

# Geoepidemiológia

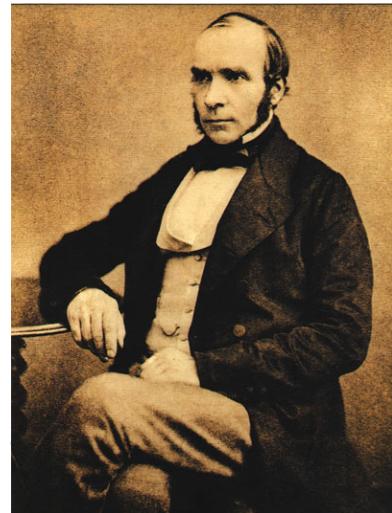
Az angol szakirodalomban *spatial epidemiology*-nak nevezik az epidemiológiának azt az ágát, amely az egészséggel összefüggő adatok földrajzi feldolgozásával, elemzésével foglalkozik. A *geoepidemiológia* megnevezést azért használjuk itt (habár nem tökéletes terminus *technicus*), mert az egyéb lehetőségek még inkább félrevezetők lehetnek. A *geo* szó helyett használhatnánk a *térbeli* előtagot is, a térinformatica magyarázathoz hasonlóan. De mint, ahogy annál is tér helyett egyértelműbb a földrajzi információs rendszer a GIS (Geographic Information System) megjelölésére, úgy a *térbeli* megnevezés a legtöbbekben nem pusztán két dimenziós megközelítést jelent, amiről itt végül is szó lenne. A geomatematika és geostatisztika elnevezése nyomán tehát addig, amíg nem találunk jobb nevet, a *geoepidemiológia* használatát javasoljuk.

A fejezet során a rövid történeti betekintés után ismertetjük a geoepidemiológia ágait: példákon keresztül a betegség-térképezést, földrajzi mintázatelemzést és összefüggés-elemzést.

## Történeti betekintés

A geoepidemiológiai megközelítés a nyugati tudomány történetében egyáltalán nem új. Annak ellenére, hogy voltak korábbi próbálkozások, a tudományterület alapítójának John Snow (1. ábra) angol sebész-aneszteziológust tekintik. Snow 1855-ben közölt cikkében a londoni Sohoban 1854-ben lezajlott kolerajárvány elemzését mutatta be. Ebben a munkájában olyan elemzést végzett, aminek során az egyes házakban a kolera okozta halálesetek számát grafikusan ábrázolta (2. ábra).

A térképen látható kumulatív mortalitási mintázatban tapasztalható heterogenitás a Broad street-en lévő közkútra irányította a figyelmét. Ennek nyomán azt vizsgálta, hogy a lakosok honnan hordják a háztartásban felhasznált vizet.<sup>1</sup> Úgy találta, hogy a magasabb incidenciájú házakban ebből a közkútból hordják a vizet. A térképen látható, hogy egész távoli épületekben is előfordult jelentős számú halálozás, ezek nagy részében is kimutatható volt a jelzett közkúttal



1. ábra. John Snow (1813-1858)

<sup>1</sup> a házakban ekkor még nem volt vezetékes ivóvíz



2. ábra. Snow térképének rekonstrukciója az 1854-ben a londoni Sohoban lefolyt kolerajárvány épületenkénti kumulatív mortalitási adataival. Az egyes telkekben előfordult halálzásokat vonalakkal vannak jelölve.

való kapcsolat. A kutak lezárása után a már amúgys csökkenő tendenciájú járvány megszűnt. Érdekességgéktel említhető meg, hogy a vizsgált területen volt egy dominikánus rendház, ahol (feltehetően Metzi Szent Arnold<sup>2</sup> javaslatát követve) egyáltalán nem ittak vizet, ehelyett csak sört fogyasztottak, és a fertőzés nem is érintette a Ház lakóit.

Snow vizsgálata és eredményei akkoriban sok ellenkezést váltottak ki, pl. a Lancet orvosi lap főszerkesztője erősen kritizálta az ivóvíz szerepét a kolera kialakulásában. Ebben a korban még a miasma-, és humorális teoriák domináltak a betegségek kórokanának meghatározásában.<sup>3</sup>

Ezt követően a geoepidemiológiai megközelítést többször használták járványügyi nyomozásokban, már az elektronikus számítógépek megjelenése előtt is. A digitális technika fejlődésével, illetve használatának széles körű elterjedésével a térinformatika az orvostudományban is egyre inkább megjelent, amivel párhuzamosan a geoepidemiológiai feldolgozások is gyakoribbá váltak.

### *A fejezetben használt adatsorok*

*Magyarországon vörös rókában diagnosztizált veszetségesetek* A veszetségi adatokat a Földművelésügyi és Vidékfejlesztési Minisztérium Állategészségügyi és Élelmiszer Ellenőrzési Főosztálya által vezetett nyilvántartásokból gyűjtöttük ki, 2001-2002-ben. minden esethez tá-

<sup>2</sup> Szent Arnold (580-640) bencés szerzeteshez kötődő legenda szerint a sörfogyasztás segített felszámolni egy pestis járványt. Arnold azt javalta, hogy „Ne igyatok vizet, igyatok sört!” A legenda szerint a Szent iránymutatásának követésével eltűnt a betegség a közösségből. Talán emiatt Arnoldot tekintik a sörfőzők védőszentjének.

<sup>3</sup> Habár Jacob Henle már az 1840-es években felvetette, hogy mikroorganizmusok okoznak betegségeket, tanítványa, Robert Koch csak 1884-ben fogalmazta meg a Koch-posztulátumokat, amelyek a mikroorganizmusok és betegségek közötti oksági kapcsolatának feltárással alapvető irányutat jelentettek hosszú évtizedekre.

roltuk az érintett állat megtalálásának, lelövésének dátumát, az eset helységét és az érintett állat faját. Az adatbázis az 1990-2001. közötti időszakban állatokban diagnosztizált összes esetet tartalmazza (Solymosi et al., 2002, 2003; Solymosi, 2009).

*Alsó-Szászországi vörös rókák *Echinococcus multilocularis* fertőzöttsége*  
 Berke (2001) az 1991 és 1997 között a németországi Alsó-Szászország tartományban kilőtt vagy elhullott vörös rókák *Echinococcus multilocularis* fertőzöttségi eredményeit dolgozta fel. A cikkben közölt eredeti adatsorhoz képest annyi módosítást végeztünk, hogy a Delmenhorstból származó adatokat Oldenburg adataival egyesítettük. Ennek az az egyszerű oka, hogy a felhasznált térképi adatbázisban<sup>4</sup> Delmenhorst úgy szerepel, mint Oldenburg része.

<sup>4</sup> <http://www.gadm.org/>

Sorszám	Járás	Vizsgáltak száma	Pozitívok száma
1	Ammerland	10	0
2	Aurich	111	15
3	Braunschweig Städte	25	1
4	Celle	255	8
5	Cloppenburg	110	8
6	Cuxhaven	73	14
7	Diepholz	143	10
8	Emden Städte	4	0
9	Emsland	242	21
10	Friesland	15	1
11	Gifhorn	158	13
12	Goslar	152	20
13	Göttingen	157	84
14	Grafschaft Bentheim	36	1
15	Hamelin-Pyrmont	99	41
16	Hanover	327	41
17	Harburg	285	17
18	Helmstedt	66	7
19	Hildesheim	202	60
20	Holzminden	60	16
21	Leer	39	2
22	Lüchow-Dannenberg	225	20
23	Lüneburg	278	22
24	Nienburg	325	20
25	Northeim	186	96
26	Oldenburg	111	16
27	Osnabrück	290	21
28	Osterholz	32	8
29	Osterode	94	23
30	Peine	115	10
31	Rotenburg	114	6
32	Salzgitter Städte	115	9
33	Schaumburg	90	12
34	Soltau-Fallingbostel	137	8
35	Stade	84	5
36	Uelzen	214	14
37	Vechta	83	2
38	Verden	107	11
39	Wesermarsch	45	9
40	Wilhelmshaven Städte	76	4
41	Wittmund	53	4
42	Wolfsburg Städte	22	6

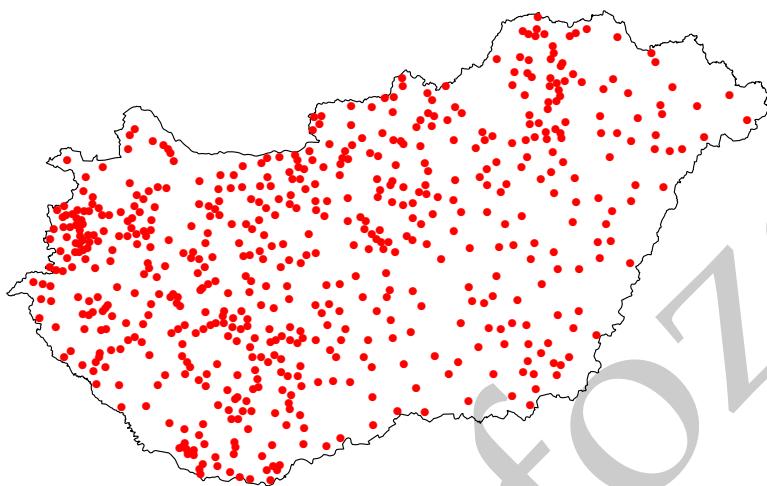
1. táblázat. Alsó-Szászországi vörös rókák *E. multilocularis* fertőzöttségi adatai, járásonként (Berke, 2001)

### Betegség-térképezés

Az egészséggel kapcsolatos információk földrajzi ábrázolásának célja a kockázat térképezése. A kockázat térképi megjelenítése jelentős

többletinformációval szolgál az egyszerű táblázatos megjelenítéshez képest. A betegség-térképezéshez számos módszert alkalmaznak az epidemiológiában. A következőkben ezek legfontosabbait tekintjük át.

A legegyszerűbb térképezési módszer, amikor a betegség előfordulásának helyét ponttal jelöljük a térképen (3. ábra). Az ábrán azokat a településeket jelöltük piros ponttal, amelyek területén találtak vagy kilőttek 1990-ben olyan rókát, amelyben később veszettséget diagnosztizáltak.

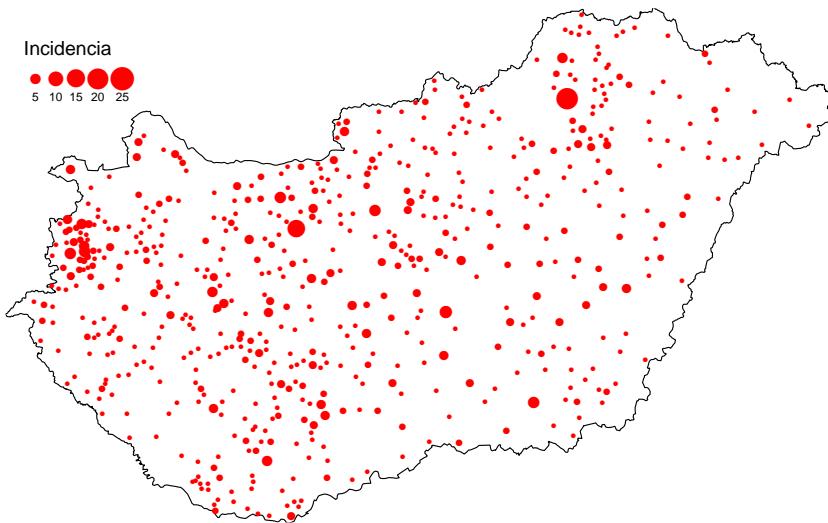


3. ábra. Veszettség-előfordulási pont-térkép. A piros pontok azokat a településeket jelölik, ahonnan származó rókákban az 1990. év során veszettséget állapítottak meg.

Az egészséggel kapcsolatos információk leggyakrabban pontszerű földrajzi referenciaival rendelkeznek nyers formájukban (pl. GPS-koordináta, lakkím, tartási hely). Azonban a pont alapú térképezésből legtöbbször csak kevés információ fogalmazható meg a kockázat földrajzi mintázatára vonatkozóan. A pont-térképek alapján kialakított benyomásokat jelentősen befolyásolhatja a térképezésnél használt felbontás, a pontok mérete, alakja és színe. Gyengesége a pont-térképeknek, hogy az esetszámot nem tudjuk megjeleníteni velük, csupán a pozitív/negatív pozíciókat.

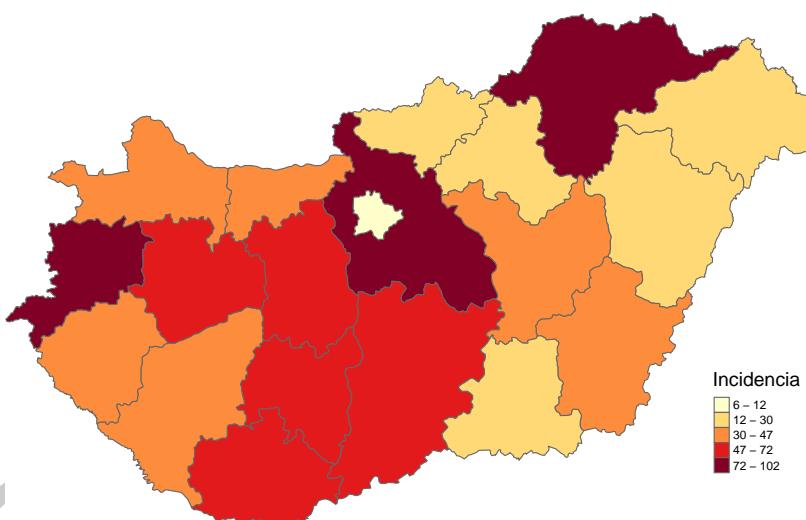
A 4. ábrán bemutatott ún. *buborék-térkép* segítségével már megjeleníthető, hogy adott helyen, településen hány esetet találtak az adott időszakban. A térképen alkalmazott jel (esetünkben korong) mérete mutatja a veszett rókák számát. Itt a pont-térképeknél említett technikai tényezők mellett további torzító mozzanat, hogy ha van olyan település, melyben sok eset volt, akkor a szomszédságában lévő pozitív helyet eltakarhatja az esetszámot megjelenítő jel.

A kockázat térképezésében további lehetőségeket jelentenek az ún. *choropleth* térképek (5-7. ábra). Az elnevezés görög eredetű, a *khora* a



4. ábra. Veszettség előfordulási buborék-térkép. Az 1990. év során érintett településeken előfordult esetek száma.

település külterületét, a *plethos* pedig a sokaságot jelenti. A betegség-terképezésben ennek megfelelően olyan térképeket jelöl, amelyek valamilyen területi egységen aggregálva mutatják be az előfordulási kockázatot. Az aggregálás alapja tetszőleges lehet, habár jelentősen befolyásolhatja a térképből levont következtetéseinket. Az 5. ábrán megyei szinten aggregáltuk a veszettség-előfordulási adatainkat, míg a 6. ábrán kistérségi<sup>5</sup> szinten.



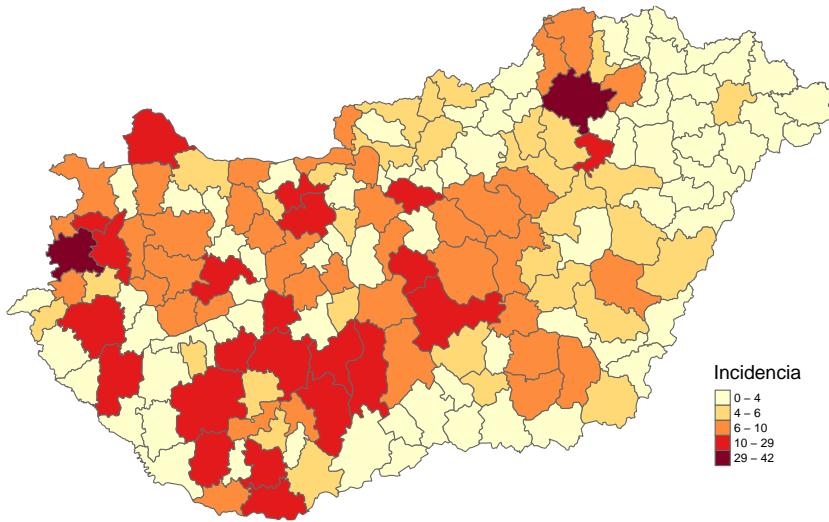
<sup>5</sup> a felhasznált <http://www.gadm.org/> 168 magyarországi kistérséget tartalmaz

5. ábra. Veszettség-előfordulási térkép. Az 1990. év során diagnosztizált esetek megyénkénti összesítése choropleth térképen.

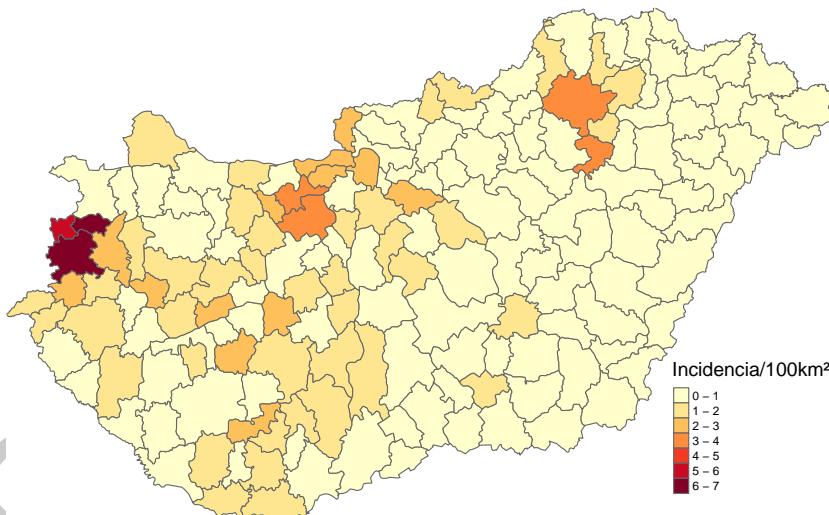
Térinformatikai szempontból a choropleth térképeken az aggregálás és ábrázolás alapegységei poligonok. Ezek a poligonok szomszédaikkal szorosan érintkezve kitöltik a vizsgált területet. Így elkerülhetővé válik a buborék-térképnél említett szomszédos adatok kitaka-

rása. Továbbá így egy határokkal ugyan megszakított, de folytonos felületen ábrázolhatjuk az adatainkat, ami megkönnyíti a kockázatra vonatkozó következtetések kialakítását.

