بنام آنکه جان را فکرت آموخت

۳۲۲۱۴

مرکز اطلاعات علمی ایران
تقدیم معرفی

دانشگاه شهید بهشتی – دانشکده علوم ریاضی
گروه آمار

۱۳۷۹ /۱۰/ ۲۰

پایان نامه کارشناسی ارشد آمار

عنوان:

پیش‌گویی مکانی میزان تراکم آلاینده‌های HC و NOx در هوا: نظریه و کاربرد

استاد راهنما:

دکتر سیامک نوربلوچی

9246

نگارش:

افسانه یزدانی

آذر ۱۳۷۸

۳۲۲۱۴

بسمه تعالی

تاریخ
ردیف
پیوست

صورتجلسه دفاع از پایان نامه

۱۳۷۹/۱۰/۲۰

جلسه هیئت داوران ارزیابی پایان‌نامه آقای/ خانم افسانه یزدانی

به شناسنامه شماره ۱۹۵ صادره از تهران متولد ۱۳۵۲

دانشجوی دوره کارشناسی‌ارشد ناپیوسته رشته آمار محض

باعنوان پیش‌گویی مکانی میزان تراکم آلاینده های HC و NO در هوا : نظریه و کاربرد

به راهنمائی دکتر سیامک نوربلوچی طبق دعوت قبلی در تاریخ ۷۸/۶/۲۴

تشکیل گردید وبراساس رأی هیأت داوری و با عنایت به ماده ۲۰ آئین‌نامه

کارشناسی‌ارشد مورخ ۷۳/۱۰/۲۵ پایان‌نامه مزبور با نمره ۱۹ نوزده تمام

و درجه عالی مورد تصویب قرارگرفت.

۱- دکتر سیامک نوربلوچی استاد راهنما

۲- " عبدالرحیم شهلایی " مشاور

۳- " محمد قاسم وحیدی اصل " داور

۴- " خلیل شفیعی " داور

۵- " محمد ذکایی " داور

تقدیم به آنان که به من آموخته‌اند

چکیده

آلودگی هوا یکی از معضلات بشر در عصر حاضر است. اثرات ویران‌گر این پدیدهٔ دنیای صنعتی بر محیط زیست و انسان تا حدی است که بشر را به تکاپویی سخت برای چاره‌جویی وا داشته است. انسان متفکر امروز در تلاش برای حل این معضل، ضمن جستجوی راهکارهایی برای کنترل کیفیت هوا، نیاز به داشتن تصویری قابل اطمینان از وضعیت آلودگی هوا را احساس می‌کند. شبکه‌های سنجش آلودگی هوا در شهرهای بزرگ دنیا به منظور پاسخگویی به این نیاز تأسیس شده‌اند.

هر شبکهٔ سنجش آلودگی هوا شامل ایستگاه‌هایی است که در هر یک از آنها میزان تراکم آلاینده‌ها اندازه‌گیری و ثبت می‌شود. گاه به دلیل بروز اشکالات فنی و یا در دست نبودن ابزار مورد نیاز، امکان اندازه‌گیری برخی از آلاینده‌ها در تعدادی از ایستگاه‌ها وجود ندارد. در این صورت لازم است میزان تراکم این آلاینده‌ها با روشی مناسب برآورد شوند. برای به دست آوردن برآوردهایی نزدیک به واقعیت، می‌توان از روش‌های آماری مناسب استفاده کرد.

مشاهداتِ ایستگاه‌های سنجش آلودگی هوا، ماهیتی مکانی دارند. به این معنا که میزان تراکم آلاینده‌ها در مکان‌هایی که به فاصلهٔ کمتری از هم قرار گرفته‌اند، مشابه است. از این رو نمی‌توان برای تحلیل آماری از روش‌هایی که مشاهدات را مستقل از هم فرض می‌کنند، استفاده کرد. «آمار مکانی» با ارائه روش‌هایی مفید، امکان تحلیل مشاهدات وابستهٔ مکانی را فراهم می‌کند.

در این پایان‌نامه ابتدا روشی برای تحلیل و پیش‌گویی مکانی معرفی می‌گردد و سپس از آن برای برآورد تراکم میزان آلاینده‌های هوا استفاده می‌شود.

کلید واژه‌ها: تحلیل مکانی، پیش‌گویی مکانی، کریگیدن، آلودگی هوا.

