

PROSTOROVÁ AUTOKORELACE V ANALYTICKÉ CHEMII

JIRÍ MILITKÝ , Katedra textilních materiálů, Technická universita v Liberci,
461 17 Liberec

MILAN MELOUN , Katedra analytické chemie, Universita Pardubice, Pardubice

1. Úvod

Autokorelace se běžně používá při analýze časových řad. Jde obecně o metodu zkoumání podobnosti nebo rozdílnosti sledované proměnné mezi zvolenými intervaly. Tyto intervaly mohou být časové i prostorové. Prostorová autokorelace umožňuje testování zda sledovaná proměnná v jedné lokalitě (oblasti, místě resp. bodu) O_i je závislá na hodnotách této proměnné v oblastech sousedících. Prostorová autokorelace tedy definuje míru a úroveň vzájemné závislosti mezi jednotlivými oblastmi v ploše nebo prostoru. Pozitivní autokorelace ukazuje na podobnost sousedních oblastí a negativní na výraznou nepodobnost. Prakticky všude, kde se studují stochastické procesy rozložené v ploše nebo prostoru se narazí na problém prostorové autokorelace. Prostorová autokorelace je typická u biologických problémů, kde jsou různé organismy rozmístěny v lokalitách na zemi. Prostorová autokorelace je zde způsobena dvěma zdroji příčin:

- lokální působení různých vnějších faktorů ovlivňujících růst a zánik živých organismů
- lokální variace vnitřních proměnných biologických systémů.

Standardně se prostorová autokorelace využívá v geografii, ekologii ekonomice, geografii a zemědělství. V analytické chemii se problémy související s prostorovou autokorelací vyskytují např. v těchto úlohách:

- vyjádření výsledků analýz s ohledem na neregularitu materiálových charakteristik resp. poruch v ploše nebo prostoru,
- hodnocení vlivu nehomogenity objemných vzorků na výsledek analýz,

Obecně vede na úlohy prostorové autokorelace aplikace analytických metod v oblasti biologie, medicíny, geologie, ale také hutnictví, výroby transistorů řízení jakosti atd.

V této práci jsou uvedeny základní myšlenky autokorelace vyjádřené přes statistiky založené na smíšených momentech. Účelem je ukázat jak jednoduše se dá postupovat při analýze tohoto typu dat. Detaily a některá zobecnění obsahuje kniha [1].

2. Základní způsoby vyjádření prostorové autokorelace

Prostorová autokorelace se využívá pro hledání závislostí (resp. prostorové blízkosti) mezi množinou N lokalit (ale také bodů, míst, poloh) $O_1, O_2 \dots O_N$ a hodnotami $p_1, p_2 \dots p_N$ nějaké proměnné p měřené na jednotlivých místech. Lokality mohou být charakterizovány svojí polohou na mapě, těžištěm atd. Pokud jde o body bývají charakterizovány svými souřadnicemi, tj. $O_i = (x_i, y_i)$. Jednotlivé lokality mohou tvořit pravidelnou nebo nepravidelnou síť. Podobně i bodové množiny mohou být pravidelné nebo jinak uspořádané. Veličina p může být kardinální nebo nominální (např. přítomnost nebo nepřítomnost stopového prvku). Omezme se nejdříve na kardinální data (např. koncentrace), i když se dá prakticky stejný aparát použít také pro nominální data..

Prostorová „blížkost“ (proximita) mezi hodnotami proměnné p se pro kardinální data definuje pomocí matice C s prvky C_{ij} vyjadřujícími různé typy vzdáleností mezi p_i a p_j . Přirozená je volba

$$C_{ij} = (p_i - p_j)^2 \quad (1)$$

Standardně lze také s využitím průměru $\bar{p} = \frac{1}{N} \sum p_i$ definovat C_{ij} ve tvaru

$$C_{ij} = (p_i - \bar{p}) * (p_j - \bar{p}) \quad (2)$$

Mezi další možnosti patří Mahalanobisova vzdálenost analogie korelace, Manhatanská vzdálenost, L_p norma atd.