A 6. ábra alapján úgy tűnhet, hogy a nagyobb területű kistérségek nagyobb kockázattal bírnak. De mielőtt ezt elfogadnánk, gondoljunk arra, hogy ha a rókák veszettsége homogén lenne az országban, akkor nagyobb területű településeken több esetünk lenne. De ez nem jelenti azt, hogy ott nagyobb is a betegség előfordulásának kockázata.



6. ábra. Veszettség-előfordulási térkép. Az 1990. év során diagnosztizált esetek kistérségi összesítése choropleth térképen.



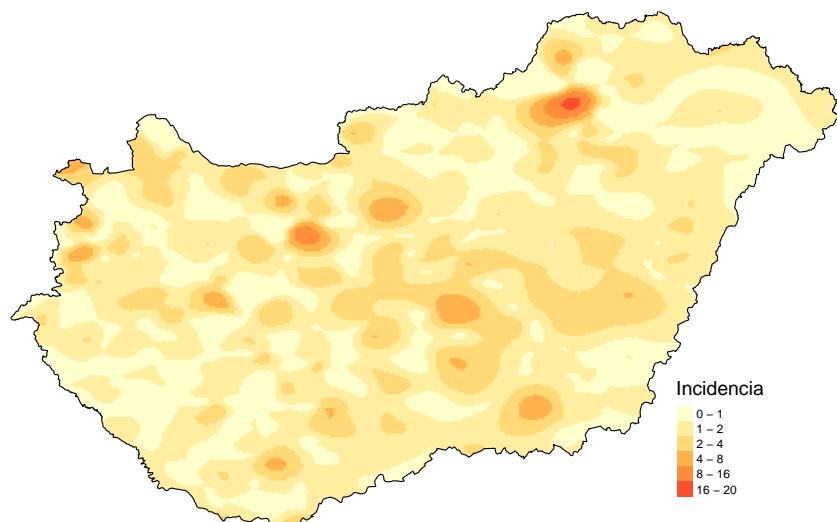
7. ábra. Veszettség-előfordulási térkép. Az 1990. évi  $100 \text{ km}^2$ -enkénti esetszámot mutatja be kistérségenként.

Mivel rókákra vonatkozó megbízható populációs adatok nem állnak rendelkezésünkre, a populáció helyett használjuk a kistérség területének méretét az esetszám normálására. A 7. ábra a  $100 \text{ km}^2$ -re

eső veszettségesetek számát mutatja be, ahol már a terület nagysága nem torzítja a kockázatra vonatkozó következtetéseinket. Ez alapján úgy tűnik, hogy egy magas esetszámú góc volt 1990-ben a szombathelyi kistérségben és környezetében, illetve néhány kisebb kockázatú góc a Dunántúlon és egy Miskolc környezetében.

A choropleth térképek értékelhetőségét jelentősen befolyásolhatja a területi egységek színezésére használt színskála és az értéktartományok megválasztása.

A kockázat földrajzi mintázatának bemutatására további lehetőséget nyújt az ún. *isopleth* térképezés (8-9. ábra). Az isopleth térképek létrehozásánál a pontszerű forrásadatokat valamilyen simítási módszerrel interpoláljuk a kétdimenziós térben.<sup>6</sup> A simítás eredményeként a kétdimenziós tér nem ismert esetszámú pontjaira is kapunk becslést, amit így már határok nélküli folytonos felületen szemléltethetünk. Ez a vizualizációs megoldás nyújt leginkább értelmezhető kockázati térképeket.

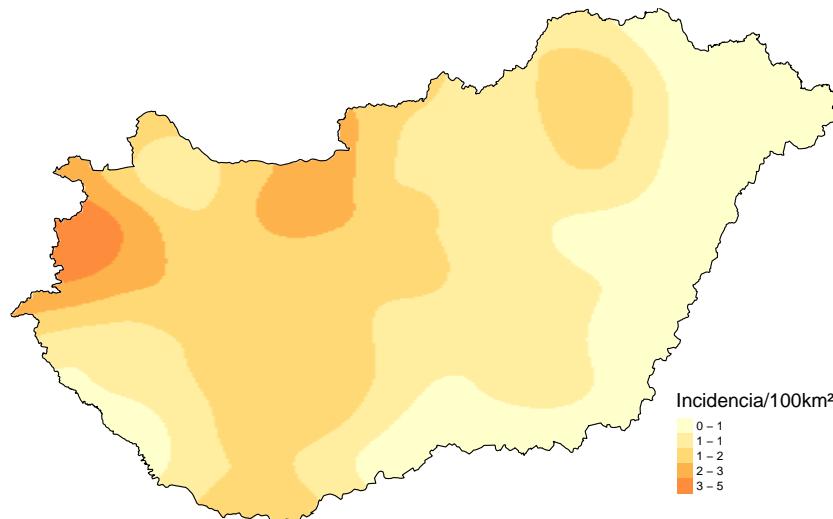


A 8-9. ábrákon bemutatott isopleth térképeket az ún. *thin plate spline* módszerrel hoztuk létre (Green & Silverman, 1993). A 8. ábrán a pont-térképen is használt településeként esetszámokat simítottuk. Ahogy a 6. ábrán a kistérség méretének, itt a település méretének torzító hatása sejthető. Ha kistérségenként  $100 \text{ km}^2$ -re jutó esetszámot simítunk ugyanazzal az eljárással, akkor a 9. ábrán látható isopleth térképet kapjuk, mely már egy ilyen szempontból torzítatlan kockázati mintázatot mutat.

Az isopleth térképek létrehozására használt simítási eljárás, illetve annak paramétereinek nagy mértékben befolyásolhatja a kockázati térképet, így az abból levonható epidemiológiai következtetéseinket.

<sup>6</sup> Fontos megjegyezni, hogy csak olyan vetületi rendszerben lévő koordináták használhatók kétdimenziós simításra, amelyekben a vertikális és horizontális dimenzióban is egyforma távolságot jelent egy egységnyi távolság. Így pl. nem használható a WGS84 vetület.

8. ábra. Veszettség-előfordulási térkép. Az 1990. év során diagnosztizált esetek települési szintű adataiból létrehozott isopleth térkép.

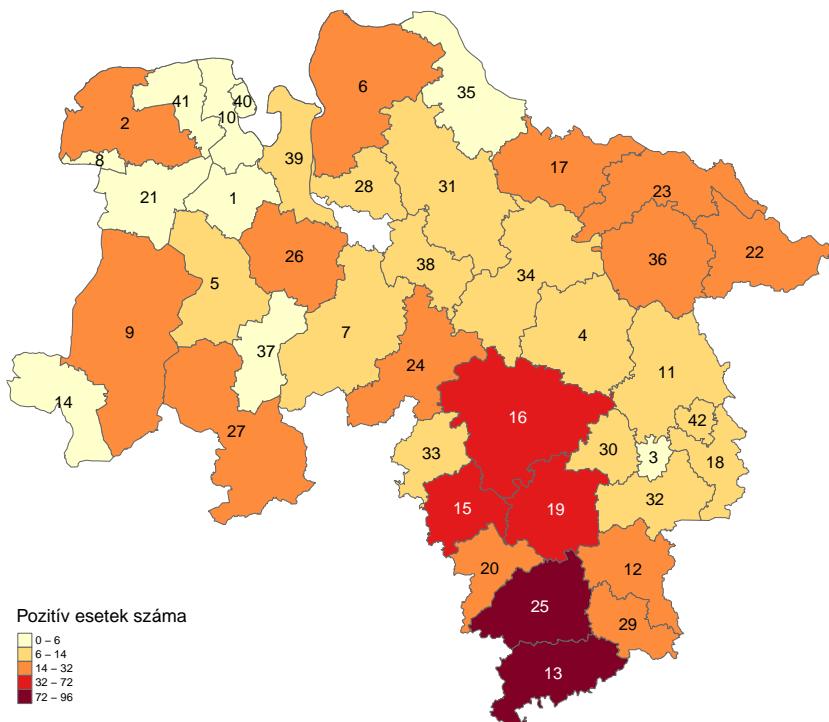


9. ábra. Veszettség-előfordulási térkép. Az 1990. év során diagnosztizált 100 km<sup>2</sup>-kénti esetek kistérségi aggregációja alapján létrehozott isopleth térkép.

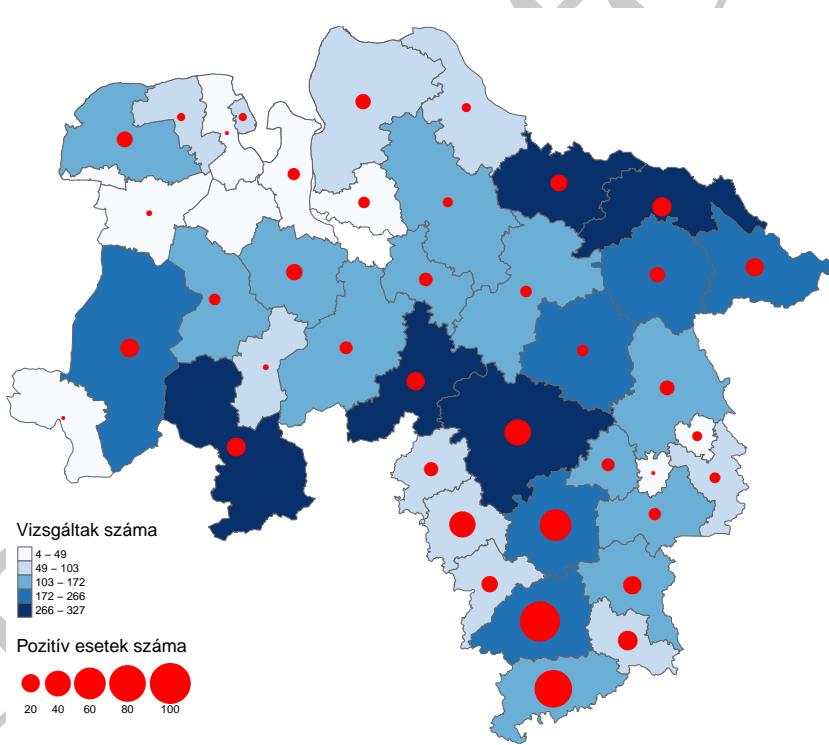
Mivel a veszettségi adatsorunkhoz nem rendelkezünk információval az egyes területek vörös róka populációjára vonatkozóan, a kockázati térképek folytatásához Berke (2001) adatait fogjuk használni. Ahogy az 1. táblázatban is látható, ebben a példában van a pozitív és a negatív egyedekre vonatkozóan is információink. A betegség-térképezés eddigi eljárásai kombinálódhatnak is, illetve egyéb információkkal egészülhetnek ki. Erre egy példa a 10. ábra, aminek segítségével összekapcsolhatóvá válnak a táblázat sorai Alsó-Szászország járásaival. A térképi poligonok színezése a járásban diagnosztizált pozitív esetek száma szerint történt, emellett a poligonokban olvashatók a járások táblázatbeli sorszámai.

További példa a térképezési eljárások kombinálására a 11. ábra, amelyen a járások a megvizsgált állatok száma szerint vannak színezve, de egyben a középpontjukba egy-egy buborék is látható, ami a területen talált pozitív esetek száma szerint van méretezve. Habár ez egy lehetőség arra, hogy mind a populációt, mind pedig az eseteket együttesen ábrázoljuk, mivel nehezen feldolgozható az általa nyújtott információ, nem szokás ennek a használata, amihez képest számoljuk a predikció hibáját.

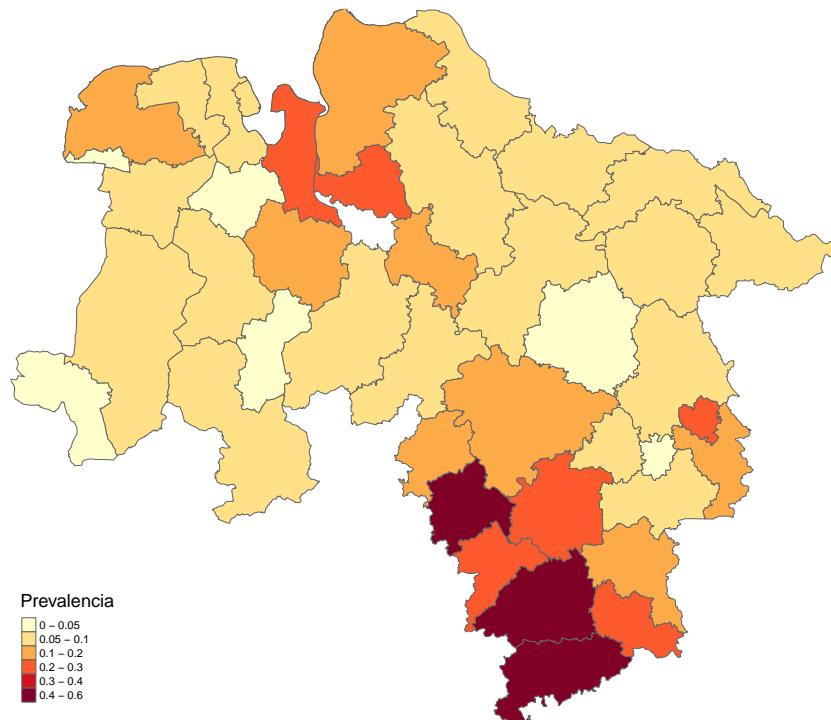
Ehelyett inkább használják a fertőzöttség kockázatának térképezésében a prevalencia ábrázolását. A prevalencia becslését végezhetjük egyszerűen úgy, hogy a fertőzöttek részarányát vesszük (12. ábra), de végezhetjük komplexebb modellek segítségével is. Ez utóbbi lehetővé teszi, hogy további információkkal (pl. a diagnosztikai eljárás szenzitivitásával, specificitásával) korrigáljuk a prevalencia becsléseket.



10. ábra. Az *E. multilocularis*-pozitív vörös rókák száma járásonként (Berke, 2001). A járásokban olvasható számok az 1. táblázatbeli sorszámukat jelzi.



11. ábra. A megvizsgált, illetve *E. multilocularis*-pozitív vörös rókák száma járásonként (Berke, 2001), kombinált térképen.



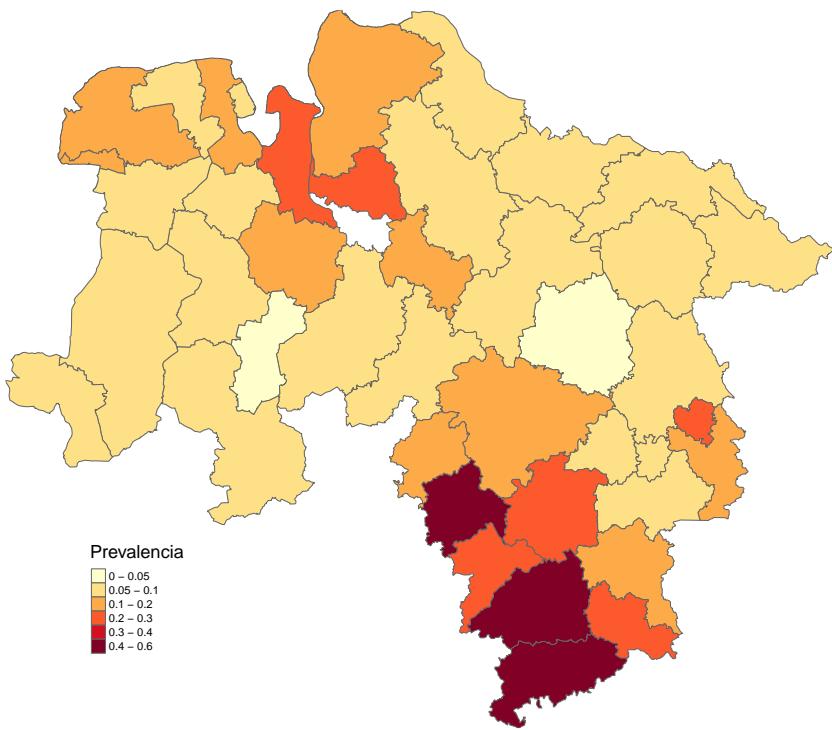
Egy ilyen, a prevalencia bayesi becslésére használható a konjugált béta-binominális modell BUGS-kódja:

```
model{
  for (i in 1:N) {
    pozitív[i] ~ dbin(prevalencia[i], megvizsgalt[i])
    prevalencia[i] ~ dbeta(1, 1)
  }
}
```

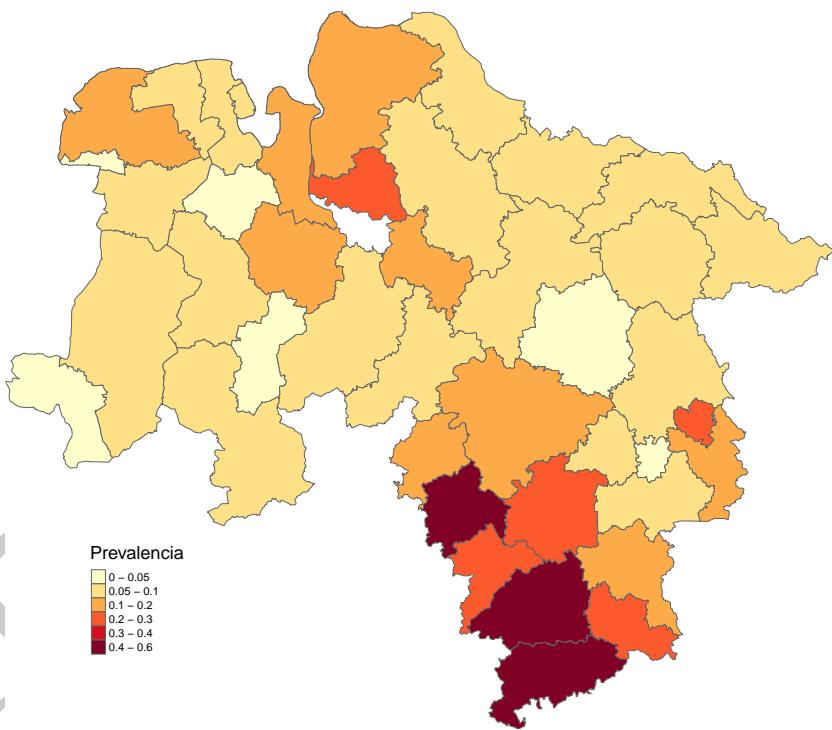
Ebben a modellben az egyes járások prevalenciájára vonatkozóan külön-külön végzünk becsléseket. Az egyes járások prevalenciájának így nincsen hatása a többi járás prevalenciájára. A prevalencia-térképet a 13. ábrán láthatjuk. A prevalencia becslésére használhatunk hierarchikus bayesi modelleket is. Egy ilyen logisztikus modell BUGS-kódja:

```
model{
  for (i in 1:N) {
    pozitív[i] ~ dbin(prevalencia[i], megvizsgalt[i])
    logit(prevalencia[i]) <- alpha + beta[i]
    beta[i] <- b[i] - mean(b[])
    b[i] ~ dunif(-10,10)
  }
  alpha ~ dunif(-10,10)
}
```

12. ábra. Az *E. multilocularis*-prevalencia járásonként, a prevalenciát egyszerűen a pozitív esetek részarányaként becsülve.



