

# **DOKTORI (PhD) ÉRTEKEZÉS**

**MIKÉNÉ HEGEDŰS FRIDERIKA**

Mosonmagyaróvár  
2006

**NYUGAT-MAGYARORSZÁGI EGYETEM**  
Mezőgazdaság és Élelmiszertudományi Kar  
Biológiai Rendszerek Műszaki Intézete

**Precíziós növénytermesztési módszerek Doktori Iskola**

Doktori Iskola vezető:  
Dr. Kuroli Géza  
egyetemi tanár, MTA doktora

**Termőhelyspecifikus precíziós növénytermesztés  
műszaki feltételrendszere program**

Program és témavezető:  
Dr. Neményi Miklós  
egyetemi tanár, MTA doktora

**A fuzzy logika és a neurális hálók alkalmazása a  
precíziós növénytermelés adatbázisának értékelésében**

Készítette:  
**MIKÉNÉ HEGEDŰS FRIDERIKA**

Mosonmagyaróvár  
2006

1. Bevezetés .....	3
2. Irodalmi áttekintés .....	6
2.1 A környezeti adatok jellemző tulajdonságai.....	6
2.2 Az adatok természete: bizonytalanság és hiba .....	9
2.3 Mintavételi stratégia keresése a precíziós mezőgazdaságban .....	15
2.4 A hozamértékek tér és időbeli stabilitása .....	22
2.5 A térbeli struktúra hatása az interpolációs módszerek pontosságára .....	25
2.6 A térbeli termésadatok becslése és azok bizonytalansága...	30
2.7 A talajtulajdonságok elemzése .....	32
2.8 Becslési modellek .....	39
2.9 Mesterséges intelligencia a környezeti adatok kezelésében	41
2.9.1 Fuzzy halmazok.....	41
2.9.2 Fuzzy osztályozás .....	46
2.9.3 Fuzzy krigelés.....	50
2.9.4 Fuzzy klaszterezés .....	52
2.9.5 Fuzzy szakértői rendszerek.....	55
2.9.6 Neurális háló az ökológiai modellezésben .....	57
2.9.7 Termésbecslés neurális hálóval .....	58
2.9.8 Neuro - fuzzy-rendszerek .....	62
3. Anyag és módszer.....	65
3.1 Mintaadatok gyűjtése.....	65
3.2 Statisztikai elemzés.....	69
3.3 Térbeli interpolálási technikák .....	71
3.3.1 Interpolálási módszerek.....	72
3. 4 A mesterséges intelligencia módszerei az adatelemzésben	77
3.4.1 Fuzzy logika alkalmazása – fuzzy számok és tagsági függvények .....	78
3.4.2 Fuzzy krigelés és variogram.....	82
3.4.3 Fuzzy rendszerek .....	85
3.4.4 Evolúciós algoritmusok- genetikus és bakteriális algoritmusok .....	91
3.4.5 Neurális háló.....	95
4. EREDMÉNYEK.....	102
4.1 Adatok elemzése.....	102
4.2 A hozam és talajadatok térbeli becslése .....	113
4.3 A talajadatok fuzzy variogramja és krigelése.....	135
4.4 A terméshozam becslése fuzzy szimulációval.....	142

4.5 Terméshozam becslése neurális hálóval.....	158
4.5.1 Az adatok előkészítése és elemzése.....	158
4.5.2 Terméshozam és a talajadatok kapcsolata.....	160
4.5.3 A kukorica terméshozamának becslése – talajadatokkal és meteorológiai tényezőkkel .....	173
5. Következtetések, javaslatok .....	182
6. Összefoglalás .....	184
7. Új kutatási eredmények (tézisek) .....	189
9. Irodalomjegyzék .....	190
10. Summary.....	200

## 1. BEVEZETÉS

A precíziós mezőgazdaságnak sokféle meghatározása van, de rendszerint tartalmazza a szántóföldi változékonyság kezelését az információtechnológia alkalmazásával. Az új információs eszközök, nemcsak az inhomogenitás kezelésére, de általános gazdálkodási feladatok, pl. szántóföldi munkák dokumentálására is alkalmasak. Az új gazdálkodási technológiák térhódításával, mint a GPS és az új érzékelési technikák, képessé teszik a helyi inputok kezelésére. A precíziós mezőgazdasági rendszer még nem teljes, különösen a döntési támogatások kezelését kell fejleszteni.

