به معنای دارا بودن خاصیت و عدم داشتن خاصیت مورد نظر است. همان‌طور که در جدول فوق مشهود است، هر یک از روش‌های مذکور دارای نواقصی هستند. روش

تحلیل حساسیت واریانس- مبنا که اساس کار ما در این تحقیق است، به‌منظور رفع این نواقص طراحی گردیده است.

۳-۴ تاریخچه روش تحلیل حساسیت واریانس - مبنا

روش تحلیل حساسیت واریانس- مبنا در دهه ۱۹۷۰ برای اولین بار توسط گروهی از شمیدانان به سرپرستی فردی به نام کیوکر^۱ (۱۹۷۳، ۱۹۷۹) معرفی شد. روش این شمیدانان تحت عنوان *FAST*^۲ مطرح و به‌کار برده شد. آن‌ها برای اندازه‌گیری تاثیر هر متغیر ورودی بر خروجی مدل لیزر شیمیایی، از واریانس شرطی برای بیان اندازه‌ای به نام اثر تکی یا اثر مرتبه اول استفاده کردند. اگرچه روش آن‌ها بسیار کارا بود اما قابلیت محاسبه اثرات متقابل ورودی‌ها بر خروجی مدل را نداشت. پس از مدتی ایمان و هورا^۳ (۱۹۸۶) اندازه حساسیت پیشنهادی خود را براساس واریانس شرطی و تحت عنوان "اهمیت عدم قطعیت"^۴ معرفی کردند. آن‌ها اندازه حساسیت پیشنهادی خود را برای متغیر i -ام به صورت رابطه

$$I_i = \sqrt{\text{Var}(Y) - E(\text{Var}(Y | X_i))}$$

بیان نمودند. این اندازه حساسیت، متوسط مقدار کاهش در واریانس خروجی مدل را هنگامی که متغیر ورودی ثابت نگه داشته شود را محاسبه می‌کند. این دو محقق در سال ۱۹۹۰ اندازه حساسیت دیگری را با استفاده از میانگین و واریانس شرطی و بر پایه لگاریتم خروجی مدل به صورت زیر معرفی کردند.

^۱ Cukier

^۲ Fourier Amplitude Sensitivity Test

^۳ Iman and Hora

^۴ Uncertainty importance

$$\frac{Var_{X_i}(E(\log(Y) | X_i))}{Var(\log(Y))} \quad (2-3)$$

اگر چه این اندازه استوار^۱ است. اما برگرداندن نتایج تحلیل حساسیت از لگاریتم به حالت عادی دشوار می‌باشد. درهمین راستا شاخص‌های حساسیت دیگری نیز توسط ایمان و هورا (۱۹۹۰)، کریکاز و هایسمن^۲ (۱۹۹۰)، سالتلی (۱۹۹۳)، هما و سالتلی (۱۹۹۶)، هما و ایشیگامی^۳ (۱۹۹۶) و مک‌کی^۴ (۱۹۹۶) تحت عنوان اندازه اهمیت و یا نسبت همبستگی معرفی شدند. بازبینی بصری نتایج حساسیت روشی بود که توسط ساکس و همکاران وی در سال ۱۹۸۹ پیشنهاد شد. آن‌ها خروجی مدل را به توابعی با بعدهای افزایشی تجزیه کردند و سپس از نمودارهای همین توابع برای تحلیل حساسیت استفاده نمودند. یک ریاضیدان بزرگ روسی به نام سُبُل^۵ که هنوز از چهره‌های شاخص در این زمینه است، در سال ۱۹۹۰ کارهای کیوکر را گسترش و به آن عمومیت بخشید و روش مونت کارلو- مبنای^۶ را برای محاسبه اندازه حساسیت هر دسته از فاکتورهای ورودی پیشنهاد داد. کار او در سال ۱۹۹۳ به زبان انگلیسی منتشر و فراگیرتر شد. همچنین این ریاضیدان در سال ۱۹۹۶، اندازه حساسیت جدیدی را بر اساس همبستگی بین خروجی مدل و امید ریاضی شرطی نیز معرفی نمود. سالتلی و همکارانش در سال ۱۹۹۹ با استفاده از نتایج کار سبل، کارآیی روش FAST را افزایش داده و با طرح روش اصلاح شده EFAST (Extended Fast) و شاخص حساسیت دیگری به نام اثر کل هر متغیر، گامی نو در این مسیر گذاشتند. این افراد،

^۱ Robust

^۲ Krzykacz-Hausmann

^۳ Ishigami and Homma

^۴ Mckay

^۵ I. M. Sobol

^۶ Monte carlo- base

گروهی از محققین بخش آمار کاربردی و اقتصادسنجی انیستیتو محافظت و امنیت شهروندان در کمیسیون تحقیقات اتحادیه اروپا^۱ هستند که کماکان فعالیت‌های زیادی در زمینه تحلیل حساسیت مدل‌ها دارند. آن‌ها همواره از یک مورد مطالعاتی به نام مدل *LEVEL E* برای ارزیابی کلیه روش‌های پیشنهادی خود استفاده می‌کنند. در مدل *LEVEL E* میزان تشعشعات ناشی از دفن زباله‌های هسته‌ای بر انسان شبیه سازی می‌شود. از آن‌جا که طول زمان اجرای مدل، پارامتری مهم جهت انتخاب روش برآورد شاخص‌های حساسیت است، سالتلی در سال ۲۰۰۲ روش جدیدی را برای کاهش محاسبات شاخص‌های حساسیت مونت کارلو- مبنا پیشنهاد کرد و آن را برای مدل *PMOD* که ساخت و دفع هیدروکربن‌ها را از یک توده صخره محاسبه می‌کند، به کار برد. در ادامه توسعه روش *FAST*، تارانتولا^۲ در سال ۲۰۰۶ برای نخستین بار روش طرح متعادل تصادفی^۳ (*RBD*) را برای محاسبه شاخص حساسیت مرتبه اول ارائه داد و آن را برای مدل *LEVEL E* پیاده‌سازی نمود. همچنین در راستای توسعه روش مونت کارلو - مبنا، مجدداً سالتلی و گروه تحقیقاتی او در سال ۲۰۱۰ ضمن مقایسه روش‌های موجود از نقطه نظر هزینه محاسبات، پیشنهادی را برای کاهش هزینه اثر کل ارائه دادند. سرانجام مارا^۴ در سال ۲۰۰۹ روش تارانتولا را بهبود بخشیده و شاخص حساسیت مرتبه دوم را محاسبه کرد.

^۱ Econometrics and Applied Statistics division, Institute for the Protection and Security of the Citizen (IPSC), Joint Research Centre of the European Commission (JRC)

^۲ Tarantola

^۳ Random Balance Design (RBD)

^۴ Mara

۳ - ۵ پیکربندی روش واریانس - مبنا

در فصل قبل مدل‌های تعیینی (مدل‌های کامپیوتری) و ویژگی‌های آن معرفی شدند. لازم به یادآوریست که در این مدل‌ها، X_i ها ورودی‌های مدل و $Y = f(X_1, \dots, X_k)$ خروجی مدل تعیینی است که ضابطه f نامشخص بوده اما امکان محاسبه مقدار Y برای هر دسته مقادیر (X_1, \dots, X_k) وجود دارد.

۳ - ۵ - ۱ شاخص حساسیت مرتبه اول

تعریف اندازه شاخص حساسیت مرتبه اول با این سوال آغاز می‌شود که اگر متغیر X_i مقدار ثابت C را بپذیرد چه تاثیری بر عدم قطعیت Y خواهد داشت. مسلماً در این حالت هر چقدر اهمیت X_i در خروجی مدل کمتر باشد، واریانس Y کمتر تحت تاثیر واقع می‌شود و برعکس. برای بیان کمی میزان تاثیر ثابت نگه داشتن X_i بر عدم قطعیت Y می‌توان از واریانس شرطی $V_{X_{-i}}(Y | X_i = c)$ استفاده نمود که در آن $V_{X_{-i}}$ به معنای واریانس بر روی همه متغیرها به غیر از X_i است. از آنجا که همواره

$$V_{X_{-i}}(Y | X_i = c) \leq V(Y)$$

و یا

$$0 \leq V(Y) - V_{X_{-i}}(Y | X_i = c)$$

لذا هر چقدر اهمیت X_i کمتر باشد مقدار $V_{X_{-i}}(Y | X_i = c)$ به $V(Y)$ نزدیکتر و تفاضل آنها مقدار کوچکی خواهد بود. همچنین با افزایش اهمیت X_i ، این تفاضل رفته رفته افزایش می‌یابد. در نتیجه می‌توان از عبارت $V(Y) - V_{X_{-i}}(Y | X_i = c)$ به عنوان اندازه تغییرات در واریانس Y وقتی که X_i در مقدار c ثابت نگه داشته شود، استفاده کرد. از آنجا که این اندازه به مقدار c وابسته است لذا با میانگین‌گیری بر روی مقادیر

ممکن c این وابستگی را از بین برده و از کمیت $E_{X_i}(V_{X_{-i}}(Y|X_i=c))$ به جای $V_{X_{-i}}(Y|X_i=c)$ استفاده می‌کنیم. بنابراین می‌توان مقدار عبارت

$$V_i = V(Y) - E_{X_i}(V_{X_{-i}}(Y|X_i=c)) \quad (۸-۳)$$

را به‌عنوان میزان تاثیر X_i بر واریانس Y دانست. از طرفی دیگر بنابر خاصیت واریانس مبنی بر اینکه

$$V(Y) = E(V(Y|X_i)) + V(E(Y|X_i))$$

می‌توان $V(E(Y|X_i))$ را به‌عنوان اثر مرتبه اول X_i بر Y معرفی نمود. این کمیت وابسته به مقیاس است که با تقسیم آن بر واریانس Y این نقیصه برطرف شده و در نهایت

$$S_i = \frac{V(E(Y|X_i))}{Var(Y)} \quad (۹-۳)$$

را به‌عنوان شاخص حساسیت مرتبه اول X_i در نظر می‌گیرند. شاخص حساسیت مرتبه اول نشان دهنده سهم اثر اصلی ورودی مدل بر واریانس خروجی است و همواره $0 \leq S_i \leq 1$ است.

به‌همین ترتیب می‌توان با ثابت نگه‌داشتن تعداد متغیرهای بیشتری، اندازه حساسیت Y نسبت به ثابت بودن آن‌ها را با استفاده از شاخص‌های حساسیت مرتبه دوم و مراتب بالاتر محاسبه کرد. در واقع شاخص‌های حساسیت مراتب بالاتر، اثرات متقابل بین متغیرهای ورودی را بر خروجی مدل اندازه‌گیری می‌کنند.

مثال ۱- مدل به‌معادله $Y = 3X_1 - 2X_2 + 6X_3 + 2X_4$ که $X_1 \sim N(0,6)$ ، $X_2 \sim U(0,5)$ ، $X_3 \sim N(2,5)$ ، $X_4 \sim N(3,3)$ و مستقل از یکدیگر هستند را با $V(Y) = 266$ در نظر بگیرید. در این صورت شاخص حساسیت مرتبه اول متغیر X_1 عبارت است از

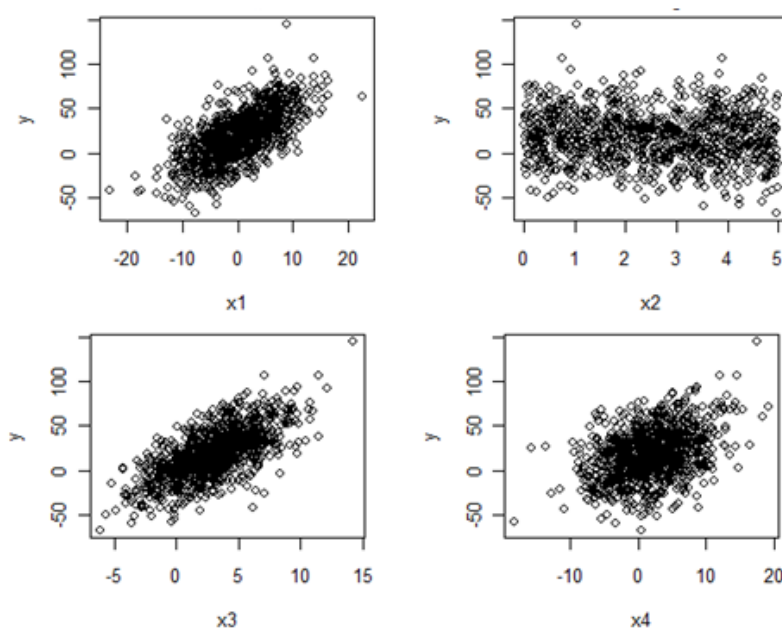
$$E(Y | X_1) = 3X_1 - 2(0) + 6(2) + 2(3) = 3X_1 + 22$$

$$V(E(Y | X_1)) = 9V(X_1) = 54$$

$$S_1 = \frac{V(E(Y | X_1))}{Var(Y)} = \frac{54}{54 + 20 + 180 + 12} = 0.2030$$

و به همین ترتیب داریم $S_4 = 0.04511$, $S_3 = 0.6766$, $S_2 = 0.0751$.

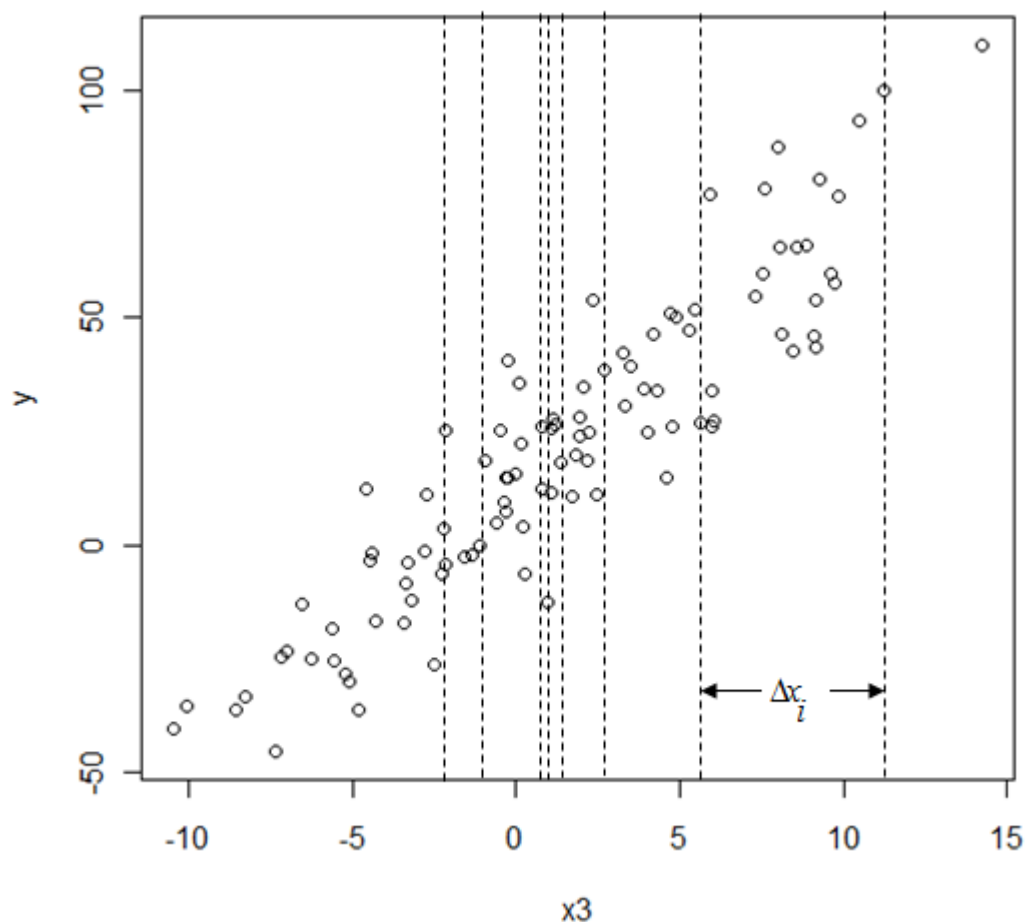
برای درک شهودی شاخص حساسیت مرتبه اول می‌توان نمودار پراکنش Y را نسبت به هر یک از X_i ها رسم نمود. هر چه این نمودار بتواند القا کننده الگویی خاص و غیریکنواخت باشد، متغیر X_i با اهمیت‌تر است، اما اگر نتوان الگوی خاصی در نمودار پراکنش بیابیم، متغیر مربوطه کم‌اهمیت و یا بی‌اهمیت به شمار می‌آید. شکل (۳-۱) نمودار پراکنش Y در مقابل متغیرهای توضیحی مثال ۱ را نشان می‌دهد.



شکل (۳-۱) - نمودار پراکنش Y در مقابل متغیرهای توضیحی

همان‌طور که در شکل مشهود است، وابستگی Y با متغیرهای X_2 , X_4 ناچیز بوده و در نمودار پراکنش این متغیرها الگوی خاصی یافت نمی‌شود. همچنین وابستگی Y به X_3 بیشتر از X_1 بوده و عملاً می‌توان نتیجه گرفت که $S_3 > S_1 > S_2 > S_4$. برای درک عمیق‌تر مفهوم واریانس‌های شرطی و نیز شاخص

حساسیت می‌توان محور X_i را به تکه‌های کوچکی تقسیم کرد تا صفحه مختصات به باریکه‌هایی مطابق شکل (۲-۳) درآید.



شکل (۳ - ۲) - تقسیم نمودار پراکنش به باریکه‌های با طول Δx_i

حال توجه خود را به این باریکه‌ها با پهنای Δx_i که در آن میانگین مقادیر Y معلوم است معطوف می‌کنیم. اگر میانگین Y ها در تمامی باریکه‌ها تغییرات کمی داشته باشد، آنگاه متغیر X_i کم اهمیت به‌شمار می‌آید و برعکس اگر این مقادیر دارای نوسانات زیادی باشند، X_i یک متغیر موثر است. به بیان دیگر میانگین Y وقتی که Δx_i به صفر میل کند همان امیدریاضی شرطی $E(Y | X_i)$ و تغییرات در میانگین Y وقتی که Δx_i به صفر میل کند مبین $V(E(Y | X_i))$ است.

۳ - ۵ - ۲ اثرات متقابل بین ورودی‌های مدل

تغییرات خروجی یک مدل ممکن است متأثر از ترکیب برخی از پارامترهای ورودی آن باشد که این مسئله دلالت بر وجود اثر متقابل بین متغیرهایی ورودی مدل دارد. در این صورت نمی‌توان تنها با استفاده از اثر اصلی، حساسیت خروجی مدل را بیان کرد. شناسایی اثرات متقابل بین متغیرهای ورودی یکی از مهمترین مسائل مدل‌های پیچیده است.

شاخص حساسیت مرتبه دوم X_i و X_j به صورت

$$S_{ij} = \frac{V_{ij}}{V(Y)} = \frac{V(E(Y | X_i, X_j))}{V(Y)} - S_i - S_j \quad (3-3)$$

تعریف می‌شود که در آن $V(E(Y/X_i, X_j))$ اندازه اثر توأم X_i و X_j بر خروجی مدل (Y) است. همچنین اثر متقابل بین X_i و X_j و X_k به صورت

$$S_{ijk} = \frac{V_{ijk}}{V(Y)} = \frac{V(E(Y | X_i, X_j, X_k))}{V(Y)} - S_{ij} - S_{jk} - S_{ik}$$

بیان شده و به همین ترتیب شاخص‌های حساسیت مراتب بالاتر نیز تعریف می‌گردند.

۳ - ۵ - ۳ اثر کل

هزینه محاسباتی اثرات متقابل ورودی‌های مدل با افزایش تعداد ورودی‌ها، به صورت نمایی افزایش یافته و عملاً محاسبه شاخص‌های حساسیت مرتبه دوم، سوم و مراتب بالاتر مقرون به صرفه نمی‌باشد. در نتیجه بسیاری از فعالان عرصه تحلیل حساسیت به دنبال شاخص جایگزینی بودند که بتواند علاوه بر کاهش هزینه محاسباتی، حاوی اطلاعات مفیدی را در خصوص تقابل اثرات باشد. هما و سالتلی (۱۹۹۶) شاخصی را تحت عنوان **شاخص حساسیت کل** یا **اثر کل** معرفی نمودند. این شاخص را با نماد S_T نشان داده و به

صورت مجموع اثر تکی و اثرات توام آن با سایر متغیرها تعریف می‌کنند. به عنوان مثال اگر خروجی مدل تابعی از ۳ متغیر باشد آن گاه اثر کل X_1 به صورت $S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$ تعریف می‌شود.

به منظور ارائه یک اندازه برای اثر کل، واریانس خروجی مدل را به نوعی دیگر و به صورت تجزیه آن به شرط ثابت بودن تمام متغیرهای ورودی به جز X_i ، یعنی $X_{\sim i}$ ، به صورت زیر در نظر می‌گیریم.

$$V(Y) = V(E(Y | X_{\sim i})) + E(V(Y | X_{\sim i})) \quad (۳-۴)$$

از آنجا که بنا به تعاریف قبل می‌توان $V(E(Y | X_{\sim i}))$ را به عنوان اثر مرتبه اول $X_{\sim i}$ در نظر گرفت لذا $E(V(Y | X_{\sim i})) = V(Y) - V(E(Y | X_{\sim i}))$ دربرگیرنده تمام اثراتی است که به نوعی متغیر X_i در آن مشارکت دارد. با تقسیم طرفین این رابطه بر $V(Y)$ شاخص حساسیت کل به صورت زیر معرفی می‌گردد.

$$S_{T_i} = \frac{E(V(Y | X_{\sim i}))}{V(Y)} = 1 - \frac{V(E(Y | X_{\sim i}))}{V(Y)} \quad (۳-۵)$$

۳-۶ تجزیه واریانس خروجی مدل بر اساس شاخص‌های حساسیت

در این بخش به بیان ارتباط بین شاخص‌های حساسیت مرتبه اول و سایر مراتب می‌پردازیم. همانطور که در بخش‌های قبل ذکر شد، تاثیر ورودی‌های چند متغیره بر خروجی مدل تعیینی می‌تواند به صورت اثرات تکی یا توام ظاهر شود. لذا طبیعی به نظر می‌رسد که بتوان $Y = f(X_1, \dots, X_k)$ را به صورت بسط سلسله مراتبی از اجزا به هم وابسته و برحسب متغیرهای تصادفی (متغیرهای ورودی) بیان کرد. رابطه (۳-۶) بسط سلسله مراتبی این اجزا را نشان می‌دهد.

$$Y = f_0 + \sum_i f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \dots + f_{12\dots k}(x_1, x_2, \dots, x_k) \quad (۳-۶)$$

که f_i مبین اثرات تکی X_i ، f_{ij} مبین اثر توام متغیرهای X_i, X_j و به همین ترتیب $f_{12, \dots, k}$ اثر توام آن‌ها است. در واقع بسط (۳-۶)، تجزیه تابع f برحسب توابعی با ابعاد افزایشی^۱ (HDMR) است. اجزا این بسط یکتا نبوده و بی‌شمار انتخاب برای آن‌ها وجود دارد. جهت تسهیل در محاسبه این اجزا، Y را تابعی k -متغیره در ابرمکعب واحد k -بعدی و X_i ها را متغیرهای تصادفی مستقل با توزیع یکنواخت در بازه (۱ و ۰) در نظر گرفته/یم. سُبُل (۱۹۹۳) با در نظر گرفتن مفروضات فوق نشان داد که اگر اجزا رابطه (۳-۶) به‌عنوان توابعی از متغیرهای ورودی مدل (متغیرهای تصادفی) دارای میانگین صفر باشند، یعنی

$$\int f_i dx_i = 0 \quad \iint f_{ij} dx_i dx_j = 0 \quad \iiint f_{ijl} dx_i dx_j dx_l = 0 \quad (۷-۳)$$

و یا در حالت کلی $\int f_{i_1, i_2, \dots, i_s}(x_{i_1}, \dots, x_{i_s}) dx_{i_k} = 0 \quad 1 \leq s \leq k$ آنگاه این اجزا یکتا خواهند بود. با برقراری شروط فوق، این اجزا برحسب امیدریاضی‌های شرطی Y به‌دست می‌آیند.

$$f_0 = E(Y)$$

$$f_i = E(Y | X_i) - E(Y) \quad (۸-۳)$$

$$f_{ij} = E(Y | X_i, X_j) - f_i - f_j - E(Y)$$

...

براساس نتیجه فوق می‌توان رابطه (۳-۶) را به صورت زیر بازنویسی کرد.

$$Y = E(Y) + \sum_i (E(Y | X_i) - E(Y)) + \sum_i \sum_{j>i} (E(Y | X_i, X_j) - E(Y | X_i) - E(Y | X_j) - E(Y)) + \dots \quad (۹-۳)$$

تجزیه فوق، بسط سُبُل نامیده می‌شود. این رابطه نشان می‌دهد که چگونه می‌توان تابعی از متغیرهای

^۱High Dimensional Model Representation

تصادفی را برحسب امیدریاضی‌های شرطی آن بیان کرد. علاوه براین، ارتباط اجزا رابطه (۳-۶) و واریانس-های شرطی Y به صورت زیر بیان می‌شود.

$$V_i = V(f_i(X_i)) = V(E(Y | X_i))$$

$$V_{ij} = V(f_i(X_i, X_j)) = V(E(Y | X_i, X_j)) - V_i - V_j$$

$$V_{ijk} = V(f_i(X_i, X_j, X_k)) = V(E(Y | X_i, X_j, X_k)) - V_{ij} - V_{jk} - V_{ik}$$

...

تمام این جملات توسط رابطه زیر به یکدیگر مرتبط می‌شوند.

$$V(Y) = \sum_{i=1}^k V_i + \sum_{1 \leq i < j \leq k} V_{ij} + \dots + V_{1,2,\dots,k} \quad (۳-۱۰)$$

که آن را نمایش تجزیه واریانس با ابعاد افزایشی ($ANOVA-HDMR^1$) می‌گویند (آرچر^۲ و همکاران، ۱۹۹۷). با تقسیم عبارات فوق بر واریانس کل، رابطه شاخص‌های حساسیت مرتبه اول و مراتب بالاتر به صورت زیر به دست می‌آید.

$$1 = \sum_{i=1}^k S_i + \sum_{1 \leq i < j \leq k} S_{ij} + \dots + S_{1,2,\dots,k} \quad (۳-۱۱)$$

$$S_i = \frac{V_i}{V}, \quad S_{ij} = \frac{V_{ij}}{V}, \quad \dots, \quad S_{ij\dots k} = \frac{V_{ij\dots k}}{V} \quad \text{که}$$

می‌توان بر اساس مطالب فوق نتیجه گرفت که خروجی مدل با استفاده از $HDMR$ به توابعی یکتا از متغیرهای ورودی و برحسب امیدریاضی و واریانس شرطی تجزیه می‌شود. همچنین مجموع تمام شاخص‌های حساسیت مرتبه اول، دوم و مراتب بالاتر باید برابر یک باشد (رابطه ۳-۱۱).

^۱ Analysis of Variance- High Dimensional Model Representations

^۲ Archer

مثال ۲- تابع $Y = f(X_1, X_2, X_3) = \sin X_1 + a \sin^2 X_2 + bX_3^4 \sin X_1$ که موسوم به تابع هما و ایشگامی (۱۹۹۶) است را در نظر بگیرید. در این تابع X_i ها دارای توزیع یکنواخت در بازه $(-\pi, \pi)$ هستند. در این صورت برای تجزیه این تابع به فرم رابطه (۳-۶) داریم

$$f(X_1, X_2, X_3) = f_0 + f_1(X_1) + f_2(X_2) + f_3(X_3) + f_{12}(X_1X_2) + f_{13}(X_1X_3) + f_{23}(X_2X_3) + f_{123}(X_1X_2X_3)$$

که این جملات به صورت زیر محاسبه می‌شوند (همه انتگرالها بر بازه $(-\pi, \pi)$ گرفته شده اند).

$$\begin{aligned} f_0 = E(Y) &= \iiint f(x_1, x_2, x_3) P(x_1)P(x_2)P(x_3)dx_1dx_2dx_3 \\ &= \frac{1}{(2\pi)^3} \iiint (\sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1) dx_1dx_2dx_3 = \frac{a}{2} \end{aligned}$$

$$\begin{aligned} f_1 = E(Y | X_1) &= \iint f(x_1, x_2, x_3) P(x_2)P(x_3)dx_2dx_3 - f_0 \\ &= \frac{1}{(2\pi)^2} \iint (\sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1) dx_2dx_3 - f_0 = \left(1 + \frac{1}{5}\pi^4 b\right) \sin X_1 - \frac{a}{2} \end{aligned}$$

$$\begin{aligned} f_2 = E(Y | X_2) &= \iint f(x_1, x_2, x_3) P(x_1)P(x_3)dx_1dx_3 - f_0 \\ &= \frac{1}{(2\pi)^2} \iint (\sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1) dx_1dx_3 - f_0 = a \sin^2 X_2 - \frac{a}{2} \end{aligned}$$

$$f_3 = E(Y | X_3) = 0$$

$$\begin{aligned} f_{12} = E(Y | X_1X_2) &= \int f(x_1, x_2, x_3) P(x_3)dx_3 - f_0 - f_1 - f_2 \\ &= \sin X_1 + a \sin^2 X_2 + \int bX_3^4 \sin x_1 dx_3 - f_0 - f_1 - f_2 = \frac{a}{2} \end{aligned}$$

$$\begin{aligned} f_{13} = E(Y | X_1X_3) &= \int f(x_1, x_2, x_3) P(x_2)dx_2 - f_0 - f_1 - f_3 \\ &= \sin X_1 + \int a \sin^2 x_2 dx_2 + bX_3^4 \sin X_1 - f_0 - f_1 - f_3 \\ &= (bX_3^4 - \frac{1}{5}b\pi^4) \sin X_1 \end{aligned}$$

$$f_{123} = E(Y | X_1 X_2 X_3) = 0$$

$$f_{23} = E(Y | X_2 X_3) = 0$$

که می‌توان نتیجه گرفت

$$f(X_1, X_2, X_3) = \left(1 + \frac{1}{5} \pi^4 b\right) \sin X_1 + a \sin^2 X_2 - \frac{a}{2} + (bX_3^4 - \frac{1}{5} b \pi^4) \sin X_1$$

واریانس و شاخص حساسیت مرتبه اول عبارات فوق نیز برابر با

$$V(f_1) = \left(1 + \frac{1}{5} \pi^4 b\right)^2 \int \sin^2 X_1 dx_1 = \frac{1}{2} \times \left(1 + \frac{1}{5} \pi^4 b\right)^2 \Rightarrow S_1 = \frac{\frac{1}{2} \times \left(1 + \frac{1}{5} \pi^4 b\right)^2}{V(Y)}$$

$$V(f_2) = V(a \sin^2 X_2 - \frac{a}{2}) = a^2 \left(\int \sin^4 X_2 dx_2 - \left(\int \sin^2 X_2 dx_2 \right)^2 \right) = \frac{a^2}{8} \Rightarrow S_2 = \frac{a^2/8}{V(Y)}$$

$$V(f_3) = V(E(Y | X_3)) = 0 \Rightarrow S_3 = 0$$

$$V(f_{13}) = V(E(Y | X_1 X_3)) = V((bX_3^4 + \frac{1}{5} b \pi^4) \sin X_1) = \frac{b^2 \pi^4}{18} - \frac{b^2 \pi^8}{50} \Rightarrow S_{13} = \frac{\left(\frac{b^2 \pi^4}{18} - \frac{b^2 \pi^8}{50} \right)}{V(Y)}$$

$$V(f_{23}) = V(E(Y | X_2 X_3)) = 0 \Rightarrow S_{23} = 0$$

$$V(f_{123}) = V(E(Y | X_1 X_2 X_3)) = 0 \Rightarrow S_{123} = 0$$

$$V(Y) = \frac{1}{2} + \frac{a^2}{8} + \frac{\pi^8 b}{5} + \frac{\pi^8 b^2}{18}$$

مجموع تمام اثرات مرتبه اول و مراتب بالاتر برابر یک می‌شود. یعنی

$$S_1 + S_2 + S_3 + S_{12} + S_{23} + S_{13} + S_{123} = 1$$

مثال ۳- تابع $Y = \prod_{i=1}^k g_i(x_i)$ که به تابع سبل یا g -function معروف است را در نظر بگیرید که در آن

$$x_i \sim u(0,1) \text{ مستقل از یکدیگر و } 0 \leq x_i \leq 1, \quad g_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i} \quad \text{از این تابع در اغلب مراجع}$$

مرتبط با تحلیل حساسیت به عنوان تابعی که با آن بتوان نتایج نظری را آزمون نمود، استفاده می‌شود.

بسط سبل (رابطه ۳-۹) برای تابع فوق به صورت زیر است.

$$\begin{aligned} \int_0^1 g_i(x_i) dx_i &= \int_0^1 \frac{|4x_i - 2| + a_i}{1 + a_i} dx_i = \int_0^{1/2} \frac{(2 - 4x_i) + a_i}{1 + a_i} dx_i + \int_{1/2}^1 \frac{(4x_i - 2) + a_i}{1 + a_i} dx_i \\ &= \frac{2x - 2x^2}{1 + a_i} \bigg|_0^{1/2} + \frac{2x^2 - 2x}{1 + a_i} \bigg|_{1/2}^1 + a_i = 1 \end{aligned}$$

$$f_0 = E(Y) = \int Y dY = \prod_{i=1}^k \int_0^1 g_i(x_i) dx_i = 1 \times 1 \times \dots \times 1 = 1$$

$$f_i(x_i) = E(Y | x_i) - f_0 = g_i(x_i) \prod_{j=1, j \neq i}^k \int_0^1 g_j(x_j) dx_j - f_0 = g_i(x_i) - 1$$

$$\begin{aligned} f_{ij}(x_i, x_j) &= E(Y | x_i, x_j) - f_i(x_i) - f_j(x_j) - f_0 \\ &= g_i(x_i) g_j(x_j) \prod_{l=1, l \neq i, j}^k \int_0^1 g_l(x_l) dx_l - f_i(x_i) - f_j(x_j) - f_0 \\ &= g_i(x_i) g_j(x_j) - (g_i(x_i) - 1) - (g_j(x_j) - 1) - 1 \\ &= g_i(x_i) g_j(x_j) - g_i(x_i) - g_j(x_j) + 1 \end{aligned}$$

و به همین ترتیب عبارات مرتبه بالاتر $HDMR$ محاسبه می‌شوند. اکنون به بررسی شرط یکتایی سبل که

در رابطه (۳-۱۶) بیان شد می‌پردازیم.

$$\int f_i(x_i) dx_i = \int_0^1 (g_i(x_i) - 1) dx_i = \int_0^1 g_i(x_i) dx_i - 1 = 1 - 1 = 0$$

$$\begin{aligned}
\int \int f_{ij}(x_i, x_j) dx_i dx_j &= \int_0^1 \int_0^1 (g_i(x_i)g_j(x_j) - g_i(x_i) - g_j(x_j) + 1) dx_i dx_j \\
&= \int_0^1 \int_0^1 g_i(x_i)g_j(x_j) dx_i dx_j - \int_0^1 \int_0^1 g_i(x_i) dx_i dx_j - \int_0^1 \int_0^1 g_j(x_j) dx_j dx_i + 1 \\
&= \int_0^1 (g_i(x_i) dx_i) \int_0^1 g_j(x_j) dx_j - \int_0^1 g_i(x_i) dx_i - \int_0^1 g_j(x_j) dx_j + 1 = 1 \times 1 - 1 - 1 + 1 = 0
\end{aligned}$$

$$\begin{aligned}
\int \int f_i(x_i)f_j(x_j) dx_i dx_j &= \int_0^1 \int_0^1 (g_i(x_i) - 1)(g_j(x_j) - 1) dx_i dx_j = \int_0^1 \int_0^1 (g_i(x_i)g_j(x_j) - g_i(x_i) - g_j(x_j) + 1) dx_i dx_j \\
&= \int_0^1 (g_i(x_i) dx_i) \int_0^1 g_j(x_j) dx_j - \int_0^1 g_i(x_i) dx_i - \int_0^1 g_j(x_j) dx_j + 1 \\
&= 1 \times 1 - 1 - 1 + 1 = 0
\end{aligned}$$

بنابراین شرایط یکتایی برای اجزای بدست آمده تابع تحلیلی سبل برقرار است. شاخص‌های حساسیت مرتبه اول و کل نیز برای تابع سبل به صورت زیر محاسبه می‌شود.

$$\begin{aligned}
V(f_i(X_i)) &= V(g_i(x_i) - 1) = V(g_i(x_i)) = V\left(\frac{|4x_i - 2| + a_i}{1 + a_i}\right) = \frac{1}{(1 + a_i)^2} V(|4x_i - 2|) \\
&= \frac{1}{(1 + a_i)^2} (E(4x_i - 2)^2 - E^2(|4x_i - 2|))
\end{aligned}$$

$$E(4x_i - 2)^2 = E(16x_i^2 - 16x_i + 4) = 16E(x_i^2) - 16E(x_i) + 4 = \frac{16}{3} - \frac{16}{2} + 4$$

$$E(|4x_i - 2|) = E_{.5 \leq x_i \leq 1}(4x_i - 2) + E_{0 \leq x_i \leq .5}(2 - 4x_i) = 1$$

$$V_i(f_i(X_i)) = V_i(E(Y | X_i)) = \frac{1}{(1 + a_i)^2} \left(\frac{-16}{6} + 4 - 1\right) = \frac{1}{3(1 + a_i)^2}$$

$$E(Y) = \prod_{i=1}^k \int_0^1 g_i(x_i) dx_i = 1 \times 1 \times \dots \times 1 = 1$$

$$E(Y^2) = \prod_{i=1}^k \int_0^1 g_i^2(x_i) dx_i = \prod_{i=1}^k V(g_i(x_i)) - E(g_i(x_i)) = \prod_{i=1}^k (V_i - 1)$$

$$\begin{aligned}
 Var(Y) &= E(Y^2) - E^2(Y) = \prod_{i=1}^k (V_i - 1) - 1 \\
 S_i &= \frac{V(E(Y | X_i))}{Var(Y)} = \frac{1}{3(1 + a_i)^2} \times \frac{1}{\prod_{i=1}^k (V_i - 1) - 1} \\
 \Rightarrow S_{T_i} &= 1 - \frac{\sum_{j=1, j \neq i}^k V_j + \prod_{j=1, j \neq i}^k V_j}{V(Y)}
 \end{aligned}$$

۳-۷ ویژگی‌های شاخص‌های حساسیت

- مجموع اثرات مرتبه اول تمام متغیرهای ورودی هر مدل همواره کمتر یا مساوی با یک است. اگر $\sum_i S_i = 1$ در این صورت مدل مورد نظر، جمعی^۱ و اثرات متقابل بین متغیرها وجود ندارد. پس $\sum_i S_i < 1$ نشان از، وجود اثر متقابل بین متغیرهای ورودی و غیرجمعی بودن مدل است.
- عبارت $1 - \sum_i S_i$ ، میزان غیر جمعی بودن مدل را اندازه‌گیری می‌کند. در واقع این عبارت مبین کسری از واریانس است که با حضور تکی متغیرهای ورودی، قابل بیان نیست. این کسر می‌تواند شامل اثرات متقابل مرتبه دوم و مراتب بالاتر باشد.
- اثر متقابل بین متغیر X_i و سایر متغیرهای ورودی مدل با $S_{T_i} - S_i$ قابل بیان است. یعنی اگر $S_{T_i} - S_i \cong 0$ آن‌گاه متغیر X_i با سایر متغیرهای ورودی اثرمتقابل ندارد. اما $S_{T_i} - S_i \neq 0$ حاکی از وجود اثر متقابل متغیر X_i با سایر متغیرهای ورودی مدل است.

^۱ Additive

- شرط لازم و کافی برای آن که i امین متغیر ورودی مدل بی‌اهمیت قلمداد شود این است که $S_{T_i} \cong 0$. اگر $S_{T_i} \cong 0$ آن‌گاه می‌توان متغیر X_i را در هر مقداری ثابت نگه داشت بدون آن‌که تاثیر قابل ملاحظه‌ای بر واریانس خروجی مدل داشته باشد.
- همواره $\sum_i S_{T_i} \geq 1$ یعنی مجموع اثر کل تمام متغیرهای ورودی مدل همواره از یک بیشتر است. در حالت $\sum_i S_{T_i} = 1$ ، مدل مورد نظر کاملاً جمعی بوده و اثرات متقابل بین متغیرهای ورودی وجود ندارد.
- تفاضل $\sum_i S_{T_i} - 1$ میزان دیگری برای غیر جمعی بودن مدل است. این اندازه بیان می‌کند که در مدل غیر جمعی، وزن اثرات متقابل مراتب بالاتر، بیشتر از وزن اثرات متقابل مراتب پایین است.

$$\sum_{i=1}^3 S_{T_i} - 1 = S_{12} + S_{23} + S_{13} + 2.S_{123} \quad \text{چراکه}$$

۳ - ۸ اهداف بنیادی یا سبک‌های تحلیل حساسیت واریانس - مبنا

موارد بسیار زیادی در مراجع مرتبط با تحلیل حساسیت می‌توان یافت که در آن برای یک مسئله مشخص، روش‌ها و آزمون‌های متفاوتی مطرح شده است. از این رو لازم است تا با تعریف اهداف بنیادی و مفهوم "اهمیت"، روش‌های تحلیل حساسیت ساماندهی شده و با آگاهی بیشتری روش مورد نیاز را انتخاب نماییم.