Prostorová souvislost mezi místy se vyjadřuje pomocí prostorové váhové matice W s prvky W_{ij} . Obecně prvek W_{ij} vyjadřuje souvislost mezi lokalitou O_i a O_j . Prostorová souvislost se často chápe lokálně a váhová matice má prvek $W_{ij} = 1$, pokud se O_i a O_j považují za blízké sousedy resp. $W_{ij} = 0$.

Pro regulární síť se lokální prostorová souvislost často definuje podle možností pohybu figur na šachovnici. Při strategii **věže (rook)** existují pro každou O_i čtyři sousedi (dva v jednom a dva v druhém směru sítě kolmo na sebe). Při strategii **střelce (bishop)** existují pro každou O_i čtyři sousedi (dva v jednom a dva v druhém směru diagonálně vzhledem k orientaci sítě). Při strategii **krále (king)** existuje pro každou O_i šest sousedů jako kombinace předchozích dvou strategií. Schematicky lze tyto typy prostorové souvislosti vyjádřit pomocí obr. 1. Zde symbol O definuje lokalitu a symboly XXX nejbližší sousedy pro které jsou váhy W jednotkové.

| Rook | | | Bishop | | | | King | | | |
|------|------|------|--------|------|---|------|------|------|------|------|
| | XXXX | | | XXXX | | XXXX | | XXXX | XXXX | XXXX |
| XXXX | O | XXXX | | | O | | | XXXX | O | XXXX |
| | XXXX | | | XXXX | | XXXX | | XXXX | XXXX | XXXX |

Obr 1. Strategie výběru nejbližšího souseda

Pro obecně neregulární uspořádání se běžně volí globální prostorová souvislost úměrná např. reciproké hodnotě vzdálenosti mezi O_i a O_j tedy např.

$$W_{ij} = \frac{d_{\min}}{d(O_i, O_j)} = \frac{d_{\min}}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}} \quad (3)$$

Zde d_{\min} je minimální hodnota daného typu vzdálenosti. Často se také provádí standardizace vah, tak aby součet všech prvků matice W byl roven jedné. Je zřejmé, že $W_{ii}=0$ a proto se často definitoricky zavádí i $C_{ii}=0$.

Všechny míry lokální autokorelace využívají definice smíšeného momentu G , pro který platí

$$G = \sum_i \sum_j W_{ij} * C_{ij} \quad (4)$$

Tento moment se často označuje jako lineární model prostorové autokorelace. Statistické chování G se obvykle sleduje na základě znáhodnění tj. simulací, kdy se počítá G pro všechny permutace sloupců a řádků matic **W** a **C**.

Různé indexy a míry autokorelace jsou standardizované varianty statistiky G. Známý Moranův index má tvar

$$I = \frac{N}{\sum_i \sum_j W_{ij}} * \frac{\sum_i \sum_j W_{ij} * (p_i - \bar{p}) * (p_j - \bar{p})}{\sum_i (p_i - \bar{p})^2} \quad (5)$$

Je zřejmé, že index I může být nulový jen pro případ, kdy $(p_i - \bar{p}) * (p_j - \bar{p}) = 0$.

Negativní prostorová autokorelace je pro $I > -1/(N-1)$ a pozitivní prostorová autokorelace je pro $I < -1/(N-1)$. Významnost indexu I se dá posuzovat na základě předpokladu přibližné normality nebo znáhodnění. V obou případech platí, že střední hodnota je rovna $E(I) = -1/(N-1)$. Pro rozptyl za předpokladu normality vyjde

$$D(I) = \left[\frac{1}{S_0^2 * (2N-1)} (N^2 * S_1 - N * S_2 + 3 * S_0^2) \right] - E(I)^2$$

Je použito označení $S_0 = \sum_i \sum_j W_{ij}$, $S_1 = \frac{1}{2} \sum_i \sum_j (W_{ij} + W_{ji})^2$ a $S_2 = \sum_i (W_{i*} + W_{*i})^2$. Zde W_{i*} označuje i-tý řádek a W_{*i} označuje i-tý sloupec matice **W**. Pro symetrické matice **W** je $S_2 = 2 * S_1$. Standardizovaná náhodná veličina

$$z = \frac{I - E(I)}{\sqrt{D(I)}} \quad (6)$$

má pak přibližně normované normální rozdělení takže lze snadno testovat významnost Moranova indexu I.