A számítógép-vezérelt eszközökkel felszerelt mezőgazdasági gépek, működésük közben GPS segítségével, a meglévő adatok alapján végzik el az aktuális területre vonatkozó feladatokat. A célok megvalósítására a precíziós mezőgazdaság 4 integrált technológiát alkalmaz, amelyek eszközei az alábbiak:

- *Globális Helymeghatározó Rendszer (GPS)*- műholdak által sugárzott jel segítségével határozzák meg a speciális vevőkészülékek az éppen aktuális földrajzi helyzetüket.
- *Távérzékelési technika*- tárgyak, vagy jellemzők közvetlen kapcsolat nélküli észlelésére, illetve beazonosítására alkalmas érzékelők.
- *Térinformatika*- számítógépes hálózat és programrendszer az adatok összegyűjtésére, tárolására, visszakeresésére, vizsgálatára és térbeli megjelenítésére

- *Termőhely specifikus beavatkozás-* mezőgazdasági gépekre szerelhető számítógép-vezérelt eszközöket alkalmaz a szükséges műveletek irányítására.

A technika mind színvonalában, mind a szolgáltatási lehetőségek tekintetében tovább korszerűsödik. A műholdas helymeghatározással egyre megbízhatóbb és egyre nagyobb információhalmazt nyerhetünk, amelyet kiegészítve a térinformatika egyéb adatgyűjtési lehetőségeivel szilárd alapot teremthetünk az összefüggések feltárásához. Az érzékelők a mezőgazdasági gépek alapvető tartozékai. A vizuális képesség segítségével információt szolgáltatnak a termés állapotáról, a fejlettségéről és a gyomfertőzöttségről. A vizuális érzékelés, vagy más érzékelők által összegyűjtött információkat használják a művelésnél, a vegyszeres kezelésnél és a betakarításnál.

A *Földrajzi Információs Rendszer (FIR)* a térinformatika eszköze, amellyel a földrajzi helyhez köthető adatokat tartalmazó adatbázisból információk vezethetők le. Technikáját tekintve olyan számítógépes rendszer, melyet a földrajzi helyhez kapcsolódó adatok gyűjtésére, tárolására, kezelésére, elemzésére, a levezetett információk megjelenítésére, megfigyelésére, modellezésére dolgoztak ki. Lehetőséget ad nagyszámú helyzeti és leíró adat gyors, együttes, integrált áttekintésére a térképhasználat, pontosabban a földrajzi adatok használatának fejlesztésére. A hálózatok terjedésével egyre nagyobb hangsúlyt kap az információk elérését, továbbítását szolgáló szerep. A FIR felépítésében, tartalmában, alkalmazási oldalról, a felhasználói

környezetet illetően nagyon eltérő formákban jelenik meg.

Hogyan kezeljük az adatokat, információkat? Alapjaiban a precíziós mezőgazdaság nem igényel különleges technikát, a FIR foglalkozik legjobban az ilyen típusú információkkal. Előnyei a számítógépes térképkészítés, a térképek összekapcsolása mellett, lehetőséget nyújt a tábla történeti adatainak (hozam, kezelés, alkalmazás) tárolására és feldolgozására. A talajvizsgálatok eredményei az egyes talajtényezők jellemzőit rögzítik, és a termelők ezek változását összehasonlíthatják. A döntéseket a FIR-ből nyert és a hagyományos források: - a kísérletek, a tanácsadók, a szolgáltatók, más gazdálkodási információk- alapján hozzák. A szántóföldek változékonyságának és alkalmasságának kutatása interdiszciplináris megközelítést igényel. Magában foglalja a különböző területekről és erőforrásokból származó adatok gyűjtését, tárolását, továbbítását elemzését és a döntési technikák kidolgozását. Ezen a területen a folyamatos agrárkutatás elvezethet a változékonyság okainak jobb megértéséhez, a több tényezős becslési technikák kidolgozásához (Tamás, 2003).