۳ - ۸ - ۱ اولویت در متغیرها^۱

هدف تحلیل حساسیت در این سبک آن است که متغیرهای ورودی، بر اساس میزان تاثیر بر خروجی مدل، اولویت بندی و یا رتبه بندی شوند. اولویت بالا مربوط به فاکتوری است که اگر در مقدار معینی ثابت نگه داشته شود، باعث بیشترین کاهش در واریانس خروجی گردد. به همین ترتیب اولویت‌های پایین‌تر نیز تعریف می‌شوند. در مورد این سبک می‌بایست به این نکته توجه کرد که فرض آن است که هر یک از متغیرهای ورودی مدل به صورت تکی (انفرادی) در مقدار معینی از دامنه تغییراتش ثابت نگه داشته شود که این عمل مانع شناسایی اثرات متقابل بین متغیرها است (سالتلی و همکاران، ۲۰۰۴). شاخص حساسیت مرتبه اول (S_j) مناسب‌ترین اندازه برای تعیین اولویت‌ها شناخته شده است.

۳ - ۸ - ۲ شناسایی فاکتورهای بی اهمیت

یکی دیگر از اهداف بنیادی آن است که با شناسایی و حذف فاکتورهای بی اهمیت، مدل‌های پیچیده به صورت ساده‌تری بیان شوند. این موضوع تحت عنوان تثبیت عوامل^۲، در مراجع برجسته شده که هدف مورد نظر در آن، شناسایی زیر مجموعه‌ای از متغیرهای ورودی مدل است که می‌توان آن‌ها را در هر مقداری ثابت نگه داشت بدون آنکه کاهش معنی‌داری در واریانس خروجی مدل رخ دهد. به عبارت دیگر، هدف دنبال شده، در راستای شناسایی مجموعه متغیرهای بی‌تاثیر در خروجی مدل است. با توجه به تعریف اهمیت، بدیهی است که شرط لازم و کافی برای بی‌تاثیر بودن متغیر X_i آن است که $S_{T_i} = 0$. بنابراین بهترین اندازه حساسیت برای برآورده شدن این هدف، شاخص حساسیت کل می‌باشد. از این رو می‌توان با شناسایی مجموعه عوامل بی‌تاثیر و کاهش بعد فضای متغیرهای ورودی، مدل را ساده‌تر نمود.

^۱ Factors Prioritisation (FP) Setting^۲ Factors Fixing (FF)

۳ - ۸ - ۳ برش واریانس^۱

هدف مورد نظر در این سبک، کاهش واریانس خروجی مدل به یک مقدار آستانه‌ای^۲ است. به عبارت دیگر چگونه می‌توان کوچکترین زیر مجموعه از متغیرهای ورودی مانند $u=(X_{i1}, \dots, X_{im})$ را یافت به‌طوری که با ثابت نگه داشتن مقادیر این مجموعه، واریانس خروجی مدل به مقدار V_r کاهش یابد. این موضوع در تحلیل ریسک و قابلیت اعتماد یک مطلوبیت است زیرا به‌عنوان مثال تحلیل‌گران مایلند تا با کنترل نمودن کمترین تعداد ورودی‌های سیستم، عدم قطعیت قابلیت اعتماد اجزاء آن سیستم، کمتر از یک مقدار مشخص باشد.

۳ - ۸ - ۴ تنظیمات تصویر متغیرها^۳

در این سبک از تحلیل حساسیت، نگاشتی از خروجی مدل نسبت به ورودی‌های مدل را مورد نظر قرار می‌دهیم. به عبارت دیگر سوال اصلی این است که به ازای چه ترکیبی از مقادیر متغیرهای ورودی، مقادیر Y در بازه مشخصی واقع می‌شوند. به عنوان مثال ممکن است تحلیلگر سیستم نیاز به مقادیر ۹۵٪ بالایی خروجی مدل در شرایط پرمخاطره در یک کارخانه داشته باشد.

^۱ Variance Cutting (VC) Setting

^۲ Threshold Value

^۳ Factors Mapping (FM) Setting

فصل چهارم

برآورد شاخص‌های حساسیت با روش

مونت کارلو - مبنای سالتلی

مقدمه

در فصل قبل، شاخص حساسیت مرتبه اول و شاخص حساسیت کل با تکیه بر مفاهیم آماری، معرفی شدند. به‌طور کلی این شاخص‌ها بر اساس امیدریاضی شرطی و واریانس شرطی خروجی مدل بیان می‌شوند که خود، انتگرال‌های چندگانه هستند. محاسبه دقیق این انتگرال‌ها بنابر دلایلی همچون بزرگ‌بودن بعد فضای انتگرال‌گیری و نیز صریح نبودن رابطه ریاضی ورودی- خروجی، امکان‌پذیر نیست. لذا موضوع تقریب انتگرال فوق و یا مسئله برآورد شاخص‌های حساسیت مطرح می‌شود. برای این منظور روش‌های گوناگونی پیشنهاد شده است که در این فصل یکی از این روش‌ها موسوم به، روش مونت کارلو- مبنای سالتلی مورد بحث قرار گرفته است. مطالبی که در خصوص این روش بیان می‌شود در قالب بخش‌های مختلف و با موضوعات گوناگونی است که شرح آن بدین قرار است. در بخش‌های ۱-۴ و ۳-۴ شاخص‌های حساسیت توسط انتگرال‌های چندگانه بیان شده‌اند. در بخش‌های ۲-۴ و ۴-۴، موضوع تقریب انتگرال برای برآورد شاخص‌های حساسیت با روش مونت کارلو به تفصیل مورد بحث قرار گرفته است. خلاصه‌ای از بخش‌های اول تا چهارم با یک دستور کار عملی و ساده، در بخش ۵-۴ ذکر شده است. موضوع عدم قطعیت برآوردگرها تحت عنوان فاصله‌های اطمینان بوت استریپی را در بخش ۶-۴ به طور مختصر ارائه نموده و سرانجام در بخش ۷-۴ به بیان اعداد شبه تصادفی که یکی از روش‌های شبیه‌سازی این پایان‌نامه را تشکیل می‌دهد پرداخته شده است.

۴ - ۱ بنیان ریاضی شاخص حساسیت مرتبه اول

همانند قبل فرض کنید $\underline{X} = (X_1, \dots, X_k)$ متغیرهای ورودی مدل و $Y = f(X_1, \dots, X_k)$ مبین

خروجی مدل تعیینی باشد. اگر X_i ها مستقل از هم بوده و هر یک از آنها دارای توزیع احتمال $P_i(x_i)$

باشد در این صورت میانگین و واریانس Y به صورت زیر محاسبه می‌شود.

$$E(Y) = \int_{\Omega^k} Y P(\underline{x}) d\underline{x} = \int \int \dots \int f(x_1, x_2, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i \quad (1-4)$$

$$Var(Y) = E(Y^2) - E^2(Y) = \int \int \dots \int f^2(x_1, x_2, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i - E^2(Y)$$

برای محاسبه صورت کسر $S_j = \frac{V(E(Y | X_j))}{Var(Y)}$ اگر j -امین متغیر ورودی را در مقدار x_j ثابت نگاه داریم،

در این صورت میانگین و واریانس شرطی Y به صورت زیر تعریف می‌شوند.

$$\begin{aligned} E(Y | X_j = x_j) &= \int \int \dots \int f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \\ Var(Y | X_j = x_j) &= E(Y^2 | X_j = x_j) - E^2(Y | X_j = x_j) \\ &= \int \int \dots \int f^2(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i - E^2(Y | X_j = x_j) \end{aligned} \quad (2-4)$$

از آنجا که $V(Y | X_j = x_j)$ به مقدار متغیر X_j وابسته است، لذا با میانگین‌گیری بر روی مقادیر X_j داریم

$$\begin{aligned} E(V(Y | X_j)) &= \int V(Y | X_j = x_j) P_j(x_j) dx_j = \int (E(Y^2 | X_j = x_j) - E^2(Y | X_j = x_j)) P_j(x_j) dx_j \\ &= \int \left\{ \int \int \dots \int f^2(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i - E^2(Y | X_j = x_j) \right\} P_j(x_j) dx_j \quad (3-4) \\ &= \int \int \int \dots \int f^2(x_1, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i - \int E^2(Y | X_j = c) P_j(x_j) dx_j \end{aligned}$$

در نتیجه صورت کسر S_j از تفاضل رابطه (۲-۴) و (۳-۴) و بنابر ویژگی واریانس مطابق زیر بدست می‌آید.

$$V(E(Y | X_j)) = V(Y) - E(V(Y | X_j)) \quad (4-4)$$

$$\begin{aligned} V(E(Y | X_j)) &= \int \int \dots \int f^2(x_1, x_2, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i - E^2(Y) - \\ &\quad \int \int \int \dots \int f^2(x_1, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i + \int E^2(Y | X_j = x_j) P_j(x_j) dx_j \\ &= \int E^2(Y | X_j = x_j) P_j(x_j) dx_j - E^2(Y) = U_j - E^2(Y) \end{aligned}$$

$$U_j = \int E^2(Y | X_j = x_j) P_j(x_j) dx_j = \int \left(\int \int \dots \int f(X_1, \dots, X_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \right)^2 P_j(x_j) dx_j$$

اکنون می‌توان شاخص حساسیت مرتبه اول را به صورت زیر باز نویسی نمود.

$$S_j = \frac{Var(E(Y | X_j))}{Var(Y)} = \frac{U_j - E^2(Y)}{Var(Y)}$$

هما و ایشیگامی در سال ۱۹۹۶ برای سهولت محاسبه U_j پیشنهاد کردند که مربع یک انتگرال، به حاصل ضرب دو انتگرال با آرگومان‌های متفاوت، تجزیه شود. یعنی

$$\int E^2(Y | X_j = x_j) P_j(x_j) dx_j \quad (5-4)$$

$$\begin{aligned} &= \int \left(\int \int \dots \int f(x_1, \dots, x_j, \dots, x_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \right) \times \left(\int \int \dots \int f(x'_1, \dots, x_j, \dots, x'_k) \prod_{i=1, i \neq j}^k P_i(x'_i) dx'_i \right) P_j(x_j) dx_j \\ &= \int \int \dots \int f(x_1, \dots, x_j, \dots, x_k) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \times f(x'_1, \dots, x_j, \dots, x'_k) \prod_{i=1, i \neq j}^k P_i(x'_i) P_j(x_j) dx_j \\ &= \int \int \dots \int f(x_1, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i \times f(x'_1, \dots, x_j, \dots, x'_k) \prod_{i=1, i \neq j}^k P_i(x'_i) dx'_i \end{aligned}$$

اما در محاسبه این انتگرال دو مشکل اساسی وجود دارد.

۱. بعد فضای انتگرال‌گیری بسیار بالا و برابر با $2k-1$ است.

۲. رابطه ریاضی ورودی- خروجی (f) به طور صریح معلوم نیست.

این چالش‌ها سبب طرح مسئله تقریب انتگرال‌ها و در نتیجه برآورد شاخص‌های حساسیت می‌شود.

۴ - ۲ برآورد شاخص حساسیت مرتبه اول

$$\hat{S}_j = \frac{\hat{U}_j - \hat{E}^2(Y)}{\hat{Var}(Y)}$$

برای برآورد اثر اصلی S_j بایستی U_j ، $E(Y)$ و $V(Y)$ برآورد شوند، یعنی

الف- برآورد U_j

ساختار انتگرال U_j در رابطه (۴-۵) یعنی

$$U_j = \int \int \dots \int f(x_1, \dots, x_j, \dots, x_k) f(x'_1, \dots, x_j, \dots, x'_k) \prod_{i=1}^k P_i(x_i) dx_i \prod_{i=1, i \neq j}^k P_i(x'_i) dx'_i$$

به گونه‌ایست که برای تقریب انتگرال با روش مونت کارلو، این ایده به وجود می‌آید که دو ماتریس مستقل A و B با N سطر توسط اعداد تصادفی تولید شوند که هر سطر این دو ماتریس معرف یک بردار k تایی از مقادیر شبیه‌سازی شده برای متغیرهای $\underline{X} = (X_1, \dots, X_k)$ باشد.

$$A = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{N1} & \dots & \dots & x_{Nk} \end{bmatrix} \quad B = \begin{bmatrix} x'_{11} & x'_{12} & \dots & x'_{1k} \\ x'_{21} & x'_{22} & \dots & x'_{2k} \\ \dots & \dots & \dots & \dots \\ x'_{N1} & \dots & \dots & x'_{Nk} \end{bmatrix}$$

از آنجایی که در عبارت $f(x_1, \dots, x_k) f(x'_1, \dots, x'_k)$ مقدار متغیر j -ام بردار $(x'_1, \dots, x_j, \dots, x'_k)$ از بردار $(x_1, \dots, x_j, \dots, x_k)$ انتخاب شده است، لذا ضروری است تا به ازای هر یک از X_j ها ماتریس جدیدی بر مبنای ماتریس B تولید شود که فقط ستون j -ام آن برگرفته از ستون j -ام ماتریس A و بقیه ستون‌ها همان ستون‌های B باشند. بدین منظور ماتریس‌های جدید C_1, \dots, C_k را طوری می‌سازیم که تفاوت ماتریس C_j و B فقط در ستون j -ام باشد. یعنی

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

بنابراین، با استفاده از روش انتگرال‌گیری مونت کارلو، برآورد U_j به صورت زیر است (سالتلی، ۲۰۰۲).

$$\hat{U}_j = \frac{1}{N-1} \sum_{r=1}^N f(x_{r1}, \dots, x_{rk}) \times f(x'_{r1}, \dots, x'_{r(j-1)}, x_{rj}, x'_{r(j+1)}, \dots, x'_{rk}) = \frac{1}{N-1} \sum_{r=1}^N f(A)_r \times f(C)_r \quad (4-6)$$

لازم به ذکر است که در برخی از روش‌های تولید اعداد تصادفی، ماتریس‌های A و B به صورت مستقل از-هم تولید نمی‌شوند. این موضوع در بخش ۴-۷ مورد بحث واقع شده است.

ب- برآورد $E^2(Y)$

به منظور برآورد $E^2(Y)$ در رابطه \hat{S}_j بنابر پیشنهاد سبل (۱۹۹۳) و یا سالتلی (۲۰۰۲) می‌توان از میانگین ماتریس A یا B و یا تلفیقی از این دو ماتریس استفاده کرد.

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

ج- برآورد $V(Y)$

برای برآورد $V(Y)$ می‌توان از ماتریس A به صورت $\hat{V}(Y) = (1/N) \sum_{j=1}^N (y_A^{(j)})^2 - \hat{E}(Y)^2$ استفاده کرد.

هزینه محاسبات برآورد اثر اصلی

ماتریس‌های A, B, C_1, \dots, C_k به کار رفته در برآورد اجزای S_j (بندهای الف-ج) هر یک دارای N سطر می‌باشند که هر سطر به عنوان یک ورودی به مدل تعیینی شناخته می‌شود. به منظور برآورد اثر اصلی باید مدل تعیینی به ازای هر ورودی اجرا شود، در نتیجه تعداد کل دفعات اجرا یا هزینه محاسبه برآورد اثر اصلی تمام متغیرهای ورودی برابر با $kN + 2N = N(k+2)$ است. تعیین مقدار N در کلیه موضوع‌های مرتبط با تحلیل حساسیت از اهمیت ویژه‌ای برخوردار است. زیرا اگر در مدل‌های تعیینی زمان اجرا برای بدست آوردن هر خروجی را برابر t فرض کنیم آن‌گاه زمان کل مورد نیاز برای برآورد اثرات اصلی برابر

$tN(k+2)$ خواهد بود که با توجه به ثابت بودن مقادیر t و k ، نقش حیاتی مقدار N بیش از پیش پدیدار می‌شود.

۴ - ۳ بنیان ریاضی شاخص حساسیت کل

نحوه محاسبه $V(E(Y | X_{-j}))$ در شاخص حساسیت کل، $S_{T_j} = 1 - \frac{Var(E(Y | X_{-j}))}{Var(Y)}$ به شرح زیر است.

در این محاسبات، نماد X_j به معنای تمام متغیرها به جز متغیر j -ام می‌باشد.

$$\begin{aligned}
 E(Y | X_{-j} = x_{-j}) &= \int f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) P_j(x_j) dx_j \\
 Var(Y | X_{-j} = x_{-j}) &= E(Y^2 | X_{-j} = x_{-j}) - E^2(Y | X_{-j} = x_{-j}) \\
 &= \int f^2(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) P_j(x_j) dx_j - E^2(Y | X_{-j} = x_{-j}) \\
 E(V(Y | X_{-j})) &= \int \int \dots \int (E(Y^2 | X_{-j} = x_{-j}) - E^2(Y | X_{-j} = x_{-j})) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \\
 &= \int \dots \left(\int f^2(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) P_j(x_j) dx_j - E^2(Y | X_{-j} = x_{-j}) \right) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \\
 &= \int \dots \int f^2(x_1, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i - \int \dots \int E^2(Y | X_{-j} = x_{-j}) \prod_{i=1, i \neq j}^k P_i(x_i) dx_i \\
 V(E(Y | X_{-j})) &= E(E(Y | X_{-j}))^2 - (E(E(Y | X_{-j})))^2 = E(E(Y | X_{-j}))^2 - E^2(Y) \\
 &= \int \dots \left(\int f(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) P_j(x_j) dx_j \right)^2 \prod_{i=1, i \neq j}^k P_i(x_i) dx_i - E^2(Y) \\
 &= \int \dots \int f(x_1, \dots, x_j, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i \times f(x_1, \dots, x'_j, \dots, x_k) P(x'_i) dx'_i - E^2(Y)
 \end{aligned}$$

با در نظر گرفتن جمله اول عبارت فوق با نماد U_j ، شاخص حساسیت کل به صورت زیر بازنویسی می‌شود.

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

۴ - ۴ برآورد شاخص حساسیت کل

از آنجا که برآوردگرهای $E^2(Y)$ و $V(Y)$ در بندهای "ب" و "ج" بخش ۴-۲ معرفی شدند، بنابراین جهت برآورد شاخص حساسیت کل کافی است که یک راهکار محاسباتی برای عبارت

$$U_{-j} = \int \dots \int f(x_1, \dots, x_j, \dots, x_k) \times f(x_1, \dots, x'_j, \dots, x_k) \prod_{i=1}^k P_i(x_i) dx_i P_j(x'_j) dx'_j$$

پیشنهاد گردد. با توجه به ماتریس‌های A و B مندرج در بخش ۴-۲، در خصوص عبارت $f(x_1, \dots, x'_j, \dots, x_k)$ لازم است تا به ازای هر یک از X_j ها ماتریس جدیدی بر مبنای ماتریس A تولید شود که فقط ستون j -ام آن برگرفته از ستون j ام ماتریس B باشد. بدین منظور ماتریس‌های جدید D_1, \dots, D_k را طوری می‌سازیم که ماتریس D_j ، همان ماتریس A باشد با این تفاوت که فقط ستون j -ام آن از ستون j -ام ماتریس B گرفته شده است، یعنی

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

بنابر این با استفاده از روش انتگرال‌گیری مونت کارلو U_{-j} به‌صورت زیر برآورد می‌شود (سالتلی، ۲۰۰۲).

$$\hat{U}_{-j} = \frac{1}{N-1} \sum_{r=1}^N f(x_{r1}, x_{r2}, \dots, x_{rk}) \times f(x_{r1}, x_{r2}, \dots, x_{r(j-1)}, x'_{rj}, x_{r(j+1)}, \dots, x_{rk}) = \frac{1}{N-1} \sum_{r=1}^N f(A)_r \times f(D)_{rj} \quad (۷-۴)$$

که در آن N حجم نمونه یا تعداد اعداد شبیه‌سازی شده برای هر متغیر است. از آنجا که $f(D)_{rj}$ در عبارت (۷-۴) معادل N بار اجرای مدل است، هزینه محاسبه هر U_{-j} برابر N و در نهایت، هزینه محاسبه

همه شاخص‌های کل برابر با $N.k$ است. لازم به ذکر است که هزینه محاسباتی $f(A)_r$ منظور نشده است زیرا قبلاً در محاسبه هزینه شاخص مرتبه اول، به حساب آمده است.

رهیافت کاهش هزینه محاسبات

بنابر بر برآوردهای ارائه شده در دو بخش قبل دیدیم که مجموع هزینه محاسبات شاخص‌های حساسیت یا تعداد دفعات اجرای مدل برابر است با $N(2k+2)$ که سهم شاخص‌های مرتبه اول، $Nk+2N$ (برای ماتریس‌های A, B, C_1, \dots, C_k) و سهم شاخص‌های کل، $N.k$ (برای ماتریس‌های D_1, \dots, D_k) است. سالتلی ایده دیگری را برای کاهش هزینه محاسباتی شاخص کل ارائه داد (سالتلی، ۲۰۰۲). وی پیشنهاد کرد که با تعویض نقش ماتریس‌های A و B در برآورد اثر کل، به جای ماتریس‌های D_1, \dots, D_k از ماتریس‌های C_1, \dots, C_k استفاده شود که در این صورت هزینه کل محاسبات به $N(k+2)$ کاهش خواهد یافت.

۴ - ۵ دستور کار برای برآورد شاخص مرتبه اول و شاخص کل

در زیر، خلاصه‌ای از آنچه تاکنون برای برآورد شاخص‌های حساسیت، مطالعه گردیده است را بیان می‌کنیم.

۱. ماتریسی به ابعاد $(N, 2k)$ از اعداد تصادفی تولید کرده و دو ماتریس A و B را که هر کدام شامل نیمی از نمونه‌های تولید شده هستند در نظر می‌گیریم. در اینجا، k تعداد متغیرهای ورودی مدل و N حجم نمونه تصادفی است که می‌تواند از ۱۰۰ تا چندین هزار تغییر کند. سبل پیشنهاد نمود

که از اعداد شبه تصادفی^۱ برای شبیه‌سازی استفاده شود. این گونه اعداد، خواص منحصر به فردی دارند که در بخش ۴-۷ به شرح آن خواهیم پرداخت،

$$A = \begin{bmatrix} x_1 & x_2 & \dots & x_i & \dots & x_k \\ x_1 & x_2 & \dots & x_2 & \dots & x_k \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_1 & x_2 & \dots & x_i & \dots & x_k \end{bmatrix} \quad B = \begin{bmatrix} x_{k+1} & x_{k+2} & \dots & x_{k+i} & \dots & x_{2k} \\ x_{k+1} & x_2 & \dots & x_{k+i} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{k+1} & x_2 & \dots & x_{k+i} & \dots & x_{2k} \end{bmatrix}$$

ماتریس‌های جدید C_i را با ترکیب دو ماتریس A و B می‌سازیم. ماتریس C_i ، همان ماتریس B است که ستون i ام آن که برگرفته از ستون i -ام ماتریس A است.

۲. خروجی مدل را برای تمام سطرهای ماتریس‌های A ، B ، C_1 ، ... و C_k محاسبه می‌کنیم تا

بردارهای خروجی N تایی $y_A = f(A)$ ، $y_B = f(B)$ ، $y_{C_i} = f(C_i)$ $i = 1, \dots, k$ بدست آیند.

۳. برآورد شاخص حساسیت مرتبه اول و شاخص حساسیت کل برای هر یک از متغیرهای مدل به

صورت روابط (۴-۸) و (۴-۹) بیان می‌شوند.

$$\hat{S}_i = \frac{\hat{Var}(E(Y | X_i))}{\hat{Var}(Y)} = \frac{(1/N)y_A \cdot y_{C_i} - f_0^2}{(1/N)y_A \cdot y'_A - f_0^2} = \frac{(1/N) \sum_{j=1}^N y_A^{(j)} \cdot y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2} \quad (۴-۸)$$

$$\hat{S}_{T_i} = 1 - \frac{\hat{Var}(E(Y | X_{-i}))}{\hat{Var}(Y)} = 1 - \frac{(1/N)y_B \cdot y_{C_i} - f_0^2}{(1/N)y_A \cdot y'_A - f_0^2} = 1 - \frac{(1/N) \sum_{j=1}^N y_B^{(j)} \cdot y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2} \quad (۴-۹)$$

$$\text{که } f_0^2 = \hat{E}^2(Y) = \left(1/N \sum_{j=1}^N y_A^{(j)} \right)^2 \text{ است.}$$

^۱ Quasi-random numbers

علاوه بر روش مذکور، روش‌های دیگری نیز برای برآورد شاخص‌های حساسیت پیشنهاد شده است. در جدول (۱-۴) فرمول محاسباتی و هزینه محاسباتی آن‌ها با یکدیگر مقایسه شده‌اند. در این پایان‌نامه از روش پیشنهادی سالتلی برای برآورد شاخص حساسیت کل و از روش پیشنهادی سبل برای برآورد شاخص حساسیت مرتبه اول استفاده شده است.

جدول (۱ - ۴) - روش‌های برآورد اثر اصلی و اثر کل

جهت برآورد عبارت U در اثر اصلی		جهت برآورد عبارت U در اثر کل	
$V_{X_i}(E_{X_{\sim i}}(Y X_i = c))$	هزینه محاسبات	$E_{X_{\sim i}}(V_{X_i}(Y X_{\sim i} = c))$	هزینه محاسبات
$\frac{1}{N} \sum_{j=1}^N f(A)_j f(C)_j - f_0^2$ سبل (۱۹۹۳)	$k.N + N$	$V(Y) - \frac{1}{2N} \sum_{j=1}^N f(A)_j \times f(D)_j + f_0^2$ هما (۱۹۹۶)	$k.N + N$
$\frac{1}{N} \sum_{j=1}^N f(B)_j (f(D)_j - f(A)_j)$ سالتلی (۲۰۱۰)	$k.N + 2.N$	$\frac{1}{N} \sum_{j=1}^N f(A)_j (f(A)_j - f(D)_j)$ سبل (۲۰۰۷)	$k.N + N$
$V(Y) - \frac{1}{2N} \sum_{j=1}^N (f(B)_j - f(C)_j)^2$ جانسن (۱۹۹۹)	$k.N + N$	$\frac{1}{2N} \sum_{j=1}^N (f(A)_j \times f(D)_j)^2$ جانسن ^۱ (۱۹۹۹)	$k.N + N$
		$\frac{1}{N} \sum_{j=1}^N f(B)_j f(C)_j - f_0^2$ سالتلی (۲۰۰۲)	$k.N + N$

^۱ Jansen

۴ - ۶ فاصله اطمینان بوت‌استرپ^۱ برای شاخص‌های حساسیت

پس از برآورد شاخص حساسیت این سوال مطرح می‌شود که میزان اطمینان به برآوردهای شاخص‌های حساسیت چقدر است؟ برای این منظور از فاصله اطمینان $100(1-\alpha)\%$ مقادیر برآوردی استفاده می‌شود که α ضریب اطمینان و معمولاً بین 0.01 تا 0.05 است. در واقع میزان دقت برآورد با استفاده از این فاصله اطمینان سنجیده می‌شود. این فاصله اطمینان بدین معنا است که اگر M بار فرآیند برآورد شاخص‌های حساسیت را با N نقطه آزمایش تصادفی (مختلف) تکرار کنیم. آن‌گاه فاصله اطمینان حاصل، $100(1-\alpha)\%$ دفعه شامل مقدار واقعی این شاخص‌ها خواهد بود. یعنی اگر حجم نمونه را برابر با N و ضریب اطمینان را 0.05 در نظر بگیریم و فرآیند برآورد شاخص‌های حساسیت را 100 بار تکرار کنیم، در این صورت 95 دفعه، فاصله اطمینان حاصل دربرگیرنده مقدار واقعی این شاخص‌ها است. می‌توان این فاصله اطمینان را با استفاده از روش بوت‌استرپ محاسبه کرد. اساس این روش به شرح زیر است. لازم به یادآوری است که مقصود از پارامتر مجهول θ همان شاخص‌های حساسیت و $\hat{\theta}$ برآورد این شاخص‌ها است.

اگر $Y=(Y_1, \dots, Y_N)$ متغیر تصادفی با تابع توزیع تجمعی F و پارامتر مجهول θ باشد. آن‌گاه می‌توان مقدار پارامتر θ را با N بار اجرای مدل برآورد کرد که آن را با نماد $\hat{\theta}$ نشان می‌دهیم. نحوه محاسبه فاصله اطمینان $100(1-\alpha)\%$ برای θ بدین شرح است. اگر توزیع $\hat{\theta} - \theta$ را بدانیم و مقادیر $\varepsilon_{\frac{\alpha}{2}}$ و $\varepsilon_{1-\frac{\alpha}{2}}$ را طوری بیابیم که $\Pr(\varepsilon_{\frac{\alpha}{2}} \leq \hat{\theta} - \theta \leq \varepsilon_{1-\frac{\alpha}{2}}) = 1 - \alpha$ باشد، در این صورت بازه $(\hat{\theta} - \varepsilon_{1-\frac{\alpha}{2}}, \hat{\theta} - \varepsilon_{\frac{\alpha}{2}})$ همان فاصله اطمینان مورد نظر است. اما در عمل توزیع $\hat{\theta} - \theta$ ناشناخته است که می‌توان این توزیع را با

^۱ Bootstrap

استفاده از روش بوت‌استرپ تخمین نمود. ایده اصلی این روش استفاده از توزیع تجربی F^* به جای F است. F^* را می‌توان از راه‌های مختلفی به دست آورد. یک راه تولید F^* تقلید کردن فرآیند تولید داده‌های بردار Y است. یعنی نمونه‌ای به حجم N از تابع F^* تولید و آن را با $Y^* = (Y_1^*, \dots, Y_N^*)$ نشان می‌دهیم. Y^* تولید شده از تابع توزیع تجمعی برآورد شده F^* را نمونه بوت‌استرپی می‌نامند. معروف‌ترین روش نمونه‌گیری بوت‌استرپ، نمونه‌گیری با جایگذاری از داده‌های Y است. با برآورد پارامتر مجهول $\hat{\theta}^*$ از نمونه Y^* ، فاصله اطمینان برای θ به صورت $(\hat{\theta}^* - \varepsilon_{\frac{1-\alpha}{2}}, \hat{\theta}^* + \varepsilon_{\frac{\alpha}{2}})$ خواهد بود که $\varepsilon_{\frac{\alpha}{2}}$ و $\varepsilon_{1-\frac{\alpha}{2}}$ با استفاده از چندک‌های توزیع نرمال قابل محاسبه است. با تکرار این فرآیند به دفعات کافی (N) و استفاده از قضیه حد مرکزی،

فاصله اطمینان برای θ به صورت $(\hat{\theta}^* - \frac{\sqrt{V(\hat{\theta}^*)}}{N} \cdot z_{1-\frac{\alpha}{2}}, \hat{\theta}^* + \frac{\sqrt{V(\hat{\theta}^*)}}{N} \cdot z_{\frac{\alpha}{2}})$ خواهد بود. در این

معادله $\hat{\theta}^* = \frac{\sum_{i=1}^N \hat{\theta}_i^*}{N}$ و عملگر $V(.)$ واریانس مقادیر $\hat{\theta}^*$ است.

بنابر مطالب فوق و با در نظر گرفتن شاخص‌های حساسیت به عنوان پارامتر مجهول θ و ویژگی

اعداد شبه تصادفی، فاصله اطمینان بوت‌استرپی ۹۵٪ برای شاخص‌های حساسیت مرتبه اول و شاخص کل

متغیر X_i عبارت است از

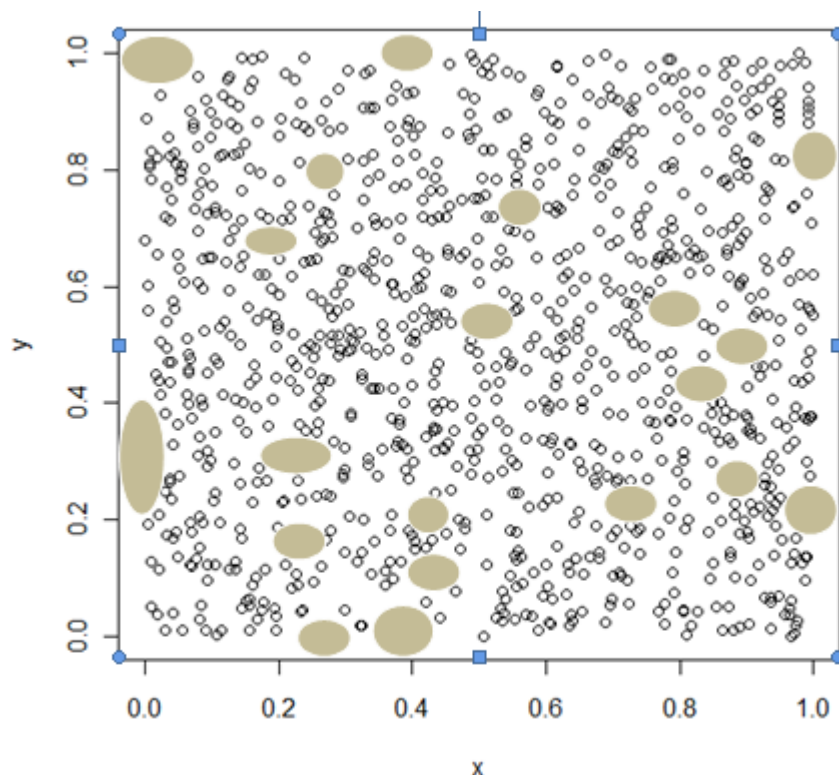
$$\begin{aligned} \hat{S}_i \pm 1.96 \times \sqrt{\frac{1}{B-1} \sum_{b=1}^B (S_i^b - S_i^{*b})^2} \\ \hat{S}_{\pi} \pm 1.96 \times \sqrt{\frac{1}{B-1} \sum_{b=1}^B (S_{\pi}^b - S_{\pi}^{*b})^2} \end{aligned} \quad (10-4)$$

که در آن \hat{S}_i و \hat{S}_{T_i} برآورد شاخص حساسیت مرتبه اول و کل در نمونه شبیه سازی شده اولیه، B تعداد نمونه‌های باجایگذاری، S_i^b و $S_{T_i}^b$ برآورد شاخص حساسیت اول و کل در نمونه با جایگذاری b -ام،

$$S_{T_i}^{*b} = \frac{1}{B-1} \sum_{b=1}^B S_{T_i}^b \quad S_i^{*b} = \frac{1}{B-1} \sum_{b=1}^B S_i^b \quad \text{است (آرچر و همکاران، ۱۹۹۷).}$$

۷-۴ اعداد شبه تصادفی

فرض کنید بخواهیم تقریب انتگرال $\iint f(x,y) dx dy$ را در ناحیه مشخصی از صفحه xy با استفاده از روش شبیه‌سازی مونت کارلو به دست آوریم. بدون از دست دادن کلیت مطلب می‌توان فرض کرد که ناحیه مورد نظر، مربعی به ضلع ۱ باشد. بدین منظور ابتدا می‌بایست برای هر دو متغیر x, y نمونه‌هایی مستقل به حجم n از توزیع یکنواخت (۰ و ۱) تولید شود و مقدار تابع f در هر نقطه شبیه‌سازی شده (x,y) محاسبه گردد. سرانجام تقریب انتگرال مذکور با میانگین‌گیری از مقادیر مختلف $f(x,y)$ یعنی $\iint f(x,y) dx dy \cong \frac{1}{1000} \sum_{i=1}^{1000} f(x_i, y_i)$ به دست می‌آید. شکل (۱-۴) نمودار پراکنش نقاط تولید شده را با $n=1000$ نشان می‌دهد. دایره‌ها و بیضی‌های رسم شده در شکل (۱-۴) نمونه‌ای از فضاهایی است که اعداد تصادفی در آنجا تولید نشده‌اند. بنابراین با از دست دادن اطلاعات رفتاری تابع در این نواحی، برآورد انتگرال با خطا همراه بود. با افزایش تعداد متغیرهای ورودی تابع و تولید اعداد تصادفی در فضاهای با بعد بالا، این نواحی خالی تبدیل به ابر فضا شده و در نتیجه برآورد انتگرال‌های چند-بعدی با خطای قابل ملاحظه‌ای همراه خواهد شد.



شکل (۴ - ۱) - نمودار پراکنش نقاط تصادفی تولید شده از توزیع یکنواخت برای X , Y

با عنایت به این‌که توابع مورد نظر در این پایان‌نامه، مدل‌های کامپیوتری هستند که زمان اجرای آن‌ها از مسایل بحرانی بوده و به‌عنوان هزینه تلقی می‌شود، افزایش تعداد نقاط شبیه‌سازی شده در صفحه یا ابر فضا نمی‌تواند راه حل درستی برای کاهش خطای تقریب باشد. بنابراین ضروری به نظر می‌رسد که الگوریتم تولید اعداد تصادفی باید به گونه‌ای طراحی شود که بتواند با تولید کمترین نقاط، بیشترین اطلاعات رفتاری تابع را در دسترس قرار داده و خطای تقریب را به کمترین میزان خود کاهش دهد. از آنجا که شاخص‌های حساسیت نیز از نوع انتگرال‌های چندگانه هستند، انتخاب مناسب نقاط شبیه‌سازی شده موجب اربیی کمتر برآوردگرهای پیشنهادی شده و مانع از بروز خطای نوع اول و دوم می‌شود. در صورت رخداد هریک از این خطاها، تحلیل حساسیت و مسائل مرتبط با آن همانند مدیریت ریسک، کنترل متغیرهای تاثیرگذار و حذف ورودی‌های کم تاثیر و بی‌تاثیر در مدل، با چالش جدی مواجه خواهد

شد. لذا مسئله طرح نقاط آزمایش در برآورد شاخص‌های حساسیت مدل‌های تعیینی که با کثرت متغیرهای ورودی و ابرفضاها سروکار دارند، بسیار مهم است.

