

Vélemény

Makra László Különböző taxonok pollenjeinek komplex statisztikai elemzése a meteorológiai elemekkel összefüggésben, különös tekintettel a parlagfű pollenjére

című MTA Doktori Értekezéséről

Az értekezés a Szeged városát veszélyeztető pollenekre vonatkozó adatsorokat dolgoz fel meteorológiai és egészségügyi adatokkal kiegészítve. Napjainkban az alkalmazott tudományokra jellemző a munkában felhasznált különböző tudományágak nagy száma. Kétféle stratégia lehetséges: nagy csoportok együttműködése avagy polihisztorok munkája. Jelen dolgozat jellegét tekintve inkább a második kategóriába sorolható, a szerző igen dicséretes módon minden érintett területen otthonosan mozog és nagy önállóságra tett szert. Végző következtetései meggyőzőek és hasznosak. (De nem idegen a szerzőtől az első stratégia sem hiszen cikkeinek többségét 3-4 társszerzővel írja.) Ezért azt javaslom, hogy az MTA fogadja el az értekezést és adja oda a szerzőnek a tudományok doktora címet.

Az alábbiakban olyan észrevételeket közlök, amelyeket kizárólag matematikai statisztikai szempontból írok és nem csökkenthetik a disszertáció értékeit.

Kiindulok a 24. ábrából (75. oldal). Ha jól értem, itt tíz év áttekintő adatsora látható valamilyen átdolgozásban, de ennek szerintem most nincs jelentősége. Apró megjegyzés, hogy ezt az ábrát jobban szeretném egész oldalon látni esetleg színesben, de nekem most a céljaimnak így is megfelel. Amit nekem ez az ábra mond az egyértelműen az, hogy egy ilyen jellegű adatsorra a klasszikus matematikai statisztika módszerei nem alkalmazhatóak. Hiszen nincs két összehasonlítható adat. Az évek mind mások, az éveken belül természetes ciklusok vannak. Valamit ilyenkor ki kell találni. Ezt természetesen nem lehet egy meteorológustól elvárni, de talán helyesebb lett volna együttműködő partnereket keresni.

Hasonló hatást mutat a 11. táblázat: a vizsgált tíz évben az összes pollenszám nagy mértékben és első ránézésre össze-vissza, rendezetlenül változott, a minimum 3857, a maximum 13 854. Itt két kérdést tartok jelentősnek, amelyekre a disszertáció nem válaszol.

1) Miért ilyen nagy az ingadozás? Laikusként azt gondolom, az általános agrár tendenciák mutatkoznak itt, de el tudom képzelni azt, hogy újabb kori politikánk összevisszasága is hatást gyakorol. A jelen támogatási rendszerben sokszor bevetik a gazdák a földet de nem várják ki a termést, hanem korábban leveszik a kultúrnövényt, ami visszahathat a gyomnövényekre.

2) Hogyan lehet ezt az igen jelentős hatást a statisztikai feldolgozásban kompenzálni? A 10. fejezet rövidtávú (1-3 napos) előrejelzéssel foglalkozik. Szerintem ezek hatékonyságát elsősorban ez a globális ingadozás magyarázza. Bő termés esetén is szűk termés esetén is a szomszédos napok egyformának mutatkoznak.

Most visszatérek a 11. ábrára (36. oldal). Szemmel láthatóan ez az egyik fő eredménye a dolgozatnak. Az első észrevételem az, hogy egy ideje már töprengék azon, mi is lehetett itt a kérdés, amire nagyon bonyolult eljárással a szerző választ keresett, de nem tudom kitalálni. Az eredményt értem: augusztus 18 és szeptember 13 között a legmagasabb a parlagfű pollenje Szegeden. Csak hát ezt nem hiszem el, mert mint láttuk, minden év más, és nem tudom, mit jelent az, hogy szignifikánsan eltérő.