Druhým známým indexem autokorelace je Gearyho c, pro které platí, že má tvar

$$c = \frac{N-1}{2 * \sum_i \sum_j W_{ij}} * \frac{\sum_i \sum_j W_{ij} * (p_i - p_j)^2}{\sum_i (p_i - \bar{p})^2} \quad (7)$$

Tento index se pohybuje v rozmezí 0 až 2. Negativní prostorová autokorelace je pro $c > 1$ a pozitivní prostorová autokorelace je pro $c < 1$. Střední hodnota je rovna 1 a pro rozptyl za předpokladu přibližné normality vychází

$$D(c) = \left[\frac{1}{S_0^2 * 2 * (N+1)} ((N-1) * (2 * S_1 + S_2) - 4 * S_0^2) \right] \quad (8)$$

Další postup testování významnosti je stejný jako u Moranova indexu.

3. Zobecněné vyjádření prostorové autokorelace

Existuje celá řada zobecnění prostorové autokorelace. Mezi jednoduché patří např. tzv. čtyřbodové statistiky $q(i, j, k, l)$ protože ke dvěma lokalitám O_i, O_j jsou k dispozici prvky W_{ij}, W_{ji}, C_{ij} . W_{ji} které nemusí být nutně stejné. Je zřejmé, že statistika $G = \sum \sum q(i, j, i, j)$.

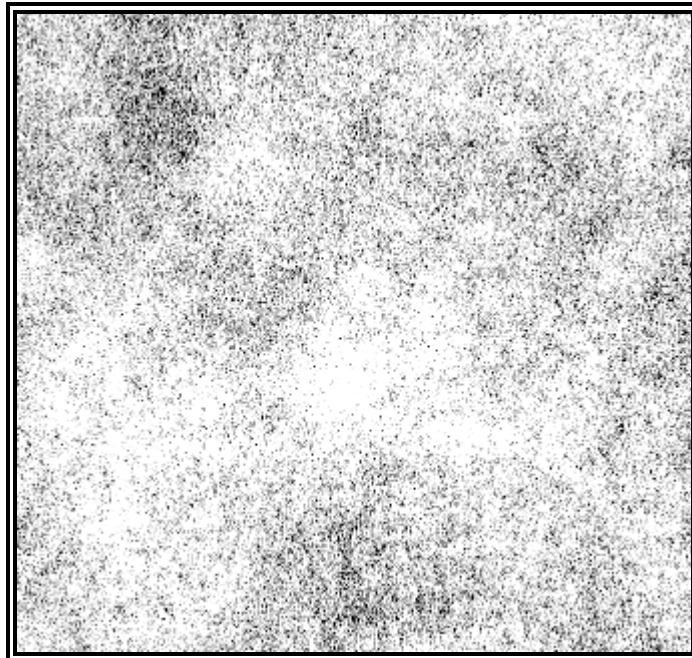
Praktické možnosti použití čtyřbodové statistiky demonstruje práce [2]. Poměrně jednoduše lze také zavést charakteristiky lokální autokorelace (viz [3]), kdy se při výpočtu momentů sumuje pouze přes jeden index. Tedy např. lokální statistika G_i má tvar

$$G_i = \sum_j W_{ij} * C_{ij} \quad (9)$$

Jednotlivá G_i mohou být použita pro indikaci lokálních prostorových shluků. Druhou možností je identifikace vybočujících a vlivných lokalit na základě porovnání G_i a G/N . Analogicky jako statistiku G lze definovat i lokální Moranovy I_i a Gearyho c_i indexy. Podrobnosti obsahuje práce [3].

