

HIVATALOS BÍRÁLAT

Ladányi Márta

Statisztikai modellezési módszerek az agrártudományi kutatásokban

című MTA doktori értekezéséről

BEVEZETÉS

Nem csekély vonakodás és belső vívódás után vállaltam el Jelölt alkalmazott statisztikai témakörben írt MTA doktori értekezésének bírálatát. Kicsit olyan érzésem volt, mint amikor egy műkedvelő olvasót felkérnek egy bonyolult vers kritikai elemzésére... A Keszthelyi Agrártudományi Egyetem Agrárkémikus Szakán (ahol én végeztem egyetemi tanulmányaimat) a statisztikai képzés abban az időben egyáltalán nem tartozott az egyetemi oktatás kiemelt súlypontjai közé. Talán éppen ahhoz volt elegendő, hogy a végzés után néhány évvel, már kezdő kutatóként, segítse az eligazodást és alkalmazni tudjam Sváb János *kék könyvének* (*Biometria módszerek a kutatásban*) módszertanát. Az adott időszakban a ma ismert gépi tanulási módszerek legtöbbje még nem volt ismert. A statisztikai modellezési módszertan területén teljesen autodidakta módon szereztem ismereteket; a dolgozatban bemutatott eljárások közül néhányat ugyan kutatásaim során alkalmaztam, azonban egyesekkel e dolgozat révén találkoztam először.

Ráadásul a dolgozat témakörében és felépítésében sem tekinthető hagyományosnak, így bírálata sem egyszerű. Az MTA doktori értekezések általában a jelölt önálló kutatási eredményeit mutatják be (egy vagy több, egymáshoz valamilyen szempont alapján kapcsolódó témakörben), mely kutatásokat ugyan legtöbbször egy csapat tagjaként folytatta, ám ebben a csapatban adott esetekben ő volt az irányító személy. Jelen dolgozat eltér a szokásostól. Jelölt nagyszámú munkatársat (köszönetnyilvánításában 360-at meghaladó társszerzőt említ) magába foglaló több kutatói csapat meghatározó „tagjaként” végzett nagyon szerteágazó, nagy szakértelmet kívánó adatelemzési, modellezési, statisztikai elemzési feladatokat, azonban a kísérletek megtervezése, beállítása, a vizsgálat adatok összegyűjtése és a szakmai következtetések levonása nem az ő feladata volt. Mint Jelölt is megfogalmazza munkája bevezetőjében: a dolgozat a statisztikai modellező munka azon részére helyezi a hangsúlyt, amely bemutatja, miként válhatnak a nyers adatok az adatgazda számára hasznos eredménnyé, a vizsgált élő rendszer üzenetévé. Arra az útra koncentrál, amelynek problémái, kihívásai és érdekes mozzanatai a publikációkban rendszerint háttérben maradnak; bepillantást enged néhány kutatói feladat adatfeldolgozási kulisszatitkaiba.

A bírálói felkérést végül három okból vállaltam el: (1) a dolgozat a talajtani kutatásokhoz is kapcsolódó fejezeteket tartalmaz; (2) olyan módszereket mutat be, amelyeket eddig nem ismertem megfelelő mélységben, így a bírálat lehetőséget kínál számomra ezek megismerésére; (3) a kezdeti elbizonytalanodás után ösztönzően hatott rám a feladat szokatlan jellege és a dolgozat formabontó megközelítése. Mindazonáltal bírálatomban elsősorban a talajtani vonatkozású fejezetekhez kapcsolódóan fogalmazom meg kérdéseimet és megjegyzéseimet.

AZ ÉRTEKEZÉS FORMAI KIVITELEZÉSE, FELÉPÍTÉSE, STÍLUSA, ADATAI

A Jelölt a *Bevezetés* fejezetben önálló alfejezetet szentel az agrárstatisztikus képzés bemutatásának, kitérve annak hazai hiányára, valamint a folyamatban lévő, általa irányított akkreditációra is. A dolgozat formai felépítése részben ezt tükrözi: csekély módosítással szinte az agrárstatisztikus mesterképzés tankönyveként is szolgálhatna.