Így jutunk vissza a bajok első gyökeréhez, a Makra-próbához (15. oldal). A kérdés az, hogy egy homogén adatsor valamely része eltér-e az egésztől. Ha különválasztjuk a részt az egésztől, kapjuk annak másik részét. Ezek összehasonlítására vonatkozik a kétmintás u- és t-próba, amelyben a két rész átlagának az eltérését vesszük. Legyen ez az átlag mondjuk az egyikben 100, a másikban 200, és legyenek a méretek egyformák, akkor a különbség 100 az átlagokra és 50 a nagy átlagnak az első átlagtól vett eltérésére. Tehát hatékonyságot veszünk azzal, hogy összevonjuk a két csoportot. Ha a szórásokat is figyelembe vesszük, egzaktul kimutatható az eltérés.

Én azzal kezdtem volna az adatok feldolgozását, hogy megkeresem a tíz év legfontosabb globális mutatóit a megvizsgált fajokra: évi összes pollenszám, annak koncentrátsága. Minden egyes fajra a napi vagy óránkénti pollenszám egy folytonos és pozitív függvény közelítései, amit az integráljával leosztva sűrűségfüggvényt kapunk. Így a valószínűségi változókra kidolgozott statisztikák (átlag, szórás, momentumok) alkalmazhatóak. Érdekes lehet a fajok közti kapcsolat, ha valamelyik szárazságtűrő másként viselkedik mint amelyik a sok csapadékot kedveli. Meg kell választani a megfelelő időszakaszt, ez talán a hét lehet, és elképzelhető, hogy a Rasch modell leírja a keletkezett tábla legfontosabb hatásait. Ez hasonló a szórásanalízisre, de pozitív mennyiségek tárgyalására alkalmas.

Túl a belső analízisen keresni kell külső globális mennyiségeket: csapadék, szél, napsütés. A mechanikus tábla módszer helyett sokkal jobb az idősoros módszer, amikor az időbeli változásokat a múlt és a külső változók hatásaként keressük. Ez továbbfejleszthető előrejelzésre és kontrollra: annak meghatározására, hogyan lehet egyrészt a lakosságot riasztani másrészt mi a legalkalmasabb védekezés. A vizsgált növényeknek jól meghatározott szezonális ciklusa van amit nem lehet figyelmen kívül hagyni. A sztochasztikus kapcsolatok keresésének két élesen eltérő módja van, az egyik az eddig vázolt időben folytonos többváltozós analízis, a másik csoportok képzése, vagy más néven keverékek felbontása. A szerző ez utóbbit használja, amit elvileg nem lehet kifogásolni, nagyon nehéz eldönteni, mikor melyik módszer hatékonyabb. A tapasztalatom szerint jobb mind a kettőt kipróbálni, ha valami alapvető szakmai érv nem segít a döntésben. Kétféle csoportosítást használ a szerző, az egyik a széljárások adatait csoportosítja először (objektív csoportok), aztán a csoportok szerint megnézi a pollenekre gyakorolt hatást. A másik egy hazai irodalmi csoportosítás (szubjektív csoportok). Meg kellene vizsgálni ennek a kettőnek a viszonyát: minden egyes napnak van a kétfajta csoportosítás szerinti beosztása. Ezekből kell keresztábrát képezni. Az első kérdés az, hogy van-e a két csoportosítás között kapcsolat. Laikusként azt várom, hogy szoros kapcsolat legyen. Miután ez kiderült, elvileg meg lehet kérdezni, melyik csoportosítás hat jobban a pollenekre avagy a kettőt egyszerre használni. A két csoportosítás hatékonyságát összehasonlítja a szerző, de mechanikusan a szignifikáns esetek számát használva. Tekintve, hogy az objektív csoportosításban 8, a szubjektívben 13 csoport van, ez így nem helyes. A szerző szemmel láthatóan nem ismeri Gulyás Ottó munkásságát, aki szintén kidolgozott időjárási csoportokat (Ambrózy Pál, Barholy Judit és Gulyás Ottó: Évszakos bontású makroszinoptikus típusok kialakítása clusteranalízissel az atlanti-európai térségre, Országos Meteorológiai Szolgálat, 1983). Dévényi Dezsővel közösen írt könyvet Matematikai statisztikai módszerek a meteorológiában címmel.