در کنار مسئله خطای تقریب انتگرال، موضوع سرعت همگرایی (روند همگرایی) برآوردها نیز بسیار حائز اهمیت است. موضوع مذکور نه تنها با طرح نقاط آزمایش در ارتباط است بلکه سبب تشدید اهمیت آن‌ها می‌شود. اغلب برای بررسی سرعت همگرایی، از واریانس مقادیر برآوردی انتگرال $f(x)$ استفاده می‌شود. چراکه کم بودن واریانس مقادیر برآوردی نشان از سرعت همگرایی بالا و برعکس، زیاد بودن واریانس مقادیر برآوردی حاکی از سرعت همگرایی پایین است. واریانس مقادیر برآوردی تابع $f(x_1, \dots, x_k)$ مطابق رابطه (۴-۱۱) محاسبه می‌شود.

$$\iint \dots \int f(x_1, \dots, x_k) dx_1 \dots dx_k \cong \bar{Y} = \frac{1}{N} \sum_{i=1}^N (f(x_{i1}, x_{i2}, \dots, x_{ik})) \quad (۴-۱۱)$$

$$V(f(x_1, \dots, x_k)) \cong \hat{V}(Y) = \frac{1}{N-1} \sum_{i=1}^N (f(x_{i1}, x_{i2}, \dots, x_{ik}) - \bar{Y})^2$$

$$\Rightarrow V(\bar{Y}) = \frac{\hat{V}(Y)}{N} \Leftrightarrow SD(\bar{Y}) \propto \frac{1}{\sqrt{N}}$$

در رابطه فوق برآورد انتگرال برابر با \bar{Y} و خطای استاندارد برآورد انتگرال $f(x_1, \dots, x_k)$ برابر با $\frac{SD(\bar{Y})}{\sqrt{N}}$

است. این بدان مفهوم است که همگرایی برآوردها رابطه عکس با جذر حجم نمونه دارد. در واقع سرعت یا نرخ همگرایی برابر با $\frac{1}{\sqrt{N}}$ خواهد بود. حال این سوال مطرح می‌شود که آیا امکان بهبود روند همگرایی وجود دارد. در پاسخ به این سوال باید گفت، اعداد تصادفی به کار رفته در روش شبیه‌سازی مونت کارلو، اعداد واقعی تابع مورد بررسی نیستند. اما اگر این اعداد واقعی براساس فرمولی خاص تولید شوند و بخواهیم از این اعداد، به عنوان اعداد تصادفی استفاده کنیم باید تصادفی بودن آن‌ها با استفاده از آزمون‌های

آماري خاصی سنجیده شود. به اعدادی که بر اساس فرمول خاصی محاسبه و در این آزمون‌های آماری صدق می‌کنند، اعداد تصادفی کاذب^۱ می‌گویند (سبل، ۱۹۹۸). یک روش بهبود یافتن سرعت همگرایی، استفاده از اعداد تصادفی کاذب برای تولید نقاط نمونه‌گیری در فضای k -بعدی است. اما اگر نقاط نمونه‌گیری را از شبکه مختصات دکارتی^۲ طوری انتخاب کنیم که در نمونه‌گیری هر نقطه شبکه دقیقاً یک بار انتخاب شود، در این صورت روش مونت کارلو به یک طرح تربیعی تعیینی^۳ کارآمدی بدل می‌شود که نرخ همگرایی با نرخ N^{-1} و یا سریع‌تر کاهش پیدا می‌کند. اعداد تصادفی که با استفاده از چنین الگوریتم‌هایی بدست می‌آید دنباله اعداد شبه تصادفی نامیده می‌شوند (لوی^۴، ۲۰۰۲). دنباله اعداد شبه تصادفی به منظور استفاده از انعطاف‌پذیری اعداد تصادفی کاذب و مزایای به کارگیری یک شبکه منظم ایجاد شده‌اند. این نوع دنباله‌ها به منظور برخورداری از سطح بالای یکنواختی (یکریختی) نقاط آزمایش در فضاهای چند بعدی طراحی شده‌اند که از نظر آماری مستقل از یکدیگر نیستند و اغلب نرخ همگرایی بهتری نسبت به دنباله اعداد تصادفی کاذب دارند. دنباله اعداد شبه تصادفی را نسخه محض اعداد تصادفی مونت کارلو نیز می‌نامند. اعداد شبه تصادفی با الگوریتم‌های متفاوتی تولید می‌شوند، دنباله‌های سبل^۵، فاور^۶ و نایدیرتر^۷ نمونه‌هایی از دنباله‌های اعداد شبه تصادفی هستند. دنباله اعداد شبه تصادفی (اعداد تصادفی محض مونت کارلو) در مقایسه با اعداد تصادفی رایج در شبیه‌سازی مونت کارلو تفاوت‌های زیادی دارند. بهترین آن‌ها عبارتند از:

^۱ Pseudo-Random Number

^۲ Cartesian grid

^۳ Deterministic quadrature scheme

^۴ Levy

^۵ Faure

^۶ Niederreiter

۱. نرخ همگرایی: یکی از مزیت‌های روش اعداد شبه تصادفی، نرخ همگرایی بالای آن است. نرخ

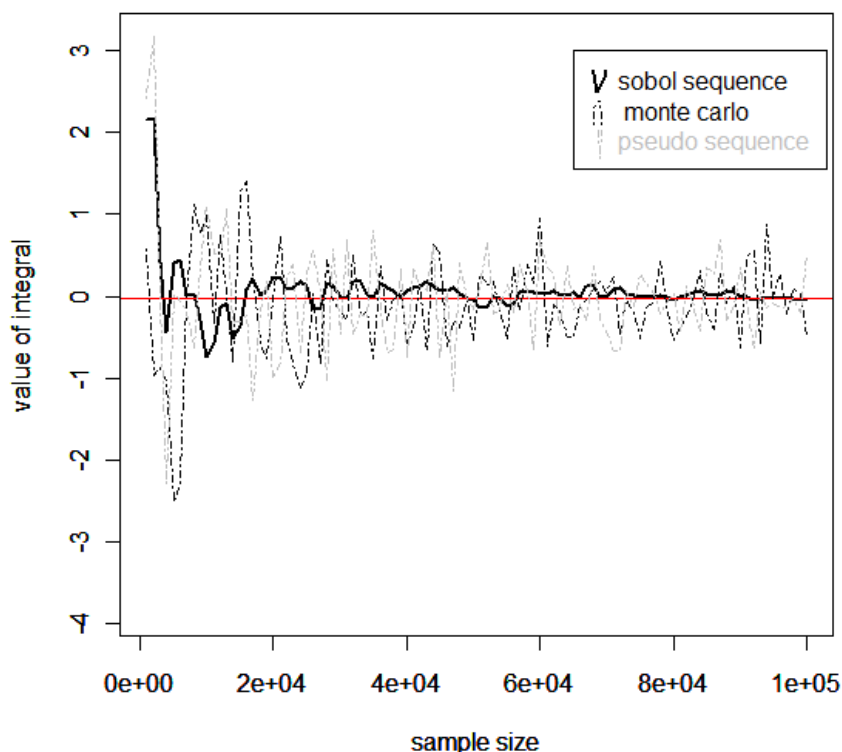
همگرایی روش‌های رایج مونت کارلو در برآورد انتگرال تابع $f(x_1, \dots, x_k)$ و یا هر تابع دیگری برابر با

$\frac{1}{\sqrt{N}}$ است. اما نرخ همگرایی اعداد شبه تصادفی N^{-c} که $c \leq 1$ است (سبل، ۱۹۹۸). شکل (۲-۴)

چگونگی همگرایی دنباله‌های شبه تصادفی سبل، یکنواخت و کاذب را در برآورد تابع

$$I = \prod_{i=1}^6 \sin(i) = -0.0219 \quad \text{برابر} \quad I = \int_{0 \dots 0}^1 \int_{0 \dots 0}^1 \int_{0 \dots 0}^1 \int_{0 \dots 0}^1 \int_{0 \dots 0}^1 \int_{0 \dots 0}^1 \prod_{i=1}^6 (i \cos(ix_i)) dx_1 \dots dx_6$$

است را نشان می‌دهد.



شکل (۲-۴) - مقایسه روند همگرایی در روش‌های سبل، کاذب و یکنواخت

همان‌طور که در شکل (۲-۴) مشهود است سرعت همگرایی دنباله سبل در مقایسه با اعداد تصادفی

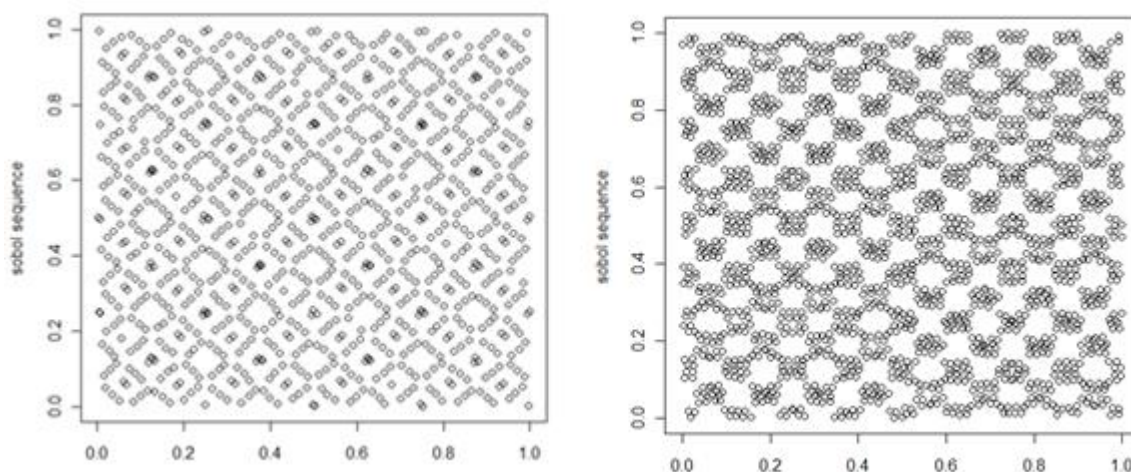
یکنواخت و کاذب بیشتر و سریع‌تر است.

۲. نحوه تولید نقاط آزمایش: تنها تفاوت نحوه تولید نقاط آزمایش این دو روش این است که در

روش‌های رایج شبه‌سازی مونت کارلو برای هر یک از متغیرها، اعداد تصادفی به صورت مجزا تولید می‌شوند. اما در الگوریتم اعداد تصادفی محض مونت کارلو تمام نقاط آزمایش برای همه متغیرها به‌طور همزمان و وابسته تولید می‌شوند.

۳. کاوش در فضای متغیرهای ورودی: مهمترین ویژگی این دنباله‌ها، کاوش در ابرفضاها است که

در مسئله‌ای با عنوان مشقت ابعادی^۱ مطرح می‌شود. مشقت ابعادی بیان‌گر این مطلب است که بیشتر نقاط آزمایش در مرکز ابر مکعب k -بعدی تجمع یافته‌اند. در نتیجه طرح آزمایشی کارآمد و مناسب است که کناره‌ها و حاشیه‌های فضای k -بعدی را نیز کاوش کند. دنباله‌های اعداد شبه تصادفی این امکان را محقق می‌سازند (سالتلی و همکاران، ۲۰۱۰). شکل (۳-۴) نمودار پراکنش دو متغیر از دنباله سبل را در فضای ۲ بعدی و ۳۰ بعدی با حجم نمونه مختلف نمایش می‌دهد. واضح است که اعداد تصادفی سبل به خوبی قادر به پوشش و کاوش در ابرفضاها هستند و با افزایش حجم نمونه این پوشش بهتر صورت می‌گیرد.



^۱ curse of dimensionality

شکل (۴-۳)- نمودار پراکنش دنباله سبل در فضای ۲ بعدی (چپ) و فضای ۳۰ بعدی برای متغیرهای ۲۵ و ۲۶ (راست).

۴. **ناهمگنی کم**^۱: ناهمگنی یک دنباله، کمیتی است که یکنواختی نقاط یک دنباله را در فضای

k -بعدی بررسی می‌کند. ناهمگنی کم، منجر به خطای کمتری در تقریب شاخص‌های حساسیت می‌شود.

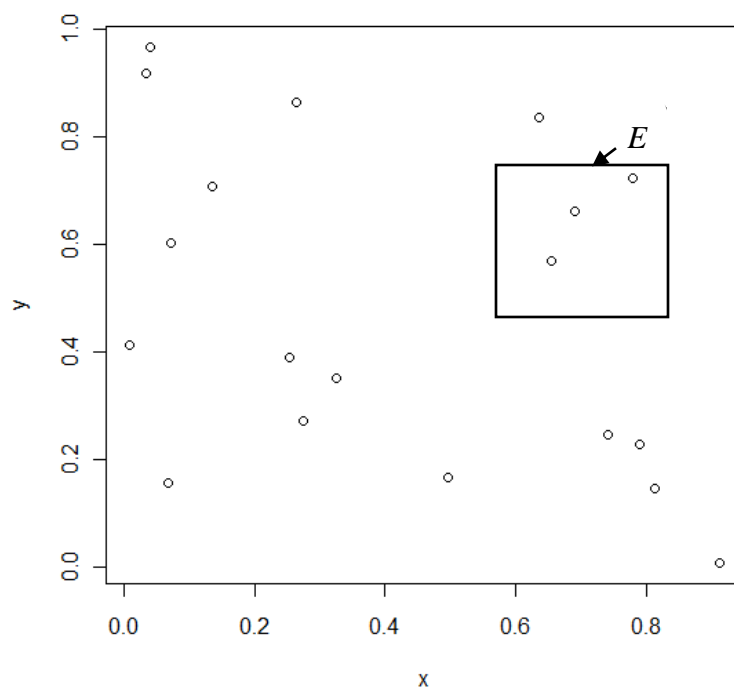
اگر مجموعه x_i نماد N نقطه در ابر مکعب واحد k -بعدی باشد، در این صورت ناهمگنی به صورت زیر تعریف می‌شود.

$$D = \sup_E \left| \frac{\# \text{ of } x_i \in E}{N} - m(E) \right| \quad (۴-۱۲)$$

که $E = [0, t_1) \times \dots \times [0, t_s)$, $0 \leq t_j \leq 1, j = 1, \dots, k$ زیرمستطیلی از ابر مکعب واحد و $m(E)$ حجم فضای E است (کارایوانوا^۲ و همکاران، ۲۰۰۰). شکل (۴-۴) نمونه‌ای از ناهمگنی در فضای دو بعدی است که می‌تواند به درک بهتر این اندازه کمک کند. تعداد نقاط در این شکل برابر با ۲۰، مربع E در برگرنده ۳ نقطه از فضا است و طول این مربع $0/۲$ است در نتیجه ناهمگنی این مربع دلخواه با توجه به رابطه فوق برابر با $0/۱۱$ است. به همین صورت مستطیل یا مربعات گوناگون بر روی نمودار پراکنش ترسیم و مقدار ناهمگنی محاسبه می‌شود. بیشترین مقدار ناهمگنی در بین ناهمگنی‌های حاصل، همان اندازه مطلوب یا D است. ناهمگنی زیاد نشان از تجمع نقاط آزمایش در قسمتی از فضا است، هرچه ناهمگنی ناچیز باشد یکنواختی و پوشش فضا بهتر انجام می‌گیرد. دنباله سبل یک طرح آزمایش مناسب و کارآمد در این زمینه است به طوری که این دنباله را به نام دنباله با ناهمگنی کم می‌شناسند. ثابت می‌شود کران بالای ناهمگنی برای N جمله اولیه دنباله شبه تصادفی برابر $D \leq C(k)(\ln(N))^k$ است که C یک مقدار ثابت می‌باشد (سبل، ۱۹۹۸).

^۱ Low discrepancy

^۲ Karaivanova



شکل (۴ - ۴) - شمای ناهمگنی در فضای دو بعدی

ویژگی‌های مذکور نشان می‌دهد دنباله اعداد شبه تصادفی سبل در مقایسه با سایر روش‌های تولید اعداد تصادفی دارای روند همگرایی سریع‌تر است. همچنین این روش پوشش و کاوش فضای متغیرهای ورودی را بهتر انجام می‌دهد. در نهایت می‌توان گفت اعداد شبه تصادفی سبل، انتخابی مناسب برای طرح نقاط آزمایش متغیرهای ورودی مدل است.

فصل پنجم

برآورد شاخص حساسیت مرتبه اول با روش

طرح متعادل تصادفی (RBD)

مقدمه

در فصل قبل، یکی از روش‌های تحلیل حساسیت واریانس - مبنا به نام روش مونت کارلو - مبنای سالتلی، معرفی و با جزئیات مورد بررسی واقع شد. اکنون روشی دیگر از کلاس روش‌های واریانس مبنا به نام طرح متعادل تصادفی را مورد مطالعه قرار می‌دهیم. در این شیوه که حالت خاصی از روش کلاسیک آزمون حساسیت بسامدی فوریه (FAST) است، برای تقریب انتگرال‌های موجود در شاخص‌های حساسیت از تجزیه طیف فوریه استفاده می‌شود. این روش برای هر متغیر ورودی مدل، یک بسامد در نظر می‌گیرد و بر اساس همین بسامدها، نقاط آزمایش، طراحی و پس از اجرای مدل کامپیوتری شاخص‌های حساسیت برآورد می‌شوند.

۵ - ۱ روش طرح متعادل تصادفی^۱

روش طرح متعادل تصادفی حالت خاصی از روش آزمون حساسیت بسامدی فوریه است. بنابراین لازم است ابتدا توضیحاتی در خصوص این روش ذکر شود.

۵ - ۱ - ۱ آزمون حساسیت بسامدی فوریه

آزمون حساسیت بسامدی فوریه نخستین بار توسط گروهی از شیمی‌دانان به سرپرستی فردی به نام کیوکر در دهه ۱۹۷۰ معرفی شد (کیوکر و همکاران، ۱۹۷۳، ۱۹۷۵، ۱۹۷۸). این دانشمندان استفاده از شاخص حساسیت مرتبه اول را برای تحلیل حساسیت پیشنهاد کردند. آن‌ها آگاهی داشتند که برای بررسی دقیق‌تر رفتار تابع نیاز به محاسبه شاخص‌های حساسیت مراتب بالاتر وجود دارد. روش این دانشمندان به

^۱ Random Balance Design Method

نام آزمون حساسیت بسامدی فوریه نام گرفت. ارتقاء محاسباتی آن نیز توسط کدا^۱ (۱۹۷۹) انجام گرفت. بنیان ریاضی این روش به شرح زیر است.

ابتدا لازم به یاد آوری است که بنابر مطالب قبل، می توان هر تابعی مانند f را به صورت بسط سلسله مراتبی مطابق رابطه (۳-۶) نوشت. یعنی

$$Y = f_0 + \sum_i f_i(x_i) + \sum_i \sum_{j>i} f_{ij}(x_i, x_j) + \dots + f_{12\dots k}(x_1, x_2, \dots, x_k)$$

همان طور که در فصل سوم بیان شد با برقراری شروط یکتایی سبل مقدار هریک از این اجزا براساس واریانس و امیدریاضی شرطی به دست می آید. این اجزا را می توان با استفاده از تجزیه طیف فوریه تابع f نیز محاسبه نمود. مقدار هریک از این اجزا براساس تجزیه طیف فوریه مطابق زیر محاسبه می شود.

$$f_0 = C_{0,0\dots 0}$$

$$f_i = \sum_{m_i \neq 0}^{+\infty} C_{0,\dots,0,m_i,0,\dots,0} \exp(-2\pi i(m_i X_i)) \quad (۵-۱)$$

$$f_{ij} = \sum_{m_i \neq 0}^{+\infty} \sum_{m_j \neq 0}^{+\infty} C_{0,\dots,0,m_i,m_j,\dots,0} \exp(-2\pi i(m_i X_i + m_j X_j))$$

...

$$f_{12\dots k} = \sum_{m_1 \neq 0}^{+\infty} \dots \sum_{m_k \neq 0}^{+\infty} C_{m_1,\dots,m_k} \exp(-2\pi i(m_1 X_1 + m_2 X_2 + \dots + m_k X_k))$$

که $C_{m_1 m_2 \dots m_k}$ ضریب فوریه است و به صورت زیر تعریف می شود.

$$C_{m_1 m_2 \dots m_k} = \int_0^1 \int_0^1 \dots \int_0^1 f(X_1, \dots, X_k) \exp(2\pi i(m_1 X_1 + m_2 X_2 + \dots + m_k X_k)) dx_1 \dots dx_k \quad (۵-۲)$$

که m_1, \dots, m_k مقادیر ثابت و مشخصی هستند. همانند رابطه (۳-۸)، واریانس های شرطی Y نیز به صورت

^۱ Koda

$$V_i = V(f_i(X_i)) = V(E(Y | X_i)) = \int_0^1 f_i^2 dx_i = \sum_{m_i \neq 0}^{+\infty} |C_{0, \dots, 0, m_i, 0, \dots, 0}|^2 = 2 \sum_{m_i \neq 1}^{+\infty} (A_{m_i}^2 + B_{m_i}^2) \quad (3-5)$$

محاسبه می‌شوند که در آن A_{m_i} و B_{m_i} به ترتیب قسمت حقیقی و موهومی ضرائب فوريه $C_{0, \dots, 0, m_i, 0, \dots, 0}$ هستند و به صورت زیر تعریف می‌شوند.

$$A_{m_j} = \int_0^1 \int_0^1 \dots \int_0^1 f(X_1, \dots, X_k) \cos(2\pi (m_j X_j)) dx_1 \dots dx_k \quad (4-5)$$

$$B_{m_j} = \int_0^1 \int_0^1 \dots \int_0^1 f(X_1, \dots, X_k) \sin(2\pi (m_j X_j)) dx_1 \dots dx_k$$

بنابراین برای محاسبه واریانس شرطی V_i می‌بایست ضرایب فوق که انتگرال‌هایی در فضای k -بعدی هستند محاسبه شود. این کار عملاً غیر ممکن بوده و بدین منظور از قضیه ارگودیک که به تبدیل ارگودیک نیز معروف است، استفاده می‌شود. قضیه ارگودیک بیان می‌کند که اگر هریک از متغیرهای ورودی را به عنوان تابعی از یک متغیر جدید مانند s در نظر بگیریم، در این صورت می‌توان انتگرال‌های k -بعدی را به یک انتگرال یک بعدی تبدیل کرد. فرم کلی این توابع تبدیل به صورت زیر است.

$$X_i(s) = G_i(\sin(w_i s)) \quad \forall i = 1, \dots, k$$

که در آن $s \in (-\pi, \pi)$ و $\{w_i\}$ یک مجموعه از بسامدهای زاویه‌ای است. بنابراین ضرایب فوريه رابطه (2-5) را می‌توان به صورت زیر بیان نمود.

$$C_{m_1 m_2 \dots m_k} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(X_1(s), \dots, X_k(s)) \exp(2\pi i (m_1 X_1 + m_2 X_2 + \dots + m_k X_k)) ds$$

که T دوره تناوب است. همچنین A_{m_i} و B_{m_i} نیز با استفاده از تابع تبدیل به صورت

$$A_{m_i} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos is \, ds$$

$$B_{m_i} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin is \, ds$$

تعریف می‌شوند. برای محاسبه شاخص حساسیت مرتبه اول کافیت $V(Y)$ را با استفاده از بسط فوریه

محاسبه کرده و سپس کسر $S_i = \frac{V_i}{V(Y)}$ را به دست آوریم. محاسبه واریانس Y در سری فوریه براساس

مفروضات و قضیه‌هایی حاصل می‌شود که نحوه محاسبه آن در سری فوریه به شرح زیر است.

اگر X_1, \dots, X_k ورودی‌های مدل و $P(X_1, \dots, X_k)$ تابع توزیع توام آن‌ها باشد، در این صورت گشتاور r -ام Y به صورت رابطه زیر خواهد بود.

$$E(Y^r) = \int \dots \int f^r(X_1, \dots, X_k) P(X_1, \dots, X_k) dX_k \dots dX_1 \quad (5-5)$$

با استفاده از قضیه ارگودیک می‌توان نوشت.

$$\bar{y}^r = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f^r(X_1(s), \dots, X_k(s)) \, ds \quad (5-6)$$

قضیه ویل^۱ بیان می‌کند که $E(y^r) \equiv \bar{y}^{(r)}$. کیوکر نشان داد که در رابطه $X_i(s) = G_i(\sin(w_i s))$ اگر

$w_i > 0$ باشند آن‌گاه $T = \pi$ خواهد بود. در این صورت با انتخاب بسامد و تابع تبدیل مناسب می‌توان

گشتاورهای Y را تقریب نمود.

$$E(y^r) = \bar{y}^{(r)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^r(s) \, ds \Rightarrow E(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \, ds \quad (5-7)$$

^۱ Weyl's Theorem

که $f(s) = f(X_1(s), X_2(s), \dots, X_k(s))$ با استفاده از خواص سری فوریه، تقریب واریانس Y به صورت زیر است.

$$\begin{aligned} \text{Var}(y) &= \bar{y}^{(2)} - (\bar{y}^{(1)})^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \right)^2 \\ &\approx \sum_{j=-\infty}^{+\infty} (A_j^2 + B_j^2) - (A_0^2 + B_0^2) \\ &\approx 2 \sum_{j=1}^{+\infty} (\Lambda_j^2) \end{aligned} \quad (۸-۵)$$

بنابراین شاخص حساسیت مرتبه اول با استفاده از بسط فوریه عبارت است از

$$S_i = \frac{V_i}{V(Y)} \approx \frac{\sum_{m_i \neq 1}^{+\infty} (A_{m_i}^2 + B_{m_i}^2)}{\sum_{m_i=1}^{+\infty} \dots \sum_{m_k=1}^{+\infty} (A_{m_i}^2 + B_{m_i}^2)} \quad (۹-۵)$$

همان طور که ملاحظه شد انگیزه اصلی برای مطالعه روش *FAST*، تبدیل انتگرال k -بعدی در فضای پارامترهای ورودی (X) به یک انتگرال یک بعدی (s) با استفاده از توابع تبدیل G است. تابع تبدیل باید طوری انتخاب شود که فضای متغیرهای ورودی را به طور یکنواخت پوشش دهد. در نتیجه مراحل برآورد شاخص حساسیت این گونه است که

(الف) ابتدا باید تابع تبدیل مناسبی انتخاب شود.

(ب) برای هر یک از متغیرهای ورودی مدل یک بسامد w_i در نظر گرفته شود و بر اساس همین بسامدها، نقاط آزمایش، تولید شده و مدل تعیینی به ازای هر نقطه یا هر ورودی اجرا گردد.

(ج) سرانجام شاخص حساسیت مرتبه اول برآورد شود.

توضیحات تفضیلی این سه مرحله به قرار زیر است.

الف- انتخاب تابع تبدیل و طرح نقاط آزمایش

از آن جایی که برآورد اثر اصلی هر متغیر براساس تابع تبدیل است، لذا انتخاب تابع تبدیل و بسامدهای مناسب در طرح آزمایش روش کلاسیک FAST بسیار مهم قلمداد می‌شود. در این راستا تبدیلات مختلفی توسط محققین پیشنهاد شده است (جدول ۵-۱).

جدول (۵ - ۱) - توابع تبدیل مختلف روش FAST

<i>Trasformation</i>	$G_i \quad i = 1, \dots, k$	<i>Reference</i>
<i>A</i>	$X_i = \bar{x}_i e^{\bar{v}_i \sin(w_i s)}$	کیوکر و همکاران (۱۹۷۳)
<i>B</i>	$X_i = \bar{x}_i (1 + \bar{v}_i \sin(w_i s))$	کدا و همکاران (۱۹۷۹)
<i>C</i>	$X_i = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(w_i s))$	سالتلی و همکاران (۱۹۹۹)

در این توابع تبدیل، \bar{x}_i یک مقدار اختیاری (اسمی) برای i امین متغیر ورودی مدل و \bar{v}_i کران بالای دامنه i امین متغیر غیرتعیینی مدل و S مقداری در فاصله $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ است. توابع تبدیل در واقع همان توابع تولید نقاط آزمایش هستند. یک تابع تبدیل مناسب باید برای هر یک از متغیرهای ورودی مدل، نقاط تصادفی رابطه‌طور یکنواخت در فضای k -بعدی تولید کند.

نمودار (۵-۱)، پانل a ، تابع تبدیل A را به ازای $\bar{v}_i = 5$ و $\bar{x}_i = e^{-5}$ و $\bar{w}_i = 11$ و پانل c ، تابع تبدیل B را به ازای $\bar{v}_i = 1$ ، $\bar{w}_i = 11$ و $\bar{x}_i = 1/2$ نشان می‌دهد. در پانل e ، تابع تبدیل سالتلی با $\bar{w}_i = 11$ به کار رفته است. همان‌طور که در پانل‌های a, c, e شکل (۵-۱) مشهود است، نوسانات متغیر ورودی (x_i) بدون وابستگی به تابع تبدیل G_i (منحنی جستجو) به صورت دوره‌ای و متناظر با بسامد w_i در سرتاسر دامنه

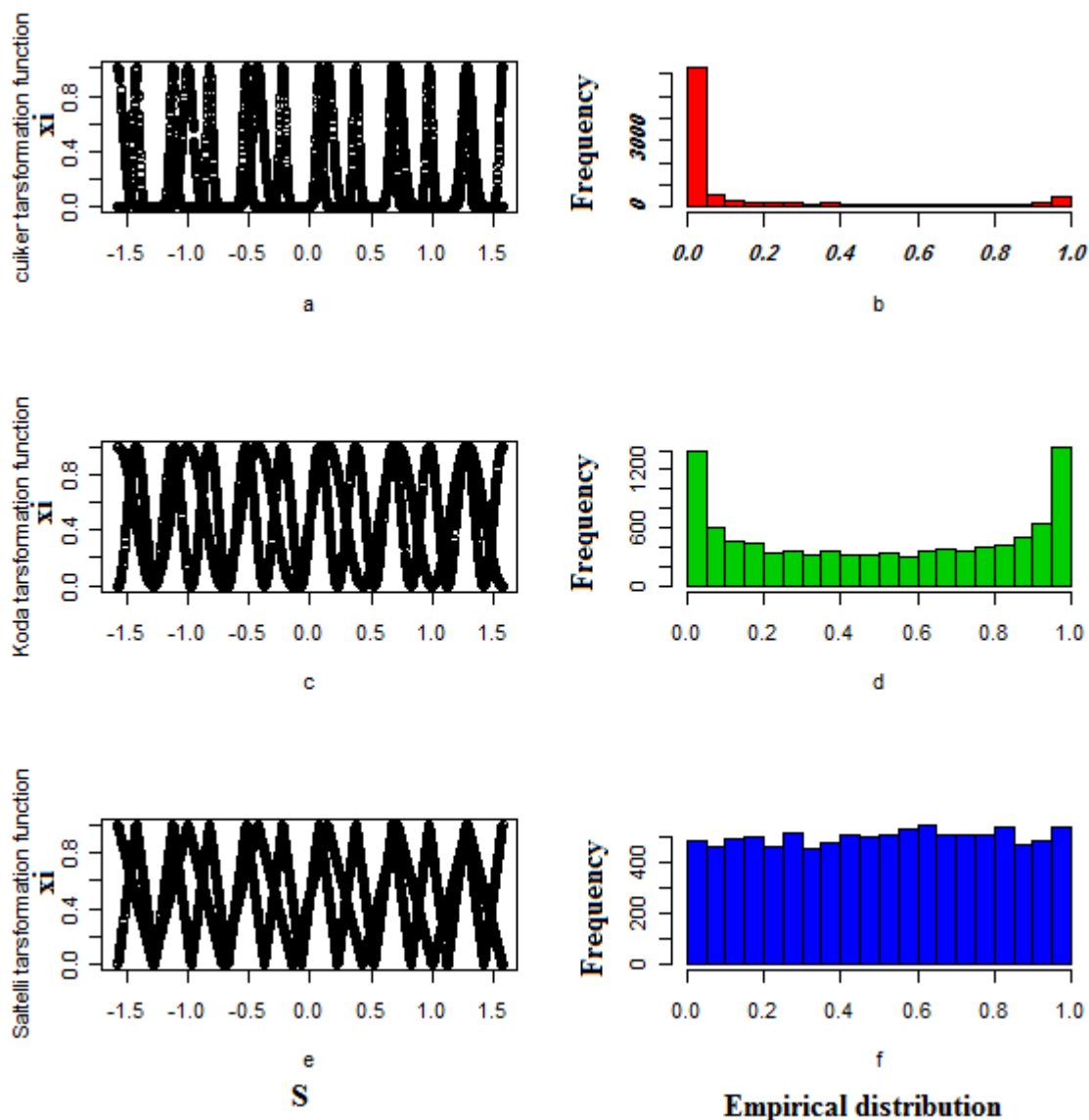
تغییرات s (محور افقی) انجام می‌گیرد. منحنی جستجو در سرتاسر دامنه s نوسان دارد به طوری که با هر تغییر در s متغیر x_i نیز به طور همزمان تغییر می‌کنند. با این عمل دامنه هر متغیر با یک اسلوب مشخصی کاوش می‌شود. منحنی جستجو می‌تواند شامل هر نقطه از فضای متغیرهای ورودی مدل باشد اگر و فقط اگر مجموعه بسامدهای $\{w_i\}$ به ازای $i = 1, \dots, k$ ترکیب خطی از یکدیگر نباشند. یعنی

$$\sum_{i=1}^k r_i w_i \neq 0 \quad -\infty \leq r_i \leq +\infty$$

تابع تبدیلی که در شرط فوق صدق کند را پوشش‌دهنده فضا^۱ گویند.

نمودار هسیتوگرام فراوانی تابع تبدیل A در پانل b ترسیم شده است. این نمودار نامتقارن است و همان‌طور که از این نمودار برمی‌آید تابع تبدیل A برای متغیری مناسب است که تابع توزیع آن چوله به راست باشد. همچنین نمودار هسیتوگرام توابع تبدیل B و C نیز به ترتیب در پانل‌های d و f نشان داده شده است. از نمودار هسیتوگرام تابع تبدیل B نیز نتیجه می‌گیریم که این تابع تبدیل برای تولید متغیرهایی مناسب است که تابع توزیع آن‌ها U شکل است. اما تابع تبدیل C برای متغیرهایی مناسب است که دارای توزیع یکنواخت هستند.

^۱ Space filling



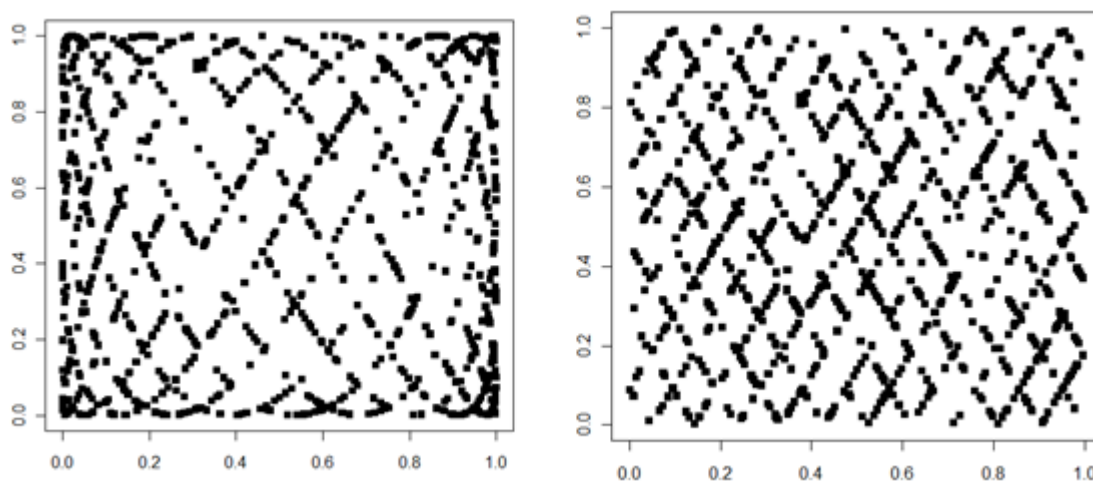
شکل (۵-۱) - نمودار توابع تبدیل و توزیع تجربی متناظر با آنها

نمودارهای هسیتوگرام متغیر x_i مربوط به سه تابع تبدیل جدول (۵-۱) نشان دهنده این است که توابع تبدیل A و B در مقایسه با C نمی‌توانند نمونه‌هایی با توزیع یکنواخت تولید کنند. تابع تبدیل C در جدول (۵-۱)، حالت خاصی از توابع تبدیل دیفرانسیلی کیوکر است. وی شرط مناسب بودن تابع تبدیل را صدق کردن در رابطه زیر بیان کرد (کیوکر و همکاران، ۱۹۷۸).

$$\pi(1 + X_i^2) P_i(G_i) \frac{dG_i(X_i)}{dX_i} = 1 \quad (5-10)$$

که P_i تابع توزیع احتمال است.

شکل (۵-۲) تفاوت نمونه‌های تولید شده با دو تبدیل B و C را در فضای دو بعدی با بسامدهای $\{w_1, w_2\} = \{11, 21\}$ نشان می‌دهد.



شکل (۵-۲) - نمودار پراکنش نقاط طرح آزمایش در فضای دو بعدی بر اساس تابع تبدیل سالتلی (راست) و تابع تبدیل کدا (چپ)

این نمودار نشان می‌دهد که نقاط تولید شده توسط تابع تبدیل C فضای متغیرهای ورودی را بهتر و یکنواخت‌تر پوشانده است.

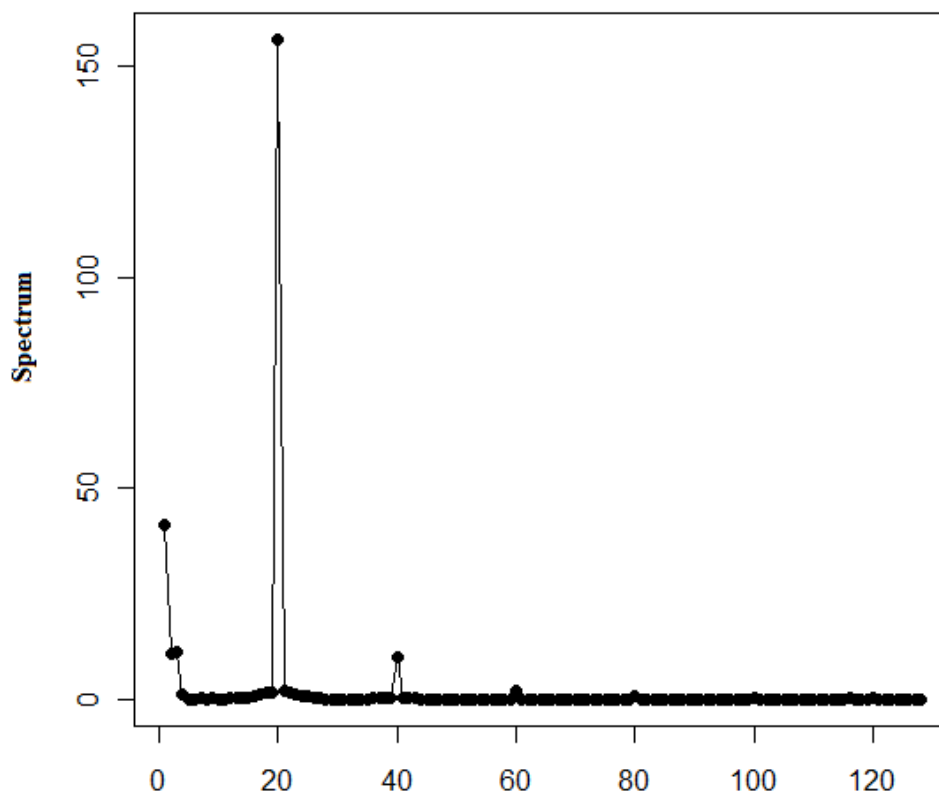
ب- انتخاب بسامد برای ورودی‌های مدل

در تخصیص بسامد به هر یک از متغیرهای ورودی محدودیت وجود دارد به‌طوری که با افزایش تعداد ورودی‌های مدل انتخاب بسامد مناسب بسیار دشوار می‌شود. به‌عنوان مثال اگر دو متغیر x_1 و x_2 با بسامدهای معین w_1 و w_2 در اختیار باشند. آن‌گاه حداقل یک ترکیب خطی از این دو بسامد وجود دارد.

یعنی ثابت‌هایی مانند p_1 و p_2 وجود خواهند داشت که رابطه $p_1 w_1 = p_2 w_2$ برقرار باشد. در این صورت طیف بسامد $p_1 w_1$ شامل ترکیبی از اطلاعات S_1 و S_2 است که راهی برای جداسازی آن وجود ندارد. این تداخل منجر به بیش برآزش یا کم برآزش در تمام مقادیر \hat{S}_i می‌شود، اما اگر تداخل، تنها در بسامدهای بالا که حاوی اطلاعات طیفی ناچیزی هستند رخ دهد، آن‌گاه خطای برآورد کمتر خواهد بود. از آنجایی که تعداد متغیرهای ورودی در آزمایش‌های کامپیوتری زیاد است لذا ممکن است ترکیب خطی w_i ها منجر به تداخل در بین متغیرهای ورودی مدل شود. برای انتخاب بسامدهای مستقل خطی، از الگوریتم بازگشتی کیوکر (۱۹۷۵) با فرم $\sum_{i=1}^k a_i w_i \neq 0$ و با شرط $\sum_{i=1}^k |a_i| \leq M' + 1$ استفاده می‌شود. در این جا a_i و M' مقادارهایی دلخواه و مثبت هستند. مجموعه بسامدهای بیشتر از مرتبه M' را آزاد از تداخل^۱ یا آزاد از پارازیت می‌گویند. به دلیل قابل اغماض بودن اثر بسامدهای بیشتر از مقادیر هارمونیک کافیتست $M' = M$ بزرگترین بسامد هارمونیک) باشد.

شکل (۳-۵) تجزیه طیف فوریه تابع $y = 2X_1 + X_1^2 + X_2 + X_3$ را نشان می‌دهد. در این تجزیه $w_1 = 20$ و $w_2 = w_3 = 1$ در نظر گرفته شده است. همان‌طور که در شکل (۳-۵) مشاهده می‌شود بیشترین نوسانات Y در بسامدهای هارمونیک ۲۰ اتفاق افتاده است. در این تجزیه بسامد انتخاب شده برای متغیر X_1 را ۲۰ در نظر گرفتیم زیرا اگر λ امین متغیر ورودی، تاثیر زیادی بر خروجی مدل (Y) داشته باشد در این صورت Y باید دارای بیشترین نوسانات در بسامد w_i باشد. نتیجه می‌گیریم که X_1 یک عامل مهم در معادله Y است.

^۱ Free of interference



شکل (۵-۳) - تجزیه فوریه تابع $y = 2X_1 + X_1^2 + X_2 + X_3$ در بسامدهای متناظر آن

ج- برآورد شاخص حساسیت مرتبه اول

تابعی با دوره تناوب $\frac{2\pi}{w_0}$ را در نظر بگیرید. اگر بسامد w_0 و تمام مقادیر هارمونیک $2w_0, 3w_0, \dots$ را با

نماد pw_i نشان دهیم که $p=1, 2, \dots$ ، آن‌گاه با محاسبه طیف فوریه و ضرائب فوریه برای بسامد w_i و

مقادیر هارمونیک آن (pw_i) می‌توان واریانس خروجی مدل و بخشی از آن که ناشی از متغیر غیر تعیینی

X_i است را مطابق رابطه (۵-۱۱) محاسبه کرد.

$$\left. \begin{aligned} D_i &\approx 2 \sum_{p=1}^{+\infty} (A_{pw_i}^2 + B_{pw_i}^2) = 2 \sum_{p=1}^{+\infty} \Lambda_{pw_i} \\ D &\approx 2 \sum_{j=1}^{+\infty} \Lambda_j \end{aligned} \right\} \Rightarrow S_i^{FAST} = \frac{D_i}{D} \quad (11-5)$$

S_i^{FAST} اندازه حساسیت خروجی مدل نسبت به متغیر ورودی i ام است. اگر فرض کنیم $\{w_1, \dots, w_k\}$ بسامدهای در نظر گرفته شده برای ورودی‌های مدل باشند. در این صورت $\{p_1 w_1, \dots, p_k w_k\}$ بسامدهای اصلی خروجی مدل هستند که $p_1, p_2, \dots, p_k \in \{1, \dots, +\infty\}$ است. بنابر ویژگی تجزیه طیف فوریه، با افزایش p_i ها مقدار اطلاعات مولفه‌های طیفی کاهش پیدا می‌کنند (شکل ۴-۸). بنابراین در محاسبات بزرگترین بسامد هارمونیک را معمولا $M=4$ یا $M=6$ در نظر می‌گیرند (کیوکر و همکاران، ۱۹۷۵). در نتیجه

$$\left. \begin{aligned} \hat{D}_i &= 2 \sum_{p=1}^M (A_{pw_i}^2 + B_{pw_i}^2) = 2 \sum_{p=1}^M \Lambda_{pw_i} \\ \hat{D} &= 2 \sum_{j=1}^{N-1/2} \Lambda_j \end{aligned} \right\} \Rightarrow S_i^{FAST} = \frac{\hat{D}_i}{\hat{D}} \quad (12-5)$$

در واقع S_i^{FAST} اثر اصلی متغیرهای ورودی i ام مدل بر واریانس خروجی است.