A benyújtott értekezés 11 decimális számozással tagolt fő fejezetből áll, 152 oldal terjedelmű. Az első fejezet előtt találjuk az értekezésben használt rövidítések, a táblázatok, az ábrák és a saját (társszerzős) publikációkból vett példák jegyzékét. A dolgozat végén háromféle irodalomjegyzékkel is találkozunk (1) a példákban idézett saját publikációk (28 db) felsorolásával, (2) az eredményekhez felhasznált saját publikációk (8 db) felsorolásával, majd (3) a felhasznált irodalmak (226 db) jegyzékével. A dolgozat mellékleteket nem tartalmaz. A saját kutatási eredményeit ismertető publikációkban 8 esetben utolsó szerzős, 28 esetben egyéb. A bemutatott publikációk minőségi megoszlása a következő: D1- 5 db, Q1- 19 db, Q2- 8 db, Q3- 4 db.

A dolgozat alábbiakban megadott fő fejezetei a tartalommal arányos, jól áttekinthető tagolást mutatnak (szükség szerinti további decimális tagolásokkal).

1. Bevezetés	(4 oldal)
2. A modellező munka	(15 oldal)
3. Lineáris modellek	(7 oldal)
4. A gépi tanulás	(9 oldal)
5. Gépi tanulási módszerek	(29 oldal)
6. Eredmények	(51 oldal)
7. Kiemelt eredmények	(2 oldal)
8. A példákban idézett saját publikációk	(3 oldal)
9. Az eredményekhez felhasznált saját publikációk ..	(1 oldal)
10. Felhasznált irodalom	(15 oldal)
11. Köszönetnyilvánítás	(1 oldal)

A dolgozat stílusa és megfogalmazása általában világos, értelemzavaró elírások és gépelési hibák nagyon ritkán fordulnak elő benne. Az ábrák és táblázatok jól szerkesztettek, feliratozásuk megfelelő, azonban az ábrák több esetben angol nyelvűek, ami a magyar nyelvű doktori dolgozatban zavaró lehet. A dolgozat megértését segítette az előforduló betűszavak és rövidítések jegyzékének beillesztése.

A dolgozathoz csatolt téziszfüzet 22 oldal terjedelmű, a „Bevezetés”, „A modellező munka”, „Lineáris modellek”, „A gépi tanulás” és „Eredmények” c. fejezetek után a „Kiemelt eredmények” c. fejezetben 8 pontban röviden (ábrák nélkül) ismerteti az értekezés téziseit. A téziszfüzet végén helyezte el a Jelölt a dolgozat témakörében megjelent fontosabb publikációinak listáját. A téziszfüzet tartalma lényegre törő, formája megfelelő.

Összességében megállapítható, hogy a dolgozat és a téziszfüzet - kissé rendhagyó szerkezetű ugyan, de - formailag megfeleltethető az MTA doktori értekezéstől elvárt követelményeknek.

BEVEZETÉS (1. fejezet)

A Jelölt a fejezetben először meghatározza az alkalmazott statisztika és az alkalmazott statisztikus helyét az agrártudományi kutatásokban. Ezt követően értelmezi az alapkutatást végző statisztikus, a biostatisztikus és az alkalmazott agrárstatisztikus feladatainak különbségeit, majd megfogalmazza (sokat segítve ezzel a dolgozat bírálójának is), hogy mit tekint az alkalmazott statisztikus saját kutatási eredményeinek. Megállapítása szerint ezek olyan specifikus hozzájárulások, amelyek az adatok előkészítéséhez, elemzéséhez, a megfelelő statisztikai módszerek kiválasztásához és alkalmazásához, a modellek fejlesztéséhez és értékeléséhez, valamint az eredmények értelmezéséhez és tudományos igényű közléséhez kapcsolódnak. (Mindezzel a bíráló is egyet tud érteni.) A fejezet ezt követően részletesen definiálja a Jelölt saját szerepét a kutatásokban (problémamegértés, konzultáció, adat-előkészítés, módszerválasztás, értelmezés és bemutatás), majd röviden ismerteti a dolgozat formabontó szerkezetét.

A MODELLEZŐ MUNKA (2. fejezet)

A fejezetben a jelölt összefoglalja a statisztikai modellező munka legfontosabb lépéseit (adatgyűjtés és -előkészítés; modellépítés, paraméterbecslés és validálás; modellértékelés; jelölt modellek összehasonlítása; továbbfejlesztés; a legjobb modell(ek) kiválasztása; eredmények közlése és vizualizáció). Minden lépést szemléletes, saját kutatásaiból vett példákkal illusztrál, ami később tankönyvi feldolgozás alapját is képezheti.

