

ADAPTATION OF STATISTICAL MATCHING IN MICRO-REGIONAL ANALYSIS OF AGRICULTURAL PRODUCTION

CS. PESTI and J. KÁPOSZTA*

Department of Farm Business Analysis, Research Institute for Agricultural Economics, 1093 Budapest, Zsil u. 3-5.

**Institute of Regional Economics and Rural Development, Szent István University, 2103 Gödöllő, Páter K. u. 1.*

KEYWORDS: micro-regional, statistical matching, income disparities, FADN
RUNNING HEAD: STATISTICAL MATCHING IN MICRO-REGIONAL ANALYSIS

ABSTRACT

A mezőgazdaságnak és az agrárpolitikának számos kérdése a térbeliséggel függ össze. Az ország különböző tájain más-más termelési szerkezet alakult ki, eltérő a jövedelmezőség, a munkaerő felhasználás, térben differenciált a termelés intenzitása, az egyes térségekben más-más kockázati tényezőkkel, piaci lehetőségekkel kell számolni. A földhasználati viszonyok, a gazdálkodás műszaki színvonala szintén nem egységes, sok helyen erősen specializált, versenyképes nagyüzemi gazdálkodás jellemző, míg máshol gyengébb koncentráció mellett vegyes termelési szerkezetű kisebb gazdaságok működnek. Az agrárpolitikának egyik fontos feladata, hogy segítse az olyan termelési szerkezet kialakulását, ami alkalmazkodik az egyes térségek természeti adottságaihoz, ami képes a versenyképesség, a tájmegőrzés, a tájvédelem és a foglalkoztatás együttes követelményeinek megfelelni.

A megyeinél alacsonyabb területi szintű elemzésekhez azonban közvetlenül nem állnak rendelkezésre adatok, hiszen a települési és kistérségi szinten is rendelkezésre álló adatok nem tartalmaznak részletes információt a gazdaságok jövedelmi helyzetéről, a jövedeleminformációs rendszerek pedig nem reprezentatívak ezeken a területi szinteken.

A megoldást a meglévő adatbázisok összekapcsolása jelenti. A Mezőgazdasági és Vidékfejlesztési Hivatal regisztrációs adatbázisa az összes közvetlen támogatást igénybe vevő gazdaság vetésszerkezeti és támogatott állatlétszám adatait tartalmazza. A tesztüzemi rendszer kevés számú üzem pénzügyi és jövedelmi adatait tartalmazza. A statisztikai megfeleltetés lehetőséget nyújt a két adatbázis összekapcsolására, így – becült adatok segítségével – részletes területi elemzésre nyílik lehetőség.

INTRODUCTION

A mezőgazdaságnak és az agrárpolitikának számos kérdése a térbeliséggel függ össze. Az ország különböző tájain más-más termelési szerkezet alakult ki, eltérő a jövedelmezőség, a munkaerő felhasználás, térben differenciált a termelés intenzitása, az egyes térségekben más-más kockázati tényezőkkel, piaci lehetőségekkel kell számolni. A földhasználati viszonyok, a gazdálkodás műszaki színvonala szintén nem egységes, sok helyen erősen specializált, versenyképes nagyüzemi gazdálkodás jellemző, míg máshol gyengébb koncentráció mellett

vegyes termelési szerkezetű kisebb gazdaságok működnek. A mezőgazdaság súlya a helyi gazdaságban szintén heterogén, az ország egyes részein nagy szerepe van mind a foglalkoztatásban, mind az alacsonyabb jövedelmű rétegek önellátásában, míg máshol szerepe az iparhoz és a szolgáltatási szektorhoz képest elhanyagolható.