سالتلی (۱۹۹۹) ثابت نمود، کمترین حجم نمونه در روش کلاسیک FAST برابر $N_s = 2Mw_{\max} + 1$ است که w_{\max} بزرگترین بسامد اختصاص یافته به متغیرهای ورودی مدل و M بزرگترین مقدار بسامد هارمونیک می‌باشد. محاسبات روش FAST نسبت به روش مونت کارلو- مبنا بسیار کمتر است، هرچند در برخی موارد این برآوردگر مشکل آریبی دارد (سالتلی و بولادو^۱، ۱۹۹۸). زیرا برای محاسبه شاخص حساسیت مرتبه اول همه متغیرهای ورودی، تنها یک نمونه به حجم N_s در نظر گرفته می‌شود.

^۱ Bolado

۵ - ۱ - ۲ معرفی روش طرح متعادل تصادفی

روش طرح متعادل تصادفی حالتی خاص از روش FAST است که برای نخستین بار توسط تارانتولا (۲۰۰۶) ارائه شد. در این روش ابتدا N نقطه آزمایش با استفاده از تابع تبدیل مناسب انتخاب می‌شود، اما برخلاف روش FAST، کاوش در فضای k -بعدی ورودی‌های مدل با تخصیص بسامد یکسان به تمام آن‌ها انجام می‌گیرد. با این عمل منحنی کاوش پوشش‌دهنده فضا نمی‌باشد اما یک بخش از فضای متغیرهای ورودی مدل را به‌طور کامل تحت پوشش قرار می‌دهد. در این روش w_i می‌تواند هر مقداری اختیار کند اما بیشترین مقدار قابل اخذ آن $\frac{N-1}{2M}$ است که N حجم نمونه و M تداخل بین متغیرها یا بزرگترین بسامد هارمونیک طیف فوریه می‌باشد. الگوریتم روش طرح متعادل تصادفی (RBD) برای برآورد شاخص حساسیت مرتبه اول است به صورت زیر است.

- مجموعه مرتب شده N تایی $A = \{m \mid m = -\pi + (i-1)(2\pi/(N-1)), i = 1, 2, \dots, N\}$ را در نظر بگیرید. این مجموعه شامل نقاطی با فواصل مساوی در بازه $(-\pi, \pi)$ است. به ازای هر یک از $X_i, i = 1, 2, \dots, k$ یک جایگشت تصادفی N تایی از مجموعه A به فرم $S_i = (s_{i1}, s_{i2}, \dots, s_{iN})$ تولید کنید.

- برای هر $S_i, i = 1, 2, \dots, k$ تولید شده، مقادیر $X_i(S_i) = G(\sin(s_{ij}))$, $j = 1, 2, \dots, N$ را محاسبه نمایید که در آن G_i تابعی است که بتوان بر اساس آن، تابع چگالی احتمال مطلوب برای X_i را یافت. این نقاط را به عنوان "طرح آزمایش"^۱ در نظر بگیرید. (به عنوان مثال اگر X_i دارای توزیع

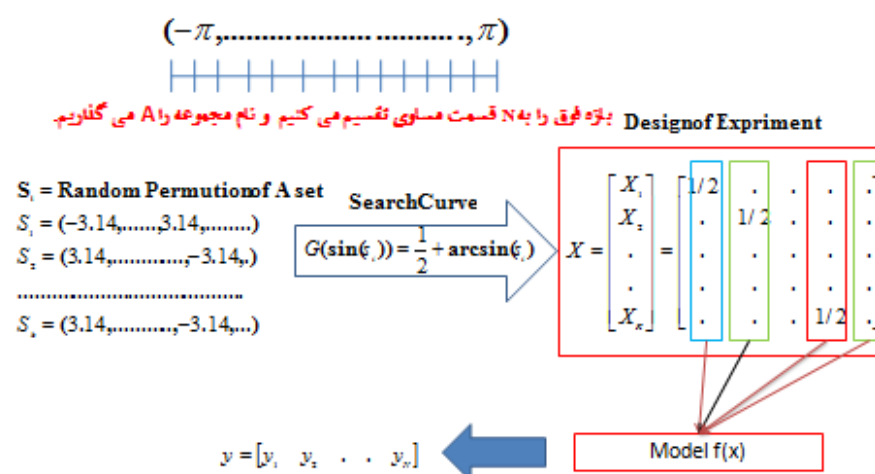
یکنواخت باشد کافی است تابع تبدیل $G(t) = 0.5 + \arcsin(t)/\pi$ باشد)

^۱ موضوع طرح و تجزیه آزمایش‌های کامپیوتری توسط سانتنر و همکاران (۲۰۰۳) به طور مبسوط مورد بحث قرار گرفته است.

- مقادیر $Y_j = f(X_1(s_{1j}), \dots, X_k(s_{kj}))$, $j = 1, 2, \dots, N$ را با N بار اجرای مدل بدست آورید.
- برای برآورد شاخص حساسیت هر یک از پارامترهای X_i , $i = 1, 2, \dots, k$ اعمال زیر را انجام دهید.
- رتبه مقادیر بردار $S_i = (s_{i1}, s_{i2}, \dots, s_{iN})$ را به صورت (R_1, \dots, R_N) در نظر گرفته و بردار (Y_1, \dots, Y_N) را متناظر با آن مرتب سازی و نام آن را $Y_i^R(s_{ij})$ بگذارید.
- تبدیل فوریه مختلط $Y_i^R(s_{ij})$ را برای بسامدهای هارمونیک $w = 1, \dots, M$ به صورت $F(w) = \left| \frac{1}{\pi} \sum_{j=1}^N Y_i^R(s_{ij}) \exp(-i w C_j) \right|^2$ محاسبه می‌کنیم. معمولاً M برابر با ۴ یا ۶ و C_j ضریب فوریه مختلط است.
- برآورد اثر اصلی عامل X_i برابر است با $\hat{S}_i = \hat{V}_i / \hat{V}(Y)$ که $\hat{V}_i = \sum_{w=1}^M F(w)$ و $\hat{V}(Y) = \sum_{w=2}^N F(w)$.

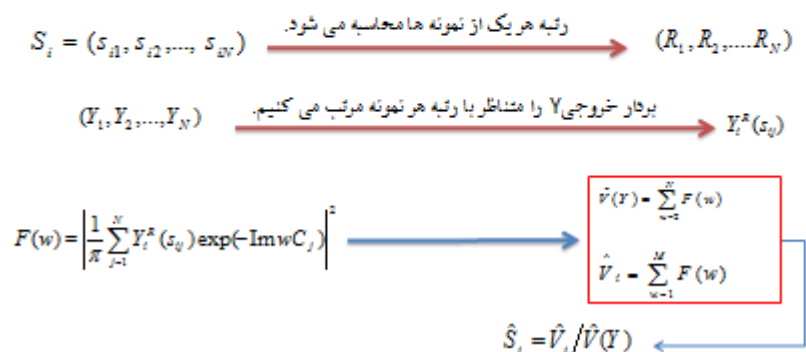
شکل (۴-۵) الگوریتم روش RBD را نشان می‌دهد.

مرحله اول



مرحله دوم

برای هر فاکتور بایستی مراحل زیر تکرار شود.



شکل (۵-۴) - الگوریتم روش RBD

۵ - ۲ مقایسه روش های تقریب شاخص های حساسیت

هر یک از تکنیک های برآورد شاخص های حساسیت دارای قابلیت خاصی هستند. جدول (۵-۲) قابلیت های برخی از این روش ها را به همراه هزینه محاسبات آن ها نشان می دهد.

جدول (۵ - ۲) - مقایسه روش های مختلف تحلیل حساسیت

روش برآورد	تعداد اجراها	امکان برآورد شاخص حساسیت مرتبه اول	امکان برآورد شاخص حساسیت کل
FAST (کیوکر و همکاران، ۱۹۷۳)	$2M w_{\max} + 1$	✓	✗
سالتلی (۲۰۰۲)	$N(k+2)$	✓	✓
RBD (تارانتولا ۲۰۰۶)	N	✓	✗

فصل ششم

ارزیابی روش‌های

مونت کارلو - مبنای سالتلی و طرح متعادل تصادفی

در یک تابع تحلیلی

مقدمه

از آن جایی که ضابطه ریاضی بین متغیرهای ورودی و خروجی در مدل‌های کامپیوتری به‌طور صریح مشخص نیست، لذا در بسیاری از موارد، برای ارزیابی روش‌های تحلیل حساسیت از توابع تحلیلی که فرم ریاضی آن‌ها، مشخص و مقادیر واقعی شاخص‌های حساسیت آن‌ها قابل محاسبه است، استفاده می‌شود. تابع سُبُل یکی از این توابع است که در فصل سوم به‌طور اجمالی معرفی و مقادیر واقعی شاخص‌های حساسیت آن محاسبه شد. در این فصل ساختار و ویژگی‌های این تابع تحلیلی را بیان نموده و مقادیر تقریبی این شاخص‌ها را براساس روش‌های مونت کارلو-مبنا سالتلی و RBD (ذکر شده در فصل چهارم و پنجم) به‌دست می‌آوریم. در انتها نیز مقادیر برآوردشده را با مقادیر واقعی، از نظر دقت و همگرایی، مورد مقایسه قرار خواهیم داد. هدف از انجام مقایسه‌های فوق انتخاب روش بهتر نیست، بلکه رویت کارایی روش‌های مذکور در تابع تحلیلی سبل است.

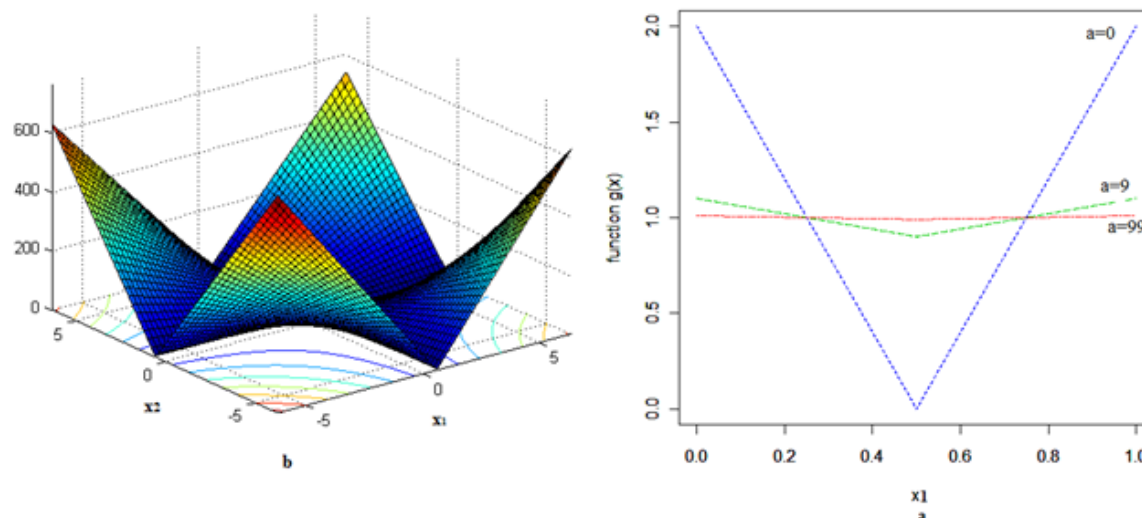
در این فصل منظور از متغیر، همان متغیر ورودی مدل و حجم نمونه نیز تعداد نقاط آزمایش است. به‌بیان دیگر، مقصود از حجم نمونه در روش مونت کارلو - مبنای سالتلی، همان تعداد سطرهای ماتریس A یا B و در روش طرح متعادل تصادفی (RBD) تعداد تقسیمات بازه $(-\pi, \pi)$ است. در ضمن برای اختصار در نوشتار، به جای عبارت‌های روش مونت کارلو- مبنای سالتلی، شاخص حساسیت مرتبه اول و شاخص حساسیت کل، به ترتیب از "روش سالتلی"، اثر اصلی " و "اثر تام" استفاده شده است.

۶ - ۱ معرفی تابع تحلیلی سبل

تابع تحلیلی سبل، موسوم به *g-function* دارای فرم کلی زیر است.

$$Y = f(x_1, \dots, x_k) = \prod_{i=1}^k G_i(X_i) \quad G_i(X_i) = \frac{|4X_i - 2| + a_i}{1 + a_i} \quad ; 0 \leq X_i \leq 1$$

که در a_i ها مقادیر ثابت و مثبتی هستند. نمودار این تابع غیر خطی و غیر یکنوا، به ازای مقادیر مختلف a_i در حالت یک و دو متغیره در شکل (۵-۱) نمایش داده شده است.



شکل (۶-۱) - نمودار تابع سبل در حالت یک متغیره (پانل راست) و دومتغیره (پانل چپ، به ازای $a_1=a_2=0$)

کران‌های بالا و پایین $g_i(x_i)$ عبارتند از

$$0 \leq X_i \leq 1 \Rightarrow 0 \leq |4X_i - 2| \leq 2 \Rightarrow 1 - \frac{1}{1+a_i} \leq \frac{|4X_i - 2| + a_i}{1+a_i} \leq 1 + \frac{1}{1+a_i} \quad (۶-۱)$$

$$1 - \frac{1}{1+a_i} \leq G_i(X_i) \leq 1 + \frac{1}{1+a_i}$$

درجه اهمیت هریک از متغیرها توسط مقدار a_i تعیین می‌شود. هر چه متغیر مورد بررسی بااهمیت‌تر باشد، نوسانات تابع Y نیز در دامنه تغییر آن متغیر بیشتر خواهد بود. در جدول (۶-۱) برخی از مقادیر a_i به همراه درجه اهمیت x_i مربوطه و مقادیر واقعی اثر اصلی این متغیرها آمده است. مقدار واقعی اثر اصلی

متغیر X_i برابر است با $S_i = v_i/V$ که در آن $V = \prod_{i=1}^k (v_i + 1) + 1$ و $v_i = \frac{1}{3(1+a_i)^2}$ است (به مثال ۲

فصل ۳ مراجعه شود). لازم به ذکر است که مقادیر a_i به کار رفته در جدول (۶-۱) به طور متداول در مراجع علمی مرتبط استفاده می‌شود.

جدول (۶ - ۱) - برد و درجه اهمیت تابع $g_i(x_i)$ به ازای مقادیر مختلف a_i

مقدار واقعی اثر اصلی	برد $g_i(x_i)$	میزان اهمیت متغیر X_i	a_i
0.716	$0 \leq g_i(x_i) \leq 2$	بسیار مهم	0
0.179	$.5 \leq g_i(x_i) \leq 1.5$	مهم	1
0.024	$0.8181 \leq g_i(x_i) \leq 1.1881$	نسبتاً مهم	4.5
0.0072	$0.9 \leq g_i(x_i) \leq 1.1$	کم اهمیت	9
7.162×10^{-5}	$0.99 \leq g_i(x_i) \leq 1.01$	فاقد اهمیت	99

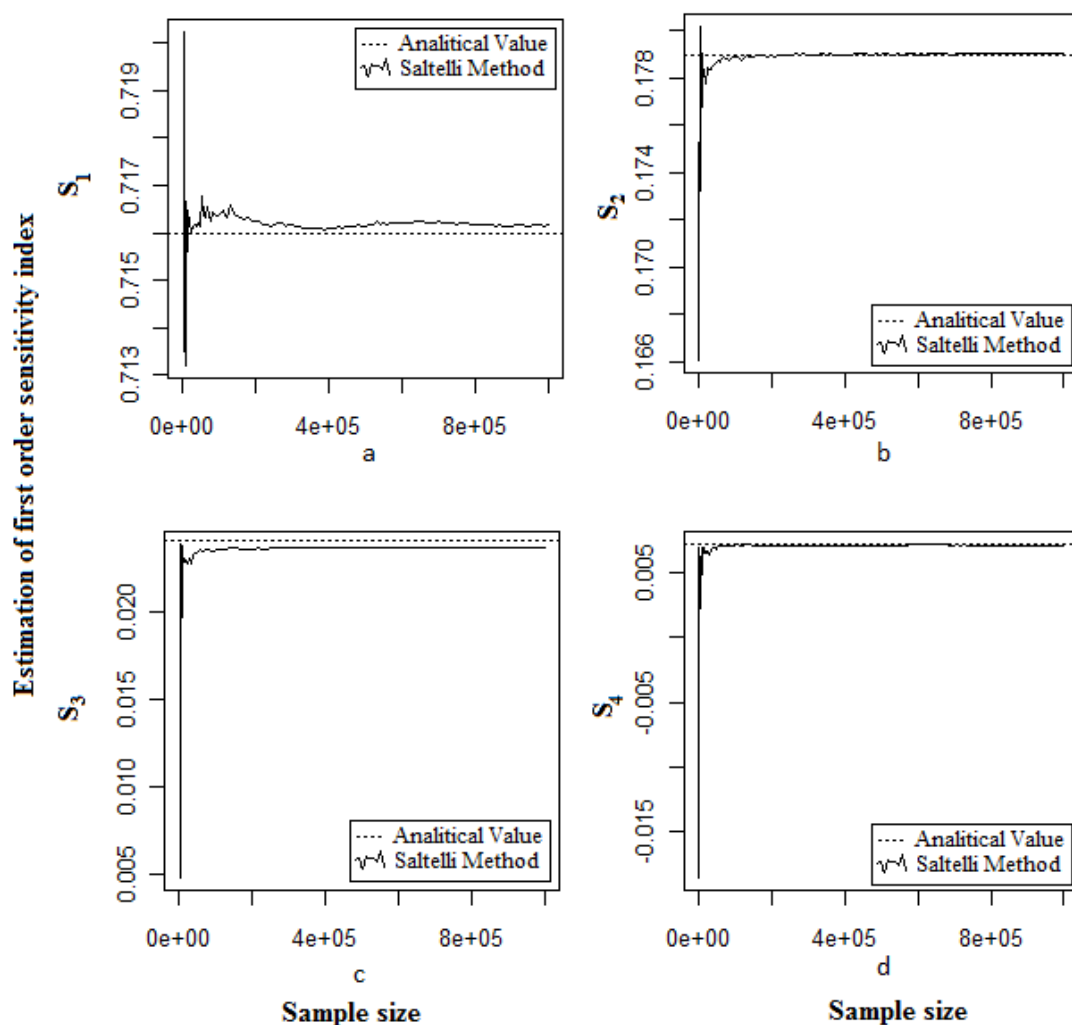
برای ارزیابی روش‌های، مونت کارلو- مبنای سالتلی و طرح متعادل تصادفی و نیز مقایسه مقادیر واقعی و برآورد شده شاخص‌های حساسیت، تعداد متغیرهای ورودی‌های تابع تحلیلی را برابر هشت ($k=8$) و میزان اهمیت متغیرها را نیز متفاوت در نظر گرفته‌ایم. جدول (۶-۲) مقادیر مختلف a_i را برای ورودی‌های تابع سبل نشان می‌دهد.

جدول (۶ - ۲) - مقادیر مفروض a_i در تابع سبل

متغیر	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
(a_i)	0	1	4.5	9	99	99	99	99

۶-۲ ارزیابی روش سالتلی توسط تابع تحلیلی

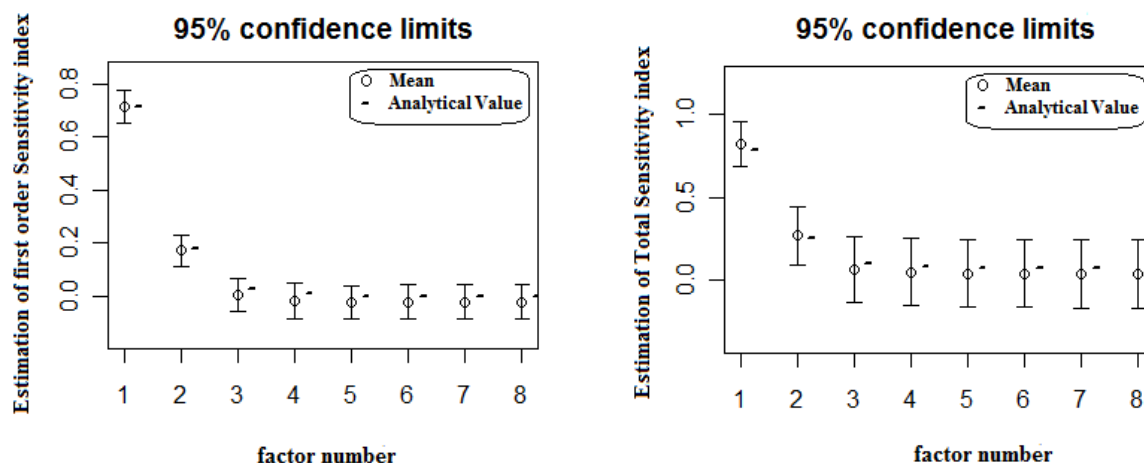
در این بخش، روند همگرایی و میزان دقت مقادیر برآورد شده روش سالتلی مورد بررسی قرار گرفته است. جهت بررسی همگرایی مقادیر برآورد شده به مقادیر واقعی، برای متغیرهای $X_1 - X_4$ نمونه‌هایی با حجم مختلف در نظر گرفته شده و اثر اصلی این متغیرها برآورد شده‌اند. شکل (۶-۲) روند همگرایی مقادیر برآورد شده را نشان می‌دهد.



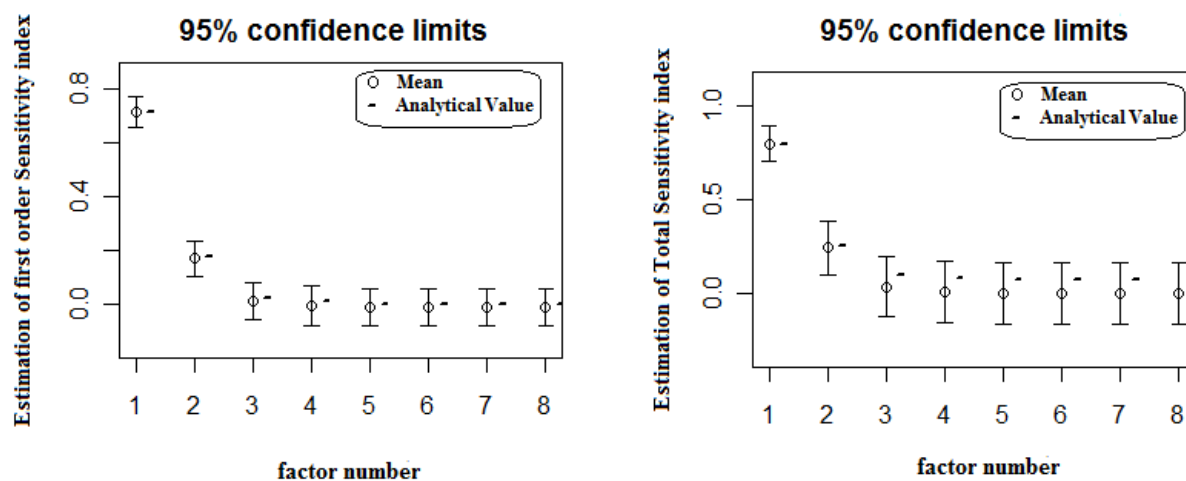
شکل (۶-۲) - روند همگرایی مقادیر برآورد متغیر بسیار مهم X_1 (پانل a)، متغیر مهم X_2 (پانل b)، متغیر نسبتاً مهم X_3 (پانل c) و بی اهمیت X_4 (پانل d)

بر اساس نمودارهای فوق می‌توان این چنین استدلال کرد که به‌طور کلی روند همگرایی برآوردهای روش سالتلی به مقادیر واقعی سریع می‌باشد. پانل a که مربوط به متغیر بسیار مهم X_1 است، نشان می‌دهد که مقادیر برآورد دارای بیش‌برازش بوده اما اریبی این برآورد بسیار ناچیز است. پانل (b) روند همگرایی متغیر مهم X_2 را نشان می‌دهد که در آن، مقدار برآورد حتی در حجم نمونه نسبتاً کم، بسیار نزدیک به مقدار واقعی بوده و پایدار است. به همین ترتیب، برآورد اثر اصلی متغیرهای نسبتاً مهم X_3 (پانل c) و کم‌اهمیت X_4 (پانل d) با افزایش حجم نمونه به پایداری می‌رسد.

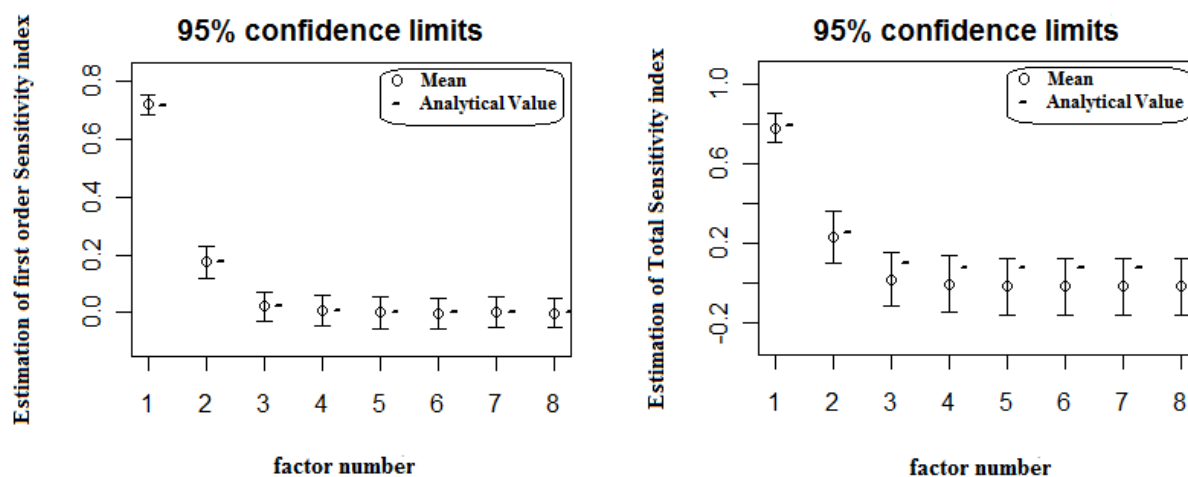
جهت بررسی دقت برآوردهای روش سالتلی، اثرات اصلی و تام هر یک از متغیرها با نمونه‌هایی به حجم ۵۰۰، ۱۰۰۰ و ۲۰۰۰ برآورد شده‌اند. اما از آن‌جاکه نمونه‌ها تصادفی هستند، قطعاً با داشتن نمونه‌های جدید، برآوردهای حساسیت متفاوتی برای هر متغیر به‌دست می‌آید. لذا جهت تعیین عدم قطعیت این برآوردها با استفاده از تکنیک بوت استرپ، فرآیند فوق را ۵۰ بار تکرار کرده و در هر تکرار، شاخص‌های حساسیت را برای هر متغیر محاسبه نموده‌ایم. نتایج حاصل از این فرآیند در قالب فاصله اطمینان ۹۵٪ بوت استرپی برای برآورد اثر اصلی و اثر تام در اشکال (۳-۶) تا (۵-۶) نشان داده شده است.



شکل (۶-۳) - فاصله اطمینان ۹۵٪ مقادیر برآورد شده اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۵۰۰



شکل (۶ - ۴) - فاصله اطمینان ۹۵٪ مقادیر برآورد شده اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۱۰۰۰



شکل (۶ - ۵) - فاصله اطمینان ۹۵٪ مقادیر برآوردی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در حجم نمونه ۲۰۰۰

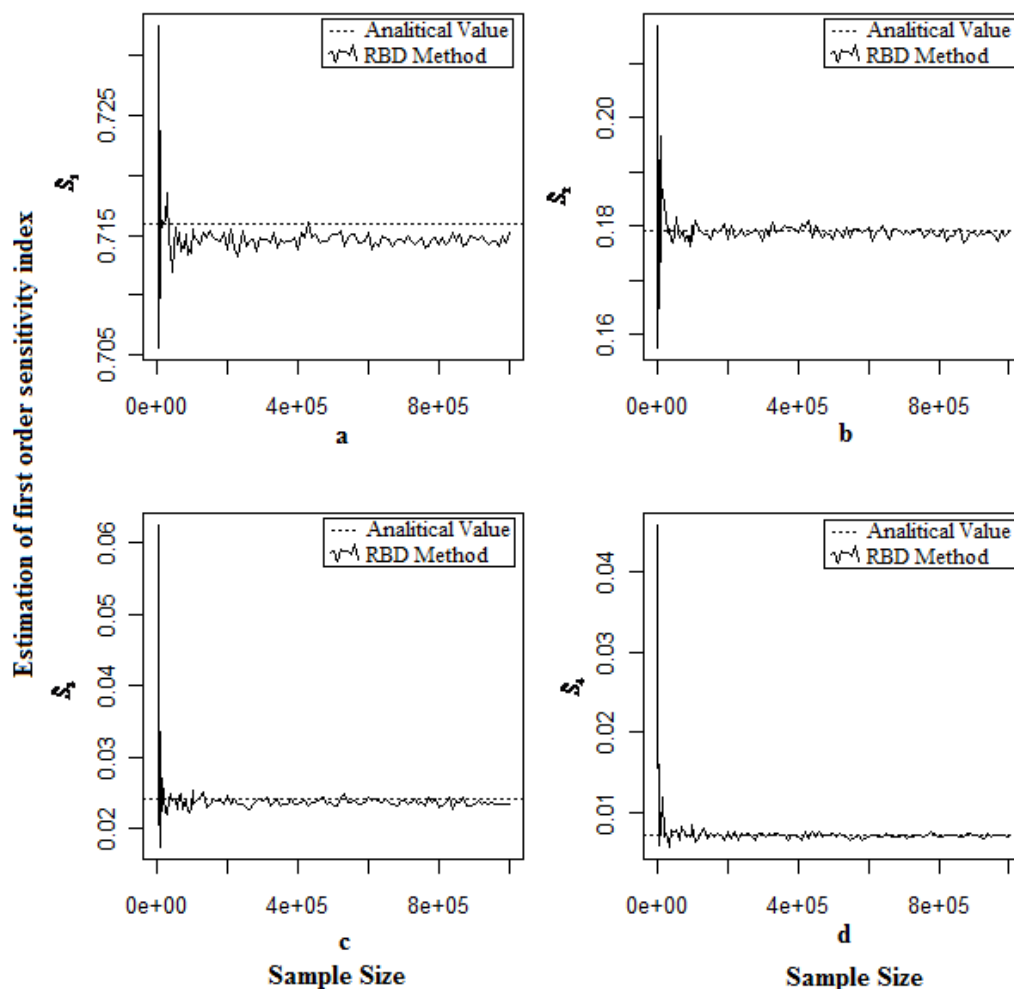
از اشکال فوق دو نتیجه زیر به دست می‌آید.

- (۱) میانگین برآورد اثر اصلی هر یک از متغیرهای ورودی مدل با افزایش حجم نمونه، به مقدار واقعی بسیار نزدیک شده و طول فواصل اطمینان نیز کمتر می‌شود. این روند در برآورد اثر تام متغیرهای X_1 و X_2 نیز به چشم می‌خورد.

۲) مقادیر برآورد شده اثر تام متغیرهای کم‌اهمیت و فاقد اهمیت (X_8-X_3) با افزایش حجم نمونه از مقدار واقعی دور می‌شوند. فواصل اطمینان این متغیرها نیز دربرگیرنده مقدار واقعی اثر اصلی نیست.

۶-۳ ارزیابی تکنیک طرح متعادل تصادفی با تابع تحلیلی

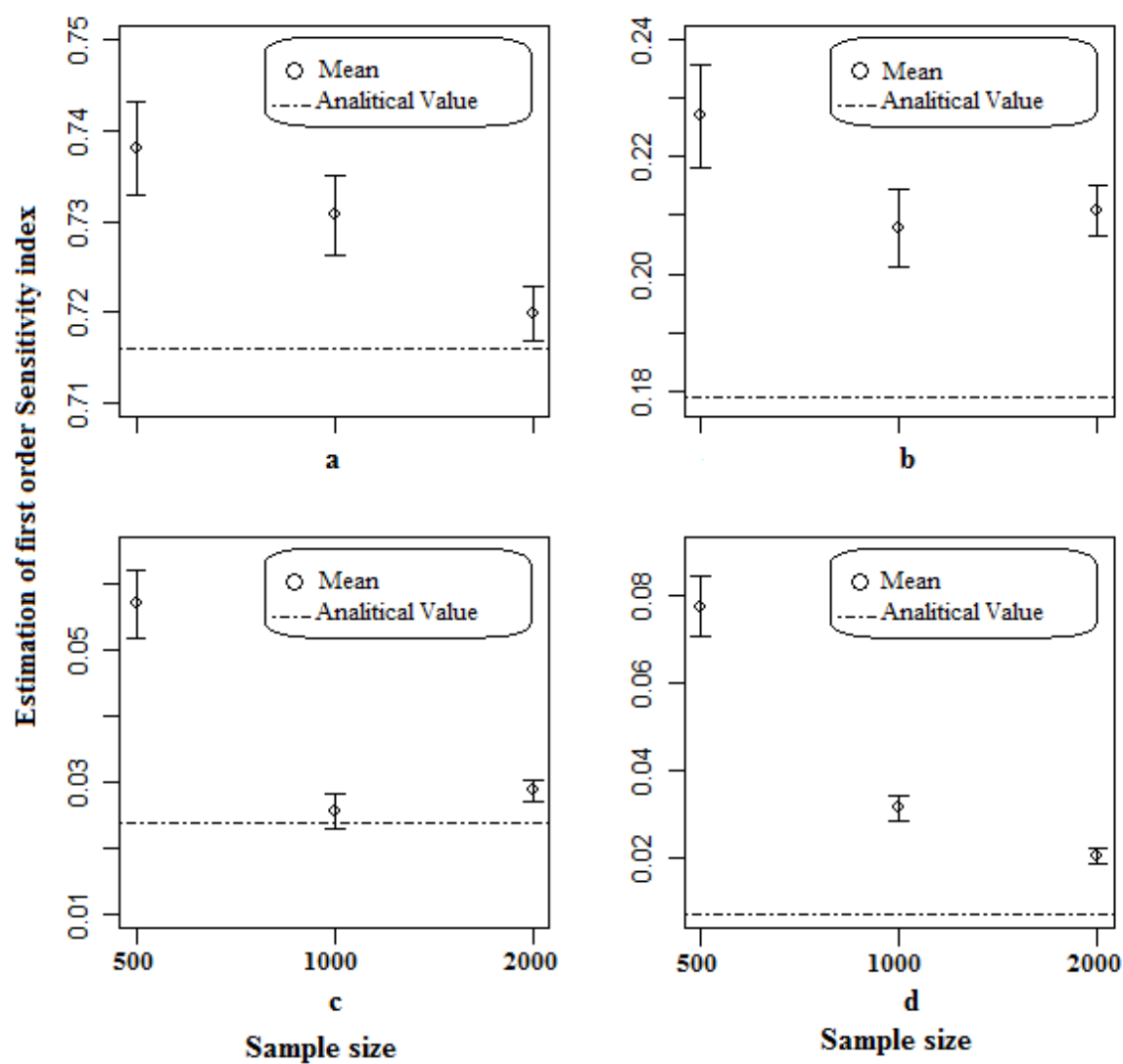
شکل (۶-۶) روند همگرایی برآورد شاخص‌های حساسیت در روش طرح متعادل تصادفی (RBD) را برای تابع تحلیلی نشان می‌دهد.



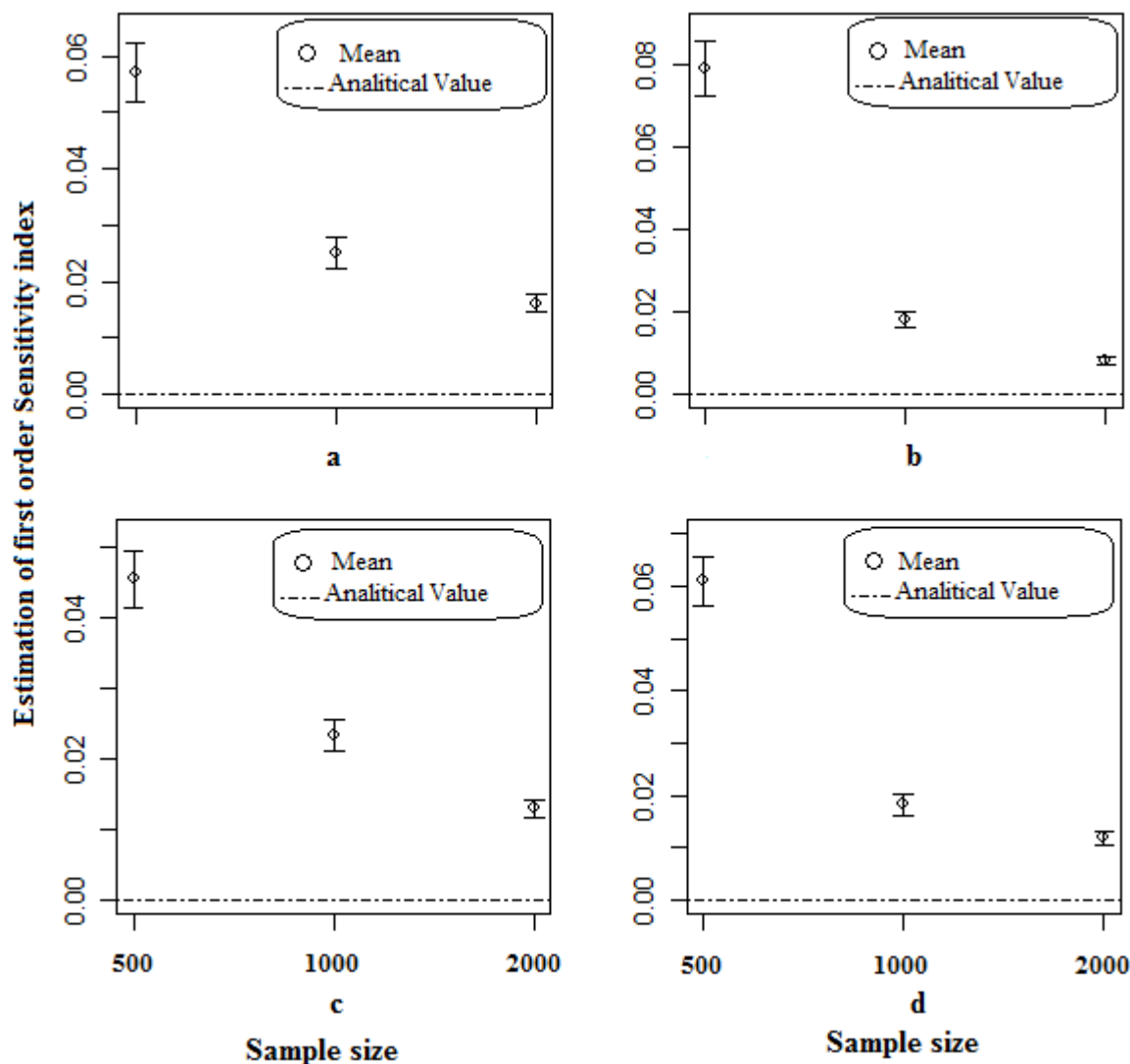
شکل (۶-۶) - روند همگرایی مقادیر برآورد متغیر بسیار مهم X_1 (پانل a)، متغیر مهم X_2 (پانل b)، متغیر نسبتاً مهم X_3 (پانل c) و بی اهمیت X_4 (پانل d)

در پانل a ، اثر اصلی متغیر بسیارمهم X_1 کمتر از مقدار واقعی برآورد شده که این اختلاف بسیارناچیز به نظر می‌رسد. برای متغیرهای مهم X_2 (پانل b)، نسبتاً مهم X_3 (پانل c) و کم اهمیت X_4 (پانل d)، روند همگرایی مقادیر برآورد شده به مقدار واقعی، سریع است به‌طوری که با افزایش حجم نمونه این برآوردها به پایداری می‌رسند. می‌توان در مجموع، روند همگرایی روش RBD را مناسب ارزیابی کرد. همگرایی این روش در برآورد اثر اصلی متغیرهای مهم، بی‌اهمیت و فاقد اهمیت بسیار سریعتر از متغیر بسیار با اهمیت X_1 است.

به منظور ارزیابی دقت روش RBD ، همانند قبل، نمونه‌گیری را به تعداد ۵۰ بار برای هر یک از حجم نمونه‌های ۵۰۰، ۱۰۰۰ و ۲۰۰۰ تایی تکرار نموده‌ایم که فاصله‌های اطمینان بوت استرپی مربوط به هر یک از متغیرهای ورودی در شکل‌های (۶-۷) و (۶-۸) نشان داده شده‌اند. در اشکال مذکور تاثیر حجم نمونه در برآورد اثر اصلی متغیرهای مدل کاملاً ملموس است. طول فاصله اطمینان برآورد اثر اصلی در تمامی متغیرهای مورد بررسی با افزایش حجم نمونه کاهش می‌یابد. همچنین مقادیر برآورد شده دارای بیش‌برازش بوده اما اریبی این برآورد بسیارناچیز است. خصوصیت بارز روش طرح متعادل تصادفی نسبت به روش مونت کارلو - مبنای سالتلی، برآورد اثر اصلی با حجم نمونه کم است. از آنجایی که در عمل، شاخص‌های حساسیت با اجرای مکرر مدل و به‌زای ورودی‌های متفاوت برآورد می‌شوند. بنابراین کم بودن هزینه اجرای مدل یک پارامتر بسیار مهم تحلیل حساسیت خروجی مدل است که روش RBD دارای این ویژگی است. به عبارت دیگر، RBD یک روش برای برآورد اثر اصلی با هزینه اجرایی بسیار کم است.