Sokat foglalkozik a szerző azzal a kérdéssel, milyen arányban vesznek részt Szeged levegőjének szennyezésében a közeli (helyi), közepesen távoli (Kárpát-medencei) és egészen távoli (európai) források. Ez valóban fontos kérdés, hiszen nagy mértékben hatása lehet a

megelőzésre. A szerző nem mond számokat, talán 80, 15 és 5 százalék lehet a véleménye. Ez valóban fontos kérdés, de szerintem a rendelkezésre álló adatok alapján erre nem lehet válaszolni. Egy lehetséges módszer a pollenek mikro-analízise, keverékfelbontása lehetne, feltehetőleg van eltérés a különböző források között.

Külön vizsgálja a szerző a csapadékos napokat. Első hallásra engem ez kicsit zavar, mert merevnek érzem a kettévágást: laikusként egy folytonos skálát jobbnak tartok (napi összes csapadék). Itt jutunk el a következő módszertani fejezethez, a regresszió analízisre. Ebben hiányolom a legjobban az idősoros szemléletet. Tapasztalatom szerint a lépésenkénti regresszió eredményei nehezen értékelhetőek ki, ezért csak akkor alkalmazom, ha nincs más.

Az objektív és szubjektív csoportosításhoz a szerző 12 meteorológiai adatot használt. A légúti megbetegedések többváltozós analízisében viszont csak ötöt (hőmérséklet, szélsébség, relatív nedvesség, globálsugárzás és légnyomás). Apró megjegyzésem az, hogy a 12 között négy hőmérséklet van, nem derül ki, az öt között melyiket használja, feltehetőleg az átlagot. De ami sokkal lényegesebb, az az, hogy a 14a táblázat (79. oldal) szerint a 3. csoport kiugróan magas napi Ambrosia koncentrátságot mutat. Ilyet nem látok sem a 4. táblázatban (50. oldal) sem a 6. táblázatban (59. oldal). Mi lehet ennek az oka? Elképzelhető, hogy az, hogy most eleve külön lett véve az Ambrosia pollenszezonja és a pollenmentes időszak? Ha igen, ez egy alapvető módszertani hibára hívja fel a figyelmet: amint láttuk, az Ambrosia szezonja enyhén ingadozik, ezt akarhatjuk a meteorológiai adatokból kiolvasni, de határozottan intervallum jellegű időszak. Ilyesmit csoportosítással megjósolni nem helyes. A pollenszennyezettség, meteorológiai adatok és a különböző korú betegek szezonális változásai elsősorban saját belső törvényszerűségeiket követik. Ha ezt nem vesszük figyelembe, akkor nem csoda, ha a regressziós együtthatók együtthatójának még az előjele is nehezen értelmezhető.

Mint mondtam, ellenvetéseim nem csorbíthatják az értekezés érdemeit. Többnyire csak lehetséges ötleteket vetettem fel, amelyek hasznossága csak akkor derülne ki, ha valaki követné azokat. Az a tapasztalatom, hogy egy komplex adatmező mondanivalóját kiolvasni nagyon sok ötlet kipróbálásával lehet csupán. Előre nem látható módon néha az egyik sikeres, a másik nem. Ez a szellem nem idegen a szerzőtől, érezhetően maga is egyes próbálkozásaival jobban meg van elégedve mint másokkal. Remélem még sokáig fog hasonló kérdéseket vizsgálni. Javaslom a nyilvános vita kitűzését és a tudományok doktora fokozat megadását.

Budapest, 2012. november 30.

Tusnádý Gábor