فهرست

پیش‌گفتار ۱

فصل ۱— آشنایی ۳

۱.۱ تاریخچه ۳

۲.۱ آشنایی ٦

فصل ۲— تحلیل مکانی یک‌متغیره ۳۰

۱.۲ مدل‌سازی داده‌های مکانی ۳۱

۲.۲ کریگیدن: بهترین پیش‌گوی خطی نااریب برای داده‌های مکانی ۳۳

۳.۲ کریگ زمین‌آماری: پیش‌گویی بر اساس نیم‌تغییرنگار ۴۲

۴.۲ خطای اندازه‌گیری و اثر قطعه‌ای ۵۰

۵.۲ مدل‌هایی برای تابع‌های کوواریانس و نیم‌تغییرنگارها ۵۸

٦.۲ برآورد تابع‌های کوواریانس و نیم‌تغییرنگار ٦٦

۷.۲ اثر کوواریانس‌های برآورد شده بر پیش‌گویی ۸۲

فصل ۳– پیش‌گویی مکانی چندمتغیره ۸۵

۱.۳ مدل‌سازی داده‌های مکانی در حالت چندمتغیره ۸۶

۲.۳ تعمیم مفهوم وابستگی مکانی در حالت چندمتغیره ۸۸

۳.۳ پیش‌گویی مکانی در حالت چندمتغیره ۹۰

فصل ۴– پیش‌گویی مکانی و آلودگی هوا ۹۷

۱.۴ معضل آلودگی هوا ۹۸

۲.۴ تحلیل مکانی داده‌های میزان آلودگی هوا ۱۰۵

پیوست‌ها ۱۲۹

پیوست ۱– آماره‌های خلاصه، بررسیِ فرض نرمال بودن متغیرهای وابسته

پیوست ۲– مراحل برازش مدل و آزمون نرمال بودن خطاها

پیوست ۳– بررسی ایستاییِ فرآیند خطا

واژه‌نامه ۱۶۰

نام اشخاص ۱۶۵

مراجع ۱۶۷

پیش‌گفتار

امروزه آمار توانسته با آشکارتر ساختن توانایی‌های گسترده‌اش، جایگاه خود را به عنوان ابزاری کارآ در تصمیم‌گیری‌ها بیابد. حضور آمار در مراحل مختلف پژوهش‌ها، از جمع‌آوری داده‌ها گرفته تا بیان یافته‌ها، به خوبی به چشم می‌خورد. اینک آمار پا را از تلخیص و نمایش داده‌ها فراتر نهاده و با تکیه بر مدل‌های آماریِ مناسب، جسورانه به بررسی، مطالعه و توصیف پدیده‌ها و گاه به پیش‌گویی رفتار پدیده‌ها در زمان و مکانی خاص می‌پردازد. از این رو است که مسألهٔ «یافتن مناسب‌ترین مدل» به یکی از مسائل تفکربرانگیز در بین آمارشناسان تبدیل شده است.

با توجه به اهمیت «مدل‌سازی آماری»، تلاش آمارشناسان همواره بر این است که با تطبیق صبورانهٔ «ماهیت مشاهدات به دست آمده از یک پدیده» با «فرض‌های لحاظ شده در مدل آماری»، مناسب‌ترین مدل را برگزینند. ساده‌ترین مدل آماری، مدلی است که با تکیه بر صحت فرض «استقلال و هم‌توزیع بودن مشاهدات» ساخته می‌شود، لیکن این مدل پاسخگوی تمام نیازهای یک آمارشناس نیست. با وجودیکه فرض استقلال راه را برای بکارگیریِ ساده‌تر بخش اعظمی از نظریه آمار و ریاضی هموار می‌کند، مدل‌هایی که وابستگی آماری را در نظر می‌گیرند، واقع‌بینانه‌تر هستند. به عنوان مثالی از مدل‌هایی از این دست، می‌توان به مدل‌های سری‌های زمانی اشاره کرد که در آنها وابستگیِ زمانیِ مشاهداتی که از نظر زمانی به هم نزدیک هستند، به خوبی لحاظ می‌شود. در محدودهٔ سه‌بعدیِ مکان و جریان تک‌بعدیِ زمان، رویارویی آمارشناس با داده‌های وابسته، تنها به داده‌هایی که از نظر «زمانی» به هم وابسته‌اند، محدود نمی‌شود. مواقعی هست که آمارشناس با داده‌هایی سروکار دارد که از موقعیت‌های مکانی مختلف جمع‌آوری شده‌اند و مفهوم همبسته بودنِ داده‌هایی که در فاصله نزدیکی از هم قرار دارند، منطقی به نظر می‌رسد. در این مواقع لازم است با توجه به ساختار همبستگی داده‌ها، مدل مناسبی که در

آن فرض‌های مناسب لحاظ شده است، ساخته شود.