A mezőgazdasági termelés területi elhelyezkedését és fejlődését számos tényező befolyásolja. E tényezők két csoportba sorolhatók: egyik részük a természeti adottságokból következik, másik részük a társadalmi és közgazdasági adottságokból fakad. Az előbbihez tartozik a domborzat, a klimatikus viszonyok, a talaj, a művelési ágak arányai, az utóbbihoz a termelés történelmi hagyományai, a tulajdonviszonyok, a munkaerő- és eszközellátottság, a termékek iránt térben differenciált kereslet, a piaci viszonyok. Az agrárpolitikának egyik fontos feladata, hogy segítse az olyan termelési szerkezet kialakulását, ami alkalmazkodik az egyes térségek természeti adottságaihoz, ami képes a versenyképesség, a tájmegőrzés, a tájvédelem és a foglalkoztatás együttes követelményeinek megfelelni.

A mezőgazdaság regionális vizsgálataiból következtetni lehet arra, hogy mely területeken érdemes korszerű, intenzív, nagyobb ráfordításokkal működő gazdálkodási rendszereket kialakítani, s hol célszerűbb a termelés diverzifikálását, extenzív állattartást, energiatermelést előirányozni.

A megyeinél alacsonyabb területi szintű elemzésekhez azonban közvetlenül nem állnak rendelkezésre adatok, hiszen a települési és kistérségi szinten is rendelkezésre álló adatok (pl. a censusok és támogatásigénylések adatai) nem tartalmaznak részletes információt a gazdaságok jövedelmi helyzetéről, a jövedeleminformációs rendszerek pedig nem reprezentatívak ezeken a területi szinteken.

A megoldást a meglévő adatbázisok összekapcsolása jelenti. VROLIJK és munkatársai (2005) az ún. statisztikai megfeleltetési eljárás segítségével végzik el egy holland önkormányzat területén lévő tejelő tehenészetek jövedelmezőségének becslését. A tejelő tehenészetekről a censusban lévő adatok rendelkezésre állnak, a becsléshez szükséges jövedelmi adatokat a tesztüzemi rendszer biztosítja. Az eljárással az üzemek árbevételét, termelési költségét, üzemi nettó jövedelmét és munkajövedelmét becsülik meg. A két adatbázist az alábbi közös változókkal kapcsolják össze:

- a gazdálkodó életkora (év);
- gyepterület (ha);
- takarmánytermő terület (ha);
- tejelő tehenek száma (db);
- üzemméret (EUME).

A közös változók felhasználásának feltétele, hogy valamilyen logikai kapcsolatban legyenek a becslni kívánt változókkal. Az elemzés során az adott település minden tejelő tehenészetéhez hozzárendelik az adott önkormányzatot magában foglaló régióban található tesztüzemi adatszolgáltató gazdaságok közül a három leghasonlóbbat. Ezek közül véletlenszerűen kiválasztják az egyiket, majd az eredményeket összegzik. Az eljárást több alkalommal megismételve ellenőrzik, hogy milyen hibával működik a megfeleltetés. A becslt területi jövedelmi adatok megbízhatóságát növeli, hogy azokra az üzemekre is elvégzik a jövedelem becslését, amelyekről az adatok rendelkezésre állnak, és bemutatják a valódi és a becslt érték közötti eltérést.

Az írországi Teagasc Egyetem kutatói statikus területi mikroszimulációs modellt fejlesztettek ki. KELLY (2004) módszert dolgoz ki arra, hogy a meglévő üzemi adatokat hogyan lehet újraszűzölni a területi statisztikai adatokhoz. A módszere során az adott körzetben található adatszolgáltató üzemeket úgy súlyozza be, hogy a súlyszámokkal felsorozva a körzet teljes mezőgazdasági területét, állatlétszámát lefedjék.

A módszert HYNES et. al. (2006) fejleszti tovább, amikor olyan statikus mikroszimulációs modellt hoz létre, ami a teljes ír mezőgazdaságot lefedi. A modell legfontosabb bemenetét két adatbázis alkotja: az ír tesztüzemi rendszer és a mezőgazdasági összeírás.

Az ír tesztüzemi rendszer évről évre mintegy 1.200 gazdaság összes termelési, munkaerő- és jövedelmi adatát tartalmazza, azonban csak országos szinten reprezentatív, tehát területi elemzésekre közvetlenül nem használható.