Az adat-előkészítéshez kilenc példát mutat be, többek között a különböző adattípusok kezelését, a folytonos adatok diszkrétizálását, a kiugró értékek különféle módszerekkel történő szűrését, az adatok normalizálását ($[0,1]$ intervallumra vagy normáeloszlásra), az adatok összevonását, valamint laboratóriumi méréseken alapuló szimulációval történő előállítását.

A modellépítés alfejezetben a Jelölt definiálja a kalibráció és a validáció fogalmát, valamint ismerteti a különféle validációs módszereket (keresztvalidáció, újramintavételezés). A modellek összehasonlításánál kitér az általánosan alkalmazott eljárásokra, köztük az Akaike- és a Bayes-féle információs kritériumra. A vizualizációt tárgyaló fejezetben három saját kutatásból származó példát mutat be: egy speciális, félig grafikus táblázatot, egy hőtérképes táblázatot és további grafikus megoldásokat.

LINEÁRIS MODELLEK (3. FEJEZET)

A fejezet áttekinti az általános és általánosított lineáris modelleket, valamint a lineáris kevert és ismételt méréses modelleket, amelyeket saját kutatási példákkal szemléltet. A végén összefoglalót ad az R környezetben elérhető kapcsolódó statisztikai csomagokról, továbbá kiegészítésként bemutatja a kis mintaszámok esetén alkalmazható válaszfelület-módszert.

A GÉPI TANULÁS (4. FEJEZET)

A fejezet áttekinti a gépi tanulás fogalmát, főbb módszertípusait és alkalmazási területeit, bemutatja a modellezés lépéseit és a legfontosabb eljárásokat. Külön foglalkozik a modellek jóságát mérő mutatókkal, majd ismerteti az R környezetben elérhető kapcsolódó csomagokat és függvényeket. Végül a fejezet a hagyományos statisztikai módszerek és a gépi tanulás közötti összefüggéseket tárgyalja, rámutatva a hasonlóságokra és a különbségekre.

GÉPI TANULÁSI MÓDSZEREK (5. FEJEZET)

A fejezet a kutatásokban leggyakrabban alkalmazott gépi tanulási algoritmusokat mutatja be, csoportosítva a felügyelt, felügyelet nélküli és metatanuló módszerek szerint:

- **Felügyelt módszerek:** K-legközelebbi szomszéd, naiv Bayes, döntési és regressziós fák, diszkriminanciaelemzés, lineáris és nemlineáris (valamint többszörös) regresszió, logisztikus regresszió, támogatóvektor-gépek.
- **Felügyelet nélküli módszerek:** klaszterelemzés (hierarchikus, illetve k-közép), főkomponens-elemzés, főkomponens-regresszió, parciális legkisebb négyzetek módszere.
- **Metatanuló módszerek:** neurális hálózatok, véletlen erdők.

Minden algoritmusnál példákkal és az R környezetben elérhető releváns csomagok, illetve függvények bemutatásával szemlélteti a gyakorlati alkalmazást.

EREDMÉNYEK (6. FEJEZET)

A dolgozat 2-5. fejezetei tulajdonképpen tekinthetők úgy is, mint egy hagyományos dolgozat „Anyag és módszer” fejezetei, már ami a Jelölt kutatásaiban használt, rendkívül szerteágazó statisztikai eszköztár bemutatását illeti. Ugyanakkor ebben a 6. fejezetben is szükségszerűen megtörténik az éppen tárgyalt kutatás „Anyag és módszer” részének bemutatása (a statisztikákat illetően bőszes visszautalással a megelőző fejezetekre).

A dolgozatban a Jelölt nyolc kiválasztott publikációból származó elemzést mutat be. A bemutatott közleményekben a Jelölt elsősorban az elemzési és modellezési munkát hangsúlyozza, míg a társszerzők szakterületéhez tartozó módszereket és eredményeket háttérben hagyja. Az ismertetés főként az adatok feldolgozására és az alkalmazott módszertanra koncentrál, a szakmai részletek az eredeti publikációkban érhetők el (sajnos nem minden publikáció volt teljes terjedelemben elérhető a bíráló számára).