شکل (۶-۷) - مقادیر واقعی و فاصله اطمینان ۹۵٪ برای متغیر بسیار مهم X_1 (پانل a)، متغیر مهم X_2 (پانل b)، متغیر نسبتا مهم X_3 (پانل c) و متغیر بی اهمیت X_4 (پانل d) در حجم‌های نمونه مختلف

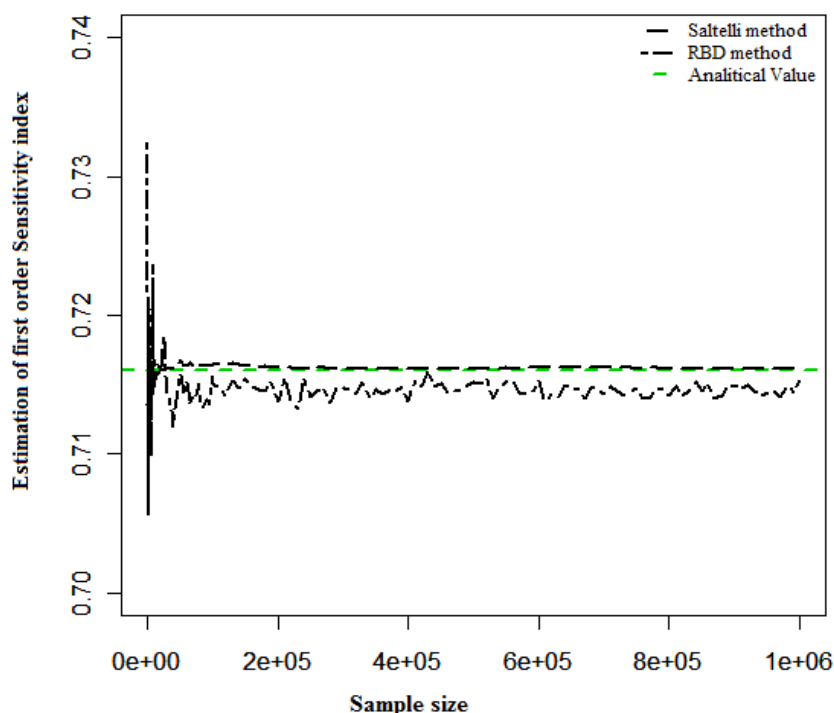


شکل (۶-۸) - مقادیر واقعی و فاصله اطمینان ۹۵٪ برای متغیرهای فاقد/اهمیت X_5 (پانل a)، X_6 (پانل b)، X_7 (پانل c) و X_8 (پانل d) در حجم نمونه مختلف

۴-۶ مقایسه روش مونت کارلو - مبنای سالتلی و طرح متعادل تصادفی

در دو بخش قبل، دقت و همگرایی مقادیر برآورد برای هر یک از دو روش به طور مجزا بررسی شد. اکنون از این نقطه نظر، آن‌ها را به طور همزمان مورد مقایسه قرار می‌دهیم. هدف در این بخش انتخاب بهترین

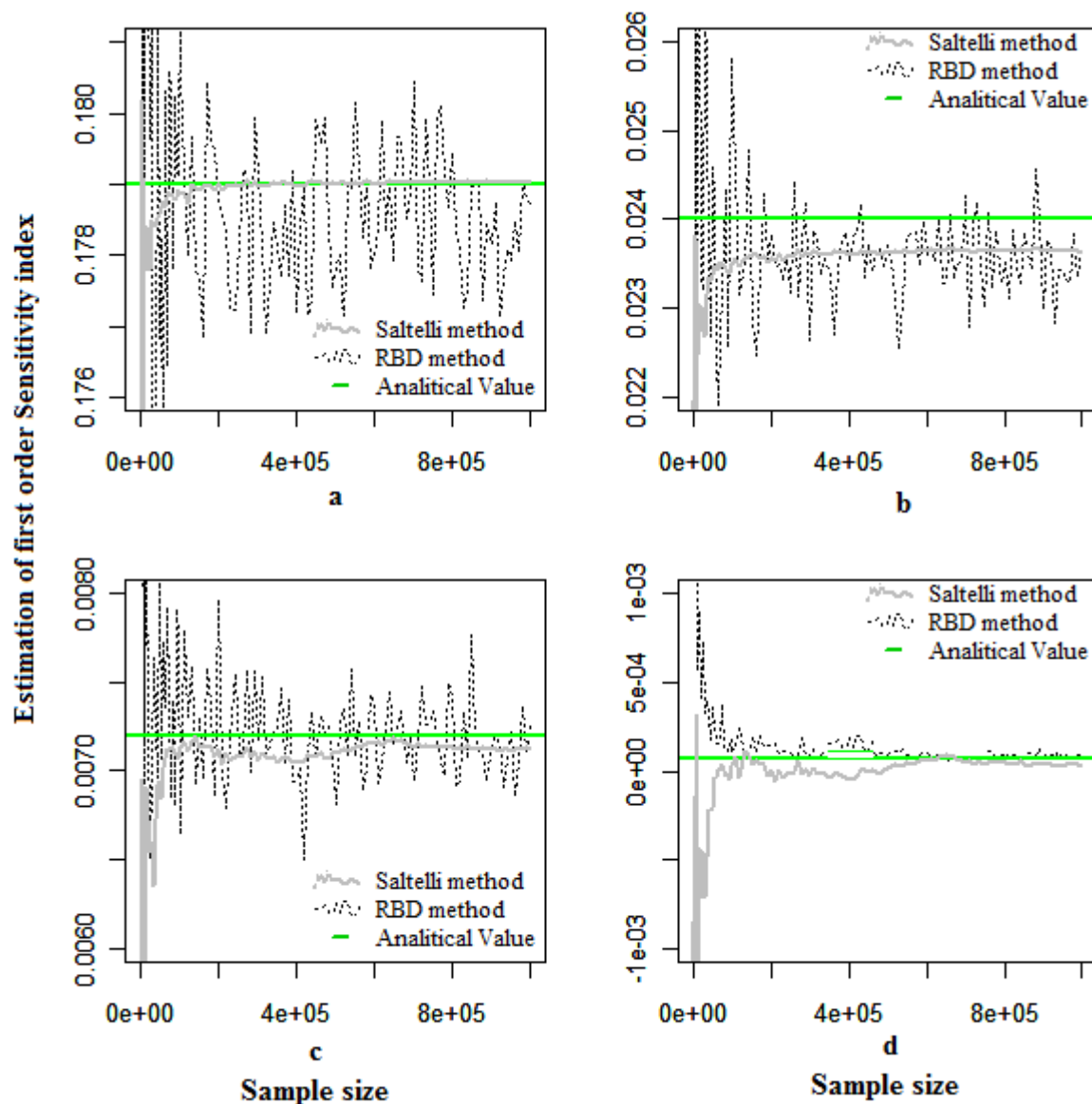
روش نیست بلکه مقصود، تنها مشاهده این دو روش در کنار یکدیگر است. از آنجایی که روش RBD قابلیت محاسبه اثر تام را ندارد، لذا تنها برآورد اثر اصلی متغیرهای ورودی در این دو روش مورد بررسی قرار گرفته است. در شکل (۶-۹) روند همگرایی برآورد اثر اصلی متغیر بسیار مهم X_1 نشان داده شده است.



شکل (۶ - ۹) - مقایسه روند همگرایی روش طرح متعادل تصادفی و روش سالتلی به مقادیر تحلیلی در متغیر X_1

واضح است که سرعت همگرایی روش سالتلی نسبت به روش طرح متعادل تصادفی بیشتر می‌باشد. اگر چه تفاوت‌ها ناچیز به نظر می‌رسد اما می‌توان چنین استنباط کرد که برآورد روش سالتلی نسبت به روش RBD با ثبات‌تر و قابل اعتمادتر است.

در شکل (۶-۱۰) روند همگرایی مقادیر برآورد برای متغیرهای مهم (پانل a)، نسبتاً مهم (پانل b)، کم اهمیت (پانل c) و فاقد اهمیت (پانل d) نشان داده شده است. سرعت همگرایی روش سالتلی در مقایسه با روش طرح متعادل تصادفی در این پانل‌ها به ترتیب سریع‌تر، هم رفتار، کندتر و کندتر است.



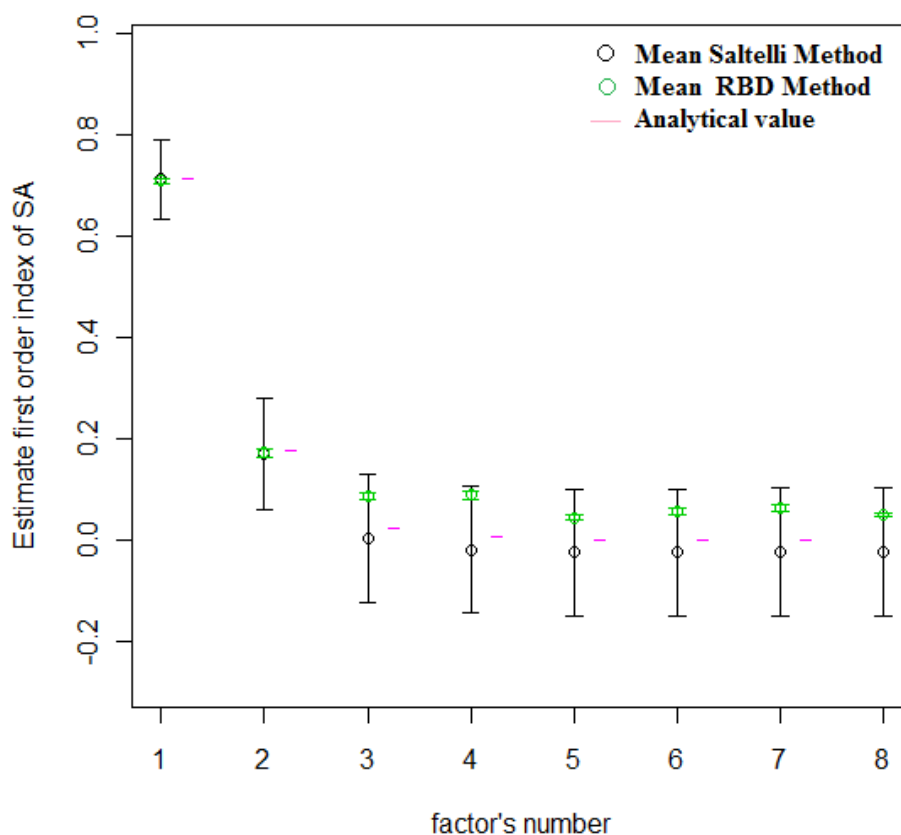
شکل (۶-۱۰) - مقایسه روند همگرایی روش RBD و روش سالتلی به مقادیر تحلیلی در متغیر X_2 (پانل a)، متغیر X_3

(پانل b)، متغیر X_4 (پانل c) و متغیر X_5 (پانل d)

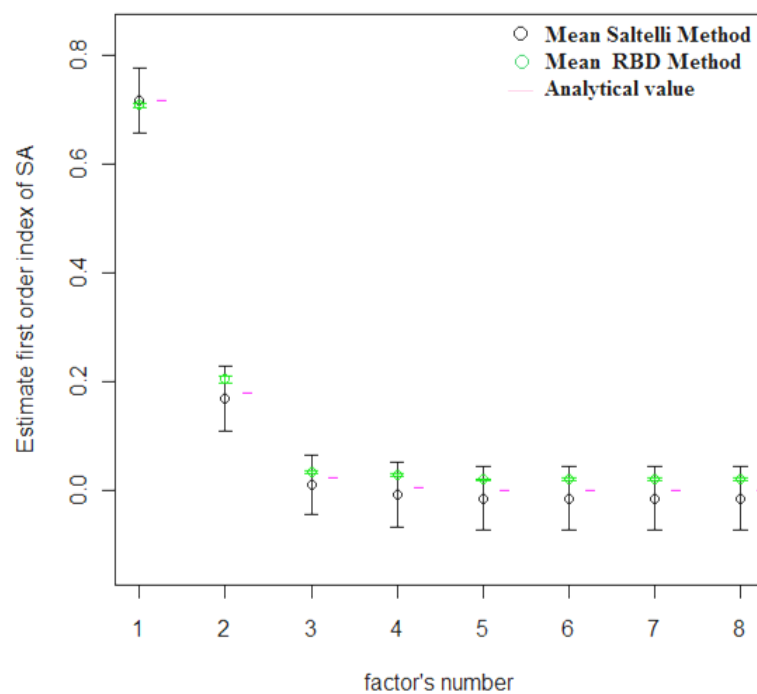
نتیجه کلی بر مبنای نمودارهای شکل (۶-۱۰) را می‌توان این‌گونه بیان نمود که در متغیرهای با اهمیت، همگرایی روش سالتلی نسبت به روش RBD بهتر و سریع‌تر است و این برتری با کم اهمیت‌شدن

متغیرهای ورودی کم‌رنگ‌تر شده به‌طوری که برای متغیرهای کم اهمیت یا فاقد اهمیت، روش طرح متعادل تصادفی ترجیح داده می‌شود.

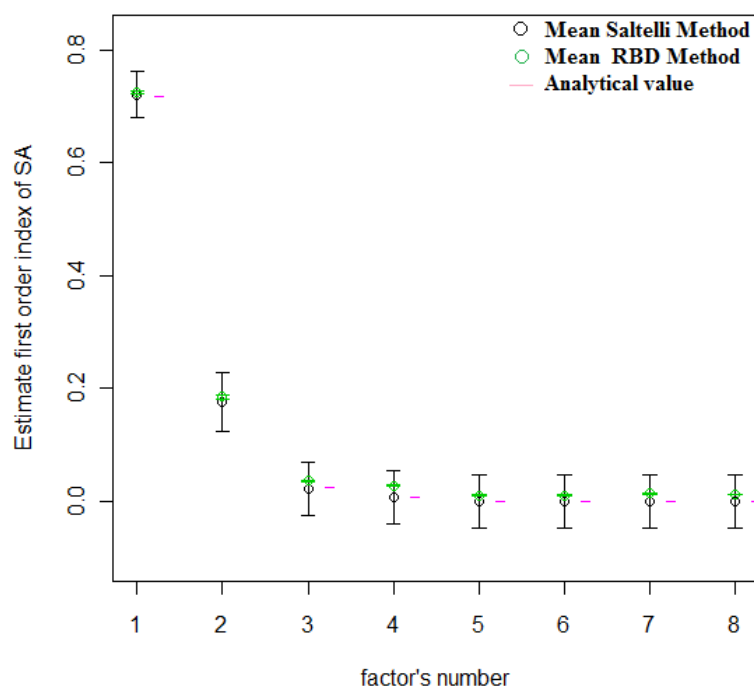
مقایسه دقت دو روش بیان شده با استفاده از فاصله اطمینان بوت‌استرپی و ۵۰ بار تکرار در نمونه-هایی به حجم ۵۰۰، ۱۰۰۰ و ۲۰۰۰ نیز انجام شده است. بر اساس هر یک از این حجم نمونه‌ها، نقاط آزمایش روش سالتلی و RBD به‌طور مجزا تولید شده و اثر اصلی متغیرهای ورودی مدل برآورد شده‌اند. نمودارهای (۶-۱۱) تا (۶-۱۳) خلاصه‌ای از نتایج را نشان می‌دهد.



شکل (۶-۱۱) - فاصله اطمینان میانگین برآورد اثر اصلی محاسبه شده با روش‌های سالتلی و RBD در حجم نمونه ۵۰۰



شکل (۶-۱۲)- فاصله اطمینان میانگین برآورد اثر اصلی محاسبه شده با روش های سالتلی و RBD در حجم نمونه ۱۰۰۰



شکل (۶-۱۲)- فاصله اطمینان میانگین برآورد اثر اصلی محاسبه شده با روش های سالتلی و RBD در حجم نمونه ۲۰۰۰

با استفاده از نمودارهای فوق می‌توان گفت، روش طرح متعادل تصادفی در برآورد شاخص حساسیت پراکندگی کمتری نسبت به روش سالتلی دارد. اما فاصله اطمینان روش سالتلی نسبت به روش طرح متعادل تصادفی بزرگتر و دربرگیرنده مقدار حقیقی اثر اصلی است. در حجم نمونه ۲۰۰۰ برآوردهای روش طرح متعادل تصادفی به مقادیر حقیقی بسیار نزدیک بوده و پراکندگی بسیار کمی دارد. به منظور درک و امکان مقایسه بهتر مقادیر برآوردی این دو روش، میانگین برآورد اثر اصلی در روش سالتلی و RBD به همراه مقدار حقیقی شاخص در جدول (۳-۶) آمده است.

شکل (۳-۶) - میانگین برآورد شاخص حساسیت مرتبه اول در روش سالتلی و RBD با حجم نمونه مختلف

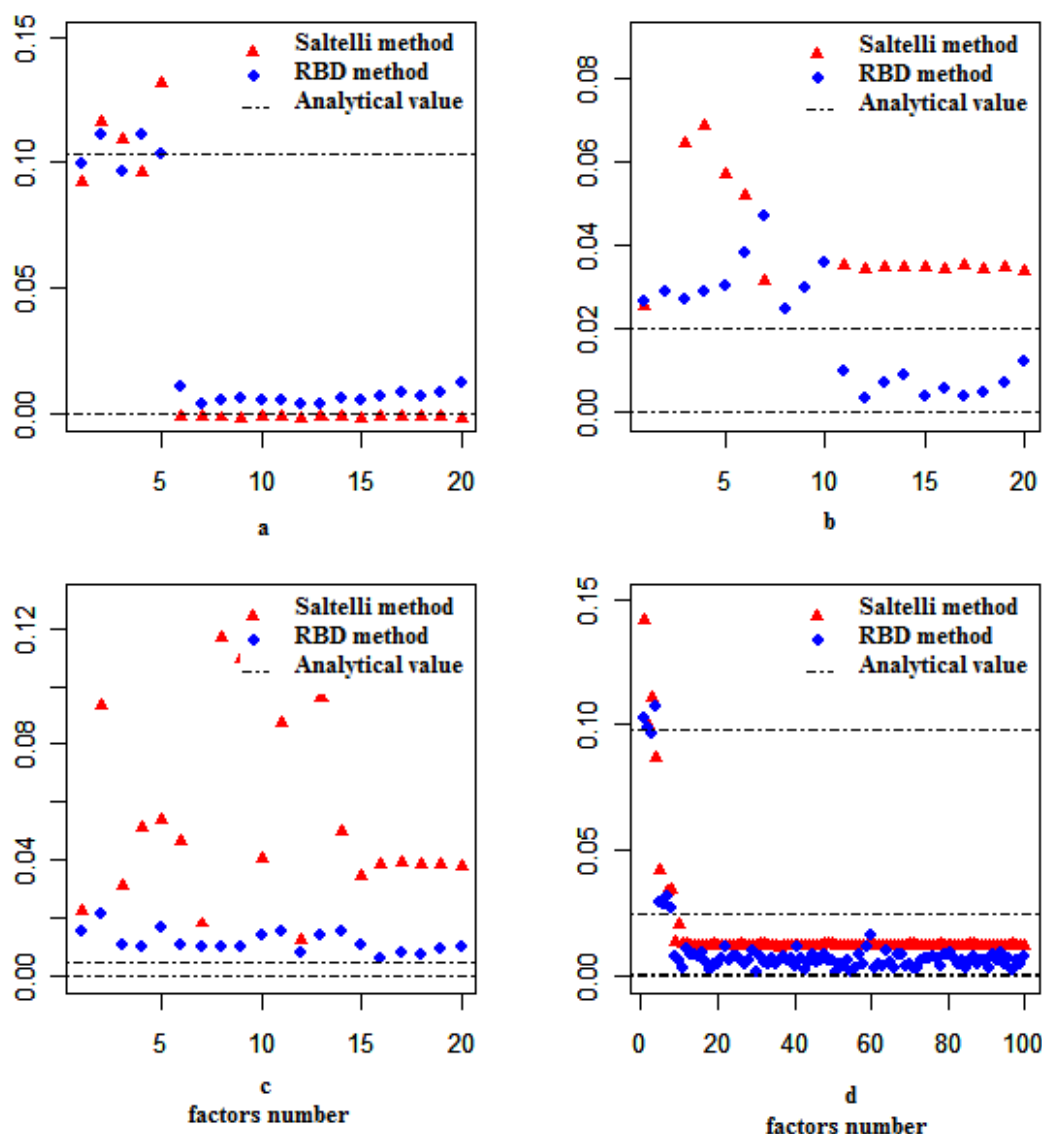
RBD method	Saltelli method	RBD method	Saltelli method	RBD method	Saltelli method	مقدار حقیقی
حجم نمونه ۵۰۰		حجم نمونه ۱۰۰۰		حجم نمونه ۲۰۰۰		
0.737932	0.713522	0.728716	0.716377	0.716542	0.72023	7.16E-01
0.214784	0.171169	0.193885	0.169216	0.159701	0.17524	1.79E-01
0.049696	0.004782	0.015162	0.010753	0.021668	0.02195	2.37E-02
0.012913	-0.01849	0.013037	-0.00773	0.01846	0.006955	7.16E-03
0.031442	-0.02415	0.011146	-0.01424	0.004724	-0.00085	7.16E-05
0.029982	-0.02385	0.014302	-0.01387	0.004577	-0.00101	7.16E-05
0.011231	-0.02372	0.007101	-0.0138	0.005799	-0.00083	7.16E-05
0.045492	-0.02363	0.015406	-0.01371	0.010557	-0.00094	7.16E-05

همان‌طور که در جدول (۳-۶) مشهود است میانگین برآوردهای روش طرح متعادل تصادفی و روش سالتلی با افزایش حجم نمونه به مقادیر حقیقی بسیار نزدیک می‌شوند. شایان ذکر است که نمونه‌های

انتخابی در روش طرح متعادل تصادفی به صورت تصادفی برگزیده شده‌اند. بنابراین ممکن است با نمونه-گیری مجدد نتایج حاصل تغییر کند.

به منظور بررسی بیشتر این دو روش، ابتدا تعداد متغیرهای ورودی تابع سبل را به ۲۰ افزایش داده و با حجم نمونه ۲۰۰۰، درجه اهمیت آنها را تغییر دادیم. در پانل‌های a ، b و c در شکل (۶-۱۴) تعداد متغیرهای بسیار با اهمیت ($a_i=0$) را بترتیب برابر با ۵، ۱۰ و ۱۵ و مابقی متغیرها را بی‌اهمیت ($a_i=99$) در نظر گرفته‌ایم. سپس فضای متغیرهای ورودی مدل را به ۱۰۰ افزایش داده و ۴ متغیر اول را بسیار بااهمیت، ۴ متغیر دوم را بااهمیت، ۲ متغیر بعدی را نسبتاً بااهمیت و مابقی متغیرهای ورودی را بی‌اهمیت تعریف نموده و روش‌های برآورد سالتلی و طرح متعادل تصادفی را اعمال کرده‌ایم. نتایج در پانل (d) نشان داده شده است.

با توجه به پانل (a) می‌توان گفت که روش طرح متعادل تصادفی در برآورد اثر اصلی متغیرهای بااهمیت، بهتر از روش سالتلی عمل کرده است و برعکس، روش سالتلی در برآورد اثرات اصلی متغیرهای بی‌اهمیت موفق‌تر بوده است. در پانل‌های (b) و (c) و (d) برتری روش طرح متعادل تصادفی در برآورد اثر اصلی متغیرهای بااهمیت و بی‌اهمیت نسبت به روش سالتلی مشهود است. در نهایت می‌توان گفت در حجم نمونه ۲۰۰۰، روش طرح متعادل تصادفی عملکرد بهتری نسبت به روش سالتلی داشته است.



شکل (۶ - ۱۳) - مقایسه مقادیر واقعی با مقدار برآورد شده با روش سالتلی و RBD با متغیرهای با درجه اهمیت مختلف

بنابر نتایج فوق انتخاب یک روش ایده‌آل برآورد شاخص یا شاخص‌های تحلیل حساسیت بسیار مشکل و حتی غیرممکن است. چراکه در تابع تحلیلی سبل در برخی از موارد روش سالتلی عملکرد بهتری داشته است و در موارد دیگر روش طرح متعادل تصادفی نسبت به روش سالتلی برتری داشته است. بنابراین به-آسانی نمی‌توان یک روش برآورد شاخص‌های حساسیت را به‌عنوان یک روش ایده‌آل مطرح کرد. در این-گونه موارد، هزینه‌های محاسباتی بسیار حائز اهمیت است.

فصل هفتم

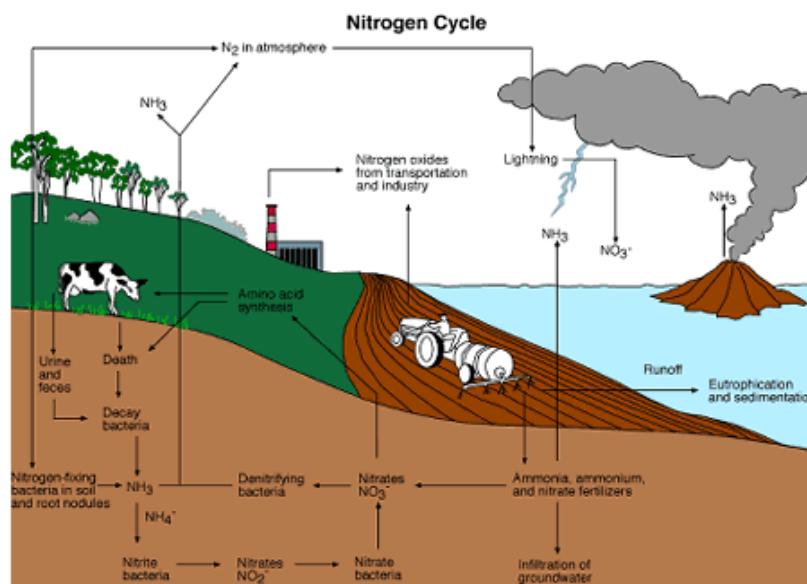
تحلیل حساسیت مدل $INCA-N$ با استفاده از

روش‌های

مونته کارلو - مبنای سالتلی و طرح متعادل تصادفی

مقدمه

گیاهان از نیتروژن برای سنتز پروتئین‌های لازم برای رشد خود استفاده می‌کنند و این ماده را با جذب مستقیم از زمین و به شکل نیترات (NO_3) یا در ترکیب با سایر مواد به دست می‌آورند. نیترات یکی از مشتقات نیتروژن است که در طی فرآیندهای تبدیلات نیتروژن تولید می‌شود. این ماده از طریق آب و مواد غذایی وارد بدن انسان شده و افزایش سطح آن موجب می‌شود که به ترکیبات دیگری چون نیتريت و نیتروسامین تبدیل شود که این ترکیبات، سمی بوده و برای سلامت بدن خطرناکند. در رژیم غذایی نوزادان و کودکان، سطح نیترات یک عامل بسیار مهم بوده به‌طوری که وجود بیش از حد این ترکیب‌های نیتروژن‌دار می‌تواند مانعی جدی انتقال اکسیژن در خون باشد. نیترات یکی از موادی است که به وفور در آب زیرزمینی مناطق روستایی یافت شده و می‌تواند سبب بروز بیماری «متاهموگلوبین» یا «بیماری کودکان آبی» شود. همچنین مقادیر بحرانی غلظت نیترات باعث بروز بیماری‌هایی مانند سرطان، ناقص الخلقه بودن نوزادان، بزرگ شدن تیروئید و التهاب غدد لنفاوی می‌شود. این آلاینده از طریق فاضلاب‌های شهری و صنعتی، تخلیه آب‌های شستشوی زمین‌های کشاورزی در رودخانه‌ها و از همه مهمتر اکسیداسیون مواد آلی ازت‌دار نظیر پروتئین‌ها وارد منابع آب می‌شود. نیترات به‌طور مستقیم و غیرمستقیم در اثر تجزیه و تغییرهای زیست - شیمیایی ترکیب‌های مختلف معدنی و آلی در خاک به وجود می‌آید و به راحتی توسط آب باران به منابع آب سطحی و زیر زمینی نفوذ می‌کند. شکل (۷-۱) چرخه نیتروژن و نحوه پدیدآمدن نیترات را نشان می‌دهد.



شکل (۷ - ۱) - چرخه نیتروژن در طبیعت

با توجه به اهمیت آب آشامیدنی و تاثیر مستقیم آن بر سلامت شهروندان، موضوع کنترل آلودگی آب و پیشگیری از افزایش نیترات به سطوح بحرانی، امری اجتناب ناپذیر است. همچنین پیچیدگی طبیعی فرآیند چرخه نیتروژن سبب شده است که انجام برخی آزمایشات فیزیکی برای شناسایی و کنترل عوامل تاثیرگذار بر آلودگی آب بنا به دلایلی نظیر داشتن هزینه‌های گراف مالی و زمانی و عدم امکان اجرا، مقدور نباشد. لذا محققین علوم مختلف به شبیه‌سازی این پدیده با مدل‌های ریاضی روی آورده‌اند. تاکنون مدل‌های متعددی برای برآورد میزان بار نیتروژن ورودی به رودخانه‌ها از طریق حوضه‌های آبریز، بسط و توسعه یافته است. به‌طور کلی این مدل‌ها تابعی از میزان بارش منطقه و مشخصات حوضه آبریز نظیر نوع خاک، عوامل شهری، رسوبات حاصل از منابع برون رفت، توپوگرافی و هیدرولوژی و غیره است. یکی از این مدل‌ها که چرخه نیتروژن و تبدیلات آن در مسیر اصلی رودخانه را شبیه‌سازی می‌کند، مدلی با نام اختصاری *INCA-N*^۱ است.

^۱ Integrated Nitrogen in Catchments

۷-۱ معرفی اجمالی مدل *INCA-N*

این مدل فرآیند-مبنا^۱ است و با استفاده از معادلات جنبشی واکنش‌ها، شبیه‌سازی سازوکارهای اصلی عملیاتی رودخانه را انجام می‌دهد. مهمترین ویژگی‌های این الگو پویایی و نیمه-توزیعی^۲ است (وایتهد^۳ و همکاران، ۱۹۹۸). این مدل تعیینی میزان جریان آب و نیتروژن را در یک حوضه آبریز تعیین می‌کند. همچنین عوامل و فرآیندهای تاثیرگذار بر میزان نترات (NO_3) و آمونیاک (NH_4) ذخیره شده در خاک و سیستم‌های آب زیر زمینی را شبیه‌سازی نموده و خروجی را با استفاده از این سیستم در یک مدل رودخانه چند ناحیه ای تولید می‌کند (وید^۴، ۲۰۰۲). ساختار این مدل در شکل (۷-۲) نشان داده شده است. مدل *INCA-N* از پنج قسمت رابط *GIS*، مدل‌های نیتروژن ورودی، هیدرولوژیکی، فرآیندهای نیتروژن حوضه آبریز و عملکردهای نیتروژن رودخانه تشکیل شده است. این مدل دارای قابلیت‌های زیاد و خروجی‌های متنوعی است که شرح مفصل آن در ضمیمه پایان نامه آمده است.

۷-۲ معرفی ورودی‌ها و خروجی‌های مدل

اطلاعات ورودی این مدل شامل عوامل طبیعی (دما، رطوبت و بارش روزانه)، عامل دخالت انسانی (نظیر رژیم کوددهی زمین‌های کشاورزی) و پارامترهای ورودی مدل است. پارامترهای ورودی در سه دسته تقسیم بندی شده و مشخصات آن‌ها در جدول (۷-۱) خلاصه گردیده است.

^۱ Process-based

^۲ Semi distributed

^۳ Whitehead

^۴ Wade

جدول (۷ - ۱) - مشخصات پارامترهای ورودی مدل *INCA-N*

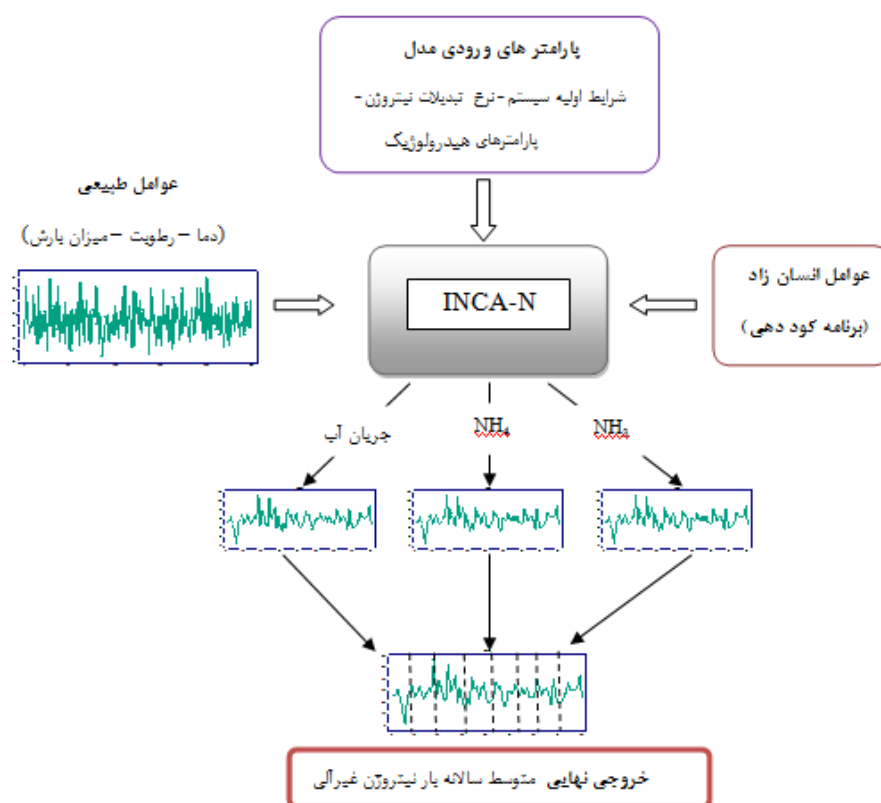
<i>Group</i>	<i>Variable name</i>	<i>Unit</i>	<i>Range</i>
<i>Initial conditions</i>	<i>X₁: Surface flow</i>	<i>m³/s</i>	<i>[0,0.01]*</i>
	<i>X₂: Surface nitrate</i>	<i>mg N/l</i>	<i>[0,10]</i>
	<i>X₃: Surface ammonium</i>	<i>mg N/l</i>	<i>[0,2]</i>
	<i>X₄: Surface drainage volume</i>	<i>m³</i>	<i>[10⁵,2×10⁷]*</i>
	<i>X₅: Sub-surface flow</i>	<i>m³/s</i>	<i>[0,0.01]*</i>
	<i>X₆: Sub-surface nitrate</i>	<i>mg N/l</i>	<i>[0,10]</i>
	<i>X₇: Sub-surface ammonium</i>	<i>mg N/l</i>	<i>[0,2]</i>
	<i>X₈: Sub-surface drainage volume</i>	<i>m³</i>	<i>[10⁵,2×10⁷]*</i>
<i>Nitrogen transformation rates</i>	<i>X₉: Denitrification</i>	<i>mol/day</i>	<i>[0 , 0.01]</i>
	<i>X₁₀: Nitrogen fixation</i>	<i>kg N/ha/day</i>	<i>[0 , 0.01]</i>
	<i>X₁₁: Plant nitrate uptake</i>	<i>mol/day</i>	<i>[0 , 0.05]</i>
	<i>X₁₂: Maximum nitrate uptake</i>	<i>kg N/ha/year</i>	<i>[80 ,140]</i>
	<i>X₁₃: Mineralization</i>	<i>kg N/ha/day</i>	<i>[0,1]</i>
	<i>X₁₄: Immobilization</i>	<i>mol/day</i>	<i>[0 , 0.1]</i>
	<i>X₁₅: Plant ammonium uptake</i>	<i>mol/day</i>	<i>[0 , 0.05]</i>
<i>Hydrogeological parameters</i>	<i>X₁₆: Soil moisture deficit</i>	<i>Mm</i>	<i>[100,170]</i>
	<i>X₁₇: Soil water residence time</i>	<i>Day</i>	<i>[0.5 , 5]</i>
	<i>X₁₈: Groundwater residence time</i>	<i>Day</i>	<i>[10 , 200]</i>
	<i>X₁₉: Maximum soil retention volume</i>	<i>M</i>	<i>[0.01 , 1]</i>

* اشاره به زیر حوضه مورد مطالعه به وسعت ۱ کیلومترمربع از زمین های کشاورزی

خروجی این مدل شامل سری‌های زمانی برآورد روزانه میزان جریان آب و غلظت‌های نیترات و آمونیاک در ایستگاه‌های مختلفی است که در طول رودخانه اصلی قرار دارند. این سه سری زمانی توسط رابطه زیر ترکیب شده و میزان بار نیتروژن غیرآلی روزانه را نتیجه می‌دهد.

$$(\text{میزان غلظت آمونیوم} + \text{میزان غلظت نیترات}) \times \text{میزان جریان آب}$$

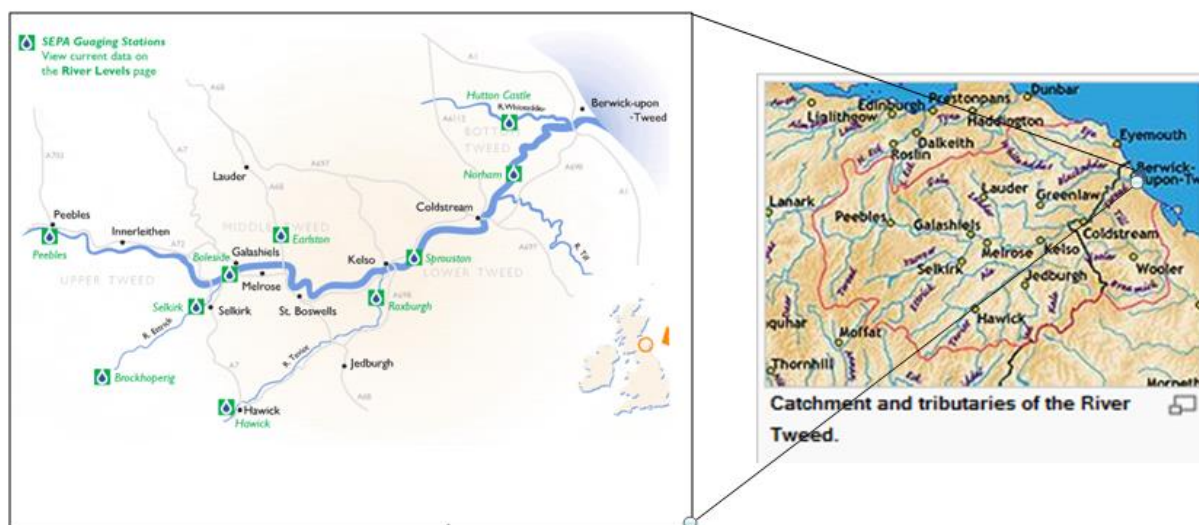
از آن‌جا که ترکیب این سه خروجی، خود نیز یک سری زمانی است لذا خروجی نهایی به صورت تک مقداری و تحت عنوان "متوسط سالانه بار ورودی نیتروژن غیر آلی به رودخانه" با معدل‌گیری از متوسط های سالیانه محاسبه شده است. شکل (۷-۲) شمای کلی مدل *INCA-N* را نشان می‌دهد.



شکل (۷ - ۲) - ساختار شماتیکی مدل *INCA-N*

۷ - ۳ منطقه مورد مطالعه

منطقه مورد مطالعه، حوزه آبریز *Tweed* به مساحت ۴۳۹۰ کیلومتر مربع با رودخانه اصلی به طول ۱۵۶ کیلومتر واقع در کشور انگلستان است. موقعیت جغرافیایی آن از عرض شمالی ۵۵° تا ۵۸° و طول شرقی ۰° تا ۱۰° است. شکل (۷-۳) حوزه آبریز رودخانه *Tweed* را نشان می‌دهد. داده‌های آب و هواشناسی شامل دما، رطوبت، بارش روزانه و دبی روزانه ایستگاه‌های واقع در طول رودخانه اصلی به مدت هفت سال (۱۹۹۴-۲۰۰۰) اندازه‌گیری و ثبت شده‌اند (جاروی^۱ و همکاران، ۲۰۰۲). جهت سهولت در تحلیل نتایج، تنها یک زیر مجموعه از حوزه آبریز به وسعت یک کیلومتر مربع از زمین‌های کشاورزی را مورد بررسی قرار گرفته است.