A mezőgazdasági összeírásban 2000-ben az összes termelő üzem szerepelt, összesen mintegy 140.000. Az adminisztratív adatbázisban ugyan 190.000 agrártámogatást igénybe vevő „gazdaság” volt, azonban a tényleges üzemek száma a censusban összeírthoz áll közelebb. A mezőgazdasági összeírás az összes üzemet tartalmazza, az üzemek területegységhez (pl. településhez) köthetők, azonban az adatok a tesztüzemi rendszerhez képest jóval szűkebb körűek, mindössze a vetésterületre, az állatlétszámokra és a munkaerőre terjednek ki.

A modell lényege abban áll, hogy a fenti két adatbázist összekapcsolja, statisztikai megfeleltetési eljárás segítségével minden censusban található gazdaságnak megfeleltet egy tesztüzemi rendszerben lévő gazdaságot, amelyről minden jövedelmi adat rendelkezésre áll. A modell megbízhatósága azon múlik, hogy a tényleges üzem adatai a valóságban mennyire hasonlítanak a neki megfeleltetett üzem adataihoz. A megfeleltetéshez az üzemmérettel, az üzemszerkezettel és a talajtípussal kapcsolatos „közös változókat” (lásd Anyag és módszer c. fejezet) használtak, üzemtípusonként különféle kombinációkban. A legmegfelelőbb változók kiválasztását automatizálták, így a modell jelentős számítási kapacitást igényel, nagy számítógépen egy futtatás két napot vesz igénybe.

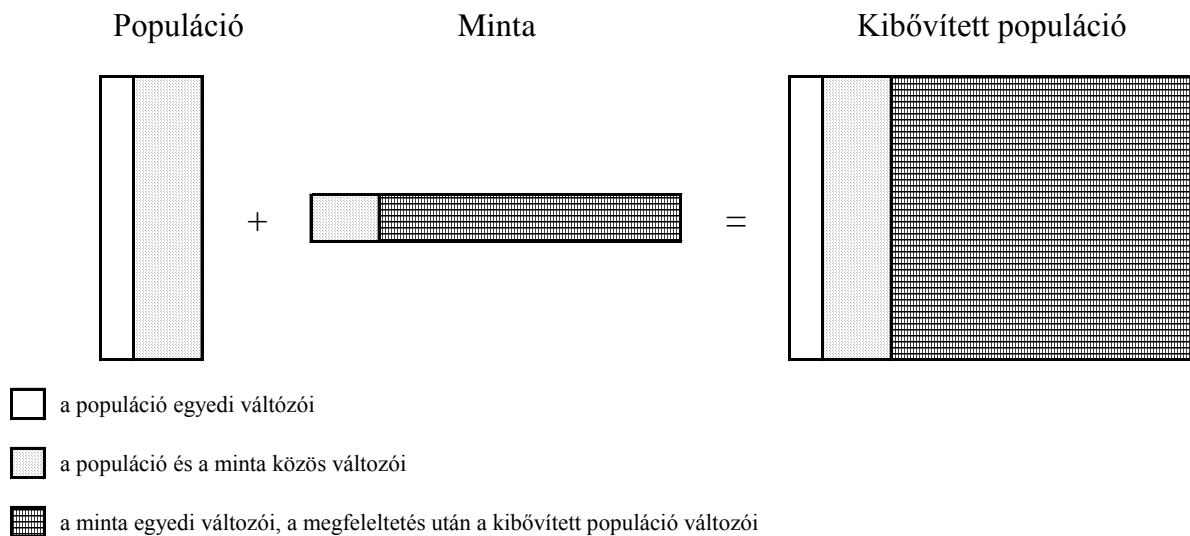
A fenti mikroszimulációs modell rendkívül árnyalt információt adhat a mezőgazdaság jövedelmi helyzetéről, így nemcsak módszertanilag érdekes, hanem eredményei a gyakorlatban is hasznosíthatóak.