13. ábra. Az *E. multilocularis*-prevalencia járásonként, a prevalenciát bayesi béta-binominális modell alapján becsülve.



14. ábra. Az *E. multilocularis*-prevalencia járásonként, a prevalenciát bayesi logisztikus modell alapján becsülve.

A modell alapján becsült prevalenciák láthatók a 14. ábrán. Ebben a modellben az egyes járásokra vonatkozó prevalencia-becslés nem független a többiből. Egész Alsó-Szászországra vonatkozó átlagos prevalenciával együtt becsüljük az egyes járások prevalenciáját. Mivel ebben a megközelítésben az egyes járások prevalencia-eloszlásai befolyásolják egymást, a bayesi modellezésben azt szokták mondani, hogy az egyes megfigyelési egységek (itt a járások) „erőt kölcsönöznek”<sup>7</sup> egymásnak.

A kockázat-térképezésben az eddig használt incidencia és prevalencia mellett gyakran használják a relatív kockázatra vonatkozó becslések földrajzi vizualizálását. A relatív kockázat becslésére egy egyszerű, konjugált bayesi modell a Poisson-gamma modell, aminek a BUGS-kódját így írhatjuk fel:

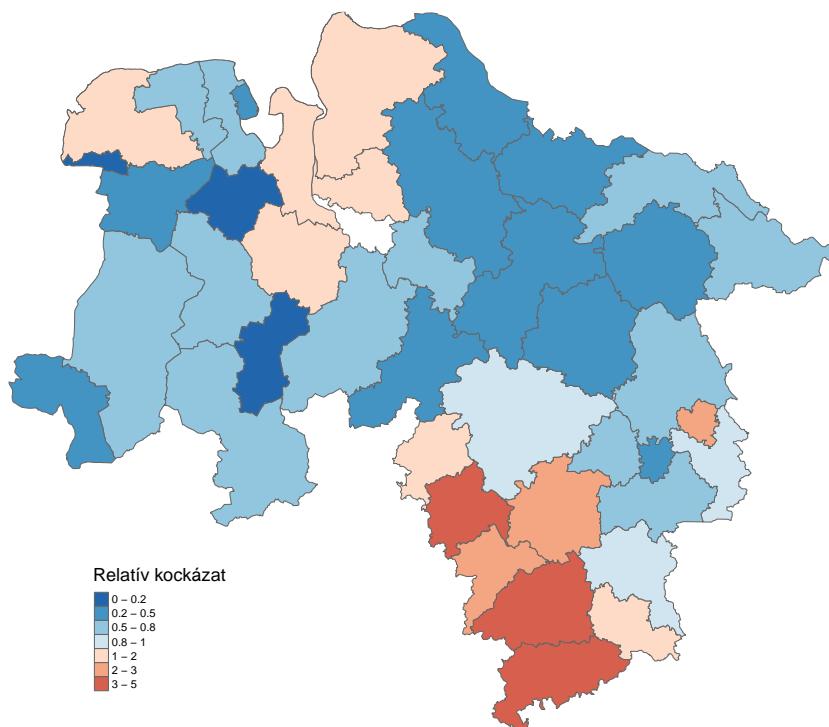
```
model{
  for(i in 1:N){
    esetszam[i] ~ dpois(mu[i])
    mu[i] <- varhato.esetszam[i] * RR[i]
    RR[i] ~ dgamma(0.01, 0.01)
  }
}
```

A kód a megfigyelt esetszámot a várható esetszám és a relatív kockázat szorzataként modellezi. A járásonkénti várható esetszámot úgy számolhatjuk ki, hogy az összes pozitív eset számát elosztjuk az összes megvizsgált róka számával, majd a kapott hányadost megszorozzuk az egyes járások megvizsgált rókáinak számával. Valójában itt a várható esetszám szorzója az SMR, ami azonban általánosan használt közelítése a relatív kockázatnak (Gelfand et al., 2010).<sup>8</sup> Ennél a modellnél ugyanúgy, mint a prevalencia becslésére használt beta-binominális modellnél, az egyes járásokra vonatkozó becsléseknek nincsen hatása egymásra. A modell futtatása alapján kapott járásonkénti relatív kockázatok becslését a 15. ábrán láthatjuk. Ebben a megközelítésben, az előbbiek alapján a relatív kockázat 2-es értéke azt jelenti, hogy a járások közti átlagos kockázathoz képest az *E. multilocularis*-fertőzöttség kockázata kétszeres.

A járásonkénti relatív kockázat becslésénél is van lehetőségünk olyan modellek alkalmazására, amelyekben az egyes járásokra vonatkozó becslések nem függetlenek egymástól, egymásnak „erőt kölcsönöznek”.

<sup>7</sup> borrowing of strength

<sup>8</sup> Annak ellenére, hogy a megközelítést széles körben használják, számos kritikája, korrekciója jelent meg a szakirodalomban (Jones & Swerdlow, 1998).



Egy ilyen modell a log-normális bayesi modell, ennek BUGS-kódja így írható le:

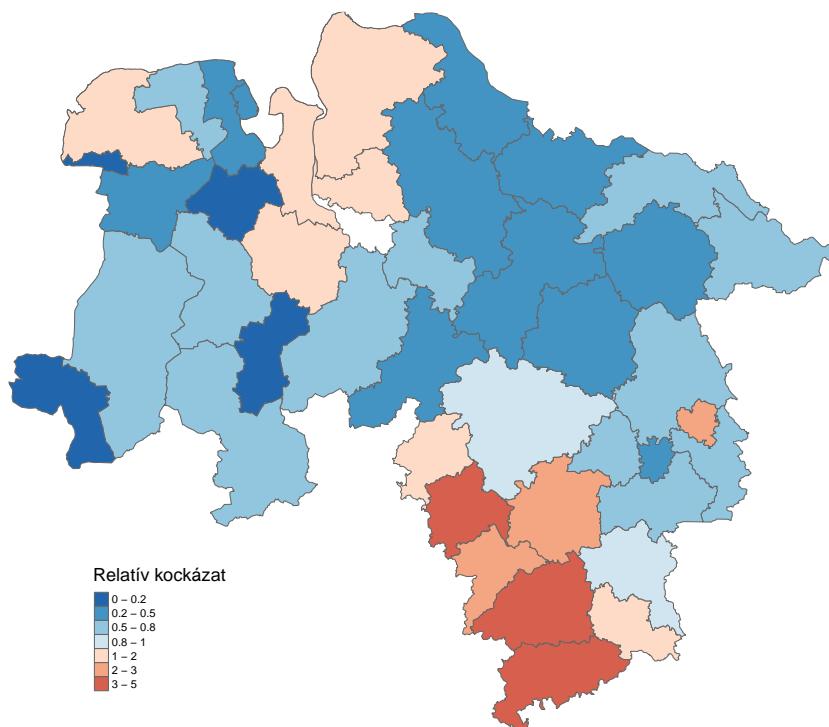
```
model{
  for (i in 1:N) {
    esetszam[i] ~ dpois(mu[i])
    log(mu[i]) <- log(varhato.esetszam[i]) + alpha + b[i]
    b[i] ~ dnorm(0, 0.001)
    RR[i] <- exp(alpha+v[i])
  }
  alpha ~ dnorm(0, 0.00001)
}
```

A modell futtatásából származó járásonkénti relatív kockázat becslések a 16. ábrán láthatók.

A kockázat-terképezésünk eddigi megközelítései a kétdimenziós simítástól eltekintve nem valódi térbeli modellek, mivel az események egymástól való földrajzi távolságát/közelségét nem kezelik. Eddig a térképeket csak vizualizációra használtuk, de a becslésekben azok szerkezetének nem volt szerepe.

Azonban könnyen belátható, hogy legalább a közvetlen szomszédságban lévő járásokban megfigyelt *E. multilocularis*-fertőzöttségi kockázatok nem lehetnek teljesen függetlenek egymástól. Gondolunk csak arra, hogy a közigazgatási határok a legtöbb esetben nem

15. ábra. Rókák *E. multilocularis* fertőzöttségének relatív kockázat térképe bayesi Poisson-gamma modell alapján becsülve.



16. ábra. Rókák *E. multilocularis*-fertőzöttségének relatív kockázat térképe bayesi log-normális modell alapján becsülve.

jelentenek fizikai korlátot, illetve nem esnek egybe olyan természeti határokkal, amelyek a rókák mozgását akadályoznák. Így könnyen lehet, hogy az egyik járásban kilött róka területuma más járásra is kiterjed, vagy valami kivételes okból ment át egyik járásból a másikba. Illetve lehetséges, hogy a fertőződése nem abban a járásban történt, amelyikben regisztrálták az esetet. Ennek következtében a relatív kockázat becsléseink torzítottak lehetnek.<sup>9</sup>

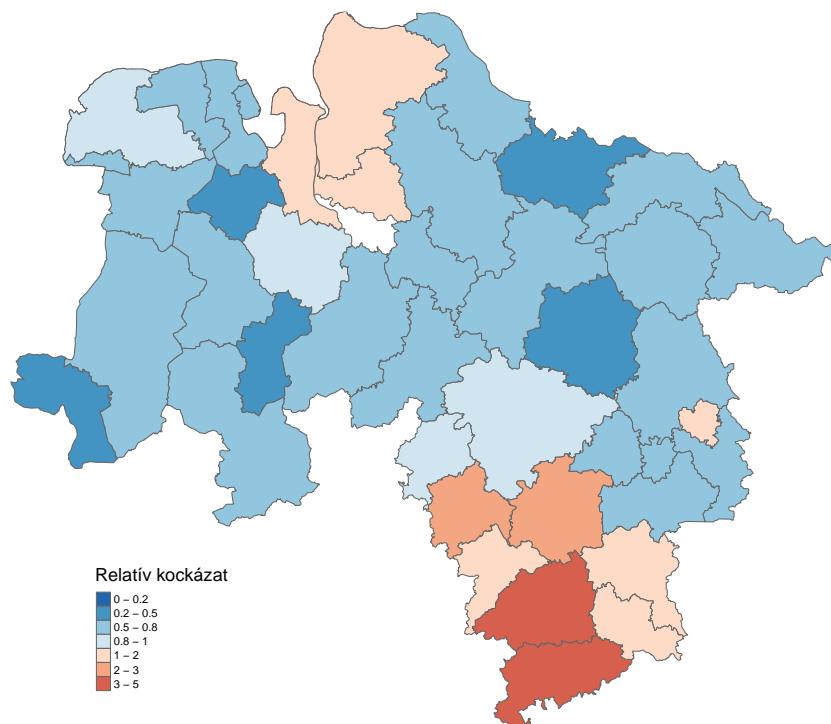
Az ilyen típusú torzítások kezelésére gyakran alkalmazzák a területi egységekre (itt a járásokra) vonatkozó becslések számítása során a szomszédaikban megfigyelt értékek (pozitív/negatív esetek, stb.) figyelembe vételét. Az előző log-normális modellünket módosíthatjuk úgy, hogy a becsléseinkben a szomszédokban megfigyelt értékeket is figyelembe vesszük. A BUGS-kód ebben az esetben így módosul:

```
model{
  for (i in 1:N) {
    esetszam[i] ~ dpois(mu[i])
    log(mu[i]) <- log(varhato.esetszam[i]) + alpha + b[i]
    RR[i] <- exp(alpha + b[i])
  }
  b[1:N] ~ car.normal(adj[], weights[], num[], tau)
  alpha ~ dflat()
  tau ~ dgamma(0.5, 0.0005)
}
```

<sup>9</sup> ecological bias

A modellben az egész Alsó-Szászországra vonatkozó átlagos kockázathoz (alpha) hozzáadódik minden járásban  $b[i]$ , amit itt random hatásként írtunk fel. Az előző, egyszerűbb log-normális modellünkben ennek a  $b[i]$  paraméternek flat priort ( $dnorm(0, 0.001)$ ) adtunk meg, ezáltal a rávonatkozó becslést csak a pozitív és várható esetszámok határozták meg.

Az újabb modellben  $b[1:N]$  priorjaként `car.normal()` eloszlást adtunk meg. A *conditional autoregressive* (CAR) modell leírja, hogy az egyes járásokban megfigyelt értékek (pozitív és várható esetszám) milyen kapcsolatban vannak a többi járásban megfigyelt értékekkel (Lunn et al., 2012). Így már a modell alapján végzett járásokra vonatkozó becslések eredményét nem pusztán az adott járásban megfigyelt értékek befolyásolják, hanem a szomszédságukban lévő járásokban megfigyelt értékek is. A modell futtatása alapján becsült járásonkénti relatív kockázatokat a 17. ábra mutatja.



17. ábra. A `car.normal()` prior alkalmazásával illesztett log-normális modell alapján becsült *E. multilocularis*-fertőzöttségi kockázati térkép.

A `car.normal()` függvény paramétereinek írásával leírjuk, hogy a járásokat reprezentáló poligonok milyen szomszédsági struktúrában helyezkednek el. Az `adj[]` vektor felsorolja minden járás szomszédainak a sorszámait. A vektor úgy kezdődik, hogy az 1. sorszámú poligon szomszédainak sorszámát csökkenő vagy növekvő sorrendben tartalmazza, majd ugyanígy a 2. poligon szomszédainak sorszámait, és így tovább. A `weights[]` vektor ugyanolyan hosszú, mint az `adj[]`

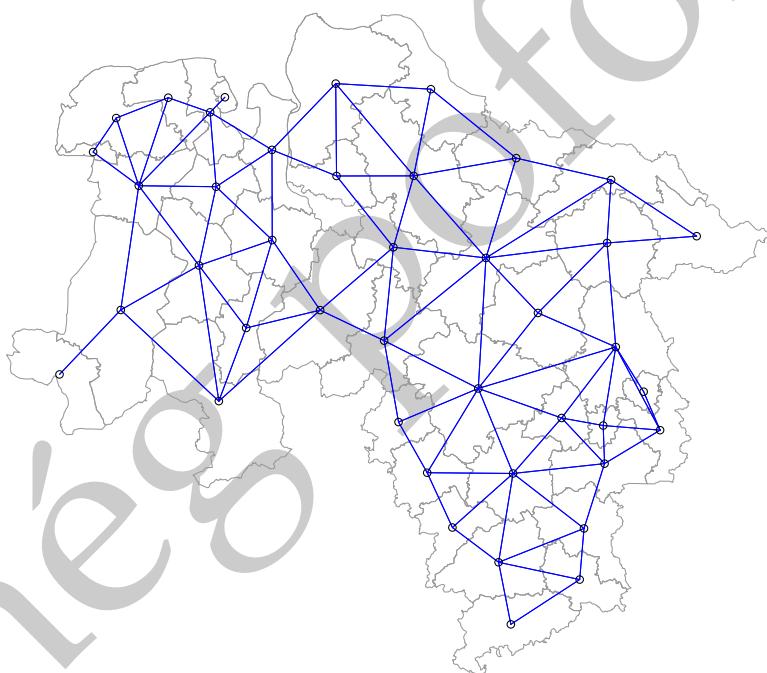
vektor és az egyes szomszédosságokhoz rendelhetünk súlyokat. Általában minden szomszédosságnak ugyanakkora érteket szokás adni, pl. egyet. A `num[]` vektor a poligonok szomszédainak számát sorolja fel.

A szomszédsági struktúra leírására több eszköz is rendelkezésre áll. Ilyen az R `spdep` könyvtárának `poly2nb()`-függvénye, aminek segítségével valamely az `sp` könyvtár `SpatialPolygons` típusú objektumából kiolvashatjuk a szomszédsági struktúrát, amit a `plot.nb()` függvénnnyel ki is rajzoltathatunk (18. ábra):

```
# R
> library(spdep)

> nb = poly2nb(lower.saxony.map, row.names=lower.saxony.map$ID)
Neighbour list object:
Number of regions: 42
Number of nonzero links: 180
Percentage nonzero weights: 10.20408
Average number of links: 4.285714

> plot(lower.saxony.map, border = "grey60")
> plot(nb, coordinates(lower.saxony.map), add=T, col='blue')
```



18. ábra. Szomszédsági térkép. A poligonok középpontjai közti kék egyenesek jelzik azt, hogy a szomszédsági struktúrában a poligonok szomszédosak. Az ábra segíthet áttekinteni, hogy a `poly2nb()` függvény helyesen hozta-e létre a szomszédsági mátrixot.