4. Příklad

Technická aplikace chemicky pojené textilie obchodního názvu Perlan, například v elektrotechnickém průmyslu pro výrobu hydroizolačních pásek, je podmíněna zaručenou spolehlivostí v podélné pevnosti a tažnosti. V upraveném stavu jsou u této netkané textilie požadovány zaručené hodnoty v elektrické průrazné pevnosti, nasákavosti a pevnosti v přetrhu. Všechny tyto vlastnosti souvisejí se stejnoměrností uspořádání vlákenných složek a pojiva. Nestejnoměrnost vzhledu je dobře patrná z obr.2.



Obr. 2 Stejnóměrnost vlákenné vrstvy v pojené textílii plošné hmotnosti 60 g m^{-2}

Účelem byl popis kolísání plošné hustoty této textilie. Vzorok pro gravimetrická měření byly odebrány ve tvaru čtverců rozměrů $100 \times 100 \text{ mm}$. Tyto vorky byly rozděleny na rektangulární síť o velikosti celý $10 \times 10 \text{ mm}$. Pro textílii o plošné hmotnosti 60 g m^{-2} má cela plošného obsahu $S_j = 100 \text{ mm}^2$ hmotnost kolem 6 mg . Kontrola přesnosti přípravy cel, byla

provedena na náhodném výběru 25 vzorků.. Relativní chyba velikosti cely se pohybovala od 0.88% do 1.22%. Hmotnost každé cely m_{ij} byla určena jako průměr z pěti vážení. Maximální relativní chyba vážení u vzorku 60 g/m^2 byla 1.606%.

V tab. 1 jsou uvedeny hodnoty m_{ij} pro vzorek pojené textilie plošné hmotnosti 60 g m^{-2}

Tabulka 1 Průměrné hmotnosti cel pojené textilie plošné hmotnosti 60 g m^{-2}

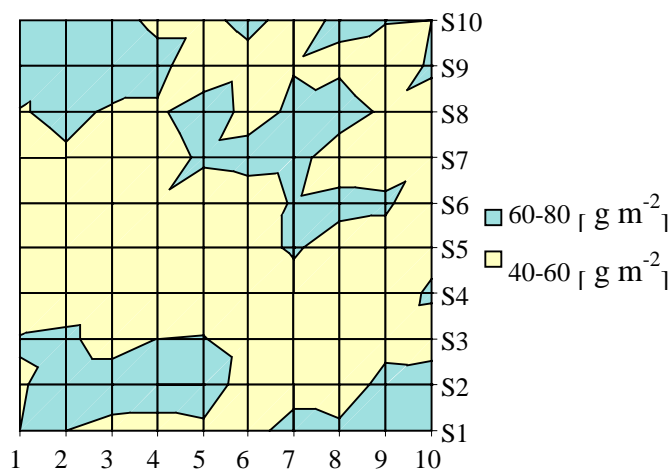
| Průměrná hmotnost m_{ij} [10^{-4} g] | | | | | | | | | |
|--|------|------|------|------|------|------|------|------|------|
| 60 | 60 | 55.7 | 56 | 57.8 | 53.8 | 67 | 62.7 | 69.2 | 63.2 |
| 58.1 | 68.8 | 68.1 | 66.1 | 66.1 | 54.9 | 52.1 | 51.8 | 64.2 | 65.3 |
| 61.1 | 63 | 53.4 | 60.1 | 60.4 | 56.1 | 56 | 57 | 55.7 | 55 |
| 51.1 | 51.9 | 53.8 | 55.4 | 56.1 | 51 | 57.1 | 54.8 | 55.4 | 61.4 |
| 55.5 | 57.1 | 53.1 | 56.8 | 59.7 | 57.2 | 61 | 51.6 | 55.8 | 57.1 |
| 54.8 | 51.2 | 60 | 59.1 | 53.1 | 54.6 | 61 | 62.7 | 61.6 | 52.1 |
| 52.4 | 58.2 | 59.2 | 53.1 | 62.2 | 63.4 | 63.2 | 54.8 | 54.8 | 58 |
| 59 | 63.9 | 58.1 | 58 | 67 | 56.3 | 61.8 | 65 | 58.1 | 53.5 |
| 70 | 63.4 | 71 | 64.3 | 51.3 | 56 | 59.5 | 58 | 51 | 62.2 |
| 69.3 | 73 | 65 | 57 | 57.2 | 63 | 56 | 62 | 61 | 60 |

Vzhledem k tomu, že byla použita stejná velikost cel o ploše $S_j = 100 \text{ mm}^2$ je plošná hustota $z_{ij} = m_{ij} / S_j$ v [g m^{-2}] číselně rovna hodnotám v tab. 1. Základní statistické charakteristiky tohoto pole plošné hustoty jsou uvedeny v tab.2.