MATERIALS AND METHODS

Az összes gazdaságról és azok ágazatairól költség-, jövedelem, illetve élők-munka-felhasználásra vonatkozó adatok nem állnak rendelkezésre. Tényadatok hiányában a rendelkezésre álló, kisebb elemszámú, de több változót tartalmazó adatbázisokra (minta) támaszkodva, statisztikai eljárással, becsléssel lehet előállítani az összes gazdaság (populáció) kibocsátással, jövedelemmel, és munkaerő-felhasználással kapcsolatos hiányzó adatait. Erre – DORGAI et. al. (2008) tanulmányához hasonlóan – az úgynevezett statisztikai megfeleltetési eljárást alkalmaztam, amelyet szematikusan a 2. ábra szemléltet. A megfeleltetés azt jelenti, hogy egy adott populáció és egy adott minta bizonyos ismérvek szerinti hasonlósága alapján, a populáció minden egyedéhez hozzárendeljük a minta hozzá leginkább hasonló elemét. Így azok a változók, amelyek a megfeleltetés előtt csak a mintában

szerepeltek, ettől kezdve a populáció minden elemére rendelkezésre állnak a becslés eredményeként.

A megfeleltetés sematikus ábrája



A statisztikai megfeleltetéshez használt közös változók kiválasztásánál VROJLIK (2004) alapján a következő matematikai szempontokat kell figyelembe venni:

- A hasonlóság alapját jelentő közös változókat úgy kell kiválasztani, hogy azok *szoros korrelációt* mutassanak a populációban ismerni kívánt, jelenleg csak a mintában meglévő megfelelő egyedi változókkal. Például, ha a megfeleltetéshez az üzemméretet (konkrétan: a földterület nagyságát) használjuk és a kimutatni kívánt változó az adózás előtti eredmény nagysága, akkor a kettő között szoros összefüggésnek, korrelációnak kell lennie.
- A *közös változók eloszlásának* hasonlóan kell lenni a populációban és a mintában, egyébként a megfeleltetés jelentős bizonytalansággal jár. Tehát teljesülnie kell annak a feltételnek, hogy a megfeleltetéshez használt változók esetében a populáció minden értékéhez legalább egy hasonló, illetve közeli érték megtalálható legyen a mintában. Ha a minta és/vagy a populáció valamely változója az eloszlásból kiugró értékeket tartalmaz, meg kell fontolni a kiugró értékeket tartalmazó esetek kizárását az elemzésből.
- A módszer jellegéből adódóan – egy eljárás során – *nem érdemes ötnél több változót* használni a megfeleltetéshez.
- Nem szabad „általánosan érvényes” megfeleltetésre törekedni, a közös változókat az elemzési célnak szerint kell kiválasztani.

A megfeleltetési eljárás a populáció adott eleméhez a mintának azt az elemét rendeli, melynél a közös változók együttes távolsága a legkisebb. A módszert korábban már többen sikerrel alkalmazták területi jellegű agrár-közgazdasági kutatásokban (DOL, 1991; VROJLIK, 2004; HYNES et. al. 2006). Az említett szerzők a módszertan területén a távolságszámítási eljárásban térnek el, jelen dolgozatban VROJLIK (2004) alapján az úgynevezett euklideszi távolságszámítást alkalmaztam.

A távolságfüggvény a mintában lévő üzemek és a populációban lévő üzemek távolságát határozza meg az alábbiak szerint:

$$D_{j,k} = \sum_{i=1}^m a_i [S_{j,i} - S_{k,i}]^{\beta_i}$$

ahol:

$D_{j,k}$ a minta j. eleme és a populáció k. eleme közötti távolság

a_i az i. változó súlyszáma

$S_{j,i}$ az i. változó normalizált értéke a minta j. elemén

$S_{k,i}$ az i. változó normalizált értéke a populáció k. elemén

j,k a minta és a populáció elemeinek azonosítói

β_i az i. változó kitevője

i a változó azonosítója

m a változók száma

A megfeleltetési eljárás populáció minden eleméhez azt a mintaelemet rendeli, amellyel a fenti függvény a minimum értéket veszi fel.