Az `spdep` könyvtár `nb2WB()` függvényével a kiolasztott szomszédsági szerkezet átalakítható olyan formába, amilyet a BUGS `car.normal()`

függvényének paramétereinek igényel:

```
> nb2WB(nb)
$adj
 [1] 5 10 21 26 39 8 21 41 11 18 30 32 11 16 34 36 1 9 21 26 27 37 28 31 35 39 24 26 27 37 38
[32] 2 21 5 14 21 27 1 21 39 40 41 3 4 16 18 30 36 42 19 25 29 32 25 29 9 16 19 20 33 4 11
[63] 15 19 24 30 33 34 23 31 34 35 3 11 32 42 12 15 16 20 25 30 32 15 19 25 1 2 5 8 9 10 41
[94] 23 36 17 22 34 36 7 16 33 34 38 12 13 19 20 29 1 5 7 37 39 5 7 9 37 6 31 38 39 12 13
[125] 25 3 11 16 19 32 6 17 28 34 35 38 3 12 18 19 30 15 16 24 4 16 17 23 24 31 36 38 6 17 31
[156] 4 11 22 23 34 5 7 26 27 7 24 28 31 34 1 6 10 26 28 10 2 10 21 11 18

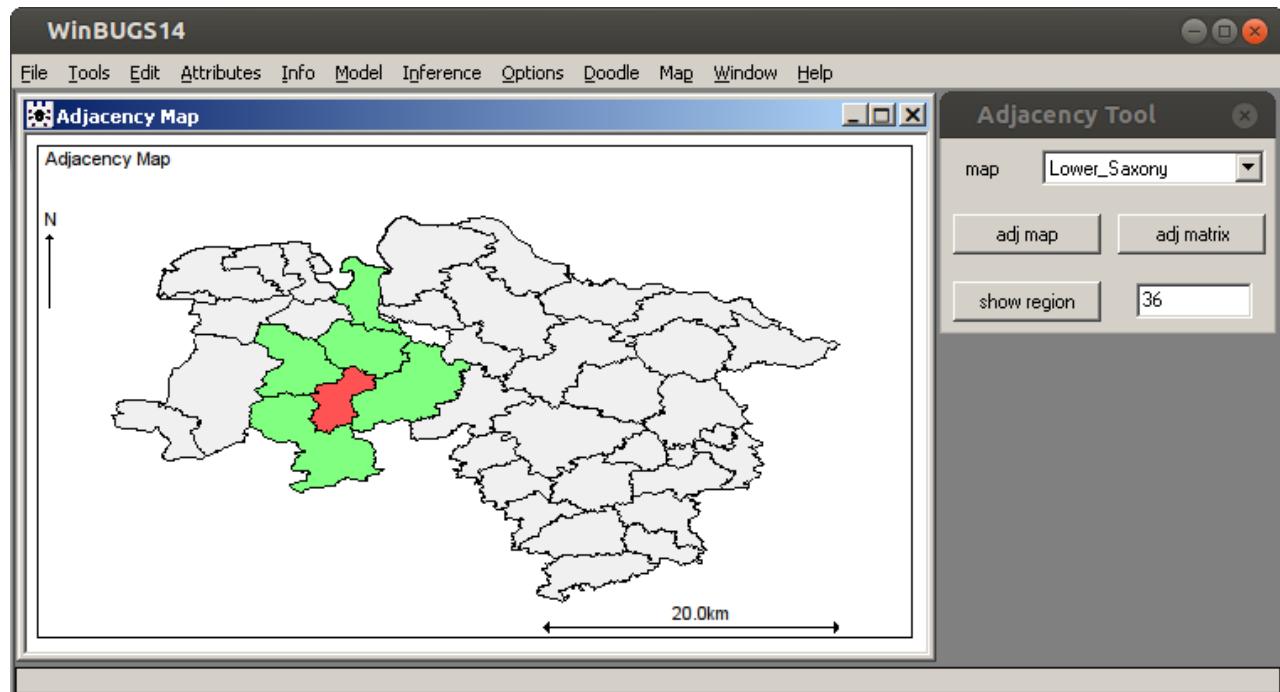
$weights
 [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[48] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[95] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[142] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

$num
[1] 5 3 4 4 6 4 5 2 4 5 7 4 2 1 4 8 4 4 7 3 7 2 4 5 5 5 4 4 3 5 6 5 3 8 3 5 4 5 5 1 3 2
```

Habár az `spdep` függvényeivel a `car.normal()` függvény paramétereinek szükséges információk kiolvashatók, a szomszédsági mátrix ( $w_{ij}$ ) szerkesztése meglehetősen körülményes. Miért lehet a kiolasztott szomszédsági mátrix szerkeszthetősége fontos? Előfordul, hogy olyan térképi állomány áll rendelkezésünkre a modellezéshez, amiben a poligonok egymáshoz való illeszkedése nem tökéletes. Ilyenkor a `poly2nb()` függvény olyan szomszédsági struktúrát eredményezhet, amelyben a valóságban szomszédos poligonok nem szomszédosként jelennék meg.

A WinBUGS-ban lehetőségünk van arra, hogy a szomszédsági mátrixot vizuálisan szerkesszük. A `Map` menü `Adjacency Tool...` menüpontjára kattintva megjelenik az `Adjacency Tool` ablak (19. ábra). Ebben az ablakban a `map` legördülő mezőből kiválasztjuk azt a térképet, amelynek a szomszédsági mátrixát szeretnénk szerkeszteni, majd használni. Az `adj map` gombra kattintva az `Adjacency Map` ablakban megjelenik a térképünk.

A szomszédsági mátrix szerkesztése során kijelölünk egy poligont, ezt megtehetjük úgy, hogy a `Adjacency Map` rákattintunk az egérrel vagy úgy, hogy a `show region` gomb melletti mezőben megadjuk a poligon sorszámát, és rákattintunk a `gombra`. Ekkor a megadott poligon pirossá válik, a WinBUGS által meghatározott szomszédai pedig zöldévé válnak. A szomszédsági szerkezetet úgy tudjuk módosítani, hogy a kontroll gomb lenyomása mellett valamely poligonra kattintunk az egérrel. Ha az így kijelölt poligon szürke, akkor zöldévé változik, ha zöld, akkor szürkevé. Ha elkészültünk a szerkesztéssel, akkor az `adj matrix` gombra kattintva a megjelenő `Adjacency Matrix` ablakban megkapjuk a szomszédsági mátrix `adj[]` és `num[]` vektorai mellé az összes szomszédossági kapcsolat számát (`sumNumNeigh`). Ez utóbbi



érték az `adj[]` vektor elemeinek a számával, illetve a `num[]` vektor értékeinek összegével egyezik meg. A példánkban használt térkép esetén az R 180, a WinBUGS 188 szomszédsági kapcsolatot állapított meg. Ha egyenként leellenőrizzük az egyes járások szomszédait, akkor azt fogjuk látni, hogy a WinBUGS hibásan azonosított olyan poligonokat szomszédként, amelyek valójában nem szomszédok, erre látunk példát a 19. ábrán. Azonban ahogy láttuk, a WinBUGS-ban legalább vizuálisan lehet ellenőrizni a szomszédsági mátrix jóságát.

De itt jön egy újabb nehézség: a WinBUGS nem tudja kezelni a térinformatikában általánosan használt térképi formátumokat (pl. ESRI shape-fájl, térbeli adatbázisok). Ehelyett importálhatunk a WinBUGS-ba három formátumba (ArcInfo, EpiMap, Splus) transzformált térképeket. Az általánosan használt térképi formátumok ilyen formátumba való átalakítására azonban nemigen vannak szoftverek. Az Epi Info<sup>10</sup> szoftver elméletileg tud olvasni különböző standard térképi formátumokat, azonban a gyakorlatban a transzformációjuk EpiMap formátumba nem működik, a legtöbb esetben.

A maps2WinBUGS<sup>11</sup> Quantum GIS<sup>12</sup> plugin abból a célból született, hogy a GeoBUGS modullal végzett elemzésekhez a szükséges térképi és egyéb adat-előkészítési feladatokat segítse (Solymosi et al., 2010). Ennek az eszköznek a segítségével a Quantum GIS által megnyitható bármely vektoros, poligon térkép átalakítható olyan formátumba (ArcInfo, Splus), amit a WinBUGS-ba importálhatunk. De emellett ge-

19. ábra. A WinBUGS GeoBUGS moduljának szomszedsági mátrix szerkesztő felülete. Az Adjacency Map ablakban bemutatott térképen látszik, hogy a pirosan jelzett poligonhoz olyan poligont is szomszédságként határozott meg a GeoBUGS, amely valójában nem érintkezik a piros poligonnal.

<sup>10</sup> <http://www.cdc.gov/epiinfo/>

<sup>11</sup> <http://solymosin.github.io/maps2winbugs/>

<sup>12</sup> <http://www.qgis.org/>

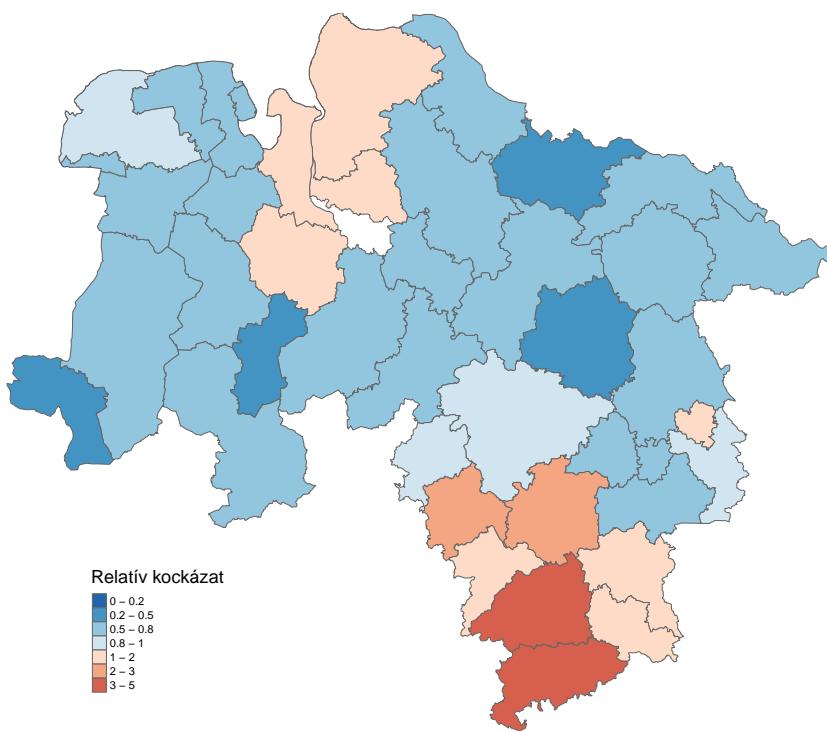
nerálhatunk szomszédsági mátrixot különböző szabályok (*Touches*, *Intersections*, *Within distance* (Allepuz et al., 2009)) alapján. Az így létrejött szomszédsági mátrix vizuálisan szerkeszthető és exportálható a WinBUGS-nak megfelelő formátumban.

A modellünkben eddig a szomszédok hatását vagy nem vettük figyelembe, vagy a modell járásoknáki hatásának priorjaként csak a szomszédok hatását használtuk. Besag et al. (1991) azt javasolja, hogy a területi egységek (itt a járások) random hatását kezeljük úgy, hogy szétbontjuk a szomszédoktól függő ( $b[i]$ ) és azuktól nem függő ( $h[i]$ ) random hatásokra. A szakirodalomban BYM (Besag-York-Mollié) modellnek nevezett modell BUGS-kódja:

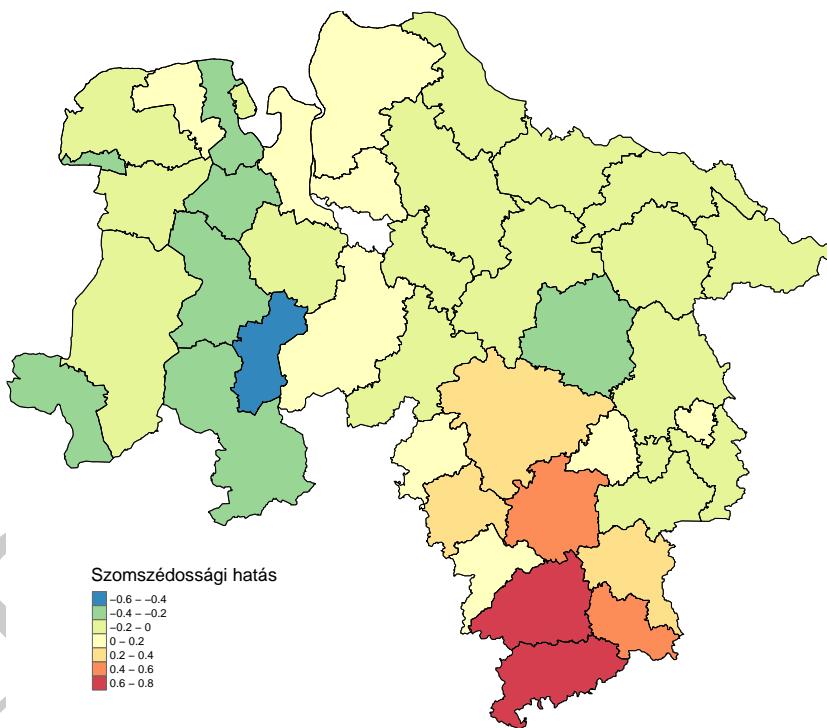
```
model{
  for (i in 1 : N) {
    esetszam[i] ~ dpois(mu[i])
    log(mu[i]) <- log(varhato.esetszam[i]) + alpha + b[i] + h[i]
    RR[i] <- exp(alpha + b[i] + h[i])
    h[i] ~ dnorm(0, tau.h)
  }
  b[1:N] ~ car.normal(adj[], weights[], num[], tau.b)
  alpha ~ dflat()
  tau.b ~ dgamma(0.5, 0.0005)
  tau.h ~ dgamma(0.5, 0.0005)
}
```

A modell futtatásából származó járási relatív kockázati becslések a 20. ábrán láthatók. A modellben a  $h[i]$  node flat prior kapott, ugyanúgy, ahogy a legelső log-normális modellünkben. Ez a modell a járások hatására vonatkozóan rugalmasabb, mivel lehetőséget nyújt arra vonatkozóan, hogy a járásoknáki reziduális kockázatok tekintetében szétválasszuk, hogy milyen mértékben származnak a szomszédsági hatásokból (21. ábra) és milyen mértékben a járás szomszédoktól független hatásából (22. ábra).

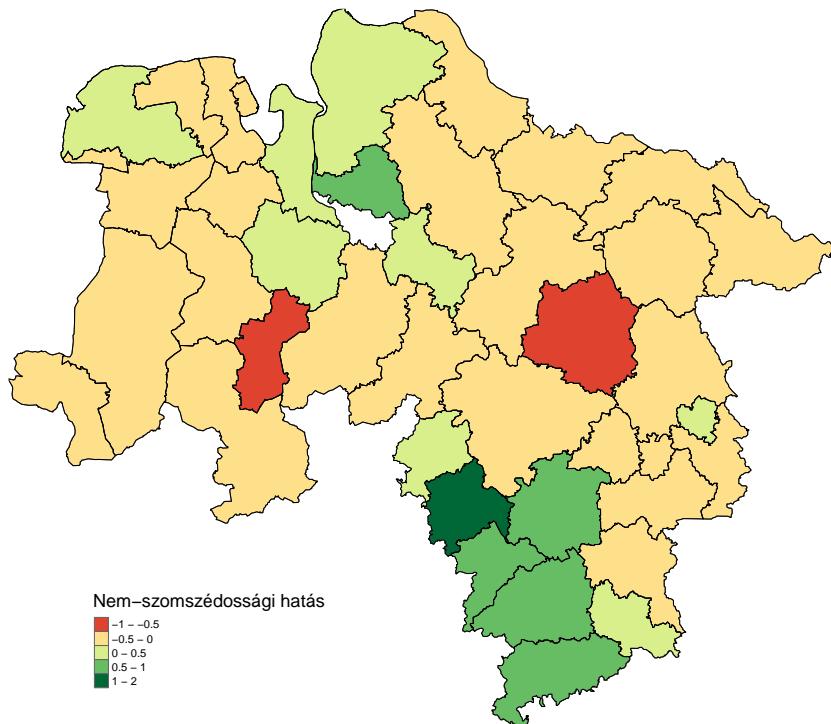
Modellilleszkedési mutatók alapján a BYM-modell illeszkedett legjobban az adatainkhoz. Amiből azt gondolhatjuk, hogy ez a modell írja le legjobban a valóságot, így az ebből származó járásoknáki relatív kockázatra vonatkozó becsléseink állnak legközelebb a valósághoz.



20. ábra. A Besag-York-Mollié (Besag et al., 1991) modell alapján becsült *E. multilocularis*-fertőzöttségi kockázati térkép.



21. ábra. A Besag-York-Mollié (Besag et al., 1991) modell alapján a *E. multilocularis*-fertőzöttségi kockázat becslésében szomszédsági hatás részesedése ( $b_{[i]}$ ).



22. ábra. A Besag-York-Mollié (Besag et al., 1991) modell alapján a *E. multilocularis*-fertőzöttségi kockázat becslésében a nem-szomszédsági viszonyokból származtatható járásoknál hatás ( $h[i]$ ).

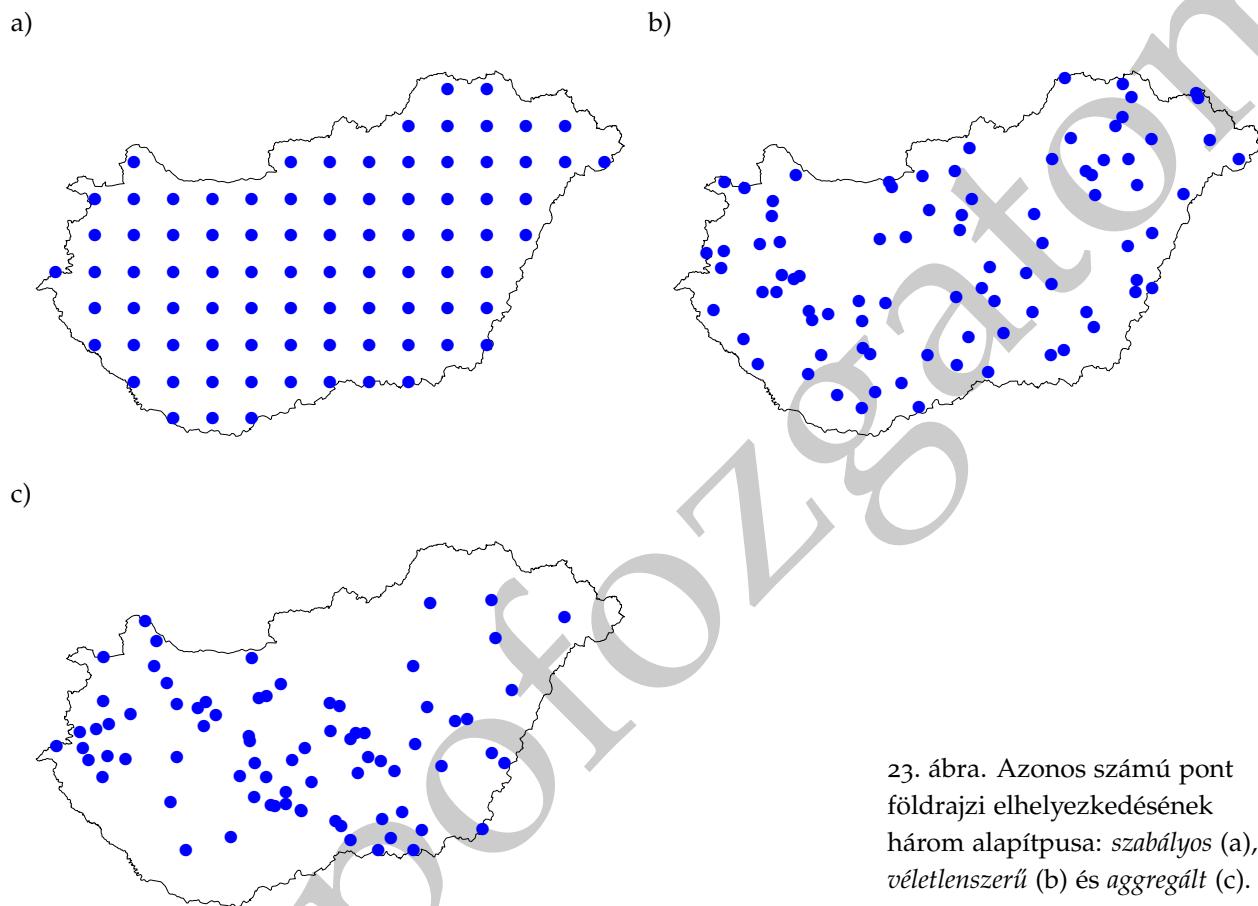
### Földrajzi mintázatelemzés

Az előző szakaszban bemutatott betegség-térképezési példákból láthattuk, hogy a térképi megjelenítést több tényező is erősen befolyásolhatja (pl. pontok mérete, színskála, simítási eljárás, illetve annak paramétereinek). Ezért a kockázat földrajzi mintázatának pusztán vizuális értékelése szubjektív következtetésekhez vezethet. Ugyanazt a kockázati térképet eltérő módon értelmezhetik különböző személyek. Ezért számos olyan módszert hoztak létre, amelyek a térképen megjelenített kockázatok értékelését objektívebbé teszi.