I. Az első publikáció az aszály hatását vizsgálja a talajban élő mikroízeltlábúakra (Flórián N., Ladányi M., Ittész A., Kröel-Dulay G., Ónodi G., Mucsi M., Szili-Kovács T., Gergőcs V., Dányi L., Dombos, M. 2019. 'Effects of single and repeated drought on soil microarthropods in a semi-arid ecosystem depend more on timing and duration than drought severity'. *PLoS ONE* 14(7): e0219975. <https://doi.org/10.1371/journal.pone.0219975 D1>)

A tanulmányban a szerzők azt vizsgálták, hogy az ugróvillások (nem pedig a dolgozatban tévesen szereplő 'rugófarkúak') és az atkák miként reagálnak különböző mértékű és ismétlődő, szimulált aszályeseményekre egy magyarországi félszáraz homoki sztyeppén végzett terepkísérletben. A hipotézis szerint az aszályt szimuláló kezelések csökkentik, míg az extra csapadék növeli a talajban élő mikroízeltlábúak aktivitássűrűségét.

Az aktivitássűrűség-adatok (AD-adatok) a fajok szezonális ingadozása és csoportos előfordulása miatt a gondos tisztítás után is nagy heterogenitást mutattak, ezért közvetlen összehasonlításuk torz eredményekhez vezethetett volna. Ennek kiküszöbölésére Jelölt a normálást (RAD) és ordinális skálázást és aktivitássűrűség-különbség (ADD) mutatók képzését javasolta. A RAD adatokat Student-féle t-próbával (2014) és MANOVA modellel (2015) hasonlította össze. Definiálta az aktivitássűrűség-különbségek irányát és nagyságrendjét kifejező aktivitássűrűség-különbség (ADD) mutatókat (0, +1, -1), melyek egy ordinális skálázáson alapultak, majd a kezelések hatását statisztikai tesztekkel hasonlította össze. Az elemzések alapján igazolható volt, hogy a közép-magyarországi száraz homoki sztyeppék talajmezofaunája rövid távon képes regenerálódni az aszályos hatások után, de tartós, extrém szárazság esetén ellenálló képessége korlátozottnak bizonyul.

Kérdésem ezzel kapcsolatban: Milyen szempontok alapján határozták meg a felszíni ugróvillás- és atkafajok aktivitássűrűségi kategóriáit, illetve a relevánsnak tekintett aktivitássűrűség-különbségek határértékeit?

II. A második publikáció a kímélő talajművelésnek a lefolyásra és a talajveszteségre gyakorolt hatását elemzi (Madarász B., Jakab G., Szalai Z., Juhos K., Kotroczó Zs., Tóth A., Ladányi M. 2021 Long-term effects of conservation tillage on soil erosion in Central Europe: A random forest-based approach, *Soil and Tillage Research*, 209(104959), ISSN 0167-1987, <https://doi.org/10.1016/j.still.2021.104959>. D1)

A kutatás célja a kímélő talajművelés (CT) lefolyásra (RO) és talajveszteségre (SL) gyakorolt hosszú távú hatásainak feltárása volt, mivel Magyarország termőföldjeinek több mint egyharmadát az intenzív mezőgazdaság okozta erózió fenyegeti. Egy 16 éves megfigyelési adatsor adatait vonta be Jelölt a statisztikai vizsgálatokba, amely során az adathiányok, az összefüggések komplexitása és a kapcsolatok nemlineáris tulajdonságai miatt speciális megoldásokat (Random forest klasszifikációs modell, hiányzó csapadékértékek k-legközelebbi szomszéd módszer segítségével történő pótlása, a kategóriabecslés pontosságának ellenőrzése konfúziós mátrix mutatókkal) kellett alkalmaznia. A statisztikai elemzésekkel kimutatható volt, hogy a RF modell hatékonyan becsüli a lefolyás és a talajveszteség kockázatát, a kímélő talajművelés (mint legfontosabb hatótényező) pedig jelentősen mérsékli az eróziót és javítja a vízmegtartást.

Kérdésem a második publikációval kapcsolatban: Mit jelent pontosan az a mondat, hogy „a lefolyás (RO) és a talajveszteség (SL) változókat négy-négy kategóriára osztottuk az eloszlásuk és a szakemberek által megítélt súlyosságuk figyelembevételével”? Statisztikai alapon történt a kategória határok kijelölése?

III. A harmadik publikáció a relatív terméshozam és a talaj tulajdonságainak összefüggéseivel foglalkozik (Juhos, K., Szabó, S., Ladányi, M. 2015. 'Influence of soil properties on crop yield: a multivariate statistical approach'. *Int. Agrophys.*, 29(4), pp. 433-440. <https://doi.org/10.1515/intag-2015-0049> Q2).