A megfelelő adatbázisok kiválasztásához először a célt, illetve annak eléréséhez legalkalmasabbnak tekinthető mutatókat kell megfogalmazni. A célt ebben az esetben az üzemi jövedelmezőség területi szintű mérése jelenti, az ehhez szükséges mutatók pedig a különféle eredménykategóriák mutatói (pl. adózás előtti eredmény, nettó hozzáadott érték stb.)

GSZÖ és az üzemi szintű tesztüzemi adatbázis összekapcsolása

Elméletileg ez a megfeleltetés lenne legalkalmasabb a mintapopulációban, vagyis a GSZÖ-ben szereplő gazdaságok fontosabb költség- és jövedelemmutatóinak becslésére, főként azért, mert a GSZÖ-ben megtalálható az üzemek teljes termelési szerkezete, valamint munkaerő-felhasználása. A munkaerő felhasználás azért fontos, mert szoros összefüggésben van a technológiával. Ha egy üzem adott termelési szerkezet mellett kevesebb munkaerőt használ fel, akkor valószínűleg fejlettebb technológiát használ, mint a több munkaerőt felhasználó, vele megegyező termelési szerkezetű üzemek. Tehát a megfeleltetés a GSZÖ minden üzeméhez – a tevékenységi irányokra bontva – a hozzá leginkább hasonló tesztüzemet rendeli. A megfeleltetés során a közös változókat az állatlétszám, a vetésterület és az üzemi munkaerő felhasználás alkotják, üzemtípusonként más-más kombinációban. A GSZÖ 2005 adatbázisban több közös változó áll rendelkezésre, mint a regisztrált gazdaságok adatbázisában, így az eredmények nagyobb biztonsággal lennének elfogadhatóak.

A hazai gyakorlat azonban a GSZÖ kialakításának módszertana miatt nem teszi lehetővé ezt a megfeleltetést, pontosabban az eredmények területi elemzésekben való hasznosítását. Az üzemszerkezeti felmérésre az úgynevezett mikro-körzetekben kerül sor, ahol minden gazdaságot felmérnek, és amelyek az ország területének csak egy részét fedik le. A mikro-körzetek eredményeiből a súlyszámok segítségével képzik a regionális (NUTS 2 szint) és az országos adatokat. A GSZÖ-ben szereplő szántóterület és az MVH regisztrációs adatbázisában szereplő szántóterület kistérségi szinten összegezve rendkívül nagy, esetenként több tízszeres eltéréseket mutatott. Tehát a GSZÖ adatbázis kistérségi szinten nem tekinthető reprezentatívnak, azonban a jövőben – az adatgyűjtés javítását, bővítését feltételezve – várhatóan lehetőség nyílik majd a következő AMÖ adatainak ilyen módon történő hasznosítására is.

Regisztrációs adatbázis és az ágazati tesztüzemi adatbázis összekapcsolása

Ezzel a megfeleltetéssel a regisztrált gazdaságok főbb ágazatainak eredményét (és az ezek összegéből álló „üzemi eredményt”) és a termelési értéket lehet számszerűsíteni és területileg elemezni. A regisztrációs adatbázis nem tartalmazza a regisztrált gazdaságok teljes termelési szerkezetét, hanem csak a támogatott ágazatokat. Ide tartoznak a növénytermelési ágazatok, a tejtermelés, a juhtartás és a hízómarha-tartás.

A megfeleltetési eljárás ebben az esetben konkrétan azt jelentette, hogy megkerestem a regisztrált gazdálkodók támogatásban részesülő ágazatainak megfelelő (a hozzájuk hasonló termelési szerkezetű, üzemméretű) gazdaságok ágazati jövedelmi adatait. Az elemzést a szántóföldi növénytermesztési ágazatokra végeztem el. A közös változók az alábbiak voltak:

- jogi státusz (egyéni/társas gazdaság)
- szántóterület nagysága (ha)
- szántóterület átlagos aranykorona (AK) értéke (a regisztrációs adatbázisban a települési átlagos AK érték)
- a fontosabb szántóföldi növények (búza, kukorica, árpa, rozs, zab, tritikálé, napraforgó, repce, silókukorica, lucerna) vetésterülete (ha)