در این پایان‌نامه ضمن معرفی مدل مناسب برای تحلیل داده‌های مکانی، روشی برای پیش‌گویی مکانی نیز ارائه می‌شود. در پایان، با بکارگیریِ روش معرفی شده بر روی داده‌های مربوط به «میزان تراکم اندازه‌گیری شدهٔ سه آلایندهٔ NO_x، SO_2 و HC در ۳۰ ایستگاه سنجش آلودگی هوا» میزان تراکم NO_x و HC برای سه ایستگاه پیش‌گویی می‌شود.

فصل ۱

آشنـــایی

۱.۱ تاریخچه

ظاهراً آمار برای نخستین بار در کِسوت «نقشه‌داده‌ها» به عرصهٔ داده‌های مکانی گام نهاد. ادموند هالی جغرافی‌دان و ستاره‌شناس انگلیسی در مقاله‌ای که به سال ۱۶۸۶ تحت عنوان «بررسیِ سابقهٔ وزش بادهای بسامان و بادهای موسمی در دریاهای مناطق استوایی و نزدیک به مدارگان؛ تلاشی برای یافتن علت فیزیکیِ وزش این بادها» نوشت، به منظور یافتن علتی فیزیکی برای وزش و جهتِ وزش بادهای موسمی و بسامان در مناطق استوایی و نزدیک مدارگان، جهت وزش بادها را بر روی نقشهٔ زمین رسم کرد و به این ترتیب نقشه‌داده‌ها را برای دستیابی به هدف خود به کار برد.

مدتها بعد مدل‌های آماری ظهور کردند، برای مثال استودنت در تحقیقی پیرامون «توزیع ذرات معلق در مایع» که نتایج آن را در مقاله‌ای تحت عنوان «خطای شمارش با

دستگاه هماسیتومتر[1]» ارائه نمود، به جای تحلیل موقعیت مکانی ذرات، تعداد ذرات موجود در هر واحد را شمارش کرد. وی دریافت که اگر هر میلی‌متر مربع را به چهار صد مربع کوچک تقسیم کند، تعداد ذرات (گلبول‌ها) در هر مربع از توزیع پوآسن پیروی می‌کند.

ر. ا. فیشر کاملاً از وجود وابستگی مکانی در آزمایش‌های کشاورزی آگاه بود زیرا برای از بین رفتن این وابستگی‌ها، کرتهای آزمایشی را از فاصله‌های بسیار دور از یکدیگر برمی‌گزید. او در دهه‌های ۱۹۲۰ و ۱۹۳۰، زمانی که در آزمایشگاه روتامستد[2] در انگلستان کار می‌کرد، قواعد «تصادفی‌سازی»، «بلوک‌بندی» و «تکرار» را وضع کرد. چنانچه پِترز به سال ۱۹۳۸ در مقالهٔ خود تحت عنوان «مقایسهٔ مزیتهای ترتیب سیستماتیک و تصادفی در طرح آزمایش‌های کشاورزی و زیستی» بیان کرد: «تصادفی‌سازی علاوه بر کنترل اریبی ناخواسته، اثر همبستگی مکانی را نیز کاهش می‌دهد، لیکن آن را کاملاً از بین نمی‌برد.»

فِیرفیلد اسمیت در سال ۱۹۳۸ در مقاله‌ای به عنوان «یک قانون تجربی برای توصیف ناهمگنی در میزان عملکرد محصولات کشاورزی» نوشت ضمن اینکه درگیر انتخاب ابعاد کرت‌های آزمایشی بوده، دریافته که هرچه اندازهٔ کرت بزرگ‌تر می‌شود، واریانس خطا به میزان کمی کاهش می‌یابد. با وجود تجربی بودنِ تحلیل وی، بیان دقیقِ مسأله، وجود همبستگی مکانی در آزمایش‌های کشاورزی را القاء می‌کند. مدتها بعد مدل‌های مناسب برای چنین پدیده‌هایی ظهور کردند و برای نخستین بار ویتل به سال ۱۹۴۵ در مقاله‌ای به عنوان «فرآیندهای ایستا در سطح مستوی» در این باره سخن به میان آورد.

امروزه می‌توان حضور آمار را در زمین‌شناسی، اکولوژی و علوم زیست‌محیطی مشاهده کرد. اینک نیاز به مدل‌ها و روش‌های آماری جدیدی که پاسخگوی پرسش‌های

[1] هماسیتومتر (Hemacytometer) دستگاهی است که برای شمارش گلبول‌های خون به کار می‌رود.