شکل (۷ - ۳) - حوزه آبریز رودخانه Tweed

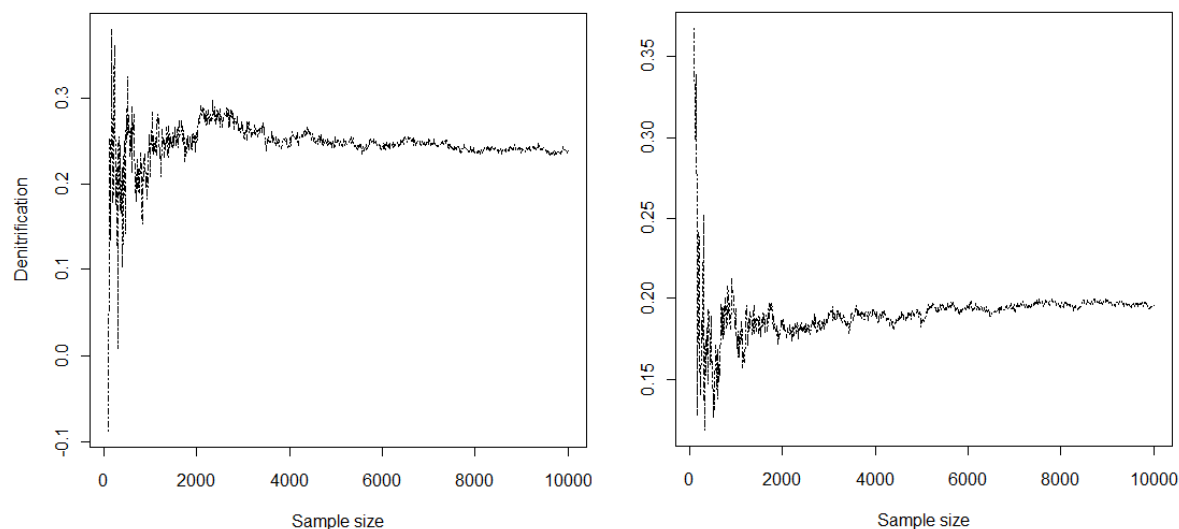
^۱ Jarvie

در این مطالعه، مدل $INCA-N$ سامانه‌ای از پارامترهای نرخ تبادلات نیتروژن ($X_9 - X_{15}$) به‌عنوان ورودی مدل، در نظر گرفته شده است. سایر پارامترهای ورودی در مقادیر میانگین دامنه تغییرات آنها ثابت نگه داشته شده‌اند. به عبارت دیگر، مدل $INCA-N$ را تابعی هفت متغیره در نظر گرفته‌ایم.

۷ - ۴ تحلیل حساسیت مدل $INCA-N$ با روش سالتلی

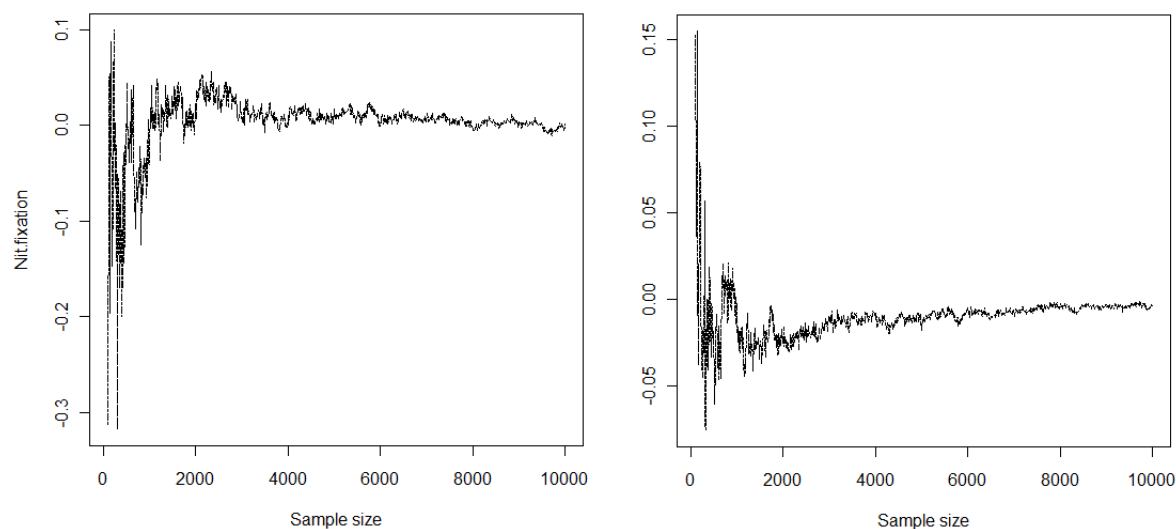
برای تحلیل حساسیت مدل $INCA-N$ با روش سالتلی حجم نمونه اولیه را برابر ۱۰۰۰۰ در نظر گرفته‌ایم که هزینه محاسبات برابر $N \times (k + 2) = 10000 \times 9$ است (بخش ۴-۴). همان‌طور که در معرفی این روش در فصل چهارم ذکر شد، نقاط آزمایش روش سالتلی با استفاده از دنباله شبه تصادفی تولید می‌شوند. در این روش، نقاط آزمایش متغیرهای ورودی به صورت غیر تصادفی انتخاب شده اما همانند اعداد تصادفی، فضای متغیرهای ورودی به طور یکنواخت پوشش داده می‌شود. در واقع تولید نقاط آزمایش در دنباله شبه تصادفی، بر اساس الگوریتم تعیینی صورت می‌گیرد. تعیینی بودن نقاط آزمایش سبب می‌شود تا اگر نمونه‌ای با حجم معین N تولید شود و دوباره بخواهیم نمونه‌ای جدید با حجم N' داشته باشیم که $N' > N$. آن‌گاه در نمونه جدید، دقیقاً همان نقاط آزمایش قبلی تولید شده و یک سری نقاط نیز به نقاط آزمایش قبلی اضافه می‌شوند. همچنین با کاهش حجم نمونه نیز یک سری نقاط از انتهای مجموعه داده قبلی حذف می‌شوند. این ویژگی دنباله شبه تصادفی سبب می‌شود تا بتوان بدون اجرای اضافی مدل، شاخص‌های حساسیت را در حجم نمونه‌های کمتر از حجم نمونه اولیه برآورد نمود. از آن جایی که حجم نمونه را برابر ۱۰۰۰۰ در نظر گرفته‌ایم، لذا برآورد شاخص‌های حساسیت در حجم نمونه‌های کمتر از ۱۰۰۰۰، بدون نیاز به تولید نقاط جدید، قابل محاسبه است. در این صورت، روند همگرایی مقادیر برآورد- شده نیز قابل رویت خواهد بود. بررسی روند همگرایی، با انتخاب دنباله اعداد ۱۰۰ تا ۱۰۰۰۰ به‌عنوان

حجم نمونه‌های روش سالتلی صورت گرفته است. اشکال (۷-۴) تا (۷-۱۰) روند همگرایی اثر اصلی و اثر تام در متغیرهای مورد بررسی مدل *INCA-N* را نشان می‌دهد.



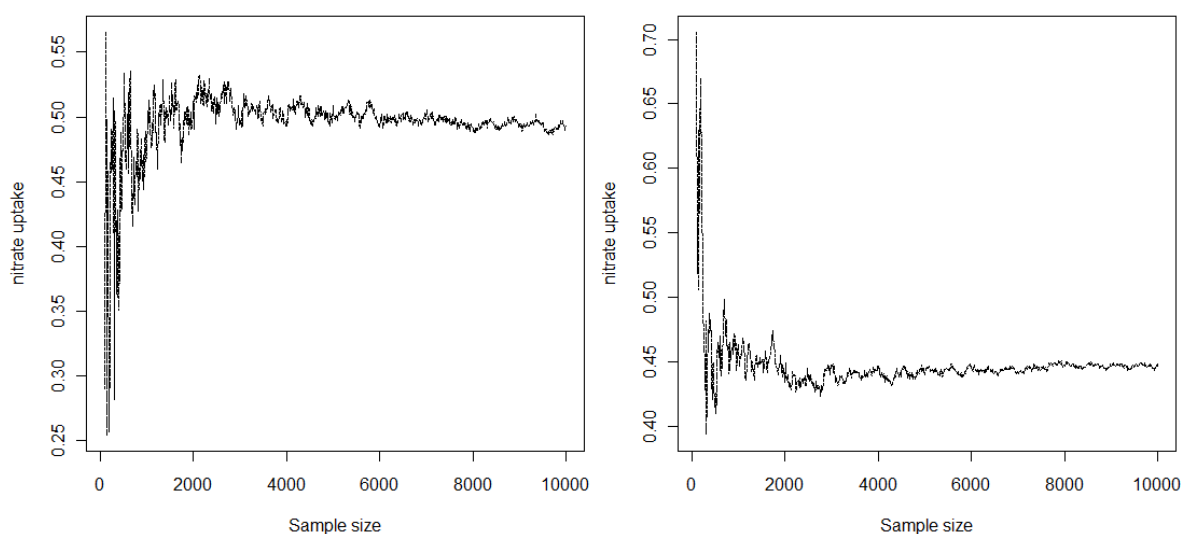
شکل (۷-۴) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) متغیر نیترات زدایی

با توجه به شکل (۷-۴) به نظر می‌رسد که برآورد اثر اصلی و اثر تام این متغیر در حجم نمونه بیشتر از ۴۰۰۰ به پایداری رسیده است. در واقع حجم نمونه مناسب برای برآورد شاخص‌های حساسیت این متغیر برابر ۴۰۰۰ است. بنابر مقدار همگرایی اثر تام که حدود ۰/۲۵ و اثر اصلی که تقریباً برابر با ۰/۲۰ می‌توان نتیجه گرفت که متغیر نیترات سازی (X_9) یک عامل تاثیرگذار بر نیتروژن ورودی به آب رودخانه است.



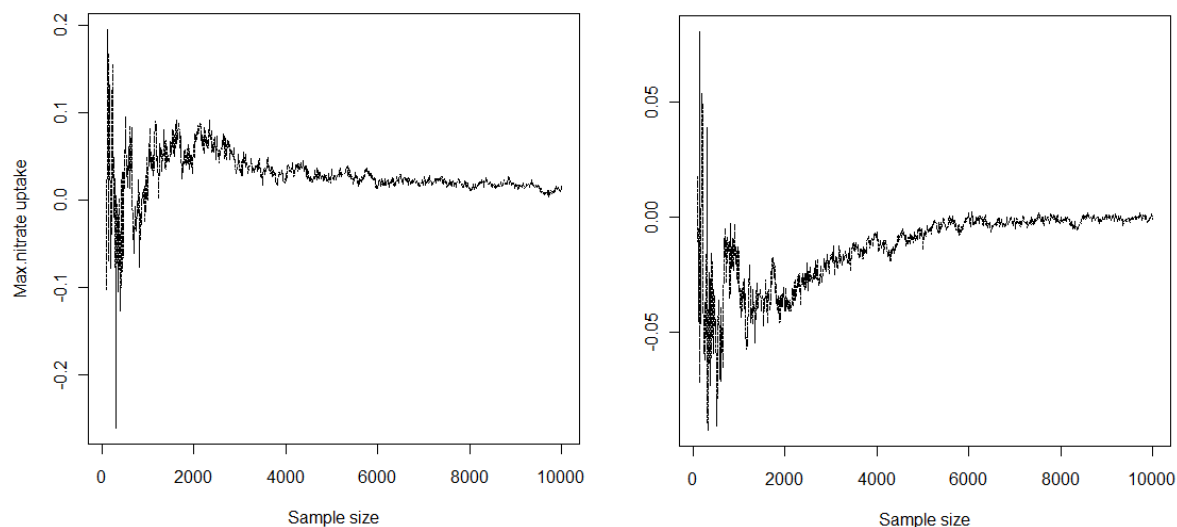
شکل (۷ - ۵)- روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر ثابت نیتروژن هوا

برآورد اثر اصلی و اثر تام با توجه به شکل (۷-۵) به مقدار صفر گرایش دارد. این بدان مفهوم است که عامل ثابت نیتروژن هوا (X_{10})، یک متغیر بی‌تاثیر در واریانس خروجی مدل است. پایداری برآورد شاخص‌های حساسیت در این متغیر پس از حجم نمونه ۴۰۰۰ رخ می‌دهد.



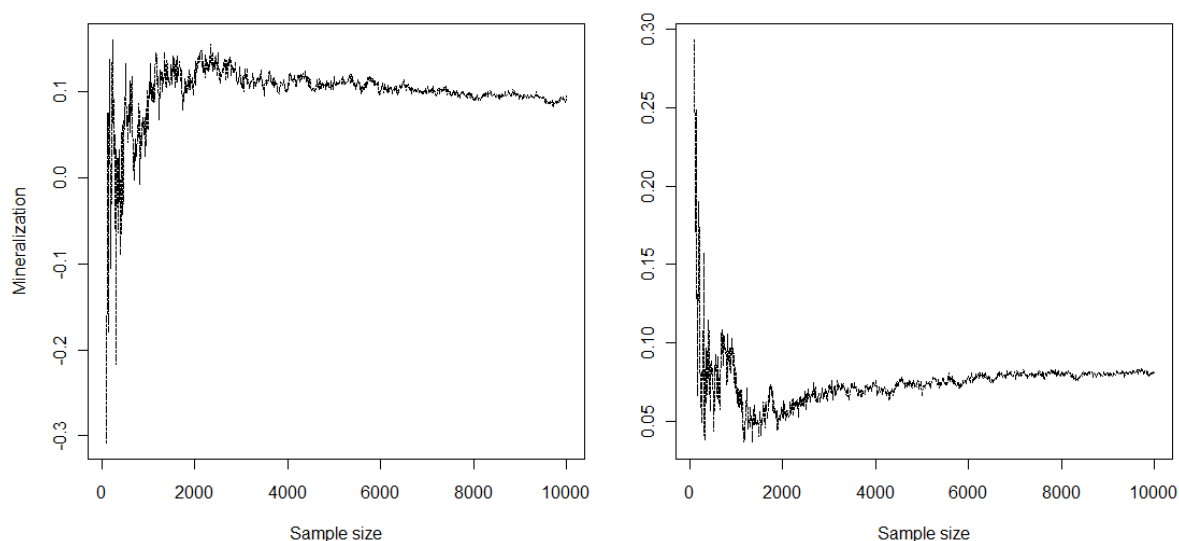
شکل (۷ - ۶)- روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر جذب نترات گیاهان

برآورد اثر اصلی و اثر تام متغیر میزان جذب نیترات توسط گیاهان (X_{11}) که در شکل (۶-۷) آمده است نیز در حجم نمونه ۴۰۰۰ پایدار می‌شود. مقدار همگرایی اثر تام که حدود ۰/۵۰ و اثر اصلی که تقریباً برابر با ۰/۴۵ می‌باشند، حاکی از موثر بودن این متغیر در واریانس خروجی مدل است..



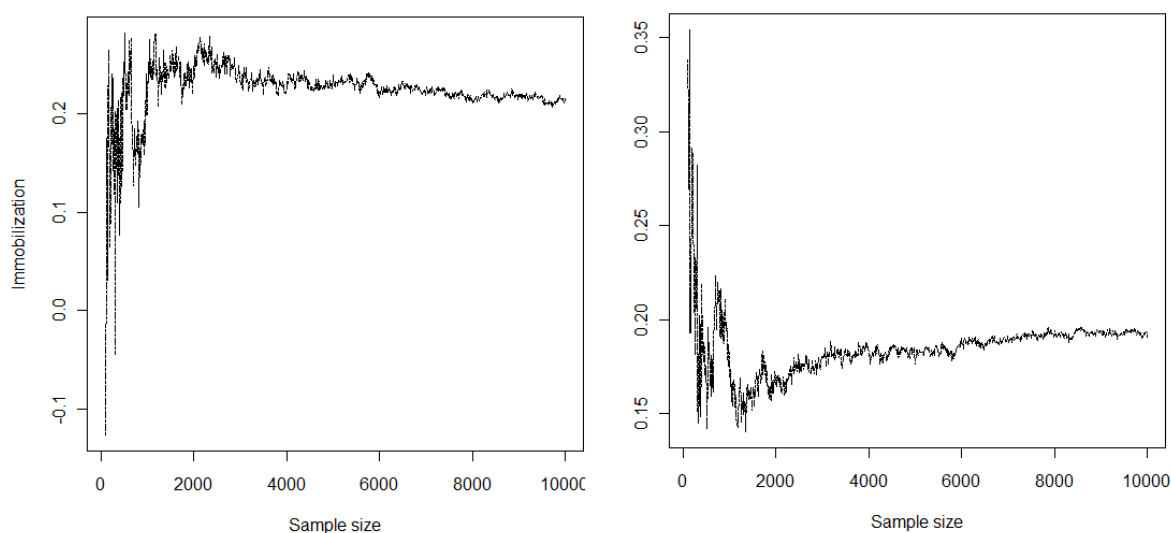
شکل (۷ - ۷) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر بیشترین میزان جذب

با توجه به شکل (۷-۷)، روند صعودی به سمت صفر در برآورد اثر اصلی و اثر تام این متغیر مشاهده می‌شود. این بدان معنی است که متغیر "بیشترین جذب نیترات توسط گیاهان" در خروجی مدل بی‌تاثیر است. برآورد اثر تام نیز رفتاری مشابه اثر اصلی دارد و پس از حجم نمونه ۴۰۰۰ به پایداری می‌رسد.



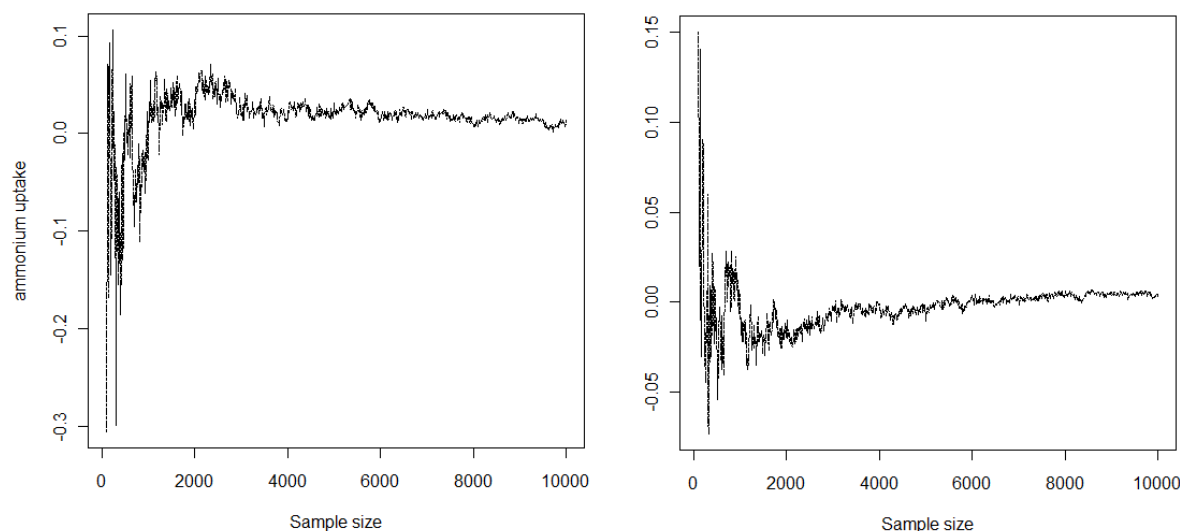
شکل (۷ - ۸) - روند همگرایی اثر اصلی (سمت راست) و اثر تام (سمت چپ) در متغیر معدنی‌سازی

پایدار شدن برآورد اثر اصلی و اثر تام در متغیر معدنی‌سازی (X_{13}) نیز پس از حجم نمونه ۴۰۰۰ اتفاق می‌افتد. روند همگرایی مقادیر برآوردی اثر اصلی و اثر تام در این متغیر به ترتیب تقریباً برابر با ۰/۰۸ و ۰/۱۰ است که نشان از تاثیرگذار بودن آن در واریانس خروجی دارد (شکل ۷-۸).



شکل (۷ - ۹) - روند همگرایی اثر اصلی (راست) و اثر تام (چپ) در متغیر آلی‌سازی

با توجه به شکل (۷-۹) به نظر می‌رسد که برآورد اثرهای اصلی و تام این متغیر نیز همانند بسیاری از متغیرهای مورد بررسی پس از حجم نمونه ۴۰۰۰ پایدار می‌شود. با توجه به مقادیر پایایی اثر اصلی (حدوداً برابر با ۰/۱۸) و اثر تام (تقریباً برابر با ۰/۲۰) نتیجه می‌گیریم عامل آلی‌سازی نیز یک متغیر مهم در واریانس خروجی مدل محسوب می‌شود.



شکل (۷ - ۱۰) - روند همگرایی اثر اصلی (راست) و اثر تام (چپ) در متغیر جذب آمونیاک

برآورد اثر اصلی و اثر تام متغیر جذب آمونیاک (X_{15}) از حجم نمونه ۴۰۰۰ به بعد، پایدار شده و به صفر میل می‌کنند (شکل ۷-۱۰). در نتیجه این متغیر یک عامل بی‌تاثیر در مدل است.

نتایج تحلیل حساسیت با حجم نمونه بهینه

از آنجایی که برآورد شاخص‌های حساسیت اکثر متغیرهای ورودی مدل در حجم نمونه ۴۰۰۰ پایدار شده است، حجم نمونه مناسب را برابر با ۴۰۰۰ در نظر گرفته و اثرات اصلی و تام را با جزئیات بیشتر مورد بررسی قرار می‌دهیم. جهت بررسی عدم قطعیت برآوردها نیز این فرآیند را ۵۰ با تکرار نموده که نتایج

حاصل در جدول (۲-۷) و به صورت فاصله اطمینان بوت استرپی در شکل‌های (۷-۱۱) و (۷-۱۲) آمده است.

جدول (۷ - ۲) - کران‌های بالا و پایین متغیرهای مدل *INCA-N* در حجم نمونه ۴۰۰۰

رتبه اهمیت متغیر	متغیر	اثر اصلی			اثر تام		
		کران پایین	میانگین	کران بالا	کران پایین	میانگین	کران بالا
۲	<i>Denitrification</i> (X_9)	0.070102	0.189265	0.308427	-0.0145	0.248923	0.512344
۷	<i>Nit.fixation</i> (X_{10})	-0.14782	-0.01067	0.126482	-0.3051	0.004372	0.313841
۱	<i>nitrate uptake</i> (X_{11})	0.32303	0.445347	0.567663	0.249247	0.49614	0.743033
۶	<i>Max.nitrate uptake</i> (X_{12})	-0.14261	-0.01006	0.122494	-0.27655	0.026633	0.329812
۴	<i>Mineralization</i> (X_{13})	-0.06845	0.072501	0.213456	-0.19377	0.105957	0.405681
۳	<i>Immobilization</i> (X_{14})	0.060969	0.18262	0.304271	-0.04722	0.227279	0.501779
۵	<i>ammonium uptake</i> (X_{15})	-0.14045	-0.00397	0.132519	-0.28743	0.018967	0.325364

با توجه به میانگین مقادیر برآورد شده اثر اصلی و اثر تام در جدول (۲-۷) مهمترین پارامترهای تاثیرگذار بر متوسط بار نیتروژن سالانه به ترتیب عبارتند از :

۱. میزان جذب توسط گیاهان (*Plant nitrate uptake*) با میانگین اثر اصلی 0.445 و اثر تام 0.496

۲. نرخ نیترات زدایی (*Denitrification*) با میانگین اثر اصلی 0.189 و اثر تام 0.249

۳. آلی سازی (*immobilization*) با میانگین اثر اصلی 0.183 و اثر تام 0.227

۴. نرخ معدنی سازی (*mineralization*) با میانگین اثر اصلی 0.073 و اثر تام 0.106

میزان اثر سایر متغیرهای الگو بسیار ناچیز است و می‌توان از اثر آن‌ها صرف نظر نمود. فواصل اطمینان اثر اصلی و اثر تام متغیرهای مدل نیز جهت درک بهتر این مسئله ترسیم شده است.

بنابر ویژگی‌های شاخص‌های حساسیت که در فصل سوم بیان شد، اثرات متقابل بین این متغیرها به صورت زیر می‌باشد.

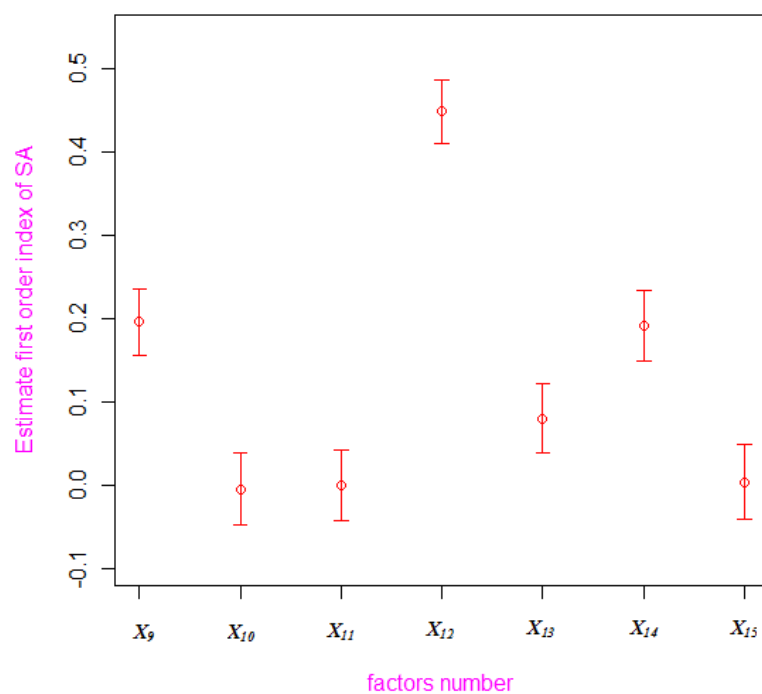
۱. اثرات متقابل متغیر میزان جذب توسط گیاهان با سایر عوامل برابر $0/051$ است.

۲. میزان اثر متقابل بین نرخ نیترات زدایی و سایر عوامل برابر $0/04$ است.

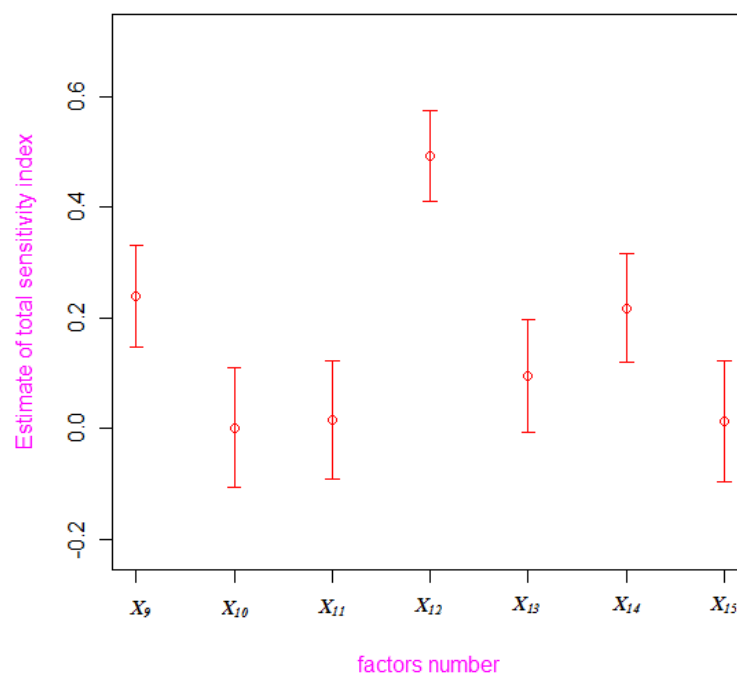
۳. میزان اثر متقابل متغیر آلی سازی با سایر عوامل برابر $0/044$ است.

۴. میزان اثر متقابل بین نرخ معدنی‌سازی و سایر عوامل برابر $0/023$ است.

همچنین مجموع اثرات اصلی تمام متغیرهای مدل تقریباً برابر با $0/90$ است. این بدان مفهوم است مدل *INCA-N* در این سامانه هفت متغیره، به‌طور کامل جمعی نبوده و اثرات متقابل بین این متغیرها وجود دارد. اما میزان آن‌ها چندان زیاد نمی‌باشد.



شکل (۷ - ۱۱) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل $INCA-N$ در نمونه به حجم ۴۰۰۰



شکل (۷ - ۱۲) - فاصله اطمینان ۹۵٪ اثر تام متغیرهای مدل $INCA-N$ در نمونه به حجم ۴۰۰۰

۷ - ۵ تحلیل حساسیت مدل *INCA-N* با استفاده از روش *RBD*

تحلیل حساسیت مدل *INCA-N* با استفاده روش طرح متعادل تصادفی در حجم نمونه‌های ۵۰۰، ۱۰۰۰ و ۲۰۰۰ انجام گرفته است. به‌منظور بررسی میزان دقت نتایج حاصل، هر یک از فرآیندهای فوق، ۵۰ بار تکرار شده است. نتایج حاصل در جداول (۷-۳) تا (۷-۶) و به‌صورت فاصله‌های اطمینان و نمودار جعبه‌ای در اشکال (۷-۱۳) تا (۷-۱۸) ارائه شده است.

جدول (۷ - ۳)- آمار توصیفی برآورد اثر اصلی متغیرهای نیترات زدایی و ثابت نیتروژن هوا

متغیر آمار توصیفی	<i>Denitrification</i>			<i>fixation Nitrogen</i>		
	حجم نمونه			حجم نمونه		
	500	1000	2000	500	1000	2000
کمترین مقدار	0.156	0.161	0.164	0.011	0.002	0.002
چارک اول	0.195	0.193	0.191	0.021	0.008	0.005
میانه	0.216	0.208	0.202	0.024	0.011	0.005
میانگین	0.214	0.208	0.201	0.027	0.013	0.006
چارک سوم	0.231	0.225	0.211	0.032	0.017	0.007
بیشترین مقدار	0.292	0.264	0.235	0.066	0.035	0.011
پراش	1.05 E -3	5.54 E -4	2.80 E -4	1.44 E -4	3.90 E -5	4.56E -6

جدول (۷ - ۴) - آمار توصیفی برآورد اثر اصلی متغیرهای جذب نیترات و بیشترین میزان جذب نیترات

متغیر آمار توصیفی	<i>nitrate uptake</i>			<i>Max.nitrate uptake</i>		
	حجم نمونه			حجم نمونه		
	500	1000	2000	500	1000	2000
کمترین مقدار	0.386	0.395	0.408	0.012	0.005	0.002
چارک اول	0.433	0.441	0.442	0.021	0.012	0.007
میانه	0.465	0.458	0.456	0.028	0.015	0.01
میانگین	0.458	0.456	0.454	0.028	0.016	0.01
چارک سوم	0.479	0.474	0.466	0.033	0.02	0.013
بیشترین مقدار	0.536	0.508	0.501	0.057	0.034	0.017
پراش	<i>1.24E-03</i>	<i>5.69E-04</i>	<i>4.01E-04</i>	<i>8.82E-05</i>	<i>4.93E-05</i>	<i>1.58E-05</i>

جدول (۷ - ۵) - آمار توصیفی برآورد اثر اصلی متغیرهای معدنی‌سازی و آلی سازی

متغیر آمار توصیفی	Mineralization			Immobilization		
	حجم نمونه			حجم نمونه		
	500	1000	2000	500	1000	2000
کمترین مقدار	0.06	0.052	0.066	0.143	0.145	0.164
چارک اول	0.094	0.08	0.083	0.187	0.188	0.191
میانه	0.108	0.095	0.089	0.208	0.205	0.201
میانگین	0.11	0.093	0.088	0.209	0.203	0.199
چارک سوم	0.127	0.104	0.093	0.232	0.218	0.21
بیشترین مقدار	0.171	0.124	0.107	0.269	0.245	0.232
پراش	<i>6.17E-04</i>	<i>2.92E-04</i>	<i>9.02E-05</i>	<i>9.15E-04</i>	<i>5.18E-04</i>	<i>2.62E-04</i>

جدول (۷ - ۶) - آمار توصیفی برآورد اثر اصلی متغیرهای جذب آمونیاک و تمام متغیرهای مدل

متغیر آمار توصیفی	ammonium uptake			مجموع اثر اصلی تمام متغیرها		
	حجم نمونه			حجم نمونه		
	500	1000	2000	500	1000	2000
کمترین مقدار	0.011	0.008	0.004	0.928	0.914	0.894
چارک اول	0.024	0.016	0.012	1.05	0.974	0.954
میانه	0.034	0.021	0.014	1.073	1.013	0.971
میانگین	0.034	0.022	0.015	1.081	1.011	0.972
چارک سوم	0.042	0.027	0.018	1.128	1.038	0.99
بیشترین مقدار	0.064	0.041	0.024	1.194	1.093	1.034
پراش	1.65 E-04	6.12 E-05	2.20 E-05	3.19 E-03	1.58 E-03	9.83 E-04

با توجه جداول فوق و بنابر میانگین برآورد اثر اصلی متغیرهای مورد بررسی، مهمترین متغیرهای تاثیرگذار بر متوسط بار نیتروژن به حوضه آبریز رودخانه در حجم نمونه‌های ۵۰۰، ۱۰۰۰ و ۲۰۰۰ به ترتیب عبارتند از:

۱. نرخ جذب نیترات گیاهان (*Plant nitrate uptake*) با میانگین اثر اصلی ۰.۴۵۸، ۰.۴۵۶ و ۰.۴۵۴

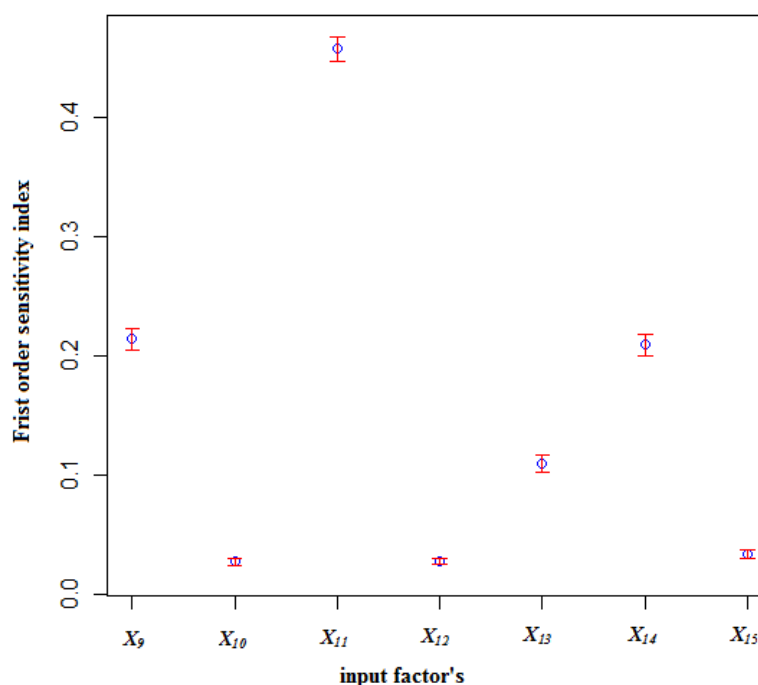
۲. نرخ نیترات زدایی (*Denitrification*) با میانگین اثر اصلی ۰.۲۱۴، ۰.۲۰۸ و ۰.۲۰۱

۳. آلی سازی (*Immobilization*) با میانگین اثر اصلی ۰.۲۰۹، ۰.۲۰۳ و ۰.۱۹۹

۴. نرخ معدنی سازی (*Mineralization*) با میانگین اثر اصلی ۰.۱۱، ۰.۰۹۳ و ۰.۰۸۸

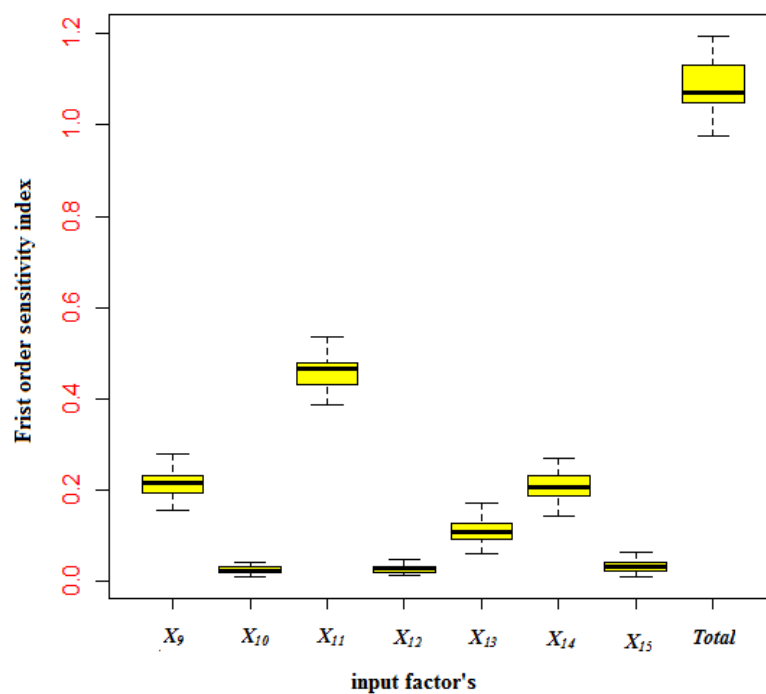
میزان اثر سایر متغیرهای مدل بر متوسط سالانه بار نیتروژن بسیار ناچیز است و می‌توان از اثر آن‌ها صرف نظر نمود. یعنی ثابت نیتروژن موجود در هوا، نرخ جذب آمونیوم و بیشترین نرخ جذب گیاهان، متغیرهای

بی‌تاثیر در مدل *INCA-N* هستند. همچنین مجموع اثرات اصلی تمام متغیرهای مدل نزدیک به مقدار ۱ است. این بدان مفهوم است که مدل *INCA-N* تقریباً یک مدل جمع‌پذیر^۱ بوده، بطوری‌که اثرات متقابل بین متغیرهای مدل اندک است. نتایج بدست آمده از این روش بسیار مشابه نتایج حاصل از اعمال روش مونت کارلو- مبنای سالتلی است. برای درک بهتر متغیرهای تاثیرگذار در جداول فوق، فواصل اطمینان و نمودار جعبه‌ای این متغیرها در حجم نمونه‌های مختلف رسم شده است. لازم به یادآوری است که نمودار جعبه‌ای، آماره‌های توصیفی هر متغیر شامل کمینه، چارک اول، میانه، میانگین، چارک سوم و بیشینه را در قالب یک نمودار به تصویر می‌کشد. از این نمودار برای درک بهتر پراکندگی مقادیر هر متغیر استفاده می‌شود.