A regisztrált (támogatott) gazdaságok száma mintegy 200 ezer, a tesztüzemek által reprezentált 2 EUME feletti gazdaságoké mintegy 90 ezer. Köztudott, hogy a regisztrált gazdaságok egy része önálló üzemként nem végez mezőgazdasági tevékenységet, csak „papíron” földhasználó, és igénybe veszi a területalapú támogatást. Ezenkívül nyilvánvaló, hogy a regisztrált gazdaságoknak legalább a fele nem éri el a 2 EUME üzemméretet, tehát nem is szerepelhetne a tesztüzemi rendszerben. Emiatt a megfeleltetés csak akkor fogadható el, ha feltételezzük, hogy nincs döntő különbség a 2 EUME alatti és a 2-4 EUME közötti üzemek ágazati eredményeiben.

Bár a cikk elkészítésének időpontjában már a 2007. évi adatok is rendelkezésre álltak, a megfeleltetéshez mind a tesztüzemi adatbázisból, mind a regisztrációs adatbázisból a 2006-os adatokat használtuk, mivel a 2007. év „rendhagyónak” tekinthető, mivel az aszály nem egyformán sújtotta az egymáshoz közel fekvő területeket sem.

A statisztikai megfeleltetést a holland LEI által kifejlesztett Statistics for Regional Studies (STARS) nevű szoftverrel, valamint MySQL adatbázis-kezelő rendszerrel végeztük.

RESULTS

Statisztikai megfeleltetéssel megbecsültük az összes területalapú támogatást igénybe vevő gazdaság szántóföldi növénytermesztő ágazatának fedezeti hozzájárulását. Az 1. ábra ennek üzemenként, majd kistérségenként összegzett értékét mutatja, 1 ha szántóterületre vetítve.