A geoepidemiológiában a kockázat földrajzi mintázatának statisztikai módszerek segítségével történő elemzését nevezzük *földrajzi mintázatelemzésnek*. Az ebben a téma körben alkalmazott statisztikai eljárások valamilyen tesztstatisztika segítségével számszerűsítik a mintázat valamely jellegzetességét. Az epidemiológiában leggyakrabban használt földrajzi klaszterelemzések mellett a *határkeresést* mutatjuk be a következőkben.

### Klaszterelemzés

A geostatistikában a pontok kétdimenziós térben való elhelyezkedésére vonatkozóan három típust szoktak megkülönböztetni: szabályos, véletlenszerű és aggregált (23. ábra).



23. ábra. Azonos számú pont földrajzi elhelyezkedésének három alapítpusa: szabályos (a), véletlenszerű (b) és aggregált (c).

A klaszterelemzés célja, hogy az esetek, kockázatok földrajzi halmozódását, klasztereződését vizsgálja. A CDC (1990) definíciója alapján *klaszternek* nevezzük az egészséggel kapcsolatos események látszólagos vagy valódi csoportosulását térben és/vagy időben.<sup>13</sup> Az állatorvosi epidemiológiai szakirodalomban számos földrajzi klaszterelemzési módszert találhatunk (Ward & Carpenter, 2000a,b). A módszereket csoportosítani szokták aszerint, hogy *pontszerű* vagy *területi* adatokra alapozott eljárások. Megkülönböztetnek továbbá *globális*, *lokális* és *fókusztált* klaszterelemzési módszereket.

A globális klaszterelemzési módszerekkel az esetek földrajzi halmozódására vonatkozóan egy tesztstatisztika értékhez juthatunk. Ez az érték arról tájékoztat pusztán, hogy a vizsgált területen az esetek

<sup>13</sup> „cluster is an unusual aggregation, real or perceived, of health events that are grouped together in time and/or space” (CDC, 1990)

halmozottan fordulnak elő vagy sem. A lokális klaszterezési módszerek lehetővé teszik, hogy ha kimutathatók klaszterek a vizsgálati területen, akkor azok hol vannak. A fókusztált klaszterezési módszerek alkalmazása esetén egy előre meghatározott földrajzi helyre vonatkozóan kapunk halmozódási információt.

A következőkben az *E. multilocularis*-adatsor felhasználásával néhány, széles körben alkalmazott eljárást mutatunk be. A bemutatott módszerek közül a Moran I és a Geary c autokorrelációra alapozott eljárások. Az autokorrelációs módszerek a megfigyelési egységek közötti szomszédsági viszonyt, illetve a vizsgált változóbeli hasonlóságot elemzik.

*Szomszédsági mátrix.* A szomszédsági kapcsolat kifejezésére a szomszédsági mátrixot használják, aminek  $w_{ij}$  elemei az  $i$  és  $j$  területi egységek kapcsolatát fejezi ki:

- $w_{ij} = 0$ , ha  $i$  és  $j$  területi egységek nem határosak
- $w_{ij} = 1$ , ha  $i$  és  $j$  területi egységek határosak
- általában  $w_{ii} = 0$ , de  $w_{ii} = 1$  alkalmazása is előfordul

A  $w_{ij}$  értékek abban az esetben számítódnak így, ha a vizsgálat területi adatok alapján történik.<sup>14</sup>

Amennyiben pontszerű adatok alapján végzünk autokorrelációs vizsgálatokat, akkor általában a pontok közötti távolság és valamilyen távolságbeli határérték alapján hozzuk létre a mátrixot:

- $w_{ij} = 0$ , ha  $i$  és  $j$  pontok távolsága nagyobb, mint  $h$ -távolság
- $w_{ij} = 1$ , ha  $i$  és  $j$  pontok távolsága kisebb vagy egyenlő, mint  $h$ -távolság
- általában  $w_{ii} = 0$ , de itt is lehet  $w_{ii} = 1$

Pontszerű adatok esetén további lehetőség, hogy a pontok köré Voronoj-cellákat<sup>15</sup> rajzolunk, majd az így létrejött poligon-struktúrát már úgy kezeljük, mint terület alapú adatállomány és annak megfelelően hozzuk létre a szomszédsági mátrixot.

*Globális Moran I.* Ez talán a leggyakrabban alkalmazott globális klaszterteszt ([Moran, 1950](#)). Ez a statisztika a területi ( $n$  egység) referenciával rendelkező  $x$  valószínűségi változó függetlenségének tesztelésére szolgál:

$$I = \frac{n}{S_0 \sum_i z_i^2} \sum_{i,j} w_{ij} z_i z_j$$

$$z_i = x_i - \bar{x}$$

<sup>14</sup> Ilyen szomszédsági mátrixot létrehozhatunk a [16.](#) oldalon kezdődő leírás alapján.

<sup>15</sup> A Voronoj-cellák olyan poligonok, amelyek területén lévő összes pont közelebb van a poligon középpontjához, mint más poligonok középpontjához.

Területi elrendeződés	Geary c	Moran I
Aggregált	$0 \leq c < 1$	$I > 0$
Szabályos	$1 < c < 3$	$I < 0$
Véletlenszerű	$c = 1$	$I = 0$

2. táblázat. A globális Geary c és Moran I tesztstatisztikák értékeinek értelmezése.

$$S_0 = \sum_{i,j} w_{ij}$$

, ahol  $x_i$  az  $i$ -edik területi egységbeli kockázatot leíró érték. Az  $I$  értéke  $-1$  és  $1$  közé esik, az értelmezéséhez szükséges határértékek a 2. táblázatban olvashatók. Az R spdep-csomagjának `moran.test()`-függvényével az *E. multilocularis* járásonkénti prevalenciájának halmozódását elemezve az alábbi eredményt kapjuk:

```
# R
> library(spdep)

> nb = poly2nb(saxony.df, row.names=saxony.df$ID_3)
> (MI = moran.test(saxony.df$prev, listw=nb2listw(nb, style = "W")))

Moran's I test under randomisation

data: saxony.df$prev
weights: nb2listw(nb, style = "W")

Moran I statistic standard deviate = 5.5582, p-value = 1.363e-08
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.54047145     -0.02439024    0.01032791
```

A Moran I statistic értéke<sup>16</sup> azt jelzi, hogy pozitív autokorreláció állapítható meg a járásonkénti prevalencia-értékek földrajzi mintázatában, vagyis klasztereződik a fertőzöttség kockázata, az eredmény szignifikáns ( $p < 0.0001$ ).

<sup>16</sup> nagyobb, mint 0

*Globális Geary c.* A folytonossági hánnyadosnak is nevezett mérték:

$$c = \frac{n - 1}{2S_0 \sum_{i,j} z_i^2} \sum_{i,j} w_{ij}(x_i - x_j)^2$$

A Geary  $c$  0 és 3 közötti értékeket vehet fel, és nem lehet negatív szám. A tesztstatisztika értékének értelmezését segíti a 2. táblázat. Az R spdep-csomagjának `moran.test()`-függvényével az *E. multilocularis* járásonkénti prevalenciájának halmozódását elemezve az alábbi eredményt kapjuk:

```
# R
> library(spdep)

> nb = poly2nb(saxony.df, row.names=saxony.df$ID_3)
> (Gc = geary.test(saxony.df$prev, listw=nb2listw(nb, style = "W")))

Geary's C test under randomisation

data: saxony.df$prev
```

```

weights: nb2listw(nb, style = "W")

Geary C statistic standard deviate = 4.5358, p-value = 2.87e-06
alternative hypothesis: Expectation greater than statistic
sample estimates:
Geary C statistic      Expectation      Variance
0.43313436      1.00000000      0.01561929

```

A Geary C statistic értéke<sup>17</sup> azt jelzi, hogy pozitív autokorreláció állapítható meg a járásonkénti prevalencia-értékek földrajzi mintázatban, vagyis klasztereződik a fertőzöttség kockázata, az eredmény szignifikáns ( $p < 0.0001$ ). A  $c$ - és az  $I$ -statisztikák eredménye ugyanúgy klasztereződésre utal.

<sup>17</sup> kisebb, mint 1

*Lokális Moran I.* Míg a *globális Moran I* az összes vizsgálati egységre egyetlen tesztstatisztika értéket és egy szignifikancia-szintet ad, a *lokális* módszer (Anselin, 1992) minden területi egységre meghatározza ezeket az értékeket, külön-külön.

$$I_i = \frac{z_i}{m} \sum_j w_{ij} z_j$$

$$m = \frac{1}{n} \sum_i z_i^2$$

Az R spdep-csomagjának localmoran()-függvényével az *E. multilocularis* járásonkénti prevalenciájának halmozódását elemezve az alábbi eredményt kapjuk:

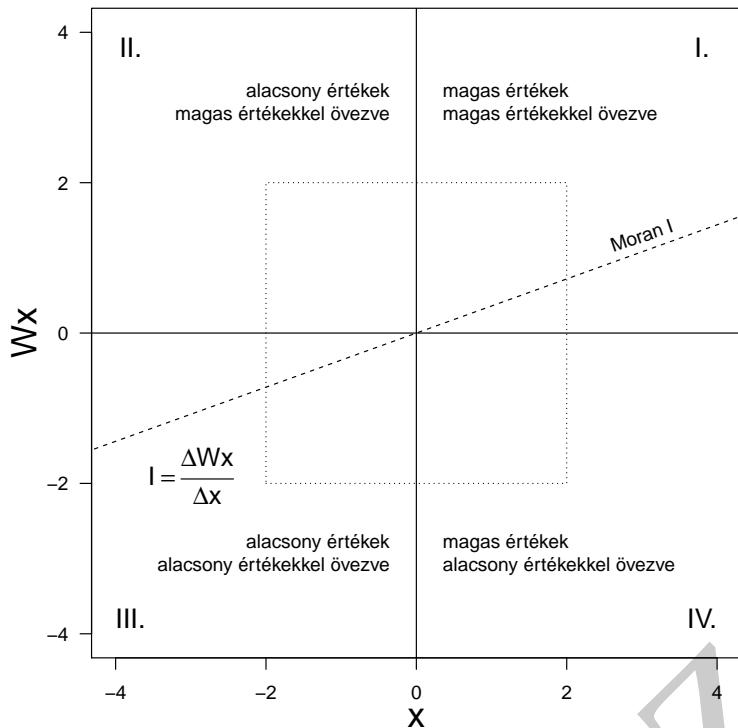
```

# R
> library(spdep)

> nb = poly2nb(saxony.df, row.names=saxony.df$ID_3)
> LISA = localmoran(as.vector(saxony.df$prev), nb2listw(nb, style="W"))
> head(LISA)
   Ii      E.Ii    Var.Ii      Z.Ii    Pr(z > 0)
23041 0.254596615 -0.02439024 0.2057601 0.61503973 2.692642e-01
23042 0.086809574 -0.02439024 0.1094258 0.33615855 3.683757e-01
23043 0.003763955 -0.02439024 0.2057601 0.06206726 4.752546e-01
23044 6.562042766 -0.02439024 0.4305401 10.03791459 5.192317e-24
23045 0.021121822 -0.02439024 0.2057601 0.10033350 4.600398e-01
23046 4.111202504 -0.02439024 0.1608041 10.31309944 3.074288e-25

```

A táblázat sorai tartalmazzák az egyes járásokra vonatkozó lokális becsléseket. Nagyszámú területi egység esetén nehezen áttekinthetők a lokális eredmények ( $I_i$ ) táblázatos formában. Ezért az egyes lokális Moran  $I_i$  értékek vizualizációjára Anselin (1992) a 24. ábrán látható, általa Moran-szórásdiagramnak nevezett megjelenítést javasolja.



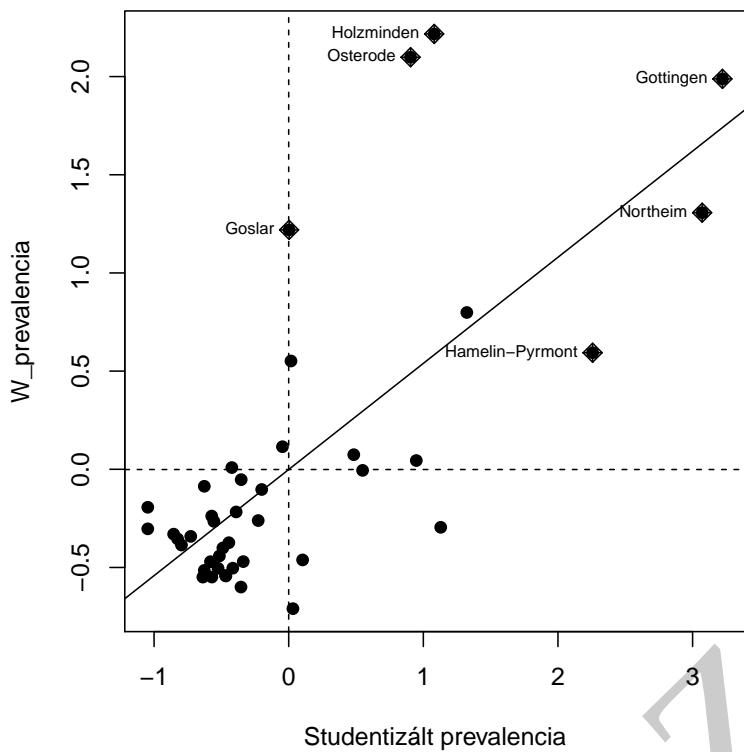
24. ábra. Moran-szórásdiagram és értelmezése. Az x-tengelyen az egységekben, az y-tengelyen pedig a szomszédaikban megfigyelt standardizált prevalenciát ábrázoljuk. A  $Wx \sim x$  lineáris modell regressziós együtthatója a globális Moran  $I$ -vel egyezik meg.

A Moran-szórásdiagramon az x-tengelyen az egységekben, az y-tengelyen pedig a szomszédaikban megfigyelt prevalenciát ábrázoljuk. Általában nem a nyers megfigyelt értékeket szokták használni az ábrához, hanem azoknak valamilyen standardizált értékét, pl. studentizált<sup>18</sup> értékeket.

Az előzőekben táblázatos formában bemutatott lokális Moran  $I$  eredmények grafikusan a 25. ábrán láthatók. Az ábrán a négyzettel jelölt járások a 24. ábrán bemutatott kvadrátorok közül az I. kvadrátba esnek. Amely kvadrátból azok a földrajzi egységek rajzolódnak, amelyekben és amelyeknél a szomszédjaiban is magas a vizsgált változó értéke (pl. prevalencia). Így az ábráról leolvasható, hogy Göttingen, Hamelin-Pyrmont, Holzminden, Northeim és Osterode járások olyan magas prevalenciájú járások, amelyek környezetében is magas prevalenciájú járások helyeződnek. Vagyis ezek a járások földrajzi klaszterként azonosíthatók a lokális Moran  $I$  statisztika segítségével.

Az így lokalizált klasztereket alkotó járásokat ábrázolhatjuk klaszter-térképen (26. ábra). Az ábrán látható szignifikáns halmozódásként azonosított terület nagyból egybevág azzal régióval, amelyet a betegség-térképezés során is magas kockázatúnak értékelhettünk. A klaszter lokalizálása a betegség-térképezésből levont következetésekkel szemben azonban teljesen objektívnek tekinthető.

<sup>18</sup> az egyes értékekből kivonjuk az átlagukat, majd elosztjuk a szórásukkal



25. ábra. Moran-szórásdiagram.  
Az I. kvadrátból lévő négyzetek azokat a járásokat jelzik, amelyekben (és környezetükben is) magas prevalenciát figyeltünk meg. Valamely tengelyen kettőnél magasabb értékekkel rendelkező klaszter-elemek szignifikánsak ( $p < 0.05$ ).

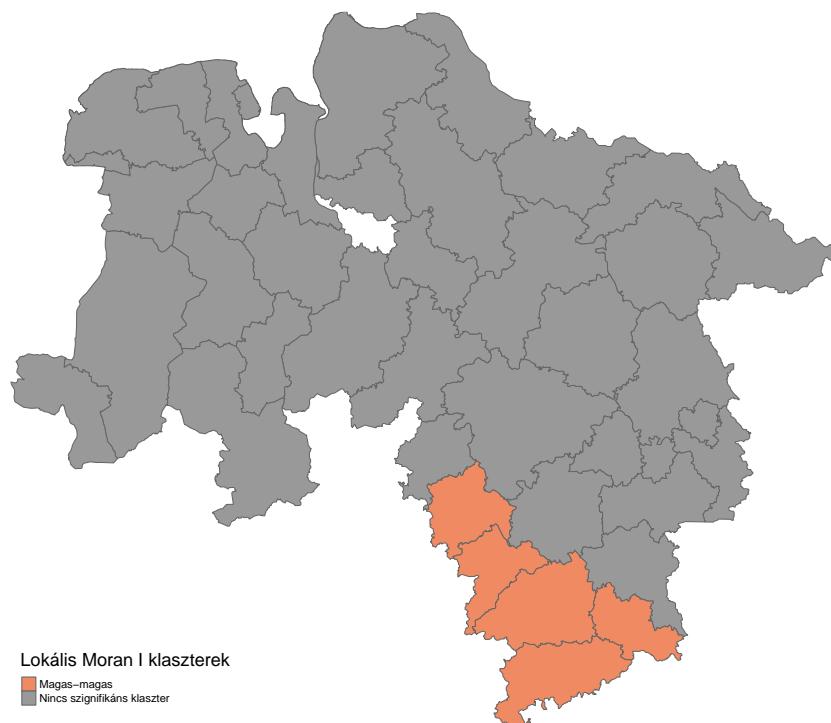
Az objektivitás azonban, ahogy a következő módszerből látható is lesz, nem jelenti azt, hogy hibátlan a módszer.