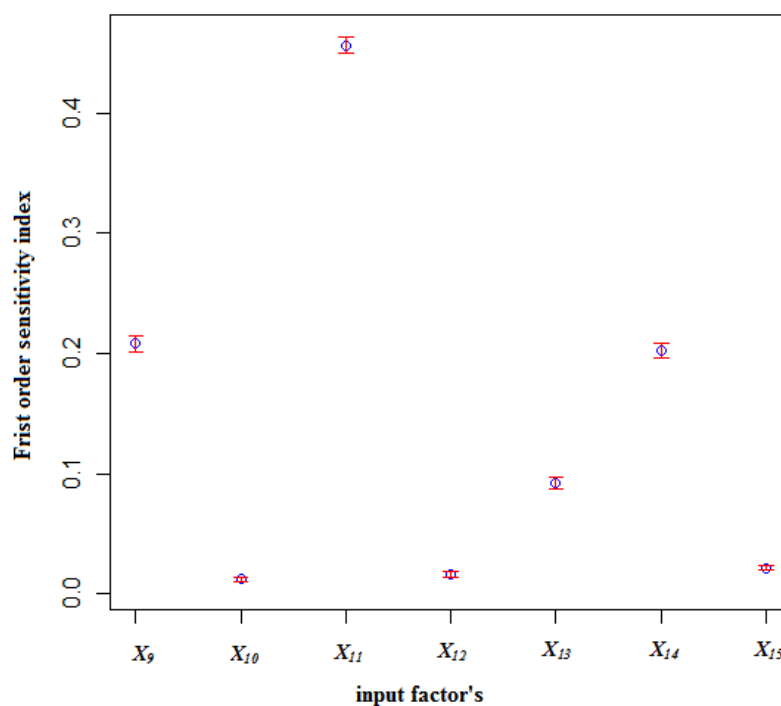


شکل (۷ - ۱۳) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل *INCA-N* با حجم نمونه ۵۰۰

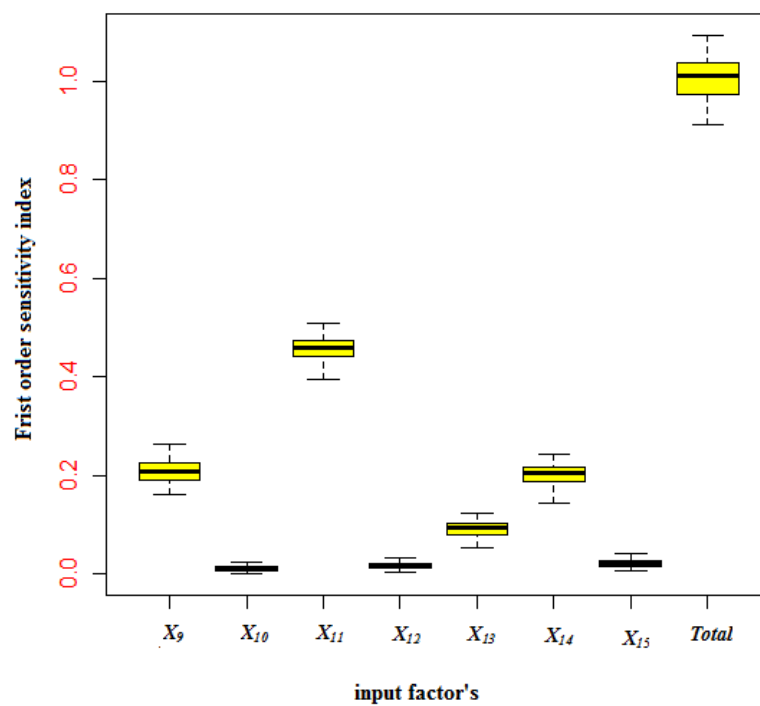
^۱ Additive



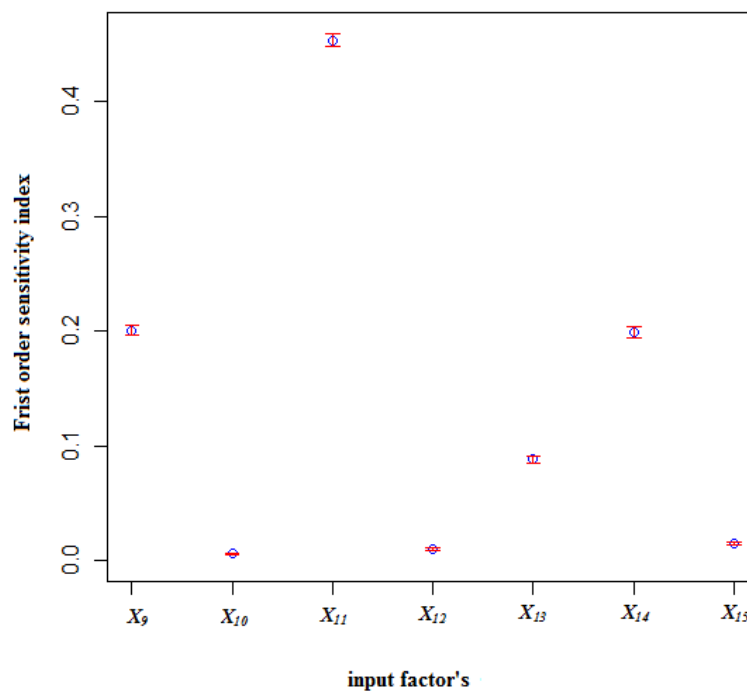
شکل (۷ - ۱۴)- نمودار جعبه‌ای اثر اصلی متغیرهای مدل *INCA-N* با حجم نمونه ۵۰۰



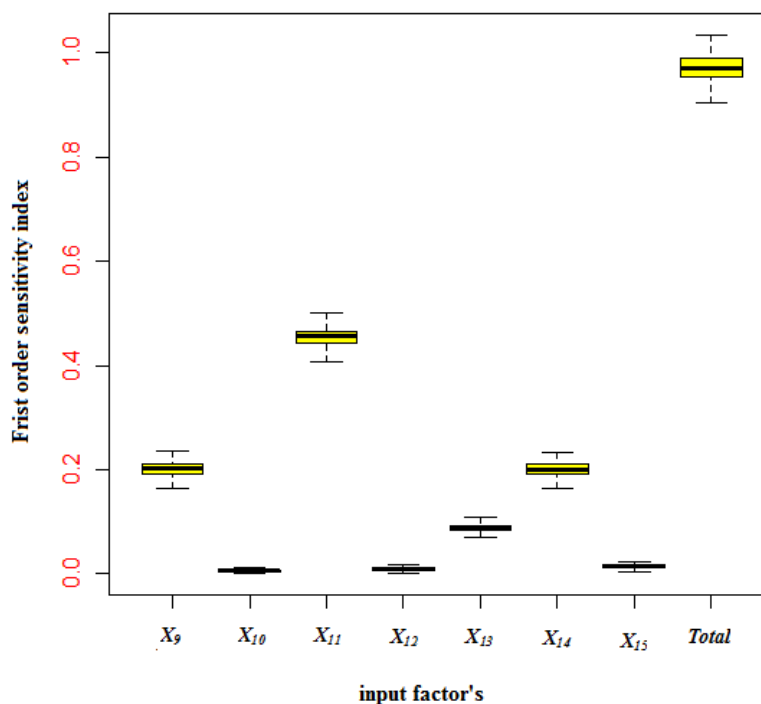
شکل (۷ - ۱۵)- فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل *INCA-N* با حجم نمونه ۱۰۰۰



شکل (۷ - ۱۶) - نمودار جعبه‌ای اثر اصلی متغیرهای مدل $INCA-N$ با حجم نمونه ۱۰۰۰



شکل (۷ - ۱۷) - فاصله اطمینان ۹۵٪ اثر اصلی متغیرهای مدل $INCA-N$ با حجم نمونه ۲۰۰۰



شکل (۷ - ۱۸) - نمودار جعبه‌ای اثر اصلی متغیرهای مدل *INCA-N* با حجم نمونه ۲۰۰۰

همان‌طور که در نمودارهای فوق مشهود است با افزایش حجم نمونه نوسانات و طول فواصل اطمینان میانگین برآورد اثر اصلی متغیرها کاهش می‌یابد.

۷ - ۶ نتیجه گیری

بر اساس برآورد شاخص‌های حساسیت با روش طرح متعادل تصادفی و روش سالتلی مشخص شد که مهمترین عوامل تاثیرگذار بر نیتروژن ورودی به آب روخانه *Tweed* در بین متغیرهای نرخ تبادلات نیتروژن عبارتند از : بیشترین میزان جذب توسط گیاهان (X_{11})، نرخ نیترات زدایی (X_9) ، آلی سازی (X_{14}) و نرخ معدنی‌سازی (X_{13}). در جدول (۷- ۷) نتایج حاصل از اعمال دو روش فوق، با یکدیگر مقایسه شده است.

جدول (۷ - ۷) - نتایج برآورد اثر اصلی و تام متغیرهای مهم مدل *INCA-N*

نام متغیر	روش سالتلی		روش طرح متعادل تصادفی
	برآورد اثر اصلی	برآورد اثر تام	برآورد اثر اصلی (N=2000)
<i>Plant nitrate uptake</i>	0.445	0.496	0.454
<i>Denitrification</i>	0.189	0.249	0.201
<i>immobilization</i>	0.183	0.227	0.199
<i>mineralization</i>	0.073	0.106	0.088
<i>Max.nitrate uptake</i>	-0.004	0.019	0.01
<i>ammonium uptake</i>	-0.01	0.027	0.015
<i>Nit.fixation</i>	-0.011	0.004	0.006

نتایج نشان می‌دهد که هر دو روش، برآوردهای تقریباً یکسانی را برای اثر اصلی ارائه می‌دهد. از آنجایی

که در روش طرح متعادل تصادفی $\sum_{i=1}^7 S_i = 0.973$ می‌توان گفت که اثرات متقابل بین متغیرهای ورودی

وجود دارد اما مقدار آن چندان زیاد نیست. همچنین در روش سالتلی مجموع اثر اصلی برابر با

$\sum_{i=1}^7 S_i = 0.90$ است که این کمیت نیز وجود اثرات متقابل بین متغیرهای ورودی مدل را تایید می‌کند اما

میزان تاکید این روش مبنی بر وجود اثرات متقابل در مقایسه با روش طرح متعادل تصادفی بیشتر است.

نتیجه گیری

تحلیل حساسیت واریانس - مبنا مدل‌ها، دریچه‌ای نو به سوی کاربرد آمار است. نتایج اخذ شده در این پایان نامه می‌تواند از دیدگاه‌های آماری، مهندسی - مدیریتی مورد توجه قرار گیرد.

از منظر آماری، نتایج حاکی از آن هستند که می‌توان شاخص‌های حساسیت را بر اساس واریانس‌های شرطی خروجی مدل اندازه‌گیری کرد. در این راستا می‌توان از دو رویکرد روش مونت کارلو، مبنای سالتلی و روش طرح متعادل تصادفی جهت برآورد این شاخص‌ها استفاده نمود. در روش اول، کلاس جدید شبیه‌سازی اعداد تصادفی، قابلیت‌های منحصر بفردی را از خود به نمایش می‌گذارد. این قابلیت باعث میشود تا بدانیم چرا پارامتر "زمان" اجرای مدل را بایستی به عنوان هزینه قلمداد کرد. روش دوم می‌آموزد که چگونه می‌توان از سری‌های فوریه برای برآورد واریانس‌های شرطی بهره گرفت. همچنین نتیجه گرفتیم که رهیافت‌های مذکور را می‌توان با استفاده از توابعی تحلیلی مانند تابع سبل ارزیابی نمود.

اعمال این دو روش بر مدل $INCA-N$ نشان داد که به طور کلی مهندسین می‌توانند با استفاده از مدل‌ها و با تکیه بر شاخص‌های حساسیت، رفتار پدیده‌های طبیعی را بهتر شناخته و آن را توصیف نمایند. شناسایی متغیرهای تاثیرگذار بر نیتروژن ورودی به رودخانه و ریشه‌یابی دلایل آن تاثیرات، که از دیدگاه مهندسی بسیار حائز اهمیت است، نیز در گرو پرداختن به روش‌های تحلیل حساسیت است. در این صورت است که مدیران می‌توانند با آگاهی از پیامدهای مخرب عوامل موثر، درصدد سیاست‌گذاری برای کنترل این متغیرها برآیند و برنامه‌هایی تخصصی مثلاً برای کوددهی زمین‌های کشاورزی و ... را تدوین نمایند.

نتایج حاصل از تحلیل حساسیت مدل $INCA-N$ نشان داد که برای کنترل نیتروژن ورودی به آب رودخانه *Tweed* می‌بایست حداقل به سه عامل "مقدار میزان جذب نیترات توسط گیاهان"، "نرخ نیترات زدایی" و "آلی سازی" توجه ویژه ای داشت. با کنترل این سه عامل که نقش تعیین کننده در

آلایندگی نیتروژنی آب دارند می‌توان از آسیب‌ها و پی‌آمدهای تاثیر آن‌ها در تهدید سلامت جامعه کاست و خسارت‌های مالی و زیست محیطی آن را تعدیل نمود. البته در کنار این سه عامل نیز می‌توان متغیر "معدنی‌سازی" را به عنوان عامل چهارم تاثیرگذار بر بار نیتروژن ورودی به آب این رودخانه در نظر گرفت.

پیشنهادهات

به منظور انجام تحقیقات و پژوهش‌های آتی در زمینه تحلیل حساسیت می‌توان به موارد زیر اشاره کرد.

۱- طراحی الگوریتم جدید جهت کاهش هزینه محاسبه برآورد شاخص‌های حساسیت در حالت کلی.

۲- یافتن متامدل یا مدل جانشین برای مدل اصلی توسط روش‌های پیشرفته آماری و یادگیری ماشین، که در این صورت شاخص‌های حساسیت می‌توانند بدون اجرای مدل اصلی و توسط مدل جانشین، برآورد شوند

۳- تعیین مقادیر بهینه متغیرهای موثر و مقایسه مشاهدات میدانی رودخانه با سری‌های زمانی شبیه‌سازی شده نیترا و آمونیوم.

فهرست منابع

- Addiscot, M.b., whitmore,A.p., computer simulation of change in soil mineral and crop nitrogen during autumn, winter and spring. **J. Agric.Sci.Camb.**Vol. **109**,p.p **141-157**.
- Archer, G. E. B. Saltelli, A. and Sobol, I.,M. (1997).Sensitivity Measures, Anova-Like Techniques and the Use of Bootstrap. **J. Statist. Comput.**, Vol. **58**, p.p **99-120**.
- Campolongo, F., J. Cariboni and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. **J. Environmental Modelling and Software**. Vol. **22**, p.p **1509–1518**.
- Cooper, D., Ragab,R., Whitehead, P.G, (1993). IHDM-TRANS a nitrat model for agricultural systems, IH report MAFF.
- Cosby, B.J, Ferrier, R.C., Jenkis,A., Emmelt, B., Tietema, A., Wright, R.F.,(1997). Model and ecosystem and loss of inorganic nitrogen (MERLIN). **J. Hydrol earth Syst. Sci.** in press.
- Cosby, B.J, Wright, R.F., Hornberger, G.M., Galloway, J.N., (1985b). modeling the effects of acid deposition:estimation a long-trem water quility response in a small forosted cachment . **J. Water Resource Res.**, Vol **21**,p.p **1591-1601**.
- Cuiker, R.I., Schaibly, J. H., Schuler, A. G. (1975), Study of the Sensitivity of coupled Reaction System to Uncertainties in Rate Coefficient. III Analysis of the Approixmation, **J. Chemical Physics**, Vol **63**,p.p **1140–1149**.
- Cukier, R. I., C. M. Fortuin, K. E. Schuler, A. G. Petschek and J. H. Schaibly (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. i theory. **J. Chemical Physics** , vol. **59**, **3873–3878**.
- Cukier, R., H. Levine and K. Shuler (1978).Nonlinear sensitivity analysis of multiparameter model systems. **J. Computational Physics**, Vol. **26**,p.p **1–42**.
- Fang, Kai-Tai., Li, Runze., Sudjianto, Agus., (2006). **Design and Modeling for Computer Experiments. 1edt., Chapman & Hall/CRC**.

- Helton, J.C., Johnson J.D., Salaberry C.J., Storlie C.B., (2006). Survey of sampling based methods for uncertainty and sensitivity analysis. **J. Reliability Engineering and System Safety**, Vol. 91, p.p 1175–1209.
- Homma, T. and A. Saltelli (1996). Importance measures in global sensitivity analysis of nonlinear models. **J. Reliability Engineering and System Safety**, Vol 52, p.p 1–17.
- Hornberger, G. and R. Spear (1981). An approach to the preliminary analysis of environmental systems. **J. Environmental Management**, Vol. 7, p.p 7–18.
- Iman, R. and S. Hora (1990). A robust measure of uncertainty importance for use in fault tree system analysis. **J. Risk Analysis** , Vol 10(3), p.p 401–403.
- Ishigami, T., and Homma, T., (1996). An importance qualification technique in uncertainty analysis for compute models, in: *Proceedings of the Isuma '90. First International Symposium on Uncertainty Modelling and Analysis*, University of Maryland, 3–5 December.
- Jansen, M.J.W., (1999) . Analysis of variance designs for model output, **J. Computer Physics Communications** , Vol 117, p.p 35–43
- Jarvie, H. P., Wade, A. J., Butterfield, D., Whitehead, P. G., Tindall, C. I., Virtue, W. A., Dryburgh, W. and McCraw, A. (2002), Modelling nitrogen dynamics and distributions in the River Tweed, Scotland: an application of the INCA-N model, **J. Hydrol. Earth Sys. Sci.**, Vol 6, p.p 443-453.
- Jenkis, A., Ferrier, R.C., Cosby, B.J., (1997). A dynamic model for assessing the impacts of coupled sulphur and nitrogen deposition scenarios on surface water acidification to two upland catchment. **J. Hhdrol.** In press
- Jishan Hu, Zheng Chen, Yiguang Ju (2007). Regularized random-sampling high dimensional model representation (RS-HDMR). **J. Mathematical Chemistry**, Vol. 43, No. 3.

- Karaivanova, A. Dimov, I. and Ivanovska, S. (2000), *A Quasi-Monte Carlo Method for Integration with Improved Convergence*.
- Koda, M., McRae, G. J., and Seinfeld, J. H. (1979), *Automatic Sensitivity Analysis of Kinetic Mechanism*, **International Journal of Chemical kinetics**, Vol 11, p.p 427-444.
- Krzykacz-Hausmann, B. (1990). *Gesellschaft fuer Reaktor Sicherheit (GRS) MbH. Technical Report GRS-A-1700, Garching*.
- Levy, G., *An introduction to quasi-random numbers* (2002). NAG Ltd, Oxford, UK
- Mara, T. A. (2009). *Extension of the RBD-FAST method to the computation of global sensitivity indices*. **J. Reliability Engineering and System Safet**, Vol. 94 ,p.p 1274–1281
- McKay, M. (1996). *Variance-based methods for assessing uncertainty importance in nureg-1150 analysis. Technical Report LA-UR-96-2695, 1, Los Alamos Laboratories*.
- Meteorological Office., (1981). *The MORCES system*, **J. Hydrological Memorandum**, No. 45.
- Morris, M. D. (1991). *Factorial sampling plans for preliminary computational experiments*. **J. Technometrics** , Vol. 33, p.p 161–174.
- Oakley, J. and A. O'Hagan, (2004). *Probabilistic sensitivity analysis of complex models: a Bayesian approach*. **J. Royal Statistical Society. B**, Vol. 66,p.p 751–769.
- Postek, K.M., Driscoll, C.T., Aber, J.D., (1995). *Application of Pnet-CN / CHESS to a spruce stand in sollig, Germany*. **J. Ecol.Model** , Vol 83, p.p 163-172.
- Sacks, J., W. J. Welch, T. J. Mitchell and H. P. Wynn (1989). *Design and analysis of computer experiments*. **J. Statistical Science**, Vol. 4,p.p 409–435.

- Saltelli A., Andres, T.H. Homma, T. (1993). *Some new techniques in sensitivity analysis of model output*, **J. Comput. Statist. Data Anal.**, Vol 15,p.p 211–238.
- Saltelli, A. (2002). *Making best use of model valuations to compute sensitivity indices*. **J. Computer Physics Communications**, Vol. 145,p.p 280–297.
- Saltelli, A., Bolado, R. (1998), *An Alternative Way to Compute Fourier Amplitude Sensitivity Test (FAST)*, **J. Computational Statistics and Data Analysis**.
- Saltelli, A., P Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S. (2010). *Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index*, **J. Computer Physics Communications**, Vol. 181. P.P 259–270.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S. (2008). **Global sensitivity analysis The Primer**, 1ed., John Wiley & Sons Ltd.
- Saltelli, A., S. Tarantola and K. Chan (1999). *Quantitative model-independent method for global sensitivity analysis of model output*. **J. Technometrics**, Vol.41, Num.1, p.p 39–56.
- Saltelli, A., T. Andres and T. Homma (1993). *Sensitivity analysis of model output: an investigation of new techniques*. **J. Computational Statistics and Data Analysis**, Vol 15,p.p 211–238.
- Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M. (2004). **Sensitivity Analysis in Practice**, Ed. 1, John Wiley & Sons Ltd.
- Santner, T. J., Williams, B. J., Notz, W. I. (2003). **The Design and Analysis of Computer Experiments**.Ed. 1, Springer, New York.
- Shahsavani, D., Grimvall, A. (2011). *Variance-based sensitivity analysis of model outputs using surrogate models* , **J. Environmental Modelling and Software**, Vol 26(6), p.p 723-730.

- Shahsavani, D., Grimvall, A., (2009). An adaptive design and interpolation technique for extracting highly nonlinear response surfaces from deterministic models. **J. Reliability Engineering and System Safety**, Vol 94, p.p 1173-1182.
- Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. **J. Mathematics and Computers in Simulation**, Vol 55 , p.p 271–280
- Sobol, I.M. (1998). On quasi-Monte Carlo integrations, **J. Mathematics and Computers in Simulation**, Vol 47, p.p 103±112.
- Sobol, I.M., (2007). Global sensitivity analysis indices for the investigation of nonlinear mathematical models. **J. Matematicheskoe Modelirovanie** 19 ,Vol 11, p.p 23–24 (in Russian).
- Sobol', I.M.(1990). Sensitivity estimates for nonlinear mathematical models. **J. Matematicheskoe Modelirovanie**, Vol 2, p.p 112–118. in Russian, translated in English in Sobol' (1993).
- Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. **J. Mathematical Modelling and Computational Experiment**, Vol 1, p.p 407–414. English translation of Russian original paper Sobol' (1990).
- Sobol', I. M. (1996). On 'freezing' unessential variables. *Vestnik Moskovskogo Universiteta*, **Journal of Serija Matematika**, Vol. 6,p.p 92–94.
- Tarantola, S. Gatellia, D., Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. **J. Reliability engineering and system safety** , Vol 91(6), p.p 717-727.
- Wade, A. J., Durand, P., Beaujouan, V., Wessel, W. W., Raat, K. J., Whitehead, P. G., Butterfield, D., Rankinen, K., Lepisto, A., 2002. A nitrogen model for European catchments: INCA, new model structure and equations. **J. Hydrology and Earth System Sciences**, Vol 6, p.p 559-582.

- Whitehead, P. G., Wilson, E. J., Butterfield, D., (1998a). A semi-distributed nitrogen model for multiple source assessments in catchments (INCA): Model structure and process equations. **J. Science of the Total Environment**, Vol 210/211, p.p 547-558.
- Whitehead, P.G., (1990). Modeling nitrate from agriculture to public water supplies. *Phil. Trans. London B329*, p.p 403-410.
- Whitehead, P.G., Beck, M.B., O'Connell, E., (1981). A system model of stream flow and water quality in the bedford- Ouse River Part 1, water quality modeling. **J. Water Res. Vol 15**, p.p 1157-1171.
- Whitehead, P.G., Williams, R., (1984). Modeling nitrate and algal behaviour in the river Thames. **J. Water Sci. Tech. Vol 16**, p.p 6521-6633.
- Whitehead, P.G., Williams, R., Lewis, D., (1997). Quality simulation along river system (QUASAR): model theory and development. **J. Sci. Total Environ. Vol 194/195**, p.p 447-456.
- Whitehead, P.G., Young, P.C., Hornberger, G., 1979. A system model of stream flow and water quality in the bedford- Ouse River Part 1, stream flow modeling. **J. Water Res. Vol 15**, p.p 1155-1169.
- Yuan, Q and Liang, D (2011). A New Multiple Sub-domain Rs-HDMR Method and Its Application to Tropospheric Alkane Photochemistry Model. **International Journal of Numerical Analysis and Modeling, Series B**, Vol. 2, Num. 1, p.p 73-90.

ضمیمه

مدل *INCA-N*

مقدمه

آب مایه حیات است. این توصیف، کامل‌ترین و ساده‌ترین عبارت برای بیان اهمیت آب است. این ماده حیاتی نه تنها مورد استفاده جانداران و آبزیان است بلکه در صنایع مختلف از جمله: کشاورزی، تولید برق، صنایع معدنی و ذوب فلزات استفاده می‌شود. مقدار آب موجود در زمین تقریباً ثابت است. تقریباً ۹۷/۲ درصد از آب‌های موجود زمین در اقیانوس‌ها و دریاها انباشته شده است و فقط ۲/۸ درصد آن به صورت رودها، یخچال‌ها، دریاچه‌ها، آب موجود در اتمسفر و آب موجود در خاک و آب‌های زیر زمینی است. آب‌های موجود در کره زمین همواره در حال تغییر شکل‌اند. این آب‌ها حالت‌های مایع، جامد و گاز به خود می‌گیرند. آب دارای یک چرخه طبیعی مشخصی است که با دخالت بشر دچار مشکلاتی در فرآیند زیستی خود از قبیل

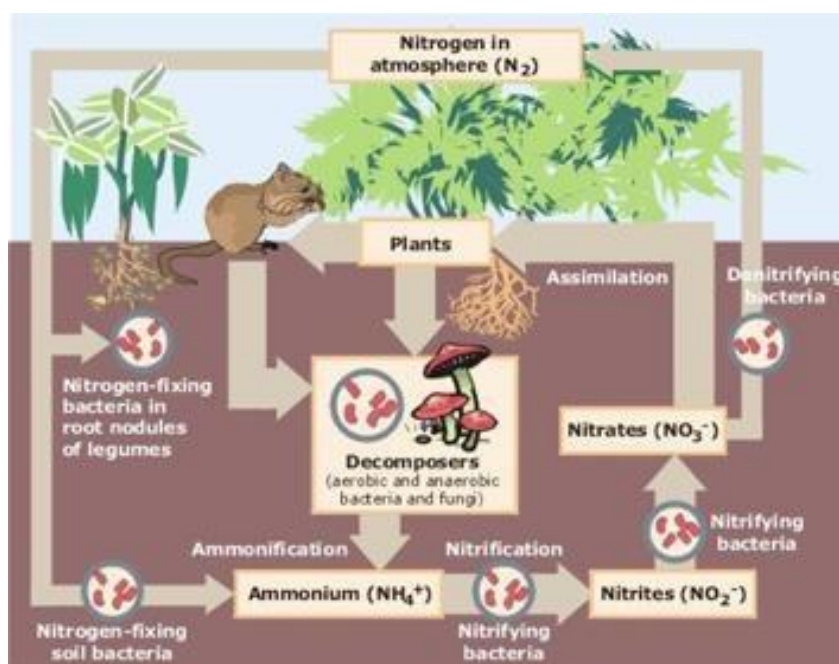
- برداشت بی رویه از آب‌های زیر زمینی
- ورود آلاینده‌های مختلف به منابع آب
- تبدیل بخش اعظم آب برداشتی به فاضلاب، زهاب و پساب

شده است. مسئله آلودگی آب با سلامت انسان و سایر جانداران و آبزیان در ارتباط است. لذا بررسی کمی و کیفی منابع آب جهت اندازه‌گیری آلودگی‌های مختلف از اهمیت ویژه‌ای برخوردار است. یکی از آلاینده‌های آب‌های زیرزمینی و آب‌های سطحی، **نیتروژن** است. گیاهان از نیتروژن برای سنتز پروتئین‌های رشد خود استفاده می‌کنند و این ماده را با جذب مستقیم از زمین و به شکل نترات (NO_3) یا در ترکیب با سایر مواد به دست می‌آورند. نترات موجود در کود شیمیایی سبب افزایش این ماده در خاک، آب‌های جاری و زیر زمینی می‌شود. نترات از طریق آب و مواد غذایی وارد بدن انسان شده و افزایش سطح آن موجب می‌شود که به ترکیبات دیگری چون نیتريت و نیتروسامین تبدیل شود. این ترکیبات، سمی بوده و برای

سلامت بدن خطرناکند. مقادیر بحرانی غلظت نیترات در آب‌های شرب باعث بروز بیماری‌هایی مانند سرطان، ناقص الخلقه بودن نوزادان، بزرگ شدن تیروئید و التهاب غدد لنفاوی می‌شود. با توجه به اهمیت آب آشامیدنی و تاثیر مستقیم آن بر سلامت شهروندان، موضوع کنترل آلودگی آب و پیشگیری از افزایش نیترات به سطوح بحرانی، امری اجتناب ناپذیر است و در حال حاضر یکی از مهمترین مسائل محیط زیست، آلودگی منابع آب به نیترات است. نیترات از طریق تماس منابع آب با فاضلاب‌های شهری و صنعتی، تخلیه آب‌های شستشوی زمین‌های کشاورزی در رودخانه‌ها و از همه مهمتر اکسیداسیون مواد آلی ازت‌دار نظیر پروتئین‌ها وارد منابع آب می‌شود. نیترات بطور مستقیم و غیرمستقیم در اثر تجزیه و تغییرهای زیست - شیمیایی ترکیب‌های مختلف معدنی و آلی، در خاک به وجود می‌آید و توسط آب باران به منابع آب سطحی و زیر زمینی نفوذ می‌کند. نیتروژن موجود در طبیعت دارای چرخه مشخصی است که در آن ترکیبات مختلف نیتروژن تحت شرایط مختلف به گونه‌های متفاوت تبدیل می‌شود. نیترات یکی از مشتقات نیتروژن است که در طی فرآیندهای تبدیلات نیتروژن و در چرخه‌ی این ماده تولید می‌شود.

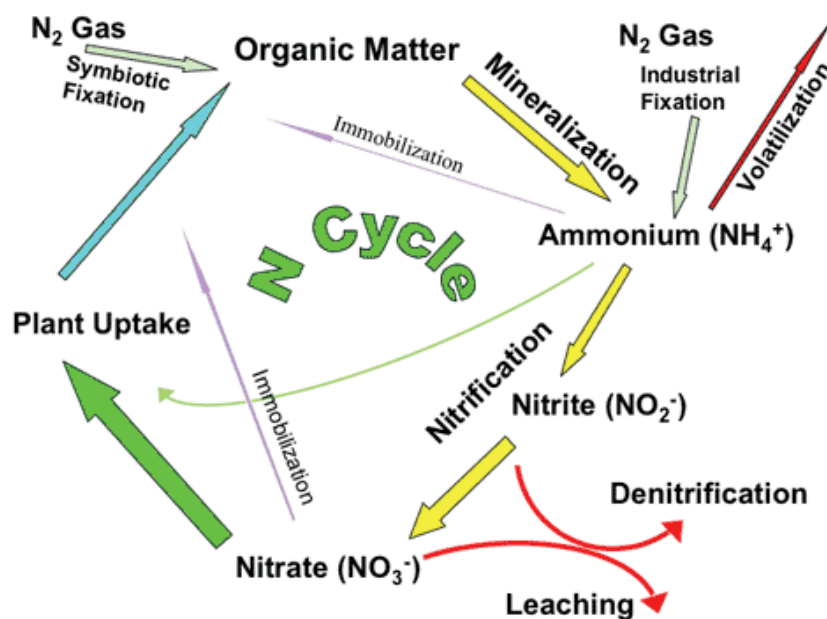
چرخه‌ی نیتروژن

حدود ۷۸ درصد از جو زمین را نیتروژن تشکیل داده است. مقدار معینی از این نیتروژن، به طور مداوم از جو گرفته شده و به آن بازپس داده می‌شود. به گردش مداوم نیتروژن بین خاک، آب، هوا و موجودات زنده چرخه‌ی نیتروژن می‌گویند. نیتروژن در پروتئین‌ها، واحدهای ساختاری و شیمیایی همه موجودات زنده و در مولکول‌های DNA یافت می‌شود. در واقع نیتروژن یکی از اجزای تشکیل دهنده‌ی پروتئین‌ها و اسیدهای نوکلئیک است که وجود هر دو برای ادامه‌ی حیات ضروری است. حیوانات و گیاهان نمی‌توانند مستقیماً از نیتروژن موجود در جو استفاده کنند اما باکتری‌های خاصی که در ریشه‌ی برخی از گیاهان مثل نخود، لوبیا، نخودفرنگی و غیره وجود دارند، نیتروژن هوا را به طور مستقیم جذب کرده و در اختیار



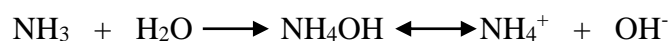
قسمتی از نیتروژن موجود در هوا هنگام رعد و برق از آن جدا می‌شود. تخلیه‌ی ناگهانی بار الکتریکی باعث می‌شود مقداری از اکسیژن و نیتروژن موجود در هوا با هم ترکیب و اکسیدهای نیتروژن به وجود آید.

اکسیدهای نیتروژن پس از حل شدن در آب با سایر عناصر ترکیب می‌شوند و ترکیبات نیتروژن‌دار تولید می‌کنند. این ترکیبات به هنگام باران جذب خاک می‌شوند. در درون خاک، باکتری‌های نیترات‌کننده این ترکیبات نیتروژنی را به نیترات تبدیل می‌کنند که جذب ریشه گیاهان می‌شوند. شکل (۲) ترکیبات مختلف نیتروژن را نشان می‌دهد.



شکل ۲ - چرخه نیتروژن

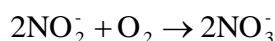
در این چرخه، آمونیاک طی یک تغییر زیست‌شناختی نیتروژن آلی که آمونیاک‌سازی نامیده می‌شود، تولید می‌گردد. آمونیوم محلول در آب، در یک معادله تعادل به صورت یونیزه شده و غیره یونیزه وجود دارد که به صورت زیر می‌باشد.



مقداری از آمونیاک مازاد مصرف گیاهان به وسیله^۱ باکتری شیمواتوتروفیک^۱ در دو مرحله که آمونیاک زدایی^۲ نامیده می شود، اکسید می شود. در مرحله اول اکسیداسیون، باکتری نیتروزوموناس^۳ به نیتريت تبدیل می شود .



نیتريت با اکسیداسیون بیشتر به وسیله^۴ باکتری نیتروباکتر^۴ به نیترات تبدیل می شود.



تبدیل نیتريت به نیترات بسیار سریع انجام می شود. وقتی گیاهان مرده و فاسد می شوند، تجزیه شده و آمونیاک دوباره از طریق عملی به نام معدنی سازی^۵ به خاک یا محلول بازمی گردد. چرخه نیتروژن تحت تاثیر ورود کودهای شیمیایی در زمین های کشاورزی و پساب های صنعتی به منابع آب، دستخوش تغییراتی شده است. شکل (۳) چرخه نیتروژن را تحت تاثیر دخالت بشر نشان می دهد. این شکل تمامی تبدیلات و فرآیندهای نیتروژن را در محیط به همراه تاثیر دخالت های بشر در این چرخه را نشان می دهد. در طی این فرآیندها مشتقات متنوعی از نیتروژن حاصل می شود که برخی از این مشتقات عامل تهدید کننده سلامت انسان و آبیان محسوب می شوند. یکی از این عوامل وجود بیش از حد میزان نیترات در منابع آب است. از این رو شناخت عوامل موثر و کنترل چگونگی ورود نیتروژن به منابع آب از اهمیت

^۱ Chemoautotrophic

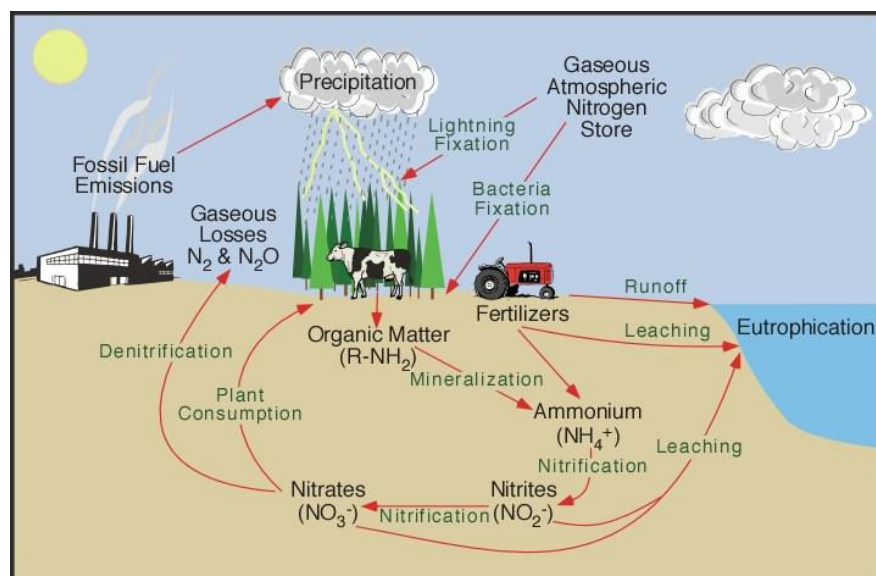
^۲ Nitrification

^۳ Nitrosomonas

^۴ Nitro bacter

^۵ Mineralization

زیادی برخوردار است. رودخانه‌ها یکی از منابع آب می باشند که نیتروژن از طرق مختلف وارد این منبع آب می شود.



شکل ۳ - چرخه نیتروژن تحت تاثیر دخالت بشر

منابع ورود نیتروژن به آب رودخانه

نیتروژن ورودی به آب رودخانه‌ها دارای منابع متعدد می‌باشد. این منابع عبارتند از:

- منشا زمینی (مناطق جنگلی، کشاورزی : کودهای شیمیایی و غیره)
- ریزش‌های جوی (باران و برف و باران اسیدی و غیره)
- فاضلاب‌های شهری و صنعتی و واحدهای متمرکز زراعی

کودهای شیمیایی یکی از منابع ورود نیتروژن به آب رودخانه محسوب می‌شوند. بسیاری از کودهای شیمیایی تحت عنوان اوره و آمونیوم به منظور باروری بیشتر زمین و رشد بهتر محصولات کشاورزی مورد

استفاده قرار می‌گیرد. نیترات (ازت) مازاد نیاز محصولات در کودهای شیمیایی با فعل و انفعالات شیمیایی و نیترات سازی تحت عنوان NO_3-N در خاک نشست می‌کند و هنگام بارندگی وارد آب‌های زیر زمینی می‌شود. کودهای شیمیایی مهمترین عامل تهدیدکننده کیفیت آب در مناطق پر قوت کشاورزی است، به طوری که میزان نیتروژن ورودی از این منبع می‌تواند ۴ تا ۴۰ برابر بیشتر نسبت به نیتروژن ورودی از ریزش های جوی باشد (وایتهد ۱۹۹۰).

منبع دیگر نیتروژن نزولات جوی است که غالبا به صورت باران، برف، باران‌های اسیدی، تگرگ و غیره می‌باشد. نزولات جوی در حوضه‌های آبریز، به دلیل عدم توانایی کنترل، منابع شهری و کشاورزی یکی از منابع مهم نیتروژن تلقی می‌شوند به طوری که در حوضه‌های آبریز مناطق کوهستانی یک عامل مهم نیتروژن محسوب می‌شوند. در این نواحی تاثیرات محیطی شستشوی نیترات بر خاک و سطح اسیدی آب بیشتر از اثرگذاری آن بر کیفیت آب آشامیدنی است. در نواحی کشاورزی حوضه آبریز، نزولات جوی سهم بسیار کمی از غلظت نیتروژن آب رودخانه را به خود اختصاص می‌دهند (وایتهد و همکاران، ۱۹۹۸).

تاسیس کارخانه‌ها، صنایع تولیدی و کشاورزی در کنار رودخانه‌ها سبب شده تا مازاد آب کشاورزی و پساب‌های اکثر صنایع مهم فلزی، شیمیایی، نفت، غذایی، سلولزی، نیروگاه‌ها و فاضلاب‌های شهرهای وارد رودخانه‌ها شوند. این امر موجب افزایش دما، مواد آلی، معدنی و ترکیبات خطرناک فلزات سنگین در آب شده و علاوه بر آلودگی محیط زیست آبریان، سبب برهم خوردن تعادل ترکیبات موجود در آب و آلودگی آب و نیز افزایش هزینه‌های تصفیه و بهسازی آب رودخانه شده است.

موارد ذکر شده، مهمترین منابع ورود نیتروژن به رودخانه‌ها تلقی می‌شوند. به منظور اندازه‌گیری نیتروژن موجود در آب رودخانه باید ضمن در نظر گرفتن تمام منابع ورودی نیتروژن، مقدار نیتروژن ورودی هر منبع محاسبه شود. این مسئله موجب پیچیدگی بیشتر محاسبات چرخه نیتروژن در رودخانه می‌شود.

پیچیدگی طبیعی این چرخه سبب شده است که انجام برخی آزمایشات فیزیکی برای شناسایی و کنترل عوامل تاثیرگذار بر آلودگی آب بنا به دلایلی نظیر داشتن هزینه های گراف مالی و زمانی و عدم امکان اجرا، مقدور نباشد. لذا محققین به شبیه سازی این پدیده ها توسط مدل های ریاضی روی آورده اند. مدل های ریاضی این پدیده با گذشت زمان تغییر و بهبود یافته اند. در ادامه روند تکاملی مدل سازی این فرایندها را به طور مختصر شرح می دهیم.

مدل سازی چرخه ی نیتروژن در حوضه آبریز

می توان نیترات رودخانه را به عنوان تابعی از آب و هوای منطقه، ویژگی های حوضه آبریز، نقشه برداری و خواص هیدرولوژی در نظر گرفت. سطح نیترات رودخانه بازتابی از مجموعه منابع ورودی نیتروژن (کودهای شیمیایی، ریزش های جوی و تخلیه فاضلاب و ...) در حوضه آبریز است. در برخی از موارد، با توجه به ظرفیت حوضه آبریز برای ذخیره سازی نیتروژن در خاک، آب های زیرزمینی و پوشش های گیاهی می توان تاثیر دخالت های انسانی (طرح های کود دهی و ...) را در کمیت نیترات رودخانه ملاحظه نمود. از آنجایی که مازاد ترکیبات طبیعی نیتروژن سهم معدنی سازی، نیترات زدایی و نیترات سازی مواد آلی نیتروژن دار خاک است و این تبدیلات متاثر از عوامل شرایط آب و هوایی، فاکتورهای فصلی مانند دما و رطوبت و غیره، مدیریت اجرایی همانند جنگل کاری، چرخاندن زمین، شخم زدن و نوع زمین های مورد استفاده است. لذا می توان از مدل های خاصی جهت شناسایی نواحی حساس نیترات و بار بحرانی ورودی به رودخانه استفاده نماییم. این مدل ها برای پیش بینی این که چگونه بار نیتروژن رودخانه ها بر اثر تغییر رسوب نیترات در خاک، نوع زمین مورد استفاده، مدیریت و شرایط آب و هوایی تغییر خواهد کرد،

طراحی شده‌اند. به‌عنوان مثال مدل‌های همانند مدل *MAGIC-WAND* که توسط کاسبی^۱ و همکاران (۱۹۹۵)، جنکیس^۲ و همکاران (۱۹۹۷)، مدل *MERLIN* که توسط کاسبی و همکاران (۱۹۹۷) و مدل *PNET-CN* که توسط پوستک^۳ و همکاران (۱۹۹۵) برای به‌دست آوردن نیتروژن در اکوسیستم‌های کوهستانی، جنگل‌ها یا فرآیندهای خاص طراحی شده‌اند. این‌گونه مدل‌ها را فرآیند- مبنای گویند. مدل‌های فرایند- مبنای دیگری نیز جهت تاثیر استفاده از کودهای شیمیایی در خاک و آب‌های زیرزمینی و آب- های سطحی بر زمین‌های پایین‌دست رودخانه‌ها توسط ادیسکات^۴ - وایتمور^۵ (۱۹۸۷) و کوپر^۶ (۱۹۹۳) بسط و توسعه یافته‌اند. مدل‌های دیگری همانند *QUASAR*، حرکت نیتروژن در مسیر اصلی رودخانه را طی فرآیندهای خاص درون رودخانه‌ای از قبیل: نیترات سازی و نیترات زدایی شبیه‌سازی می‌کند (وایتهد و همکاران، ۱۹۹۷، وایتهد و ویلیامز^۷ ۱۹۸۴). مدل‌های تجربی نیز برای حرکت نیتروژن در آب رودخانه طراحی شده‌اند. این‌گونه مدل‌ها بر مبنای مشاهدات و سوابق زمین مورد استفاده طراحی شده‌اند. این تکنیک یک روش بسیار موفق برای برآورد میزان بار نیتروژن سالانه و فصلی به آب رودخانه است. اما قادر به محاسبه تغییرات غلظت روزانه نیست. علی‌رغم تمام این توسعه‌ها تنها تعداد محدودی از این مدل‌ها، توانایی ترکیب فرآیندهای درون رودخانه‌ها و حوضه‌های آبریز دارند. با توجه به قابلیت‌های

^۱Cosby^۲Jenkis^۳Postek^۴Addiscott^۵Whitmore^۶Cooper^۷Williams

مدل‌های طراحی شده و ایرادات وارد شده به این مدل‌ها می‌توان نتیجه گرفت مدلی برای نیتروژن ورودی به آب رودخانه‌ها مناسب است که بتواند:

۱. علاوه بر بررسی و اندازه‌گیری نیتروژن ورودی از طریق ریزش‌های جوی و کودهای شیمیایی توزیع شده در حوضه آبریز، قادر به اندازه‌گیری میزان این ماده از طریق تغییرات خاص حوضه مانند: زمین مورد استفاده، خواص هیدرولوژیکی و پوشش گیاهی تاثیرگذار باشد. بطور خلاصه مدلی مناسب است که بطور همزمان نفوذ نیتروژن از طریق منابع مختلف را به‌طور افقی و عمودی در محیط بررسی کند.
۲. ضمن برآورد میزان بار نیتروژن سالانه و فصلی به آب رودخانه قادر به محاسبه تغییرات روزانه غلظت نیتروژن نیز باشد.