[2] Rothamsted Experiment Station

جدید در این زمینه باشند، به خوبی احساس می‌شود. بسیاری از مسائل و مشکلات مطرح شدهٔ جدید (مانند حفظ منابع طبیعی و حفاظت از محیط زیست) ماهیتی مکانی دارند. حال «تحلیل داده‌های مکانی» با گرد هم آوردن تمام روش‌هایی که در حیطهٔ مطالعهٔ متغیرهایی با ماهیت مکانی، سخنی برای گفتن دارند، به یکی از مباحث مهم آمار بدل شده است. یکی از زیربخش‌های «تحلیل داده‌های مکانی» (که این پایان‌نامه از آن مدد می‌جوید)، «زمین آمار» است.

پیدایش و گسترش کاربرد عملیِ زمین‌آمار به دور از چشم آمارشناسان صورت گرفت. از این‌رو عجیب نیست که زمین‌آمار اصطلاحاتی را به کار می‌برد که به گوش آمارشناسان کلاسیک ناآشنا است. هوِبر و واتِرمایِر نخستین افرادی بودند که از روش‌های آماری در علوم زمین استفاده کردند و با مطالعات خود توانستند الگوهای توزیع طلا در معادن آفریقای جنوبی را شناسایی کنند. این دو نتایج تحقیقات خود را به سال ۱۹۱۹ در مقاله‌ای منتشر کردند. بعدها تروسکات در سال ۱۹۲۹ روشی برای محاسبه تخمین میانگین ذخیره ارائه کرد. سی‌شِل در سال ۱۹۴۷ مدلی برای توزیع طلا معرفی کرد و مسأله خطاهای سیستماتیک در نمونه‌گیری را مطرح ساخت. وی همچنین فرمول و جدولی را برای محاسبه دقت میانگین محلی متغیرها و فاصله اطمینان متغیرها ارائه کرد. در همین راستا مطالعات تکمیلی گسترده‌ای توسط راس، کریگ، دویس و دیگران انجام شد. مدتی بعد با مشاهدهٔ نوعی رابطه بین عیار یک کانی در بخش‌های مختلف یک منطقه معدنی، تلاش‌هایی برای یافتن «ارتباطِ مکانی» نمونه‌ها شروع شد که می‌توان در این زمینه به کارهای ویتن، کرومباین، گریفیث، کوچ، لینک و هاربو اشاره کرد. این تلاش‌ها در دهه‌های ۱۹۵۰ و ۱۹۶۰ منجر به ابداع و تکمیل روش «تجزیه و تحلیل سطوح روند» شد. گام تکمیلی بعدی برای گسترش استفاده از روش‌های آماری در علوم زمین را ماترون در دهه ۱۹۶۰ برداشت و شاخه جدیدی از علم آمار را به نام

«زمین‌آمار» پایه گذاشت. وی در سال ۱۹٦۵ بحث متغیرهای ناحیه‌ای و برآورد آنها و در سال ۱۹٦۹ بحث کریگ عام را مطرح ساخت و در این زمینه از کمک‌های دی. جی. کریگ بهره جست. دیوید در سال ۱۹۷۷ در کتاب «برآوردهای زمین‌آماریِ ذخیره طلا» و ژورنل و هیوجبرگتز در سال ۱۹۷۸ در کتاب «زمین آمار در معدن» به طور مفصل روشهای زمین آمار را با اصطلاحات خاص زمین‌آمار شرح دادند. ریپلی به سال ۱۹۸۱ در کتاب «آمار مکانی» دریچهٔ دیگری را گشود و با زبانی آماری سخن گفت. وی نظرات مربوط به پیش‌گویی فرآیندهای تصادفی را که ارتباط نزدیکی با سری‌های زمانی دارند، به کار گرفت.

اینک پس از گذشت سال‌ها و به مدد تلاش بسیاری از پژوهشگران، زمین‌آمار به عنوان زیر بخشی از «روش‌های تحلیل داده‌ها مکانی» و به عنوان ابزاری کارآ برای مطالعه برخی داده‌های مکانی —که لزوماً در گسترهٔ علوم زمین قرار ندارند— مطرح شده است.