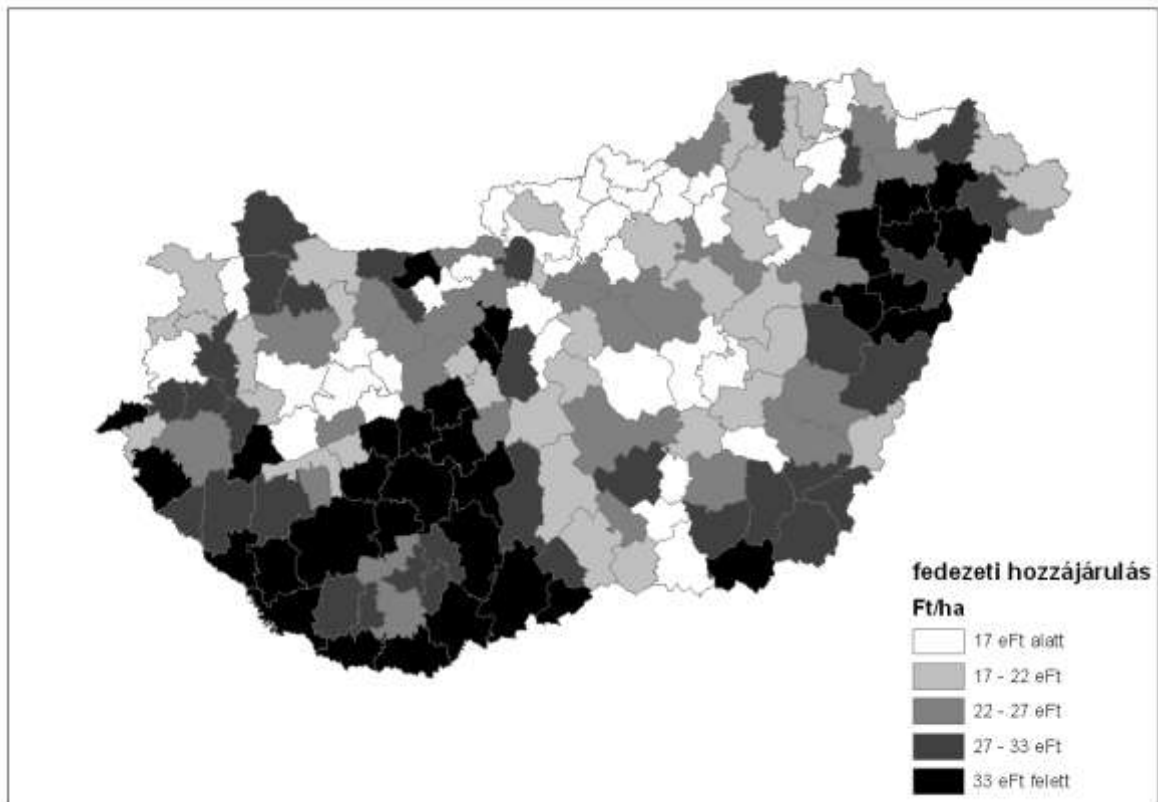


Figure 1: Gross margins of field crop production in the Hungarian micro-regions, 2006

Megfigyelhető, hogy a szántóföldi növénytermesztés a legnagyobb jövedelmezőséget a dél-dunántúli területeken, Békés megyében és a Debreceni-lőszhát területén érte el. Ezeknek a területeknek a legnagyobb részén a klimatikus és a talajviszonyok kedveznek a növénytermesztésnek, valamint az üzemszerkezet és a termelési hagyományok is a jövedelmező gazdálkodást vetítik előre. A középhegységeken, a Homokhátságon, valamint a belvizes és szikes talajokkal rendelkező közép-tiszai területeken lehet a legkisebb jövedelmezőséggel folytatni a növénytermesztést.

Érdemes összevetni az 1. ábrát a 2. ábrán található földbérleti díjakkal valamint a 3. ábrán található földminőséggel.