مدل *INCA-N* به منظور برطرف کردن ایرادات وارد شده بر سایر مدل‌های نیتروژن ورودی به آب رودخانه طراحی شده است. این مدل دربرگیرنده چرخه نیتروژن در حوضه آبریز یک رودخانه است.

معرفی مدل *INCA-N*

مدل *INCA-N* به منظور بررسی وضعیت و توزیع نیتروژن در محیط‌های خاکی و آبی طراحی شده است. *INCA-N* یک مدل فرآیند-مبنا است که با استفاده از معادلات جنبشی واکنش‌ها به شبیه‌سازی مکانیزم‌های اصلی عملیاتی رودخانه می‌پردازد. این مدل یک مدل شبیه‌ساز روزانه است که اطلاعات را به صورت یک سری زمانی روزانه در ایستگاههای اصلی مسیر رودخانه نمایش می‌دهد. این مدل دارای ویژگی‌های زیر است (وایتهد و همکاران، ۱۹۹۸).

۱. دینامیکی: غلظت و جریان نیتروژن در رودخانه را به عنوان یک سری زمانی تولید می‌کند. تغییرات مدام خواص هیدرولوژی در مسیرهای مختلف جریان آب و تبدیلات نیتروژن در حوضه آبریز رودخانه در طی زمان محاسبه می‌کند.
۲. تصادفی: این ویژگی سبب می‌شود تا خروجی مدل احتمالی باشد و حساسیت پارامترهای مختلف مدل قابل اندازه‌گیری باشد.
۳. نیمه – توزیعی: عبارت نیمه-توزیعی یعنی در مدل جزئیات رفتار سطح زمین حوضه آبریز لحاظ نشده است. اما همزمان زمین‌های حوضه آبریز را دسته‌بندی و حدود و مرز زیر حوضه‌های آبریز مشخص شده و در مدل به کار رفته، سپس این اطلاعات به مدل رودخانه چند ناحیه‌ای تزریق شده است.

قابلیت‌های مدل *INCA-N*

این مدل دارای توانایی‌های متعددی است. از جمله

- شبیه‌سازی همزمان ۶ نوع از زمین‌های مورد استفاده
- شبیه‌سازی جریان آب در مسیرهای اصلی رودخانه
- ردیابی جریان نیتروژن موجود در نیترات و آمونیوم در سطح زمین و داخل رودخانه
- در نظر گرفتن پارامترهایی مانند میزان جذب نیتروژن توسط گیاهان، سطح و زیر سطح نواحی قرار گرفته در مسیر اصلی رودخانه
- بررسی تغییرات روز به روز جریان و غلظت نیتروژن با تغییر در منابع ورودی نیتروژن همانند ریزش‌های جوی، تخلیه فاضلاب و استفاده از کودهای شیمیایی

- تخصیص سهم به نیتروژن ورودی از طریق آتش سوزی، کشاورزی، منابع شهری و طبیعی به نیتروژن حوضه های آبریز و غلظت نیترات رودخانه

- استفاده آسان و خروجی های گرافیکی مناسب و داشتن فهرست انتخاب برای کاربر جهت مشخص کردن نوع حوضه های آبریز نیمه توزیعی و نیز اعمال تغییر در پارامترهایی مانند حدود مرزهای هر ناحیه، نسبت ضرائب، روابط سرعت جریان آب پایه، نوع زمین مورد استفاده، دما، بارندگی، رسوب نیتروژن

- خروجی متنوع مدل که شامل

۱. مقدار شار روزانه و سالانه نیتروژن در تمام تبدیلات درون زمین های مورد استفاده
۲. محاسبه جریان آب، غلظت های نیتروژن موجود در نیترات و آمونیوم و بیان آن به صورت سری زمانی روزانه در ایستگاههای مختلف مسیر رودخانه
۳. ارائه تصویری کلی از جریان آب و غلظت های نیتروژن در طول مسیر رودخانه بر اساس زمان انتخاب شده

۴. توزیع فراوانی تجمعی جریان آب، نیتروژن و جدول آماری در ایستگاه های مختلف

می توان از این مدل برای نشان دادن تاثیر منابع مانند: شهرنشینی، تغییرات آب و هوایی و مدیریت اجرایی و نیز سیاست های روستایی و بومی مانند: افزایش یا کاهش زمین مورد استفاده، تغییر در مقدار یا توزیع NO_2 و NH_3 بر میزان نیتروژن ورودی به آب رودخانه استفاده کرد. قابلیت بالا و خروجی های متنوع این مدل به دلیل وجود یک ساختار یکپارچه و پیچیده در مدل است که امکان شبیه سازی چرخه نیتروژن را در حوضه آبریز فراهم می کند.

ساختار مدل INCA-N

این مدل از ۵ مولفه اصلی تشکیل شده است. ابتدا به معرفی اجمالی این ۵ مولفه سپس به شرح و نحوه عملکرد این مولفه ها می پردازیم. مولفه های تشکیل دهنده مدل INCA-N عبارتند از

- **GIS interface** یا رابط GIS، این قسمت مدل حدود و مرزهای هر زیر حوضه آبریز را مشخص کرده و مساحت ۶ کلاس دسته بندی زمین های مورد استفاده در هر زیر حوضه را محاسبه می کند.
- **مدل نیتروژن ورودی**، کل نیتروژن ورودی از تمام منابع در هر زیر حوضه آبریز، بکارگیری کود شیمیایی با توجه به نوع زمین مورد استفاده و رسوبات خشک را محاسبه می کند.
- **مدل هیدرولوژی**، شامل مدل های جریان آب حاصل از بارندگی تاثیر گذار بر نواحی واکنش پذیر خاک و آب های زیر زمینی هر زیرحوضه آبریز و درون رودخانه است. این مولفه مقدار شار نیتروژن در حوضه آبریز را محاسبه می کند.
- **مدل فرآیند نیتروژن حوضه آبریز**، تبادلات نیتروژن در خاک و آب های زیر زمینی را شبیه سازی می کند. این مدل شامل پارامترهای، میزان جذب نیتروژن توسط گیاهان و فرآیندهای میکروبی همانند: معدنی سازی، نیتروژن دهی و نیتروژن گیری و غیره است.
- **مدل عملکرد نیتروژن رودخانه**، تبادلات نیتروژن (نیتروژن دهی و نیتروژن گیری) در داخل رودخانه را شبیه سازی می کند. مدل عملکرد نیتروژن رودخانه، نیتروژن خالص به دست آمده از هر زیرحوضه برای محاسبه میزان شار نیتروژن متناظر با هر ناحیه رودخانه محاسبه می کند. این مولفه پس از بدست آمدن نتایج مولفه ۴ شروع به کار می کند (وایتهد و همکاران، ۱۹۹۸).

GIS interface یا رابط *GIS*

GIS interface حدود و مرزهای هر زیر حوضه آبریز را مشخص کرده و مساحت ۶ کلاس دسته بندی زمین‌های مورد استفاده در هر زیر حوضه را محاسبه می‌کند. زمین بر اساس پوشش گیاهی به ۱۸ کلاس تقسیم می‌شود. به منظور اجرای مدل و ساده‌تر شدن ارتباط بین نیتروژن ورودی و مدل‌های فرآیند نیتروژن (مولفه دوم و چهارم) این نواحی به ۶ دسته تقلیل یافته است. ۶ نوع زمین مورد استفاده عبارتست از: جنگل، زمین‌های زراعی، زمین شهری، زمین‌ها با پوشش‌های گیاهی کم و دست نخورده، زمین با پوشش‌های گیاهی کم دست‌خورده عاری از کودهای شیمیایی و زمین با پوشش‌های گیاهی کم دارای کودهای شیمیایی. جزئیات بیشتر هر دسته در جدول (۱) ذکر شده است.

جدول ۱ - کلاس دسته بندی پوشش زمین در بریتانیا و مدل *INCA-N*

INCA-N land use class	ITE land cover map class
Forest	Coniferous woodland (4) Deciduous woodland (5)
Short vegetation (ungrazed)	Bog (1) Dense shrub/heath moor (6) Rough/marsh grass (12) Saltmarsh (13)
Short vegetation (ungrazed not fertilazer)	Braken (2) Head/ moor grass (7) Open shrub heath / moor (11)
Short vegetation (fertilazer)	Managed grassland (10)
Arable	Tilled land (16)

Urban	Inland bare (8) Suburban (15) Unclassified (17) Continuous ueban (18)
-------	--

حوضه آبریز رودخانه توسط این ۶ کلاس تحت پوشش قرار می‌گیرد. سهم هر یک از این کلاس‌ها در هر زیرحوضه محاسبه می‌شود. همچنین مرزهای هر زیر حوضه آبریز بنا بر تحلیل هیدرولوژی استاندارد که بر اساس نقشه‌های توپوگرافی یا بطور خودکار با استفاده از مدل‌های دیجیتالی و الگوریتم‌های موجود GIS مشخص می‌شود. مدل *INCA-N* از اطلاعات زمین مورد استفاده جهت محاسبه نیتروژن ورودی به هر زیر حوضه از طریق استفاده از کودهای شیمیایی، رسوبات خشک و تبدیلات نیتروژن که متناسب با نوع زمین مورد استفاده است، بهره می‌گیرد.

مدل نیتروژن ورودی

در مدل نیتروژن ورودی، نیتروژن ورودی بر اساس منبع ورودی و به تفکیک فرم نیتروژن محاسبه می‌شود. یعنی به‌منظور محاسبه مقدار نیتروژن ابتدا تمام منابع ورودی نیتروژن شناسایی و فرم‌های (گونه‌های) نیتروژن نیز معین شود سپس مقدار نیتروژن ورودی بر اساس این دو فاکتور محاسبه شود. جداسازی منابع ورودی نیتروژن و فرم‌های نیتروژن (اکسایش-کاهش) امکان بررسی تغییرات انتشار نیتروژن در منابع دیگر همانند: تبادلات، قدرت تولید و کشاورزی را فراهم می‌کند. مدل نیتروژن ورودی، مقدار نیتروژن ورودی بوسیله تمامی منابع نیتروژن در هر زیر حوضه آبریز، بکارگیری کود شیمیایی با توجه به نوع زمین مورد استفاده و رسوبات خشک را محاسبه می‌کند. در کشاورزی، نحوه توزیع کودهای شیمیایی به نوع زمین مورد استفاده بستگی دارد و برای هر محصول نیز دارای مقدار معینی است که توسط شرکت سازنده کودهای شیمیایی مشخص می‌شود. چگونگی توزیع کودهای شیمیایی یک مسئله

مهم در مدل نیتروژن ورودی است. در مدل *INCA-N* فرض شده است کودهای شیمیایی در نیمه اول فصل رشد محصولات و بطور یکسان توزیع شده‌اند.

منابع نیتروژنی که با اکسیژن ترکیب شده‌اند یک منبع نیتروژن محسوب می‌شوند. نیتروژن ورودی از طریق این منابع بوسیله مدل *MATADRON*^۱ برآورد می‌شود (موسسه قدرت طبیعی ۱۹۹۷). مدل *MATADRON* مقدار سالانه رسوب مرطوب نیتريت (NO_2) و آمونیوم (NH_3) و ذرات ریز نیترات تا شعاع بیست کیلومتری هر منطقه را با استفاده از میانگین غلظت سالانه و میانگین بارش محاسبه می‌کند. این مدل برای محاسبه رسوبات خشک، از خروجی‌های *GIS* که شامل: کاهش رسوب نیتروژن خشک و مرطوب، رطوبت و خشکی ترکیبات نیتروژن‌دار در هر زیرحوضه آبریز و زمین مورد استفاده، بهره می‌گیرد. شار رسوب خشک آمونیوم و هر یک از فرم‌های اکسید شده نیتروژن بر اساس نوع زمین مورد استفاده و پوشش گیاهی حوضه آبریز با استفاده از بهترین برآوردهای رایج ضرائب سرعت ته نشست خشک (v_g)، سنجش می‌شود. مقادیر این ضرایب در جدول (۲) ذکر شده است (وایتهد و همکاران، ۱۹۹۸). مقدار رطوبت کل و خشک سالانه ترکیبات دارای نیتروژن در بالادست و پایین دست رودخانه و منابع اقلیمی محاسبه می‌شود و کاهش کل رطوبت و خشک ته نشست نیتروژن برای هر زیر حوضه محاسبه می‌شود. بار نیتروژن روزانه (ته نشست رطوبت) بوسیله میانگین چندگانه غلظت سالانه نیتروژن از بارندگی‌های روزانه حاصل می‌شود.

جدول ۲ - ضرائب سرعت رسوب خشک برای زمین‌های مختلف مورد استفاده برای هر یک از فرم‌های نیتروژن

Dry deposition velocity mms^{-1}		
Forest	Arable	Moor

^۱ Model of Atmospheric Transport and Deposition of Reacting Nitrogen

	(INCA-N class 1)	(INCA-N classes 4,5)	(INCA-N clases 2,4)
NO ₂	5	3	2
HNO ₃	40	20	20
HNO ₂	14	10	10
NO ₃	6		1
NH ₃	20	5	15

مدل هیدرولوژی

مدل هیدرولوژی از سه قسمت تشکیل شده است. قسمت اول مدل *MORECS* که رطوبت خاک و مقدار تبخیر را محاسبه می‌کند. از این مدل برای تبدیل داده‌های بارندگی روزانه به سری‌های زمانی بارندگی تاثیرگذار استفاده می‌شود. منظور از بارندگی موثر، آبی بارانی است که در سطح خاک بعد از تبخیر و تعریق باقی می‌ماند و در خاک نفوذ می‌کند. این مدل یک سری زمانی روزانه از کاهش رطوبت خاک تولید می‌کند. این سری زمانی نه تنها به عنوان بارندگی موثر هیدرولوژی تلقی می‌شود بلکه اطلاعات مهمی برای مدل سازی تبدیلات نیتروژن وابسته به رطوبت خاک فراهم می‌کند (موسسه هواسنجی ۱۹۸۱).

قسمت دوم مدل به شبیه سازی تاثیر سطح زمین (توپوگرافی) بر جریان آب می‌پردازد که از آن برای مدل جریان آب در مسیر اصلی رودخانه استفاده می‌شود. هیدرولوژی درون حوضه آبریز با آب‌های ذخیره شده در نواحی واکنش پذیر خاک و آب‌های زیرزمینی با استفاده از جعبه ساده در شکل (۴) نمایش داده شده است. از آنجا که نرخ جریان آب مناسبی بین ناحیه واکنش پذیر خاک و آب‌های زیر-زمینی وجود دارد، فرآیندهای پیچیده تبدیلات نیتروژن در ناحیه واکنش پذیر خاک از طریق حرکت افقی

جریان آب از بین لایه‌های خاک وارد رودخانه یا با جابه جایی عمودی جریان آب وارد آب‌های زیرزمینی شود. رابطه جریان آب و با ناحیه‌های واکنش‌پذیر خاک و آب‌های زیرزمینی به صورت

$$\frac{dx_1}{dt} = \frac{1}{T_1}(U_1 - x_1) \quad \text{نواحی واکنش‌پذیر خاک}$$

$$\frac{dx_2}{dt} = \frac{1}{T_1}(U_8 x_1 - x_2) \quad \text{نواحی آب‌های زیر زمینی}$$

است. x_1, x_2 خروجی جریان آب در خاک و آب‌های زیرزمینی، U_1 خروجی بارندگی موثر هیدرولوژی، T_1 و T_2 زمان ماندگاری تبدیلات نیتروژن در خاک و آب‌های زیرزمینی و U_8 شاخص جریان پایه است. اطلاعات شاخص جریان پایه را می‌توان از کتاب‌های سالانه هیدرولوژی بدست آورد.

سومین مولفه مدل هیدرولوژی، مدل جریان آب رودخانه است. این مدل بر اساس تعادل حجم جریان آب یک رودخانه عمل می‌کند (وایتهد و همکاران، ۱۹۸۱، ۱۹۷۹، ۱۹۹۷). در این مدل تغییرات جریان آب درون هر ناحیه بوسیله یک مدل غیرخطی مشخص می‌شود. در هر ناحیه رابطه بین جریان ورودی (I)، جریان خروجی (Q) و جریان ذخیره (S) به صورت زیر است.

$$\frac{dS(t)}{dt} = I(t) - Q(t)$$

که $S(t) = T(t) * Q(t)$ و T : زمان حرکت پارامتر که به صورت زیر بیان می‌شود.

$$T(t) = \frac{L}{v(T)}$$

که L : طول حوضه و v : متوسط سرعت حوضه (ms^{-1})، رابطه سرعت (v) با جریان خروجی (Q) بصورت

$v(t) = aQ(t)^b$ است. در این معادله a, b مقادیر ثابتی هستند که بوسیله ردیاب‌های تجربی یا بررسی‌های تئوری برآورد می‌شوند.

مدل هیدرولوژی اطلاعاتی در خصوص جریان آب در لایه‌های خاک و آب‌های زیرزمینی و سیستم رودخانه را فراهم می‌آورد. در حالی که معادلات جریان آب حل می‌شود، لازم است معادلات دیفرانسیل تعادل حجم برای نیتروژن موجود در نیترات و آمونیوم هم در لایه‌های خاک و آب‌های زیر زمینی بطور همزمان حل شود. این مدل مقدار شار نیتروژن حوضه آبریز را محاسبه می‌کند.

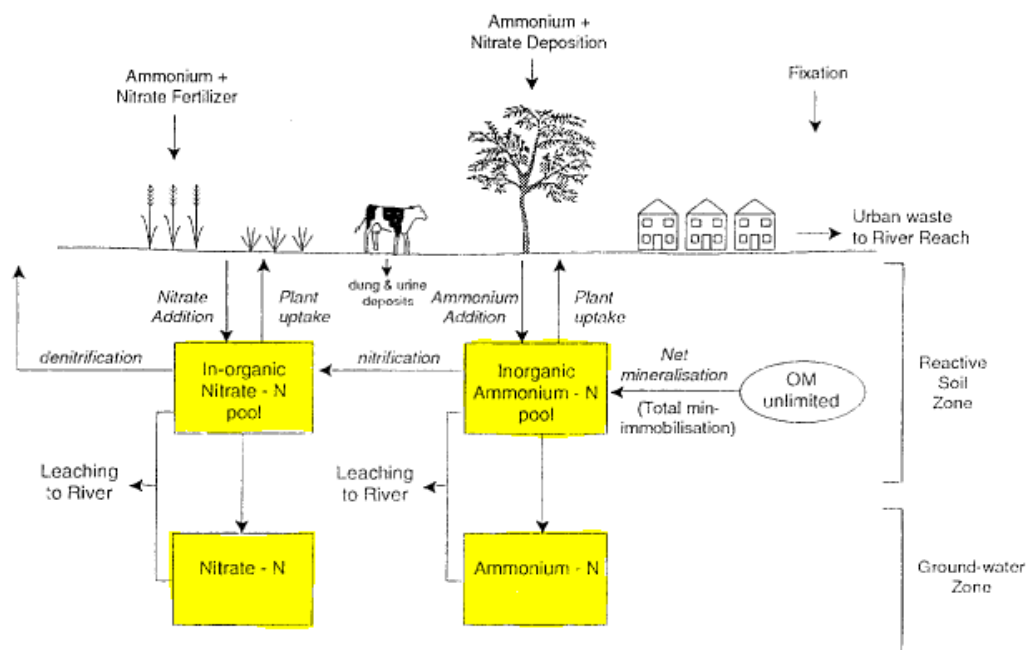
مدل فرآیند نیتروژن حوضه آبریز

همانطور که بیان شد تبدیلات نیتروژن و جریان‌های مختلف موجود در رودخانه به‌وسیله یک مجموعه معادلات در مدل فرآیندهای نیتروژن حوضه آبریز شبیه‌سازی می‌شوند. فرآیندهای کلیدی و تبدیلات نیتروژن نیز در سطح زمین صورت می‌گیرند (شکل ۴). برای این‌گونه تبدیلات نیتروژن پارامترهای میزان جذب نیتروژن موجود در نیترات و آمونیوم، نیتروژن‌گیری و نیتروژن‌دهی، آلی‌سازی و معدنی‌سازی برای هر یک از زیر حوضه‌ها مدل $INCA-N$ مد نظر قرار گرفته است.

مدل فرآیند چرخه نیتروژن حوضه آبریز از یک دسته معادلات کلی به همراه مجموعه پارامترهای خاصی که از کلاس‌های مختلف دسته‌بندی زمین به‌دست می‌آیند بهره می‌گیرد. با تغییر این پارامترها می‌توان شار نیتروژن هر یک از این تبدیلات را برای زمین مورد استفاده با توجه به تجربیات و اطلاعات در دسترس زمین زراعی تنظیم مجدد نمود. فرآیندهای خاص مانند: میزان جذب نیتروژن توسط گیاهان بسیار وابسته به نوع زمین مورد استفاده است. از طرفی نرخ جذب و الگوی جذب نیتروژن توسط گیاهان به صورت فصلی تغییر می‌کند. به‌عنوان مثال: محصولات زراعی که در فصل بهار کشت می‌شوند رشد

فصلی بسیار کمتری نسبت به درخت‌های کاج دارند در عین حال نیتروژن بیشتری نسبت به این نوع جنگل‌ها جذب می‌کنند. همچنین تبدیلات میکروبی نیتروژن در خاک وابسته به دما و رطوبت است که این دو عامل خود، به نوع زمین وابسته هستند. دمای خاک در فصل تابستان توسط پوشش گیاهی تعدیل و کاهش می‌یابد و یا دمای خاک در زمین‌های زراعی بیشتر از جنگل‌های انبوه است. رطوبت خاک در حوضه‌های آبریز مختلف متفاوت است به‌طوری که در نواحی کوهستانی که خاک دارای زغال سنگ و پوسته خاکی است ماندگاری آب نسبت به خاک‌های زه‌کشی شده نواحی پایین دست کشاورزی بیشتر است.

مدل نیتروژن ورودی حوضه آبریز، خروجی **مدل نیتروژن ورودی** که شامل : نیتروژن رطوبت و خشک موجود در نیترات و آمونیوم که از طریق ریزشهای جوی وارد می‌شوند، به همراه ثابت نیتروژن (مقدار ثابت که در هوا وجود دارد) و کود شیمیایی استفاده شده در زمین‌های زیرحوضه آبریز را می‌گیرد. از طرفی شرایط ابتدایی برای غلظت‌های در نظر گرفته شده در آب‌های سطحی و زیر زمینی و نیتروژن موجود در آب‌های جاری خاک و آب‌های زیر زمینی لازم است. می‌توان شرایط آغازین را اندازه‌گیری و یا با استفاده از برآورد مدل و اطلاعات موجود زمین تنظیم مجدد نمود. مدل فرآیند چرخه نیتروژن حوضه آبریز با در نظرگرفتن این شرایط آغازین، گرفتن خروجی‌های **مدل نیتروژن ورودی** و لحاظ کردن پارامترهای میزان جذب نیتروژن موجود در نیترات و آمونیوم، نیترات سازی و نیترات زدایی، آلی سازی و معدنی سازی برای هر یک از زیر حوضه‌ها به شبیه‌سازی تبدلات نیتروژن در خاک و آب‌های زیرزمینی می‌پردازد. فرض شده است در آب‌های زیر زمینی واکنش‌های بیوشیمی رخ نمی‌دهد و تعادل بین نیتروژن موجود در نیترات و آمونیوم برقرار است (وایتهد و همکاران، ۱۹۹۸).



شکل ۴- ورودی‌ها، خروجی‌ها و فرآیندهای نیتروژن در خاک و آب‌های زیر زمینی

در معادلات مدل فرآیند نیتروژن حوضه آبریز تمام فرآیندهای کلیدی کنترل رفتار نیتروژن گنجانده شده است. هر یک از فرآیندهای نیتروژن به صورت یک معادله بیان وحل می‌شود. معادله‌های این فرآیندها به شرح زیر است.

۱. نیتروژن موجود در نیترات در لایه‌های خاک و آب‌های زیر زمینی

$$\frac{dx_3}{dt} = \frac{1}{V_1} (U_3 - x_1 x_3) - C_3 U_7 x_3 + C_6 x_5 - C_1 U_5 x_3 + C_2$$

nitroژن لایه‌های خاک
plant uptake
nitrification
denitrification
fixation

$$\frac{dx_4}{dt} = \frac{1}{V_2} (U_3 x_3 x_1 - x_2 x_4)$$

نیتروژن سفره‌های آب زیرزمینی

۲. نیتروژن موجود در آمونیوم در لایه‌های خاک و آب‌های زیر زمینی

$$\frac{dx_5}{dt} = \frac{1}{V_1} (U_4 - x_1 x_5) - \underset{\text{plant uptake}}{C_{10} U_7 x_5} - \underset{\text{nitrification}}{C_6 x_5} + \underset{\text{mineralisation}}{C_7 U_6} + \underset{\text{immobilisation}}{C_8 x_5} \quad \text{نیتروژن لایه‌های خاک}$$

$$\frac{dx_6}{dt} = \frac{1}{V_2} (U_8 x_5 x_1 - x_2 x_6) \quad \text{نیتروژن سفره‌های آب زیرزمینی}$$

x_3 و x_4 غلظت‌های روزانه نیتروژن موجود در نیترات در خاک و آب‌های زیرزمینی، x_5 و x_6 نیتروژن موجود در آمونیوم خاک و آب‌های زیر زمینی بر حسب mg/l است. V_1 و V_2 حجم آب در خاک و آب‌های زیر زمینی است که از طریق رابطه $V_1 = T_1 x_{1m}$ و $V_2 = T_2 x_{2m}$ به دست می آید. T_1 و T_2 زمان ماندگاری تبدیلات نیتروژن در خاک و آب‌های زیرزمینی و اندیس m به میانگین جریان اشاره دارد. U_8 شاخص جریان پایه و C_3 نرخ جذب روزانه نیترات توسط گیاهان، C_6 نیتروژن‌گیری، C_1 نیتروژن‌دهی، C_2 ثابت نیتروژن غیر بیولوژیکی، C_{10} نرخ جذب آمونیوم توسط گیاهان، C_7 معدنی سازی، C_8 آلی سازی است. U_3 و U_4 بار روزانه نیتروژن نیترات و نیتروژن آمونیوم درون خاک به همراه ترکیبات اضافی ته نشست خشک و مرطوب و پارامترهای کشاورزی (مانند اضافه کردن کود شیمیایی) می‌باشد. تمام پارامترهای مذکور وابسته به دما هستند و از معادله زیر پیروی می‌کنند.

$$C_n = C \cdot 1.407^{(\theta_s - 20)}$$

θ_s : دمای خاک است. این پارامتر نیز به دمای هوا بستگی دارد و از طریق فرمول زیر برآورد می شود.

$$\text{دمای خاک} = \text{دمای هوا} + C_{16} \sin\left(\frac{3}{2}\pi \frac{\# \text{ of day}}{365}\right)$$

که C_{16} بیشترین مقدار اختلاف دمای هوا تابستان و زمستان و مقصود از تعداد روزها، تعداد روزهایی است که فرآیند مورد نظر انجام می‌شود. در مدل $INCA-N$ از متوسط دمای روزانه حوضه آبریز به عنوان

دمای روزانه استفاده شده است. در معادلات فوق فرض شده است فرآیندهای تبادل و تجزیه تنها در خاک صورت می پذیرد و در آبهای زیر زمینی هیچ فرآیندی صورت نمی گیرد.

۳. میزان جذب توسط گیاهان

بیشترین نرخ جذب بر اساس نوع درختان، طول عمر و نوع محصول متفاوت باشد. فرم رایج و عمومی نرخ جذب با استفاده از معادله $Up\ take = C_n U_7 x_n$ محاسبه می شود. C_n نرخ جذب روزانه و x_n غلظت های در دسترس نیتروژن یا آمونیوم یا نیتروژن نیترات و U_7 شاخص رشد فصلی گیاهان است که با استفاده از فرمول زیر بیان می شود.

$$U_7 = .66 + 0.34 \sin\left(2\pi \frac{[\# of\ day - C_{11}]}{365}\right)$$

C_{11} تعداد روزهای مشترک با فصل بهار و مقصود از تعداد روزها، روزهایی است که جذب توسط گیاه انجام می شود.

۴. آلی سازی و معدنی سازی

جریان آمونیوم حاصل از معدنی سازی مواد آلی بصورت $Net\ mineralisation = C_7 U_6 - C_8 x_5$ محاسبه می شود. که C_6 و C_7 نرخهای آلی سازی و معدنی سازی، x_5 غلظت نیتروژن موجود در آمونیوم و U_6 نیز رطوبت خاک که با استفاده از رابطه زیر محاسبه می شود.

$$U_6 = \frac{SMD - SMD_{max}}{SMD_{max}}$$

که SMD مقدار کاهش روزانه رطوبت خاک است.

۵. نیترات سازی

مقدار شار نیترات بوسیله نیترات سازی با استفاده از معادله $Nitrification = C_6 \times x_5$ برآورد می شود که C_6 نرخ نیترات سازی روزانه (day^{-1}) و x_5 غلظت آمونیوم (mg/l) است.

۶. نیترات زدایی

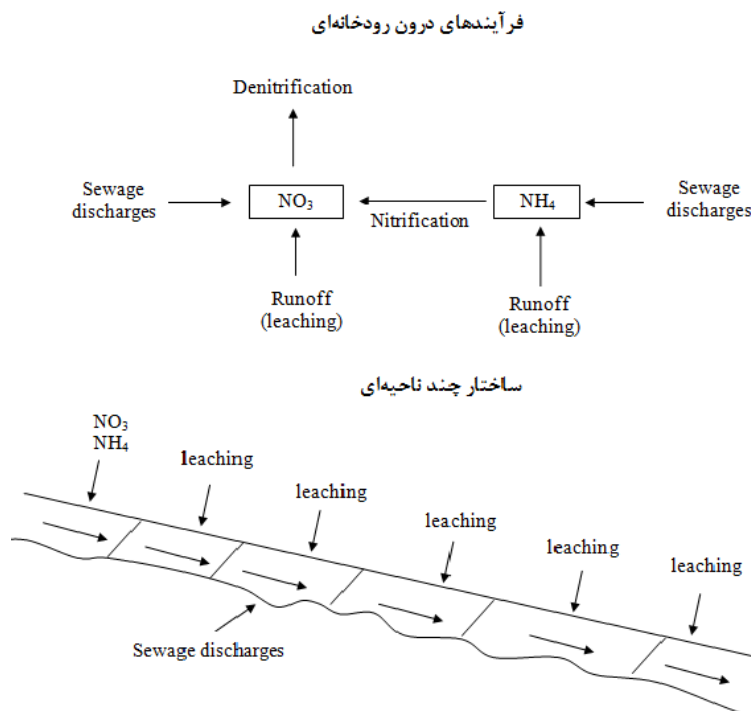
نیترات زدایی تنها در نواحی مرطوب خاک صورت می گیرد. نیترات زدایی از طریق معادله $Denitrification = C_1 U_5 \times x_3$ محاسبه می شود. C_1 نرخ نیترات زدایی (day^{-1}) و U_5 مقدار ۰ یا ۱ را بسته به میزان رطوبت خاک اختیار می کند (۰ برای عدم رطوبت در خاک و ۱ برای وجود شرایط مرطوب) است.

۷. ثابت نیتروژن

ثابت غیر زیستی نیتروژن در جو به صورت N_2 (تخلیه بار الکتریکی) وجود دارد. فرض می شود با یک نرخ ثابت C_2 بر حسب ($mg /l/day$) رخ می دهد. این نرخ در مدل بر اساس ($kg N/ ha/year$) محاسبه می شود.

۹ - ۱ - ۱ مدل عملکرد نیتروژن رودخانه

مدل عملکرد نیتروژن رودخانه به شبیه سازی تبدلات نیتروژن (نیترات زدایی و نیترات سازی) در درون رودخانه می پردازد. این مدل با نیتروژن حاصل از مدل فرآیندهای نیتروژن حوضه آبریز (ناحیه واکنش- پذیر خاک و نواحی آب های زیر زمینی) و نیتروژن ورودی بوسیله تخلیه مستقیم فاضلاب و سیلاب شهری سرو کار دارد. فرآیندهای کلیدی هر ناحیه از رودخانه شامل : نیترات زدایی و نیترات سازی است که در شکل زیر نشان داده شده است.



شکل ۵- ورودی‌ها، خروجی‌ها و فرآیندهای نیتروژن درون رودخانه

در تعادل حجم هر ناحیه باید مازاد نیتروژن موجود در نیترات بالا دست رودخانه و نیتروژن موجود در آمونیوم حوضه آبریز به همراه عوامل ورودی خارجی در نظر گرفته شود. معادلات جریان، شار نیتروژن نیترات و نیتروژن آمونیوم در نواحی رودخانه بوسیله معادلات زیر بیان می شود.

Folw
$$\frac{dx_7}{dt} = \frac{1}{T_3} (U_9 - x_7)$$

Nitrate
$$\frac{dx_8}{dt} = \frac{1}{V_3} (U_{10}U_9 - x_7x_8) - C_{17}x_8 + C_{14}x_9$$

Ammonium
$$\frac{dx_8}{dt} = \frac{1}{V_3} (U_{11}U_9 - x_7x_8) - C_{14}x_9$$

که $V_3 = T_3x_7$ و $T_3 = \frac{L}{v} = \frac{L}{aQ^b}$ است.

T_3 مدت زمان نگهداری است و از مدل غیرخطی شدت جریان بدست می‌آید. U_9 مجموع، جریان بالا دست رودخانه از نواحی قبلی رودخانه، سیلاب زیر حوضه آبریز و تخلیه فاضلاب می‌باشد. U_{10} جریان وزنی نیتروژن موجود در نیترات بالا دست رودخانه بر اساس نیتروژن نیترات نواحی قبلی رودخانه، سیلاب زیر حوضه‌های آبریز و تخلیه فاضلاب بر حسب (mg/L) است. U_{11} جریان وزنی نیتروژن موجود در آمونیوم بالا دست رودخانه بر اساس نیتروژن آمونیوم نواحی قبلی رودخانه، سیلاب زیر حوضه‌های آبریز و تخلیه فاضلاب بر حسب (mg/L) است. x_7 برآورد نرخ جریان پایین دست رودخانه (m^3/s) و x_8 و x_9 غلظت‌های نیترات و آمونیوم در پایین دست رودخانه است. C_{17} و C_{18} پارامترهای نیترات‌سازی و نیترات‌زدایی که وابسته به دما هستند. تاثیر دمای هوا (θ_w) بر دمای آب رودخانه بوسیله معادله زیر بدست می‌آید.

$$C_n = C1.407^{(\theta_w - 20)}$$

شکل (۵) سیستم‌های عملکرد رودخانه و ساختار چند ناحیه‌ای مدل رودخانه را نشان می‌دهد. جریان آب از طریق زیر حوضه‌های آبریز وارد نواحی رودخانه می‌شود و معادلات تعادلی رودخانه به منظور حفظ تعادل حجمی طی مسیر رودخانه حل می‌شود. ورودی‌های اضافی، فاضلاب یا تخلیه مواد صنعتی بطور همزمان وارد ساختار تعادل حجمی می‌شود که این ورودی‌ها از قبل تعیین می‌شوند. در نهایت یک مجموعه ترکیبی از ۹ معادله دیفرانسیل و پارامترهای مرتبط با آن‌ها خواهیم داشت. این مجموعه ترکیبی با استفاده از روش رایج راج کاتا^۱ مرتبه چهار، بصورت همزمان حل می‌شود تا خروجی‌های مورد نظر به دست آید.

^۱ Runge Kutta

با توجه به توضیحات داده شده در خصوص ساختار مدل *INCA-N*، واضح است این مدل در زمره مدل‌های کامپیوتری قرار می‌گیرد. معادلات دیفرانسیل و مشتق‌های و غیره بیان شده در این مدل با استفاده از روش‌های عددی حل می‌شود.

از آنجا که در قاره سبز میزان بارندگی بسیار زیاد است و رودخانه‌های متعددی در این قاره وجود دارد. از طرفی دیگر در کنار بسیاری از این رودخانه‌ها کارخانجات و شهرک‌های صنعتی متعددی مشغول فعالیت هستند. پساب‌های صنعتی این کارخانه‌ها، آب‌های شستشوی زمین‌ها و مراتع کشاورزی مدرن که غالباً از کودهای شیمیایی استفاده می‌کنند، وارد رودخانه‌ها می‌شود. از آنجایی که این رودخانه‌ها مهمترین منبع تامین آب اروپا محسوب می‌شود توجه خاصی به رودخانه و بررسی آلودگی‌های آن می‌شود. مدل *INCA-N* نیز توسط متخصصین همین قاره ساخته شده است و به نحو بسیار گسترده‌ای برای رودخانه‌های متعددی به کار گرفته شده است. از جمله این رودخانه‌ها می‌توان به *Thames*، *Kennet*، *Tywi* در انگلستان، حوضه آبریز *Fuirosos* واقع در ایالت کاتانیا اسپانیا، رودخانه *Mures* رومانی، حوضه آبریز *Cuhna* سائوپائولوی برزیل *INCA-N* به کار رفته است. مدل *INCA-N* نیز در طی سال‌های اخیر ارتقاء یافته و برخی از معادلات فرآیندهای این مدل تغییر و بهبود یافته‌اند (وید و همکاران، ۲۰۰۲). در نسخه های جدید مدل *INCA-N* این اشکالات رفع شده است.

Abstract

The study many of scientific phenomena is impossible in the laboratory. Hence this phenomenon explains by mathematical model and simulate by using computer code. This model consists of ordinary differential equations or partial differential equations set that solved by using numerical methods this is done. The computer program or code that is able to solve numerically this system of equations, say computer models and run repeatedly this model with different values of inputs say computer experiment. The structure of these models has a special complex that one of its agents, is the large number of inputs. The input variables can be single or interaction with other inputs can affect in outputs of model. Thus, identify the influence of inputs and the inputs are less important or insignificant is necessary. This problem was studied In format as a sensitivity analysis of computer model. various methods have been proposed for sensitivity analysis of models. One method them is sensitivity analysis of variance - based. The sensitivity analysis in this method is done by using of sensitivity measure that famous first-order sensitivity index and sensitivity index total. These indicators are based on the concepts of multiple integrals and conditional variance and conditional expectation is provided. The real value of this integral equation due to the absence of explicit input - output model, it is unknown. The problem arises in the estimation or approximation of integrals. For this purpose, several approaches have been proposed by researchers. Variance based method sensitivity analysis represent tow general approach yclept Monte Carlo based method Saltelli and Random balanced design for estimated sensitivity index's. After introducing these two methods, calibration (validation) was performed using the analytic function Sobol. The results showed that these two methods have sufficient accuracy and suitable performance for estimating the sensitivity indices. These methods was used to analyze sensitivity of environmental the INCA-N. The computer model INCA-N simulated the flow of nitrogen input river water. Nitrogen is pollution of water resources that an adverse effect on human health and the animals. In this study, the model as a function of the seven input variables (Nitrogen transformation rates) and a variable output (average annual load of nitrogen entering the river) was considered. Sensitivity analysis of the

Nitrogen Input in Tweed River England, using the Monte Carlo based method of Saltelli and Random balanced design was showed that four variables, the nitrate uptake by plants, the rate of Denitrification, Immobilization and mineralization were the most important factors affecting on This pollutant the water river teewd .



Shahrood University of Technology

Faculty of mathematical

*This thesis submitted in part fulfillment of the degree of Master Science in
mathematical statistics*

*Variance-based methods for sensitivity analysis of deterministic
model output*

By:

Majid Janfada

Supervisor:

Dr. D. Shahsavani

Advisor:

Dr. N. Arghami

Desmber 2011