*Kulldorff-féle szken-statisztika.* Számos ún. szken-statisztikát<sup>19</sup> használunk földrajzi és/vagy időbeli klaszterek lokalizációjára. Az epidemiológiában talán leggyakrabban alkalmazott szken-statisztika a **Kulldorff & Nagarwalla (1995)** által közölt megoldás.

A módszer főbb mozzanatai a következők: A kiindulási térkép előkészítése során általában olyan vetületi rendszerbe transzformáljuk azt, amelyen a hosszúsági és szélességi egységek azonos távolság-nak felelnek meg. A vizsgált területre egy rácsot illesztünk, aminek rácspontjain végigmenve mindegyikből, mint középpontból, köröket képzünk.<sup>20</sup> A körök sugarát folyamatosan növeljük, aminek felső korlátját általában úgy állapítják meg, hogy az azzal kialakított kör területe ne legyen nagyobb, mint az egész vizsgálati terület nagyságának a fele. Más megközelítésben nem a *terület* nagyságának a fele, hanem a vizsgálati terület *populációjának* a fele a kör méretének felső határa. minden létrehozott körhöz összeszámoljuk, hogy a körön belül és kívül hány eset fordul elő, illetve kiszámoljuk a körön belül és kívül várható esetek számát. A megfigyelt esetszámokat összevetjük elméleti eloszlások alapján várható esetszámokkal, és azt a kört, amiben az esetek száma a körön kívüli esetszámokhoz viszonyítva

<sup>19</sup> scan statistic

<sup>20</sup> A kör helyett használhatunk ellipszeket is.



26. ábra. A lokális Moran *I* segítségével azonosított *E. multilocularis* prevalencia alapján lokalizált klaszterek.

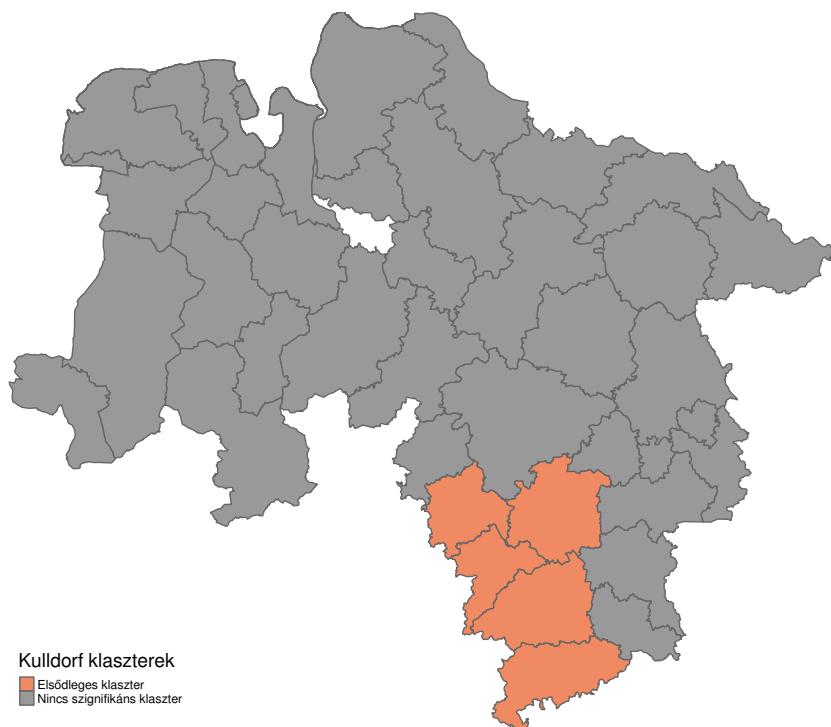
a legkisebb valószínűségű, azonosítjuk, mint elsődleges klasztert.<sup>21</sup> A szken-statisztikákban különböző valószínűségi modelleket használhatunk, attól függően, hogy milyen típusú adatokat vizsgálunk. Ha pl. ha olyan adatsort elemzünk, amelyben az egyes földrajzi helyekről csak azt tudjuk, hogy pozitívak vagy negatívak voltak, akkor Bernoulli-eloszlást használunk. Ha olyan adataink vannak, amelyek leírják az egyes helyeken előfordult esetek számát is, akkor Poisson-eloszlást használunk.

AZ R-ben több csomag is elérhető olyan függvényekkel, amelyek lehetővé teszik a Kulldorff-féle klaszter-lokalizációt. A SpatialEpi csomag `kulldorff()`-függvényének segítségével, Poisson-modellt alkalmazva a 27. ábrán látható elsődleges klasztert azonosítottuk a járási *E. multilocularis*-prevalencia adatok alapján.

A Kulldorff-eljárással a következő járások alkotnak magas prevalenciájú klasztert: Göttingen, Hamelin-Pyrmont, Holzminden, Lüneburg, Northeim. A lokális Moran *I* módszerhez képest nemileg eltérő klaszter azonosítható. Az ottaniban szereplő Osterode kiesett ebből a klaszterből és bekerült új járásként Lüneburg.

Ahogy más esetekben is ajánlatos több oldalról, több módszerrel megvizsgálni adatainkat, ugyanez áll a földrajzi klaszterelemzésre is. Mivel minden egyes statisztikai eljárásnak vannak korlátai, ezek egy részét nem is találjuk meg részletesen leírva a szakirodalomban,

<sup>21</sup> valószínűséghányados-próba (likelihood ratio test) segítségével vetjük össze



27. ábra. Kulldorff-féle eljárás-sal lokalizált klaszterek.

ha párhuzamosan használunk kettőt vagy többet, árnyaltabb képet kaphatunk a vizsgált jelenséggel kapcsolatban.

### Határkeresés

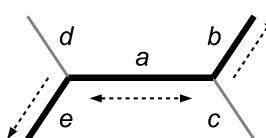
Míg a klaszterelemzések arra a kérdésre keresik a választ, hogy a kockázat vonatkozásában földrajzilag meg lehet-e állapítani halmozódásokat, a határkeresési megközelítések a szomszédok közti különbségek alapján húznak meg határokat (Solymosi et al., 2005a,b).

*Monmonier-féle legnagyobb különbsségek módszere.* A módszer (Monmonier, 1973) számítási lépései a következők: A vizsgált terület minden  $i$  és  $j$  szomszédos területi-egység-párjára vonatkozóan kiszámítjuk az  $x$  kockázatbeli<sup>22</sup> különbségeket ( $d_{ij}$ ):

$$d_{ij} = |x_i - x_j|$$

A  $d_{ij}$ -értékeket sorba rendezve kiválasztjuk a legnagyobb értéket, és mint legnagyobb különbséget jelentő határt jelöljük meg. Ezután a kijelölt határral szomszédos határok felé indul el az algoritmus. Monmonier (1973) azt javasolja, hogy először balra indulunk el, és addig húzzuk tovább a határt, amíg egy másik határvonalba nem ütközünk, illetve amíg el nem érjük a vizsgálati terület szélét. Ezek után indul-

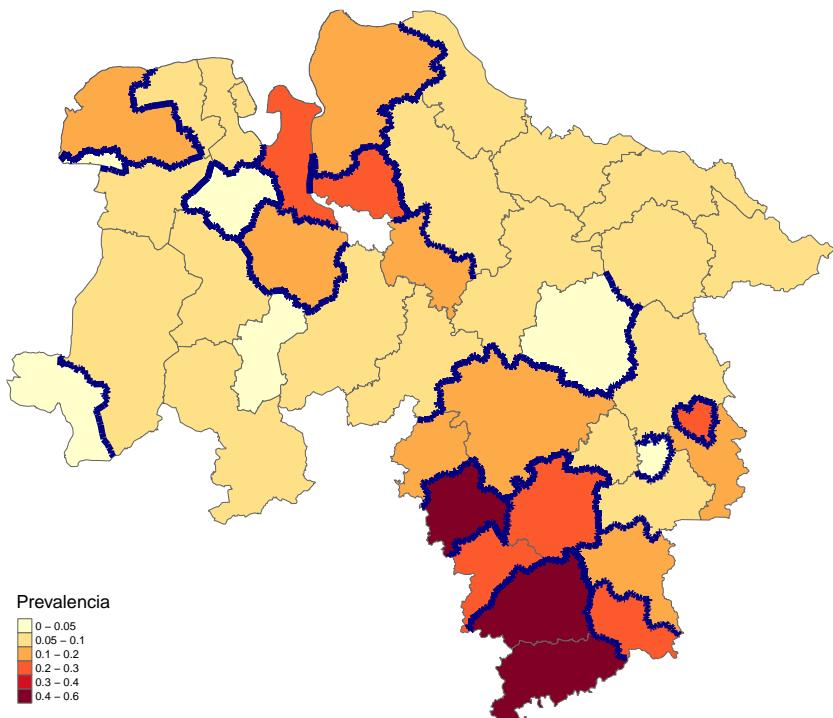
<sup>22</sup> esetünkben a prevalence



Monmonier (1973) határkeresés. Az  $a$  egyenes jelzi a legnagyobb különbsséget mutató két szomszédos poligon közötti határt. Ennek végpontjaiból a nagyobb különbségeket jelentő egyenesek mentén rajzolódik tovább a határ.

junk el jobbra, és tegyünk hasonlóan. Monmonier (1973) azt javasolja, hogy ha olyan szakaszhoz jutunk, amely által elválasztott területi egységek közti különbség nem nagyobb, mint a  $d_{ij}$ -értékek mediánja, akkor azt a szakaszt ne vonjuk be a határok közé és a határövezést abban az irányban állítsuk le. Ha így kialakítottunk egy több szakaszból álló határt, akkor a  $d_{ij}$ -sorban a következő legnagyobb különbségeket választva, a hozzáartozó  $l_{ij}$ -szakasznál indulva folytassuk a határépítést.

A 28. ábrán az *E. multilocularis*-prevalencia adatok alapján, a Monmonier-féle legnagyobb különbség módszerével létrehozott határok láthatók.



28. ábra. Az *E. multilocularis* járási prevalencia adatai alapján a Monmonier-féle legnagyobb különbőg módszerével létrehozott határok (kék)

## Összefüggés-elemzés

A betegség-terképezés és mintázatelemzés alapján egyértelműnek tűnik, hogy Alsó-Szászország délnyugati járásaiban jelentősen magasabb a fertőzöttség kockázata, mint a tartomány nagy részében. Állat- és humán-egészségügyi szempontból is fontos lenne tudni, hogy ez a kockázattöbblet miből származhat.

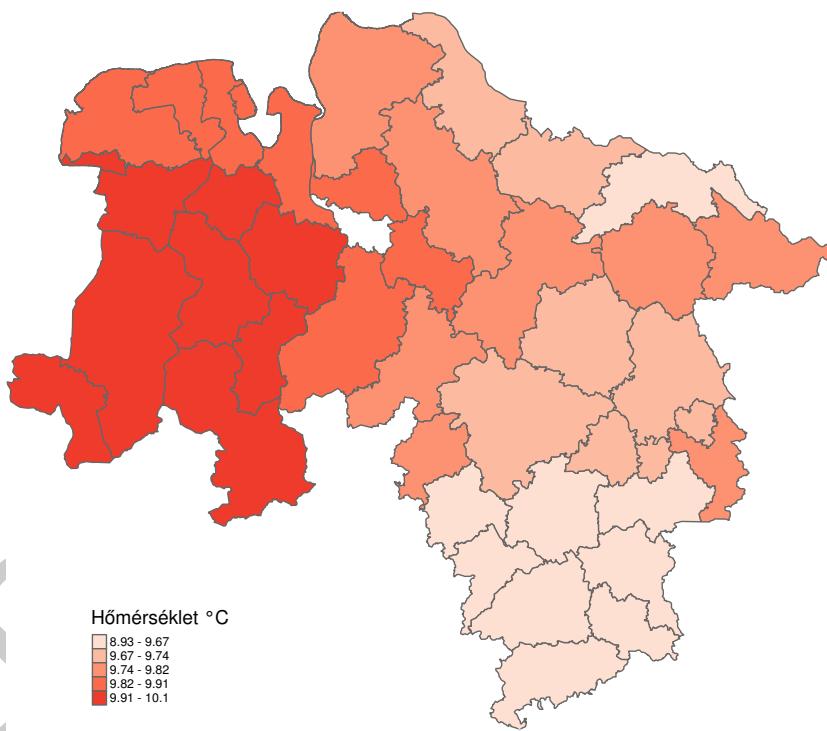
A *E. multilocularis* fejlődésmenetében van olyan szakasz, amikor a pete közvetlenül ki van téve külső környezeti tényezőknek. Így számos, a fertőzöttségi kockázatot befolyásoló tényező között a külső

környezeti feltételeknek is szerepe lehet. Ha valakinek alapos ismeretei vannak az alsó-szászországi környezeti tényezők földrajzi heterogenitására vonatkozóan, akkor könnyen lehet, hogy egyfajta szubjektív, de a valósághoz közelí magyarázattal szolgálhat a kockázatok eltérésének miértjeire.

A szakirodalomban számos szerző állapította meg, hogy az *E. multilocularis*-fertőzöttség mértéke pozitív kapcsolatot mutat az éves csapadék mennyiséggel, illetve negatív kapcsolatot a környezeti hőmérséklettel (Miterpáková et al., 2006; Burlet et al., 2011; Takeuchi-Storm et al., 2015). Emellett Miterpáková et al. (2006) úgy találta, hogy a talaj nedvességtartalma és a fertőzöttség prevalenciája között szintén pozitív kapcsolat mutatható ki. Abból a célból, hogy az alsó-szászországi, járásonkénti prevalenciák e környezeti változókkal való összefüggését vizsgálhassuk, be kell szereznünk ezeket az adatokat a vizsgálati területre vonatkozóan. A napi hőmérsékleti<sup>23</sup> és csapadék-adatokat a European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalízis adatbázisból kérdeztük le, 0,1-os felbontású rácson, az 1991–1997. közötti időszakra vonatkozóan (Dee et al., 2011).<sup>24</sup> Mindegyik járás esetén a rácsnak azt a pontját használtuk, amelyik a járás középpontjához legközelebb esett. A hőmérséklet esetén az éves átlaghőmérsékletek átlagát a 29. ábra mutatja be.

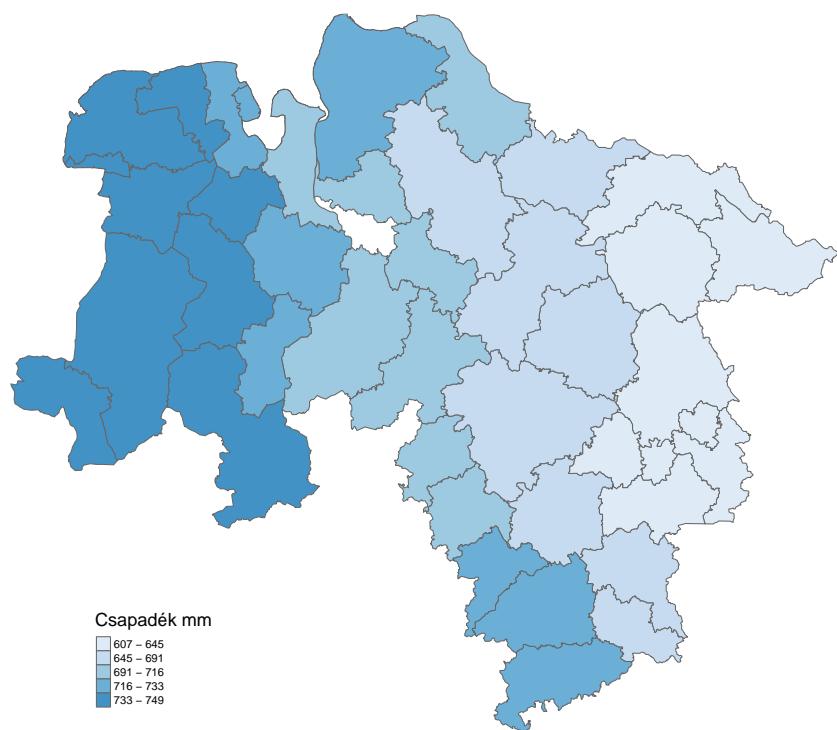
<sup>23</sup> 2 m magasságra vonatkozó hőmérséklet

<sup>24</sup> <http://apps.ecmwf.int/datasets/data/interim-full-daily>



29. ábra. Az 1991–1997. időszakban az évi átlaghőmérsékletek átlaga az ERA-Interim reanalízis adatbázis alapján.

A 29. ábrán a legalacsonyabb átlaghőmérsékletű járások nagy része a legmagasabb *E. multilocularis* prevalenciájú járásokkal azonos. Ami ugyan csak vizuális, szubjektív következtetés, de a hivatkozott közlemények eredményeit támogatja. Az ERA-Interim adatok alapján számított évi összes csapadék átlagát a 30. ábra mutatja be.



Habár a térkép alapján a legmagasabb prevalenciájú járások nem a legtöbb csapadékot kapó járások, ha a hőmérsékleti térképpel összevetjük, akkor azt láthatjuk, hogy a legtöbb csapadékot kapó járáskban az átlaghőmérséklet is a legmagasabb tartományba esik. Ez utóbbi pedig kedvezőtlen a parazita életciklusának külső környezetben töltött szakaszában. Ez értelmezhető úgy is, hogy a kevesebb csapadék mellett is kedvezőbb az alacsonyabb hőmérséklet a parazitapata fertőzőképességének megtartása szempontjából.

A talaj nedvességre vonatkozó adatokat a European Soil Data Centre (ESDAC) adatbázisból szereztük be (Panagos, 2006; Panagos et al., 2012).<sup>25</sup> Az ESDAC egész Európára és Oroszország ázsiai területeire vonatkozóan tartalmaz talajtani információkat, ennek megfelelően nagy állományban<sup>26</sup> tárolható. Ahhoz, hogy a csak a vizsgálati területre vonatkozó adatokat kinyerhessük érdemes fájlban való tárolás helyett olyan adatbázisba tölteni, ami az OGC-szabványnak<sup>27</sup> megfelelően lehetővé teszi térbeli objektumok tárolását, illetve azokon különböző műveletek végrehajtását.