A kutatás célja az volt, hogy egy, a talajparaméterek esetenként nemlineáris összefüggéseit is figyelembe vevő, jól értelmezhető módszerrel feltárja a talajtulajdonságok és a hozam kapcsolatát egy kelet-magyarországi régióban az időjárási viszonyok függvényében, és igazolja a terméshozam régióspecifikus, pontosabb előrejelzésének lehetőségét. A Jelölt egy 28 parcellát magában foglaló, 225 ha-os terület 10 éves termés-, talaj- és meteorológiai adatsorát elemezte. Alaphipotézise szerint az egyszerű talajtani mutatók önmagukban nem elegendők a talaj tulajdonságai és a terméshozam közötti kapcsolat feltárásához, ezért a mélyebb összefüggések megértéséhez összetettebb módszerek szükségesek. Ennek érdekében a vizsgált 11 talajparamétert - a termés nagyságára való hatásuk alapján - transzformálta, a termésadatokat normálta, az adatokkal főkomponens regressziót végzett. Elemzése alapján kimutatta, hogy az aszályos években a terméshozamot főként a szikesedés, a talaj textúrája és a tápanyagtartalom, míg a nedvesebb években a domborzati helyzet, a szervesanyag- és tápanyagtartalom befolyásolta.

Megjegyzéseim a harmadik publikációval kapcsolatban: (1) A 11 talajtulajdonság bemutatása során számos elírás, illetve hibás fordítás került a dolgozatba (pl. CaCO_3 CaCO_3 helyett; az AL-Na ammon-laktát oldható vagy a növények számára könnyen felvehető nátriumot jelent; ugyanez áll az AL- P_2O_5 -re és a AL- K_2O -ra, ezek nem „elérhető”, hanem a növények számára könnyen felvehető mennyiségei a foszfornak, káliumnak; az OC az nem humusz, hanem a szerves C; és a Hargitai-N is definiálásra szorulna). (2) „Szürke erdőtalajok” nincsenek a jelenleg elfogadott hazai talajosztályozásban. Az eredeti cikkben lévő Gleysol-ok leginkább a réti talajainknak feleltethetők meg. (3) Megítélésem szerint a statisztikai elemzések azt mutatták, hogy inkább az aszályos években volt fontos a domborzati helyzet (a mélyebb fekvésű területek előnyösebbek lehettek talán). A nedvesebb években viszont fontos lehetett még a talaj mésztartalma is.

Kérdésem a harmadik publikációval kapcsolatban: A dolgozatban azt írja a Jelölt: „Annak érdekében, hogy jól interpretálható főkomponenseket kapjunk, a talajparamétereket először a (0,1) intervallumra képeztem úgy, hogy a magasabb értékek a talajfunkció szempontjából előnyösebb tulajdonságot képviseljenek”. Mi alapján döntötte el, hogy pl. a csökkenő pH egyre előnyösebb (a nagyon alacsony nem az!) vagy a csökkenő agyagtartalom előnyösebb (a homok területeken nem az!) vagy a növekvő mésztartalom előnyösebb (a nagyon nagy nem az!)? (Egyébként meg a következő publikációban sokkal árnyaltabb, függvényszerű hatását írja le az egyes talajparamétereknek...)

IV. A negyedik publikáció a talajminőségi indikátorok kiválasztásának módszertanával foglalkozik (Juhos K., Czigány Sz., Madarász B., Ladányi M. 2019. 'Interpretation of soil quality indicators for land suitability assessment – A multivariate approach for Central European arable soils, Ecological Indicators, 99 pp. 261–272. <https://doi.org/10.1016/j.ecolind.2018.11.063> Q1).

A kutatás célja egy helyspecifikus talajminőségi és ökológiai indikátorrendszer kidolgozása, valamint a növényi növekedést korlátozó tényezők azonosítása volt, a közép-európai szántóföldekre jellemző magyarországi talajok adatbázisa alapján. Jelölt a vizsgálathoz 1045 parcella (5000 ha) talajvizsgálati adatait használta fel. A transzformált talajadatokat főkomponens-elemzést végzett, majd a meghatározó talajtulajdonságok és a talajtípusok (mint válaszváltozók) bevonásával diszkriminanciaelemzést végzett. A talajparaméterek földhasználati értékeit különféle lineáris és nemlineáris függvényekkel fejezte ki, esetenkénti talajkategória-bontásban. Az elemzések alapján Jelölt a 25 hazai talajtípuson azonosította a növénytermesztést leginkább korlátozó tényezőket (textúra, talajvízszint mélysége, SOM, pH, Na, K, P, Zn), amelyek a talajminőség értékeléséhez szükséges minimális adatkészletet jelzik. Az eredmények egy komplex, hierarchikus rangsoroláson alapuló talajminőségi index kidolgozásának alapját képezhetik.