Tabulka 2. Základní statistické charakteristiky plošné hustoty

| | | |
|---------------------|-------|-----------------------|
| Počet hodnot | 100 | rozměr |
| Průměr | 58.92 | [g m^{-2}] |
| Maximální hodnota | 73 | [g m^{-2}] |
| Minimální hodnota | 51 | [g m^{-2}] |
| Směrodatná odchylka | 5,12 | [g m^{-2}] |
| Variační koeficient | 8.68 | [%] |

Graficky je kolísání plošné hustoty znázorněno na obr. 3.



Obr. 3 Kolísání plošné hustoty pojené textilie 60 g m^{-2}

Informace o náhodném poli lze získat na základě sekvence povrchových hustot $z(i,j)$ určených

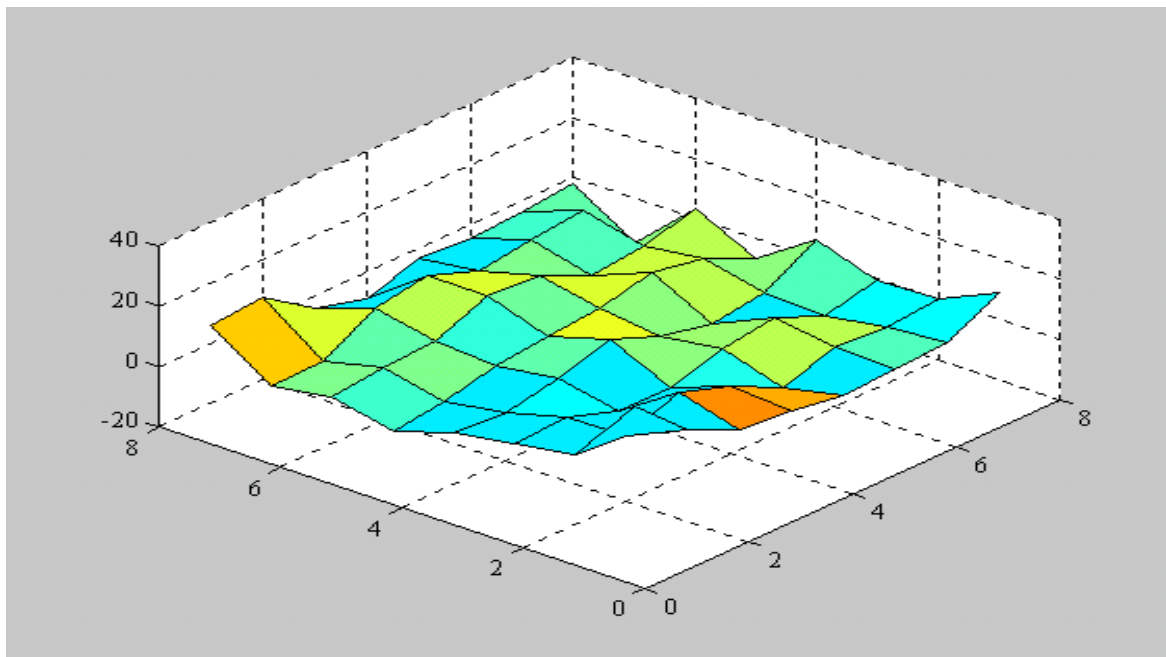
na pravoúhlé síti, kde i, j ($i = 1 \dots m, j = 1 \dots n$) definuje i, j -tou celu. Pro odhad korelační funkce pak platí

$$R(K, L) = \frac{1}{(m-K)(n-L)-1} \sum_{i=1}^{m-K} \sum_{j=1}^{n-L} (z(i+K, j+L) - \bar{z})(z(i, j) - \bar{z})$$

kde

$$\bar{z} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n z(i, j)$$

je průměrná hodnota plošné hustoty. Je zřejmé, že korelační funkce $R(0,0) = D(z(i,j))$. Zde symbol $D(z(i,j))$ označuje výběrový rozptyl plošné hustoty. Na obr 4 je znázorněn odhad korelační funkce $R(K,L)$ pro $K=0,1,\dots,7$ a $L=0,1,\dots,7$.