30. ábra. Az 1991–1997. időszakban az éves összes csapadék átlaga az ERA-Interim reanalízis adatbázis alapján.

<sup>25</sup> <http://esdac.jrc.ec.europa.eu/>

<sup>26</sup> 115 MB

<sup>27</sup> <http://www.opengeospatial.org/>

Jelen példánkhoz az adatokat a többi eddig használt térképpel együtt egy PostgreSQL-PostGIS<sup>28</sup> adatbázisban tároltuk el. Ebben egy egyszerű SQL-utasítással kivágható a talajtani adatoknak az a része, amelyik már csak Alsó-Szászországra vonatkozik. Az RpostGIS könyvtár<sup>29</sup> wkt2sp()-függvényével az SQL-utasítás közvetlenül lefuttatható az R-környezetben, aminek eredményeként egy az sp-könyvtárban definiált térkép-objektumot kapunk (Solymosi et al., 2006a,b).<sup>30</sup>

```
# R
> library(RpostGIS)

sql = "select map.ID as id, talaj.gid, talaj.hg, astext(ST_Intersection(talaj.the_geom, map.the_geom)) as
geom from sgdb_ptr talaj, lower_saxony map where ST_Intersects(talaj.the_geom, map.the_geom)"

saxtalaj.dt = dbGetQuery(con, sql)
saxtalaj.dt$sid = 1:dim(saxtalaj.dt)[1]
saxtalaj.map = wkt2sp(geoms=saxtalaj.dt, gcol='geom', idcol='sid')
```

Az ESDAC adatbázisban többek között a hidrogeológiai típus mező (hg) írja le a talaj nedvességét. Ebben a mezőben négy fő típust tárolnak:

*HG1*: áteresztő altalajú talaj, a talajvíz mélyen van: ritkán nedves

*HG2*: alföldi talaj, amire a talajvíznek hatása van, szezonálisan vagy folyamatosan nedves, vagy mesterségesen csatornázott

*HG3*: olyan talaj, amiben 80 cm-en belül vízzáró réteg van, szezonálisan vagy folyamatosan nedves

*HG4*: felföldi vagy hegyvidéki talaj

Az alsó-szászországi talajok hidrogeológiai típusait a 31. ábra mutatja be. Látható, hogy a fehérrrel jelölt járáshatáron belül küllönböző talajtípusok fordulnak el. Érdekes módon az *E. multilocularis*-fertőzöttség vonatkozásában magas kockázatú területeken a HG1 típus dominál. Ami első ránézésre ellentmond Miterpáková et al. (2006) eredményeinek, mivel ennek a talajtípusnak a legalacsonyabb a nedvességtartalma. Ha jobban megnézzük, akkor a magas kockázatú járásokban vannak területek, amelyek HG2 típusúak, de mivel a fertőzöttségre vonatkozóan nem rendelkezünk pont-szerű nyers adatokkal, így nem tudjuk azt vizsgálni, hogy vajon ezekről a nedvesebb talajú területekről származtak-e nagyobb számban a fertőzött egyedek.

Az eddigiekben szubjektív, vizuális alapon fogalmaztunk meg lehetséges összefüggéseket<sup>31</sup> a fertőzöttség kockázata és néhány környezeti tényező között. Az összefüggés-elemzésnek ez a lépése nagyon fontos, mint ahogy minden adatfeldolgozást, statisztikai elemzést érdemes a nyers adatok grafikus megjelenítésével, illetve azok értelmezésével kezdeni.

<sup>28</sup> <http://postgis.net/>

<sup>29</sup> <https://github.com/solymosi/RpostGIS>

<sup>30</sup> A fejezetben bemutatott összes térképet ugyanígy importáltuk az R-be.

<sup>31</sup> kvalitatív következtetések



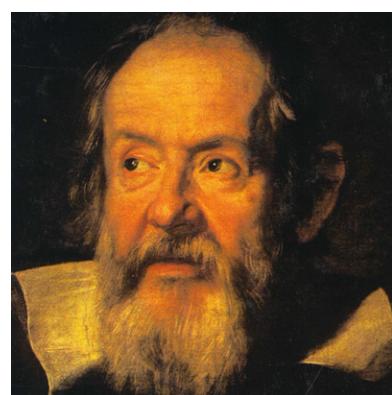
31. ábra. A ESDAC-adatbázisból származó hidrogeológiai típusok (HG1-HG4) mintázata Alsó-Szászországban.

Azonban ez csak a bevezetése az összefüggések matematikai statisztikai leírásának.<sup>32</sup> Ebben a megközelítésben a függőváltozónak (pl. prevalencia) a független változóktól (pl. hőmérséklet, csapadék, talajnedvesség) való függését valamelyen matematikai modell segítségével írjuk le. Ezek a modellek egyrészt lehetővé teszik, hogy számszerűsítsük a független változók hatásának mértékét a függő változóra, másrészt ezeknek a hatásoknak a számszerű értékei lehetővé teszik, hogy predikciókat adhassunk.

Mit jelent a geoepidemiológiában a predikció? Egyrészt jelenti azt, hogy a modellünk azokra a helyekre (pl. járások) vonatkozóan, amelyekből van információink a függő változóra (pl. prevalencia) milyen függőváltozó értékeket becsül, prediktál. Másrészt, mivel azt hisszük, hogy a fizikai világban vannak szabályok és ezek a szabályok matematikailag leírhatók, azt is hisszük, hogy ezek a szabályok nem csak egy adott helyen érvényesek, hanem más területekre is kiterjeszthetők. Vagyis, ha egy földrajzi területen erős matematikai kapcsolatot tudunk kimutatni a függő és független változók között, akkor annak a kapcsolatnak egy másik földrajzi területen is működnie kell.

Az erős matematikai kapcsolat többféleképpen határozható meg. Azokon a területeken, ahol fontos, hogy a matematikai modellekben származó számszerű eredmények mennyire megbízhatók (a gyakorlati alkalmazás szempontjából) a matematikai kapcsolatot a predikció

<sup>32</sup> kvantitatív következtetések

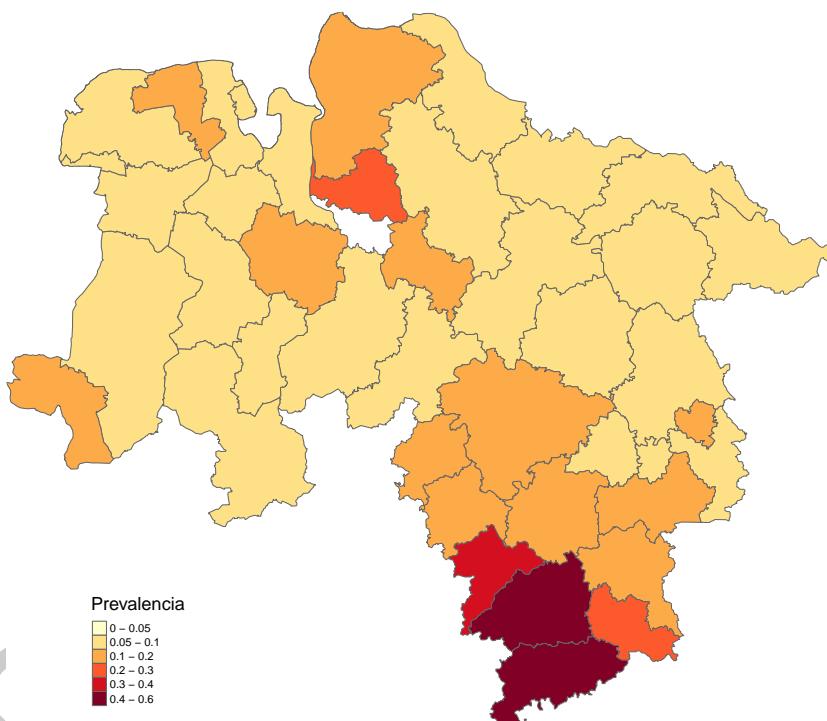


32. ábra. Galileo Galilei (1564-1642)

pontossága alapján minősítik. Az orvosi tudományok, így az állatorvosi epidemiológia is e körbe tartozik. A szakirodalomban számos módszer található a predikció megbízhatóságának számszerűsítésére vonatkozóan (Tofallis, 2015). Itt a MAE-t<sup>33</sup> használjuk, ami a megfigyelt és prediktált értékek abszolút különbségének átlaga:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

A célunk leírni egy olyan modellt,<sup>34</sup> amelyből származó predikció a legkisebb eltérést mutatja a megfigyelt értékektől. Ehhez három független változónkból (hőmérséklet, csapadék, talajnedvesség<sup>35</sup>) az összes lehetséges kombináció szerint létrehozunk modelleket és azokat illesztjük. minden illesztett modell alapján prediktáljuk az egyes járásokban várható prevalenciát. Ezen prediktált prevalenciák felhasználásával mindeneket modellre kiszámítjuk a MAE értékét. Azt a modellt tekintjük a legjobbnak, amelyik a legkisebb MAE-értéket<sup>36</sup> adó prediktált prevalenciát eredményezte. A példánkban használt modellek közül a legkisebb MAE-értékű (0.043) alapján prediktált prevalenciákat mutatja a 33. ábra.



A korábbi prevalencia-térképekkel összevetve a 33. ábrán látható mintázat helyenként magasabb, máshol alacsonyabb értékeket mutat. A megfigyelt és a legkisebb MAE-rtékű modell alapján becsült

<sup>33</sup> mean absolute error

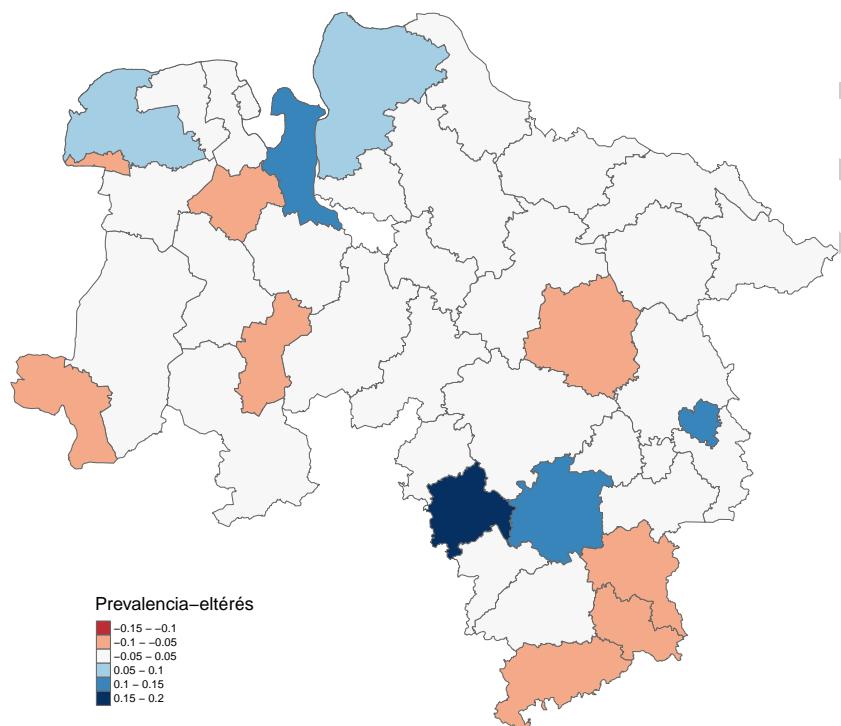
<sup>34</sup> logisztikus modell

<sup>35</sup> A modellekben a talajnedvesség változó értékeiként az egyes járásokra legjellemzőbb kategóriát használtuk.

<sup>36</sup> modell-szelekció

33. ábra. A legkisebb MAE-értékű modell alapján prediktált járásonkénti prevalencia.

járásonkénti prevalencia-különbségeket a 34. ábrán láthatjuk. Ennek alapján megállapítható, hogy a kiválasztott, a három független környezeti változónkat tartalmazó modellünk a járások többségében kevesebb, mint 0.05-nyi hibával képes prediktálni a prevalenciát. Azonban azt is látjuk, hogy vannak területek, ahol jelentősen alul (kék skála), illetve felülprediktálja (piros skála) a megfigyelt prevalenciát. Ezt értelmezhetjük úgy is, hogy a kiválasztott modell ugyan meglehetősen jól megragadja a prevalencia mértékére hatást gyakorló szabályokat, azonban vannak a felhasznált független változókon túl is hatótényezők. Azokban a járásokban, ahol a megfigyelttől nagyobb mértékben tér el a prediktált prevalencia, ezek a további hatótényezők a többi járástól eltérő mértékben hathatnak.



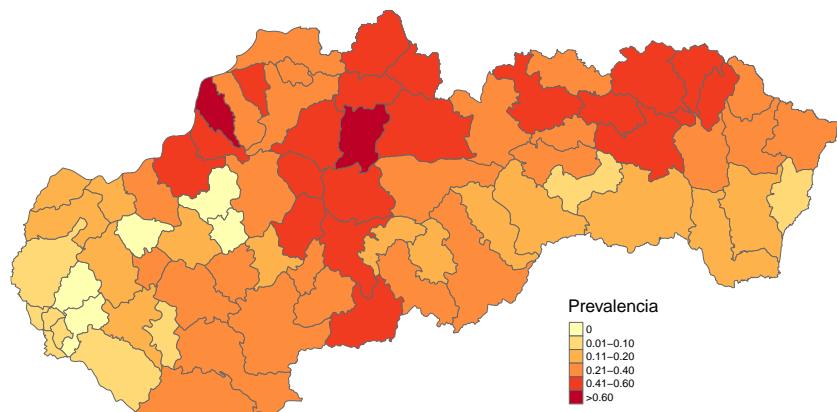
34. ábra. A megfigyelt és a prediktált prevalencia különbsége. A felülprediktált járások a piros skálával, az alulprediktáltak pedig a kék skálával vannak színezve.

A modellünkbeli a predikció mellett kiolvashatjuk a regressziós együtthatókat, amelyek a prevalencia és a környezeti tényezők kapcsolatát számszerűsítik. Ezek alapján a prevalencia negatív kapcsolatban van a hőmérséklettel, és pozitív kapcsolatban a csapadékkal. A talaj hidrogeológiai típusainak hatását a HG1 típushoz hasonlítjuk a modellben. Ehhez képest a HG2 csökkenti, a HG3 és a HG2 növeli a fertőzöttség valószínűségét.

Ennél a pontnál felmerülhet bennünk a kérdés, hogy ha ennyire jól használható a modellünk alsó-szászországi adatokon, akkor vajon milyen megbízhatósággal használható földrajzilag eltérő területen?

A kérdés megválaszolásához szükségünk van egy másik földrajzi területen végzett felmérés adataira. Mégpedig azért, hogy összehasonlíthatunk az alsó-szászországi adatokon illesztett modellünk alapján az új területre prediktált prevalenciákat az ott megfigyelt értékekkel.

Miterpáková et al. (2006) adatait használjuk fel. A szerzők Szlovákiában 2000–2004 között 3096 vörös rókát vizsgáltak meg *E. multilocularis*-fertőzöttségre vonatkozóan. Sajnos kistérségenkénti prevalencia-értékeket nem közölték, azonban a cikkben szereplő térképről leolvasható, hogy az egyes kistérségek milyen prevalencia-tartományba esnek (35. ábra).

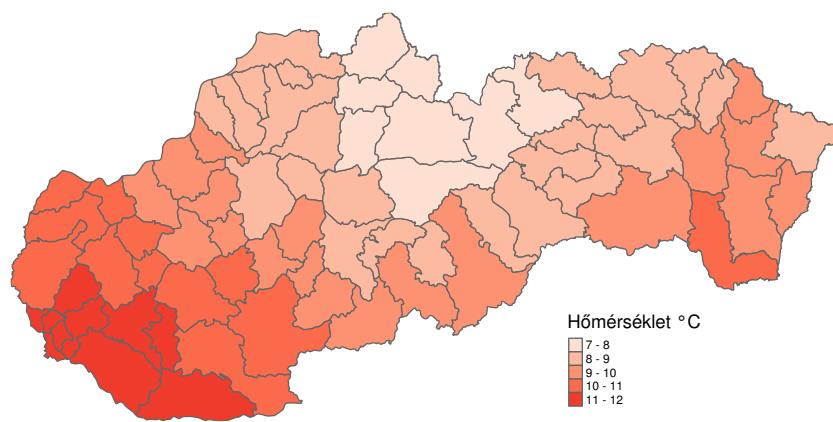


35. ábra. Szlovákiai megfigyelt *E. multilocularis*-prevalencia térkép (Miterpáková et al., 2006).

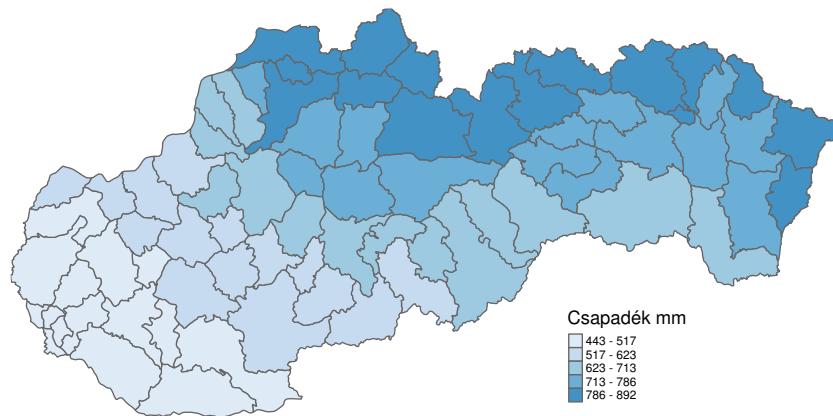
Az ERA-Interim adatbázisból letöltöttük a 2000–2004 közötti szlovákiai napi hőmérséklet- és csapadék-adatokat, illetve a ESDAC adatbázisból lekérdeztük a kistérségek legjellemzőbb hidrogeológiai típusát. Ezeket az adatokat mutatja be a 36–38. ábra. Az alsó-szászországi adatokra illesztett modell és a szlovákiai adatok alapján prediktált kistérségenkénti prevalencia-értékek a 39. ábrán láthatók.

Már ránézésre is látszik, hogy nagyon eltér a prediktált és a megfigyelt prevalencia-térkép. Számszerűsítve, a kistérségek 89%-ban másik prevalencia-tartományba esnek, mint amit Miterpáková et al. (2006) közölt. Ebben számos tényező játszhat szerepet.