Megjegyzéseim a negyedik publikációval kapcsolatban: (1) A vizsgálatba vont 25 „talajtípus” számomra definiálatlan: ismét megjelenik a „szürke erdőtalaj” (ami nem létezik a hazai talajosztályozásban), az „agyag” és a „homok” (ami szintén nem létező talajosztályozási kategória), illetve a csernozjom, a barna erdőtalaj és a réti talaj (amik meg nem típusok, hanem főtípusok). Ráadásul az angol nyelvű közleményben WRB referenciacsoportok szerepelnek (Chernozems, Phaeozems, Luvisols, Cambisols, Gleysols, Solonetz, Arenosols), amik csak elvétve feleltethetők meg az itt ismertetett kategóriáknak. Az sem világos az olvasó számára (és a 18. táblázat sem segít ezt tisztázni), hogy miként lesz mindebből 25 talajtípus kategória? (2) A talajparaméterek fordítási hibáiról („rendelkezésre álló” helyett „könnyen felvehető” a helyes) már az előzőekben szóltam. A konduktivitás helyett talán megfelelőbb lenne a sótartalomról beszélni, hiszen az inkább megfogható talajtulajdonság és a vezetőképességből számolható. És „talajszerkezetet” sem mértek, még csak textúraadatok sincsenek az adatbázisban, hanem (valószínűsíthetően, a 17. táblázat „textúrahátáraiból” kikövetkeztethetően) Arany-féle kötöttségi szám értékek. (3) A Na- és S-tartalmak földhasználati érték leírását nem találom a 17. táblázatban, sem a Mn telítődési modelljének pH szerinti megkülönböztetését. (4) A „talajvízszint mélység” a leggyakoribb korlátozó tényezők közt szerepel a „Következtetések” alfejezetben, ám az elemzésekben az input adatok közt nem akadtam a nyomára.

Kérdéseim a negyedik publikációval kapcsolatban: (1) Honnan és miként származtatta az elemzésekhez használt földhasználati érték adatokat (ez nekem a dolgozatból nem derült ki, a feldolgozásnak ez a részét nem is igazából értem)? (2) A 17. táblázatban bemutatott „talajkategória határok” (textúra, pH, CaCO₃, talajvízszint) meghúzása mi szerint történt, illetve miért nem pl. a talajterképezési útmutató alapján lett megállapítva?

V. Az ötödik témakör a kanyargós szillevéldarázs fagyűrűsével és hőmérséklettől függő fejlődésével foglalkozik (Papp V., Ladányi M., Vének G. 2016. 'Temperature-dependent development of *Aproceros leucopoda* (Hymenoptera: Argidae), an invasive pest of elms in Europe' *Journal of Applied Entomology*, 42(6) pp. 589–597. <https://onlinelibrary.wiley.com/doi/10.1111/jen.12503> Q1; Vének G., Fekete V., Ladányi M., Cargnus E., Zandigiacomo P., Oláh R., Schebeck M., Schopf A. 2020. 'Cold tolerance strategy and cold hardiness of the invasive zigzag elm sawfly *Aproceros leucopoda* (Hymenoptera: Argidae)'. *Agricultural And Forest Entomology* 22(3) pp. 231–237. <https://doi.org/10.1111/afe.12376> Q1)

Ennek a kutatásnak a célja a kanyargós szillevéldarázs (*Aproceros leucopoda*) lárváinak hidegtűrési stratégiájának feltárása volt, SCP- (szuperhűlési pont) értékeik becslésével és elemzésével, fagyűrűsi típusuk meghatározásával, valamint fejlődésük és túlélésük hőmérséklet-függésének vizsgálatával. Az *A. leucopoda* gubóit magyarországi és olaszországi mintákból gyűjtötték, SCP-értékeiket Bécsben mérték, hidegtűrésüket több hőmérsékleti beállításban, fejlődésüket pedig szibériai szilfán, különböző hőmérsékleteken és fotoperiódus mellett vizsgálták. A Jelölt az SCP-értékeket robusztus ANOVA-val és Games–Howell post-hoc teszttel hasonlította össze, valamint öt szakirodalmi modellt parametrizált és ezek alapján becsülte az *A. leucopoda* fejlődésének hőmérsékleti küszöbértékeit. A fagyponat alatti hőmérsékleteken a telelő egyedek túlélése 89–100% között alakult, ami arra utal, hogy a faj nem fagyérzékeny, így a téli hideg nem jelent korlátozó tényezőt, és Magyarországon évente akár négy-öt generációja is kialakulhat. A fejlődés modellezésére a Lactin-2-modell bizonyult legalkalmasabbnak. A modellek eltérései további, szélesebb hőmérsékleti beállításokat igényelnek. Az eredmények fontos új ismereteket adnak az Európában invazív faj biológiájának megértéséhez.