۲.۱ آشنایی

۱.۲.۱ داده‌های مکانی

در برخی از رشته‌های علوم مانند زمین‌شناسی، کشاورزی، همه‌گیرشناسی، ستاره‌شناسی، علوم جوّی، محیط زیست و ... پژوهشگران با مجموعه‌ای از داده‌ها سروکار دارند که مقادیر متغیر (یا متغیرهای) آن از موقعیت‌های مکانی مختلف جمع‌آوری شده و «مکان مشاهده» به عنوان یکی از عوامل تغییر در مقدار متغیر (یا متغیرها) مطرح است. این

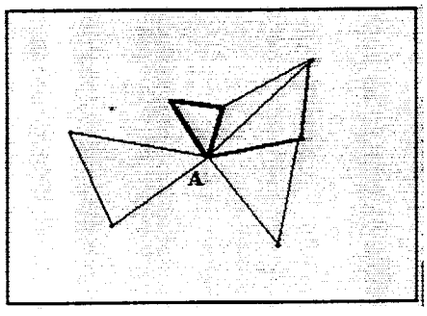
داده‌ها که آنها را «مجموعه داده‌های مکانی» می‌نامند، علاوه بر مقادیر مربوط به متغیر (یا متغیرهای) مورد بررسی، اطلاع مربوط به «مکانِ مشاهده» را نیز شامل می‌شوند.

کاملاً منطقی به نظر می‌رسد که مشاهدهٔ مربوط به یک مکان به نوعی به مشاهدات سایر مکان‌ها وابسته باشد. این امر در داده‌های زمانی نیز به نوعی دیده می‌شود، با این تفاوت که در داده‌های زمانی به دلیل جریانِ تک‌بعدیِ زمان، مشاهدات به طور متوالی و به ترتیب جمع‌آوری می‌شوند و در نتیجه وابستگی به صورت «پیاپی» بروز می‌کند و وابستگی مشاهداتی که بر روی خط زمان در فاصلهٔ نزدیکی نسبت به هم قرار دارند، بیشتر است. لیکن در داده‌های مکانی، مکان‌های مشاهده در فضایی دو یا سه بعدی واقع شده‌اند و در نتیجه مشاهدات لزوماً به طور منظم و متوالی روی یک خط قرار ندارند، مفهوم ترتیب مشاهده برای داده‌های مکانی معنایی ندارند و وابستگی بین مشاهدات نه در یک جهت بلکه در تمام جهت‌ها وجود دارد.

در داده‌های مکانی (مشابه داده‌های زمانی)، مشاهداتی که از مکان‌های نزدیک به هم جمع‌آوری می‌شوند، وابسته‌تر هستند. فاصلهٔ مکان‌های مشاهده را می‌توان با معیارهای متفاوتی مانند: فاصلهٔ اقلیدسی، زمان یا مقدار پول صرف شده برای رفتن از یک مکان به مکان دیگر و ... اندازه گرفت. گاهی نیز از روش‌های خوشه‌بندی برای تعیین اینکه کدام مکان‌ها به هم نزدیک‌تر هستند، استفاده می‌شود. در این مواقع لازم است خوشه‌بندی با استفاده از معیاری مناسب انجام گیرد تا وابستگی مکانی موجود بین مشاهدات نادیده گرفته نشود.

برای روشن‌تر شدن آنچه که دربارهٔ «وابستگی مکانی مشاهداتِ جمع‌آوری شده از چند مکان» گفته شد، به شکل (۱—۱) توجه کنید. در این شکل نقاط، نشان‌دهندهٔ «مکان‌های مشاهده» هستند. وابستگی مشاهدات چند مکان، با رسم خطوطی بین مکان‌ها نشان داده شده است که خطوط پررنگ‌تر بیانگر وابستگی بیشتر و خطوط

کم‌رنگ‌تر بیانگر وابستگی کمتر هستند. مشاهده می‌شود که مکان‌های مشاهده لزوماً بر روی یک خط و به طور متوالی قرار ندارند و وابستگی در تمام جهت‌ها وجود دارد. (به نقطه A توجه کنید)



شکل ۱.۱ وابستگیِ مکانی

۲.۲.۱ انواع داده‌های مکانی

هر مجموعه دادهٔ مکانی با توجه به نوع متغیر (یا متغیرهای) آن و ویژگی‌های مکان‌های مشاهده و ثبت داده‌ها در یک گروه قرار گرفته، بررسی و تحلیل می‌شود. متغیر ممکن است پیوسته یا گسسته و مکانِ مشاهده نیز ممکن است پیوسته یا گسسته، نقطه‌ای یا منطقه‌ای، منظم یا نامنظم باشد. وقتی مکانِ مشاهده از مجموعه‌ای پیوسته انتخاب می‌شود، مکان «پیوسته» و وقتی از مجموعهٔ چند نقطه انتخاب می‌شود، مکان «گسسته»