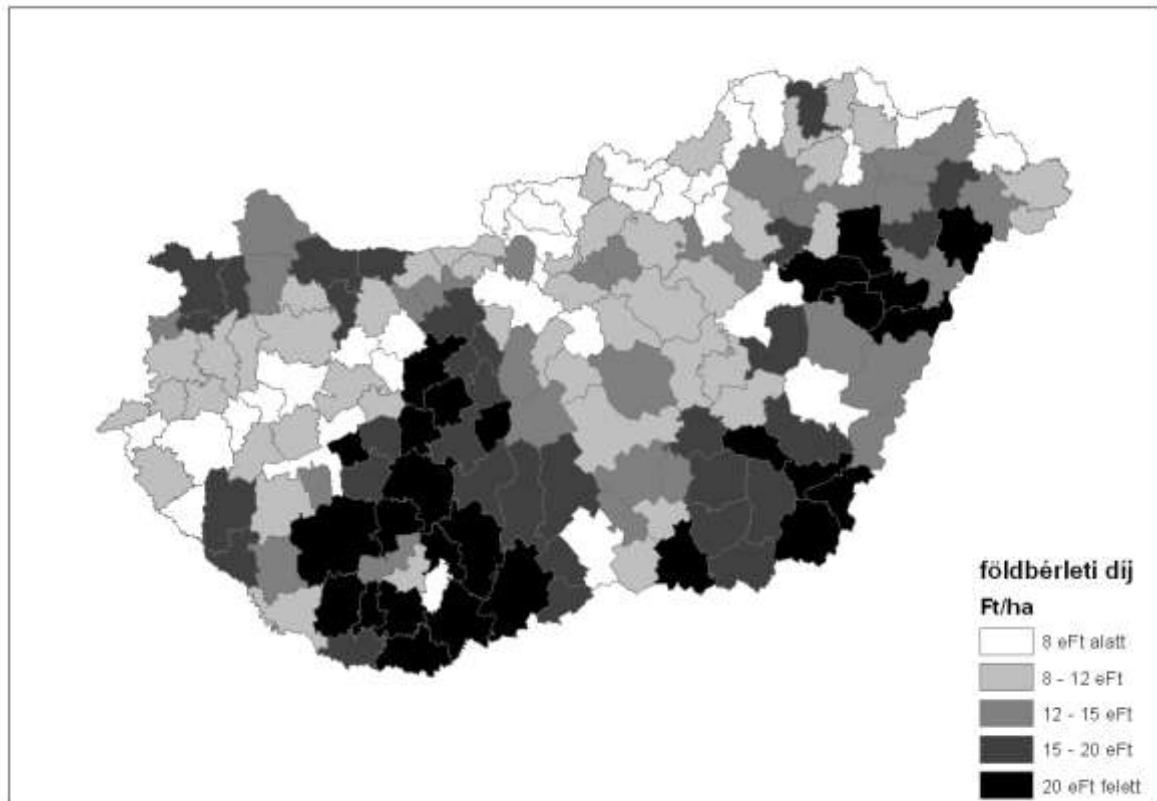


Figure 2: Land rent prices in the Hungarian micro-regions, 2004-2006 (on 2006 prices)

Figure 3: Soil quality in the Hungarian micro-regions

Az ábrák nagyfokú hasonlósága arra utal, hogy a szántóföldi növénytermesztés jövedelmezősége szoros összefüggésben van a talajviszonyokkal és ezt a földtulajdonosok a bérleti díjakon keresztül be is árazták.

DISCUSSION

A statisztikai megfeleltetés sikeresen használható a mezőgazdasági termelés területi elemzésére is. Területi mikroszimulációs modellek segítségével lehetőség nyílik a cenzusban vagy a regisztrációs adatbázisban található üzemek pénzügyi, jövedelmi helyzetének elemzésére és előrejelzésére.

A növénytermesztés fedezeti összegének területi vizsgálata – természetesen a kockázatelemzéssel és a termelés szerkezetének elemzésével kiegészítve – rámutat arra, hogy az Alföld nagy területein hozam és jövedelem biztonságot elősegítő beruházásokra lenne szükség, például öntözés-fejlesztésre és meliorációra. Ezzel szemben a hegyvidéki területeken

előnyben kellene részesíteni az extenzív gazdálkodási formákat, például a legelőre alapozott húsmarhatartást és a juhtenyésztést.

A cikk szerzőinek véleménye szerint a statisztikai megfeleltetést a területi elemzések során számos gyakorlati probléma megoldásához is sikeresen lehet használni. Ilyen alkalmazási terület például egy tervezett biomassza erőmű közelében lévő üzemek hozamainak és jövedelmi helyzetének becslése vagy egy gyümölcsfeldolgozó térségében található potenciális beszállítók feltérképezése. Ebben az esetben a helyspecifikus szempontokat figyelembe véve lehet paraméterezni a megfeleltetési eljárást, kiválasztani a közös változókat, a távolságszámítás módszerét, az ellenőrzési eljárást.

REFERENCES

1. VROLIJK H., DOL W., KUHLMAN T. (2005): Integration of small area estimation and mapping techniques. Tool for Regional Studies. LEI, The Hague 60 pp.
2. KELLY D. (2004): SMILE Static Simulator Software User Manual. Teagasc: Teagasc Athenry Publication.
3. HYNES S., MORRISSEY K., O'DONOGHUE C. (2006): Building a Static Farm Level Microsimulation Model: Statistically Matching the Irish National Farm Survey to the Irish Census of Agriculture. 46th Congress of the European Regional Science Association, <http://www.ersa.org/ersaconfs/ersa06/papers/431.pdf>
4. DORGAI L. (Szerk.), LUDVIG K., MÁRKUSZ P., MOLNÁR A., PESTI CS., SZÉKELY E., TÓTH E., UDOVECZ G. (2008): A közvetlen támogatások feltételezett csökkentésének társadalmi-, gazdasági- és környezeti hatásai (első megközelítés). Agrárgazdasági Kutató Intézet: Agrárgazdasági Tanulmányok (megjelenés alatt) 137 pp.
5. VROLIJK H. (2004): STARS: statistics for regional studies. Proceedings of Pacioli 11: New roads for farm accounting and FADN. Report 8.04.01. LEI, The Hague