VI. A hatodik kutatási témakör a bor NMR spektroszkópiával mért beltartalmi értékeit elemezte gépi tanulási módszerekkel (Nyitrai Á. D., Ladányi M., Varga Z., Szövényi Á.P., Matolcsi R. 2022. 'The Effect of Grapevine Variety and Wine Region on the Primer Parameters of Wine Based on ¹H NMR-Spectroscopy and Machine Learning Methods'. *Diversity*. 14(2):74. <https://doi.org/10.3390/d14020074> Q1).

A kutatás 861 borminta (¹H NMR-spektroszkópia) vizsgálati eredményei alapján négy magyar szőlőfajta és két borvidék modellalapú osztályozását végezte el, és meghatározta az osztályozás legfontosabb paramétereit. Először a mért 53 komponensből a fontosabb 22 ismérvet válogatták le, majd Jelölt felügyelt gépi tanulási osztályozási módszereket (lineáris diszkriminanciaelemzés, mesterséges neurális hálózatok, támogatóvektor-gépek és véletlen erdő) tesztelt a négy fajta és két származás (összesen nyolc csoport) elkülönítésére. Értékelte az egyes változók fontosságát és a legfontosabb transzformált változók hatásait a fajták és borvidékek közötti különbségek vizualizálására pókhálóaábrán mutatta be. Az ¹H NMR-alapú metabolizmus adatok elemzése véletlen erdő és támogatóvektor-gép módszerekkel sikeresen osztályozta a vizsgált borokat fajták és régiók szerint, valamint azonosította a legfontosabb elkülönítő jellemzőket (pl. szikimisav, kaftársav, metanol, borkósav, etanol, 2-feniletanol, szukcininsav).

VII. A hetedik kutatási témakör végül a rózsagyökérben (*Rhodiola rosea*) való metabolit-felhalmozódás és génexpresszió időbeli változásának hasonlósági struktúráját vizsgálta (Mirmazloun I., Ladányi M., György Zs. 2015. 'Changes in the Content of the Glycosides, Aglycons and their Possible Precursors of *Rhodiola rosea* during the Vegetation Period'. *Nat Prod Commun*. 10(8) pp. 1413–1416. <https://doi.org/10.1177/1934578X150100082> Q2).

Vizsgálat célja volt a *Rhodiola rosea* L. gyógynövények növekedésének, csökkenésének és stagnálásának vegetációs mintázatait jellemezni, valamint összehasonlítani a négy vizsgált növényegyed két szervén (levél és gyöktörzs) mért génexpressziós megfigyeléseket. Jelölt a vizsgált négy gén mindegyikére egy-egy háromdimenziós kódot képzett, majd a kódok alapján euklideszi

távolságokat definiált, mely értékeket az ún. „különbözőségi index” számításokhoz használt fel. Ezután 7 növényegyed gyökér és gyöktörzsmintájából HPLC-vel mért metabolitok mennyiségét elemezte hasonló indexképzéssel. A kidolgozott dimenziómentes távolság-alapú módszer hatékonyan alkalmazható volt a növények különböző szerveiben mért génexpressziós és metabolitváltozások dinamikájának összehasonlítására.

Megjegyzések a fejezethez: a vizsgálatban szereplő (stresszoldásra is szolgáló) gyógynövénynek (*Rhodiola rosea* L.) a köznyelvben többféle neve is elterjedt (aranygyökér, sarki gyökér, sarkvidéki gyökér, illatos varjúháj, rózsás varjúháj, rózsás gyökerű varjúháj, orosz aranygyökér, rózsás kötőrfű, rózsagyökér), de a dolgozatban egy helyen megemlített „rózsatövis” nem szerepel ezek közt.