Obr 4 Odhad korelační funkce $R(K,L)$

Pro vyjádření prostorové autokorelace bylo použito jak Moranova I tak i Gearyho c indexu. Vzhledem k regularitě sítě bylo použito lokální definice nejbližších sousedů podle obr 1. Výsledky získané s využitím doplňku (add-in) pro EXCEL (viz [4]) jsou uvedeny v tab. 3.

Tabulka 3. Indexy prostorové autokorelace

| Typ souseda | I | $D(I)$ | $Z(I)$ | c | $D(c)$ | $Z(c)$ |
|-------------|-------|--------|--------|-------|--------|--------|
| Rook | 0,255 | 0,0053 | 3,662 | 0,721 | 0,0057 | 3,677 |
| Bishop | 0,14 | 0,0059 | 1,952 | 0,819 | 0,0078 | 2,03 |
| Queen | 0,20 | 0,0027 | 4,06 | 0,768 | 0,0037 | 3,79 |

V tab. 3 jsou $Z(I)$ a $c(I)$ standardizované náhodné veličiny s přibližně normálním rozdělením. Jejich velikost (podstatně výše než 2) svědčí o výrazné pozitivní autokorelaci.

Z uvedeného je patrné, že pole povrchové hustoty je mírně anizotropní a vykazuje lokální neregularitu. Existuje zde výrazná prostorová autokorelace. To bylo potvrzeno také analýzou založenou na vizuální nestejnomyšlnosti [6].

5. Závěr

V příspěvku bylo ukázáno použití prostorové autokorelace na příkladu pole povrchové hustoty pro hodnocení nestejnomyšnosti netkaných textilií. Byl použit doplněk k EXCELU, který umožňuje jednoduché i komplikovanější výpočty související s prostorovou autokorelací. Tento doplněk lze získat na adrese:

<http://www.uottawa.ca/academic/arts/geographie/lpcweb/sections1/software/frmsoft.htm>

Některé další programy pro vyjádření prostorové autokorelace jsou k dispozici u autorů této práce.

Poděkování

Tato práce vznikla s podporou grantu GAČR č. 106/99/1184, grantu MŠMT č. VS 97084 a výzkumného záměru MŠMT J11/98:244101113

6. Literatura

- [1] Cliff A. D., Ord J.K.: Spatial autocorrelation, Pion, London 1973
- [2] Hubert L. J. a kol. Geographical Analysis **13**, 224 (1981).
- [3] Anselin L.: Geographical Analysis **27**,93 (1995).
- [4] Sawada M.: Bull. Ecol. Soc. Amer. **80**, 231 (1999)
- [5] Klička. V.,: Disertační práce, 1998, Liberec
- [6] Militký J., Rubnerová J.,Klička V.:Vzhledová nestejnomyšnost netkaných textilií , Sborník z konference STRUTEX 98, Liberec 1998

Název souboru: autoko~1
Adresář: E:\Pom
Šablona: D:\Program Files\Microsoft Office\Sablony\Normal.dot
Název: autoKorelace
Předmět:
Autor: katedra textilních materiálů
Klíčová slova:
Komentáře:
Datum vytvoření: 14.09.00 13:38
Číslo revize: 2
Poslední uložení: 14.09.00 13:38
Uložil: Milan Meloun
Celková doba úprav: 0 min.
Poslední tisk: 14.09.00 13:43
Jako poslední úplný tisk
Počet stránek: 7
Počet slov: 1 816 (přibližně)
Počet znaků: 10 355 (přibližně)