Az egyik fontos mozzanat, hogy azok a meteorológiai változók, amelyeket használtunk a modellben Szlovákiában tágabb tartományban vesznek fel értékeket, mint Alsó-Szászországban. Márpedig a modellek becslései csak azokban a tartományokban érvényesek, amelyekben azokat illesztettük. Csak azokra a szlovákiai kistérségekre megvizsgálva a modell predikcióját, ahol a paraméterek beleesnek az alsó-szászországi tartományokba, valamivel jobb eredményt kapunk



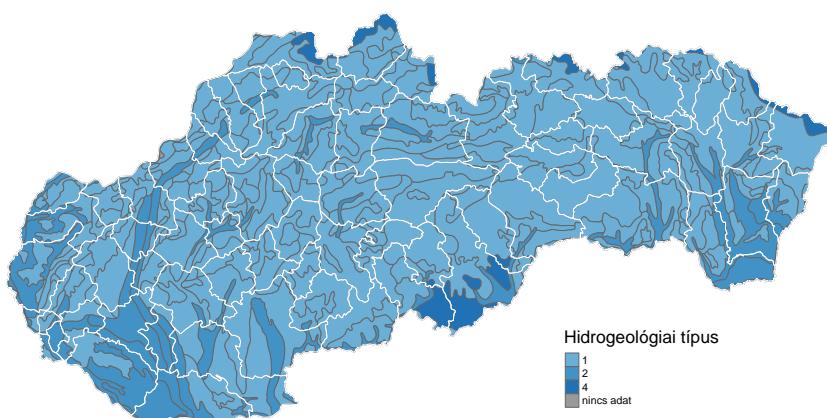
36. ábra. A 2000–2004. időszakban az évi átlaghőmérsékletek átlaga az ERA-Interim reanalízis adatbázis alapján.



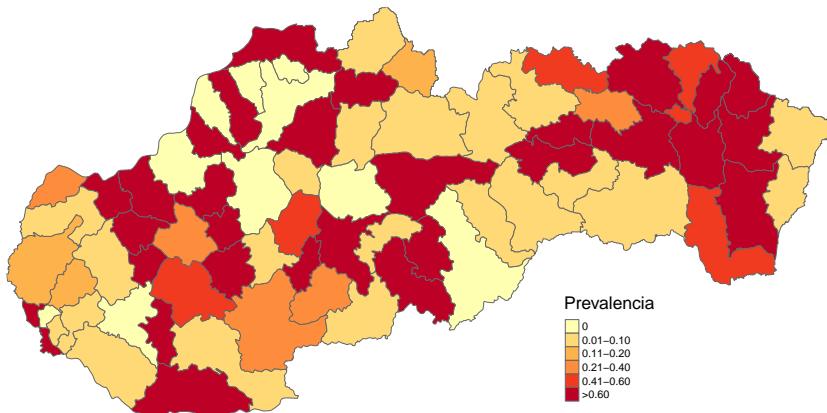
37. ábra. A 2000–2004. időszakban az éves összes csapadék átlaga az ERA-Interim reanalízis adatbázis alapján.

(83%-os besorolási hibát), azonban ez még mindig használhatatlan predikció.

A szlovákiai prediktált prevalenciák eltérése – mivel Miterpáková et al. (2006) kistérségi tartományokat közölt – csak durvább skálán értékelhető. Így ha csak néhány századnyira tért is el a megfigyelt prevalenciától, akkor is eshet egy másik prevalencia kategóriába, ezzel jelentősebben befolyásolva a predikció minőségét, mintha a pontos különbségeket használhatnánk annak megtételére. Ha megnézzük, hogy a Miterpáková et al. (2006) kategóriáit használva az alsó-szászországi predikciók milyen besorolási hibát jelentenek, akkor azt látjuk, hogy 29%-ban más kategóriába esnek a járások, mint a megfigyelt prevalencia alapján meghatározott kategória. Ha ezt a



38. ábra. A ESDAC-adatbázisból származó hidrogeológiai típusok (HG1-HG4) mintázata Szlovákiában.



39. ábra. Az alsó-szászországi adatokon illesztett modell és szlovákiai környezeti adatok alapján prediktált prevalenciaterkép.

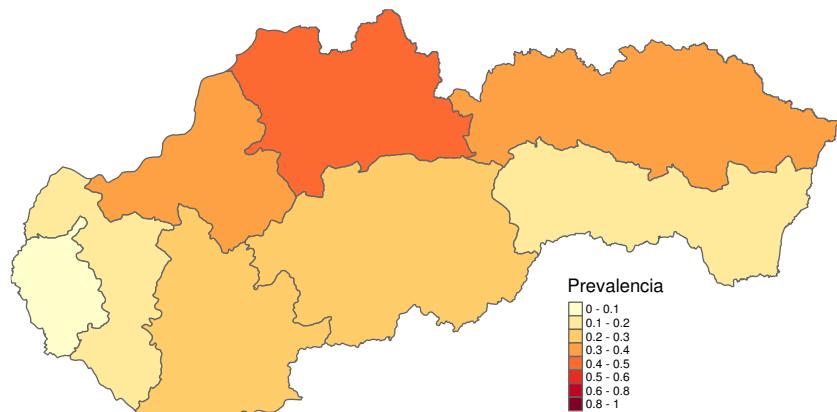
$MAE = 0.043$  értékkel vetjük össze, akkor látható, hogy a kategóriákba való besorolási hiba jelentősen meghaladja az átlagos eltérést.

Ugyanakkor az is lehetséges, hogy a forrásadatokban van valamilyen hiba. Ez származhat a környezeti adatok létrehozásának hibájából. Előfordul, hogy ha a különböző területekre különböző adatforrásokból származó adatokat használunk, akkor azok generálása során különböző módszereket alkalmaztak a mérések vagy az adatok normálása, homogenizálása során. Esetünkben ez kizártatott, mert a két vizsgálati területre vonatkozóan minden a talajtani, mint a meteorológiai adatok azonos módon lettek létrehozva. Az eltérés háttérében lehet továbbá az is, hogy a két vizsgálatban az *E. multilocularis*-fertőzöttségre vonatkozó adatgyűjtésben van jelentős eltérés,

pl. mintavételezési, diagnosztikai. A prediktált kockázati térképeken látható jelentős eltérések rávilágíthatnak a prevalencia-, incidencia-adatok hibájára is (King et al., 2004).

Természetesen az itt felsorolt torzító tényezőkön kívül számos biológiai (pl. róka-, kis rágcsáló populáció méretek), környezeti tényező létezik, amit nem kezeltünk a modellben, pedig hatással van a fertőzöttség prevalenciájára.

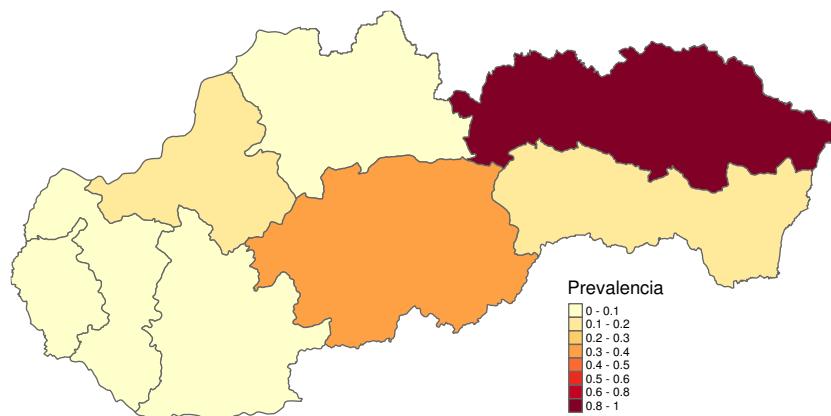
További torzítást jelenthet a predikció vonatkozásban, hogy az alsó-szászországi járások átlagosan kétszer nagyobb területűek, mint a szlovákiai kistérségek. A kisebb területen valószínűsíthető kisebb mintaelemszám alapján nagyobb hibával lehet becsülni a prevalenciát. Mivel Miterpáková et al. (2006) megyei szintű prevalencia-adatokat is közölt (40. ábra) megvizsgáltuk, hogy ezekre a nagyobb területű egységekre vonatkozóan nem tudunk-e jobb predikciót adni.



40. ábra. A 2000–2004. időszakra vonatkozó megfigyelt megyei prevalenciák.

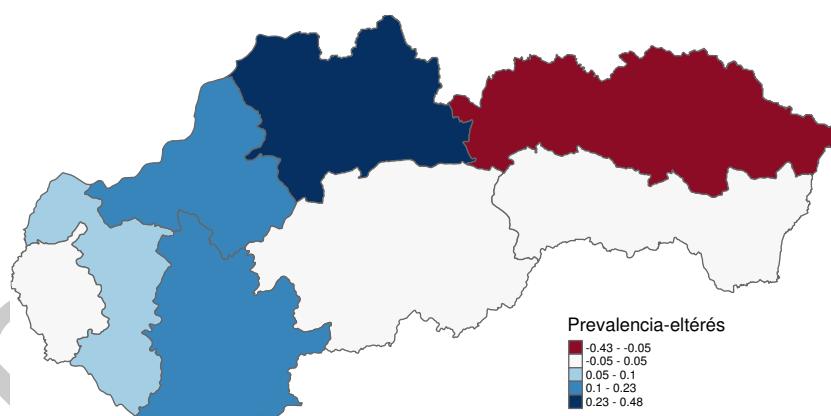
A szlovákiai megyékre kigyűjtött környezeti adatok alapján az alsó-szászországi modell-paraméterek felhasználásával a 41. ábrán látható prediktált prevalenciákat kapjuk. A megyei predikciókra kiszámítható a  $MAE = 0.19$ , ami jóval rosszabb mint az alsó-szászországi  $MAE$ , de jobbnak tűnik, mint a szlovákiai kistérségekre adott predikció.

A megyénkénti megfigyelt és prediktált prevalenciák különbségét a 42. ábra mutatja be, amelyen látható, hogy a megyék felében kevesebb mint 0.1-el hibázik a modell a prevalencia predikciójában. Viszont egy-egy megyében több mint 0.23-al felül- vagy alulbecsli a modell a prevalenciát. Zilina megyében a modell majdnem 0.5-el alulbecsli a megfigyelt prevalenciát. Miterpáková et al. (2006) adatai alapján ebben a megyében a prevalencia szórása majdnem a



41. ábra. A 2000–2004. időszakra vonatkozó prediktált megyei prevalenciák.

fele volt a prevalenciának, ami nagy bizonytalanságot jelez arra vonatkozóan, hogy valóban 0.475 lehetett-e a fertőzöttség kockázata. Presov megyében viszont több mint 0.4-el magasabb a prediktált prevalencia, mint a megfigyelt. Ebben a megyében a prevalencia egyharmada volt annak szórása, ami szintén arra utal, hogy a megfigyelt érték is pontatlan. A megyei és kistérségi skálán párhuzamosan végzett vizsgálat arra mutat rá, hogy az aggregáció szintjének nem csak a betegség-térképezésben van jelentősége, de hatása lehet az összefüggés-vizsgálatra és a predikcióra is.



42. ábra. A megfigyelt és a prediktált prevalencia különbsége. A felülprediktált járások a piros skálával, az alulprediktáltak pedig a kék skálával vannak színezve.

### Összefüggés-vizsgálat és földrajzi kapcsolatok

Az itt bemutatott példában olyan logisztikus modellt mutattunk be, amely nem használta a földrajzi szomszédsági, közelségi/távolsági összefüggéseket.

Vagyis úgy tűnhet – ahogy a betegség-térképezésnél is láttuk –, hogy csak vizualizációs szerepe van a térképeknek. Valójában azonban a geoepidemiológiában alkalmazott összefüggés-elemzések esetén a földrajzi viszonyok az elemzést megelőző adat-előkészítésben nagyon is fontos szerepet játszanak.

Ahogy láttuk ebben a példában is, a fertőzöttségi, meteorológiai és talajtani információk különböző térbeli felbontásban, struktúrákban (rács, poligonok) voltak elérhetőek. Az ilyen, eltérő felbontású, típusú adatok csak a földrajzi elhelyezkedés alapján társíthatók, amihez különböző geográfiai feldolgozó eljárásokat kell használnunk. Ezek az eljárások a különböző térbeli objektumok (pont, vonal, poligon) közös metszetei, átfedettsége, érintkezése alapján kapcsolnak össze az objektumokhoz tartozó adatokat (Solymosi et al., 2004; Allepuz et al., 2009). A kapcsolat kialakításához térinformatikai szoftvereket használhatunk, ezek közül néhány ingyenesen használható eszköz: QGIS, PostGIS, SpatiaLite,<sup>37</sup> GRASS GIS.<sup>38</sup>

<sup>37</sup> <http://www.gaia-gis.it/gaia-sins/>

<sup>38</sup> <https://grass.osgeo.org/>

# Irodalomjegyzék

- Allepuz, A., Saez, M., Solymosi, N., Napp, S., & Casal, J. (2009). The role of geographical factors on the success of Ajeszky's disease eradication programme in a high pig density area (northeast of Spain, 2003 – 2007). *Prev Vet Med*, 91(2-4), 153–160.
- Anselin, L. (1992). Local indicators of spatial association – LISA. *Geogr. Anal.*, 27(2), 93–115.
- Berke, O. (2001). Choropleth mapping of regional count data of *Echinococcus multilocularis* among red foxes in Lower Saxony, Germany. *Prev Vet Med*, 52(3), 119–131.
- Besag, J., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.
- Burlet, P., Deplazes, P., & Hegglin, D. (2011). Age, season and spatio-temporal factors affecting the prevalence of *Echinococcus multilocularis* and *Taenia taeniaeformis* in *Arvicola terrestris*. *Parasites & Vectors*, 4(6), 1–9.
- CDC (1990). *Guidelines for investigating clusters of health events*. Technical Report RR-11.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Källberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., & Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597.
- Gelfand, A. E., Diggle, Peter and Fuentes, M., & Guttorp, P. (2010). *Handbook of Spatial Statistics*. CRC Press.
- Green, P. J. & Silverman, B. W. (1993). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman and Hall/CRC.
- Jones, M. E. & Swerdlow, A. J. (1998). Bias in the Standardized Mortality Ratio when Using General Population Rates to Estimate Expected Number of Deaths. *American Journal of Epidemiology*, 148(10), 1012–1017.
- King, R. J., Campbell-Lendrum, D. H., & Davies, C. R. (2004). Predicting Geographic Variation in Cutaneous Leishmaniasis, Colombia. *Emerging Infectious Diseases*, 10(4), 598–607.
- Kulldorff, M. & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14, 799–810.

- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS Book - A Practical Introduction to Bayesian Analysis*. Texts in Statistical Science. Boca Raton, Florida, USA: Chapman and Hall/CRC. ISBN 978-1-58488-849-9.
- Miterpáková, M., Dubinsky, P., Reiterová, K., & Stanko, M. (2006). Climate and environmental factors influencing *Echinococcus multilocularis* occurrence in the Slovak republic. *Ann Agric Environ Med*, 13, 235–242.
- Monmonier, M. S. (1973). Maximum-difference barriers: An alternative numerical regionalization method. *Geogr. Anal.*, 3, 245–261.
- Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1-2), 17–23.
- Panagos, P. (2006). The European soil database. *GEO: connexion*, 5(7), 32–33.
- Panagos, P., Van Liedekerke, M., Jones, A., & Montanarella, L. (2012). European Soil Data Centre: Response to European policy support and public data requirements. *Land Use Policy*, 29(2), 329–338.
- Solymosi, N. (2009). *Spatial and spatio-temporal studies on pathogens*. PhD thesis, Juhász-Nagy Pál Doktori Iskola, Debreceni Egyetem.
- Solymosi, N., Harnos, A., & Reiczigel, J. (2003). A Tiszántúlon, állatok között előfordult veszettség esetek térbeli halmozódásának vizsgálata, 1990-2001. In *VIII. Geomatematici Ankét Szeged*, Hungary.
- Solymosi, N., Harnos, A., Reiczigel, J., & Speiser, F. P. (2006a). RpostGIS an R-library for using PostGIS spatial structures and functions. In *Book of Abstracts of The second international R user conference useR! 2006* (pp. 154). Vienna, Austria.
- Solymosi, N., Molnár, D. L., & Mészáros, J. (2005a). Választási eredmények térbeli-statisztikai elemzése. In J. Mészáros & I. Szakadát (Eds.), *Magyarország Politikai Atlasza 2004*. Gondolat, Budapest. ISBN: 963-9510-01-1.
- Solymosi, N., Reiczigel, J., Berke, O., Harnos, A., Szigeti, S., Fodor, L., Szigeti, G., & Bódis, K. (2004). Spatial risk assessment of herd sero-status of Aujeszky's disease in a county in Hungary. *Prev Vet Med*, 65(1-2), 9–16.
- Solymosi, N., Reiczigel, J., Edélyi, K., Báñhidy, J., Földi, Z., & Bódis, K. (2002). *Spatial analysis of rabies cases in foxes in Hungary between 1990 and 2001: a preliminary report*, volume 90 of *Studies in Health Technology and Informatics*, (pp. 770–773). Budapest, Hungary. ISBN: 978-1-58603-279-1.
- Solymosi, N., Reiczigel, J., Harnos, A., Abonyi-Tóth, Z., Speiser, F. P., Csabai, I., & Rubel, F. (2006b). A Multitask PostGIS based vet GIS framework. In *1st OIE International Conference. Use of GIS in Veterinary Activities Silvi Marina, Abruzzo*, Italy.
- Solymosi, N., Reiczigel, J., Harnos, A., Mészáros, J., Molnár, L. D., & Rubel, F. (2005b). Finding spatial barriers by Monmonier's algorithm. In *International Society for Clinical Biostatistics Conference Szeged*, Hungary.
- Solymosi, N., Wagner, S. E., Maróti-Agóts, Á., & Allepuz, A. (2010). maps2WinBUGS: a QGIS plugin to facilitate data processing for Bayesian spatial modeling. *Ecography*, 33(6), 1093–1096.
- Takeuchi-Storm, N., Woolsey, I. D., Jensen, P. M., Fredensborg, B. L., Pipper, C. B., & Outzen Kapel, C. M. (2015). Predictors of *Echinococcus multilocularis* prevalence in definitive and intermediate Hosts: A meta-analysis. *Journal of Parasitology*, 101(3), 297–303.

- Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66, 1352–1362.
- Ward, M. P. & Carpenter, T. E. (2000a). Analysis of time-space clustering in veterinary epidemiology. *Prev. Vet. Med.*, 43(4), 225–237.
- Ward, M. P. & Carpenter, T. E. (2000b). Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev. Vet. Med.*, 45, 257–284.