KIEMELT EREDMÉNYEK (7. fejezet)

Jelölt ebben a fejezetben foglalja össze tézisszerűen legfontosabb eredményeit. Ezeket kicsit lerövidíteni javaslom:

1. A Jelölt 28 saját (társszerzős) publikációból (D1:3; Q1:15; Q2:8; Q3:2) vett 22 példán keresztül mutatta be az agrártudományi kutatások adatelemzésen alapuló háttér munkáját, amely a publikációkban gyakran nem látható, de nélkülözhetetlen az adatok értelmezéséhez és a tudományos kérdések megválaszolásához.
2. Az aszály mikroizotlábúakra gyakorolt hatását két megközelítéssel elemezte: (1) a relatív aktivitássűrűségen (RAD) alapuló, dinamikus faktorú MANOVA-val és (2) az aktivitássűrűségkülönbséget (ADD) értékelő speciális ordinális skálázással. Mindkét módszer alkalmasnak bizonyult a populációk aszályhoz való (súlyosságtól és időtartamtól függő) alkalmazkodásának jellemzésére.
3. 16 éves szántóföldi kísérlet adatai alapján véletlen erdő modellekkel sikeresen becsülte a lefolyás- és talajvesztés-események súlyosságát, és azonosította a folyamatokat meghatározó fő tényezőket.
4. Főkomponens-regresszióval azonosította az aszályos és kevésbé száraz években a relatív terméshozamot és annak változékonyságát leginkább meghatározó talaj- és meteorológiai tényezőket egy kelet-magyarországi, 28 parcellás kísérletben.
5. Kidolgozott egy új módszert a talajminőségi indikátorhalmaz kiválasztására és értékelésére, amely lineáris és nemlineáris modellekkel írja le a talajparaméterek földhasználati értékét, és lehetővé teszi a növénytermesztést korlátozó tényezők helyspecifikus azonosítását Magyarországi legjellemzőbb talajtípusain.
6. A Jelölt kimutatta, hogy az invazív kanyargós szilvéldarázs fejlődésének hőmérsékletfüggését leíró modellek közül a Lactin-2 teljesít a legjobban, és a kritikai értékelés alapján javaslatot tett a további kísérletek beállítására.
7. Igazolta, hogy a véletlen erdő és a támogatóvektor-gépek hatékonyan alkalmazhatók a borok ^1H NMR-alapú metabolomikai adatainak fajták és régiók elkülönítésére, és azonosította a megkülönböztetést meghatározó hét kulcsoösszetevőt.
8. A Jelölt egy olyan módszert dolgozott ki, amely a *Rhodiola rosea* L. különböző szerveiben mért génexpressziók és metabolitok vegetációs időszak alatti dinamikájának összehasonlítására alkalmas, és megfelelő módosítással más biológiai folyamatok időbeli lefutásának vizsgálatára is alkalmazható.

A megfogalmazott tézisek különböző témakörűek, így nem összevonhatók. A Jelölt által 8 pontban ismertetett adatelemzési, modellezési eredmények publikálása megfelelő szinten megtörtént. Új tudományos eredményként elfogadhatók az 5. pont kivételével (amennyiben a „földhasználati érték”

származtatását Jelölt a védés során hitelesen elmagyarázza, akkor elfogadom ezt a tételt is, egyelőre bizonytalan vagyok benne...).

ÖSSZEGEZÉS, NYILATKOZAT

A dolgozat széles tematikus spektrumon, gazdag példatárra támaszkodva mutatja be az agrártudományi kutatásokhoz kapcsolódó statisztikai és modellezési módszerek alkalmazásának lehetőségeit. A jelölt következetesen és kritikai érzékkel értelmezi a módszerek előnyeit, hátrányait, valamint gyakorlati korlátaikat, és rámutat azokra a pontokra, ahol az adott eljárások hatékonyan segíthetik a kutatási kérdések megválaszolását.

Összességében a dolgozat hiánypótló módon, sok példával szemlélteti az új statisztikai és modellezési módszerek lehetőségeit. Hasznosan hozzájárulhat az új kutatói nemzedék statisztikai ismereteinek bővüléséhez, és értékes ötleteket adhat a kutatási eredmények feldolgozásához és értelmezéséhez.

Összegezve megállapítható, hogy **Ladányi Márta doktori műve megfelel az MTA doktori nyilvános vitára bocsátás követelményeinek. Javaslom a nyilvános vitára bocsátását, és sikeres védés esetén támogatom számára az MTA doktora cím odaítélését.**

Keszthely, 2025. augusztus 29.



Dr. Makó András

az MTA doktora