*A kutatás célja a hozamtérképek bizonytalansági tényezőinek, hibáinak elemzése, kiszűrése, a vizsgált időszakban termesztett növények és talajjellemzők heterogenitásának, térbeli szerkezetének meghatározása. Ezek ismeretében és a többtényezős becslési modellek kidolgozásával, valamint az elemző és előrejelzési eszközök alkalmazásával döntések előkészítése a precíziós mezőgazdasági termelés számára.*

## 2. IRODALMI ÁTTEKINTÉS

### 2.1 A környezeti adatok jellemző tulajdonságai

Magyarország egyik legfontosabb megújítható természeti erőforrása a talaj (Várallyay, 1998). A talaj tulajdonságait és ezen keresztül a környezet minőségét nagymértékben befolyásolja az alkalmazott agrotechnika szintje. A talaj fizikai, kémiai és biológia jellemzői, így a talaj termékenysége nem állandó, térben és időben is változik, mely változás léptéke is eltérő lehet, így akár egy mezőgazdasági táblán belül is elkülöníthetünk különböző termékenységű talajfoltokat.

A környezeti adatok heterogenitását és bizonytalanságát jellemző tulajdonságok összefoglalhatók az alábbiakban (Salski, 1999):

Nagy adathalmazok (hosszú idősorok, nagy felbontású térbeli adatok) jellemzik,

*Heterogenitás*, amely az alább felsoroltakból származik:

- az adatforrások különböző típusúak; az ökológusok minden lehetséges adat- és ismeretforrásból gyűjtik és használják az adatokat
- az objektív információforrások (főleg kvantitatív) előnyben részesítik a méréseket,
- a szubjektív információ forrásai – a mérési adatok helyett kedveli a szakértői ismereteket és szubjektív becsléseket (nem minden ökológiai paraméter mérhető).
- különböző típusú adatszerkezetek és adatformátumok (idősorok, térbeli adatok) jellemzik



- különböző minőségű (kvantitatív és kvalitatív) adatokat használnak.

*Bizonytalanság* – az információkban rejlő bizonytalanságot okozzák:

- a véletlen változók megjelenése,
- a hiányos és pontatlan adatok (mérési pontatlanság),
- adatok összehasonlíthatatlansága (változatos mérések vagy megfigyelési feltételek),
- mérések helyett közelítő becslések (technikai és pénzügyi problémák következtében),
- kvantitatív információk helyett kvalitatív,
- hiányos vagy bizonytalan szakértői ismeretek,
- a szakértőtől nyert információk szubjektivitása.

Az adatelemzést és az ökológiai modellezést támogató rendszerek követelményei a fentebb említett tulajdonságokból következnek. Az adatelemzés és modellezés speciális módszerei használhatók a környezeti adatok heterogenitásának és bizonytalanságának kezelésére.

A kutatásokban foglalkoztak mind az objektív mérési adatokkal, mind a szakértőtől származó szubjektív információkkal. Megvizsgálták, hogyan befolyásolta a térbeliség az adatokat, hogyan lehet elemezni a térbeli hatásokat, de a bizonytalansági probléma kezelésénél statisztikai vagy sztochasztikus szempontokat nem vettek figyelembe.

Adatelemzésen adatok olyan feldolgozását értjük, amelynél meghatározott kérdésekre információt kapunk az adatokról. A

feldolgozás valamilyen statisztikai vizsgálat, osztályozás, klaszteranalízis, de akár mintaillesztés, vagy függvényközelítés is lehet. Számos szoftver támogatja magát az adatanalízis folyamatát és pl. hagyományos statisztikai módszerekkel, vagy fuzzy-technikával elemzi az adatokat, állapít meg összefüggéseket közöttük.

Az adatanalízis folyamata négy egymásra épülő lépcsőre bontható, és minden lépcsőben összetettebb feladatok fogalmazhatók meg (Höppner et al. 1997):

- Az első lépcsőbe a legegyszerűbb feladatok tartoznak: gyakoriságvizsgálat, megbízhatósági vizsgálat, vagy az adatok szelektálása különböző szempontok alapján.
- A második lépcsőt a mintafelismerés jellemzi, amely az adatok különböző csoportosítását, ill. finomabb belső struktúra alapján való további bontását adja. Általában távolságok, hasonlóságok definiálásával, felhasználásával alakítunk ki csoportokat.
- A harmadik lépcsőben akár több matematikai modellel is vizsgálhatók az adatok, melyek más-más szempontú megközelítéseket jelentenek. Így bizonyos eloszlásokat feltételezve, ellenőrzéseket kérhetünk, vagy függvénykapcsolatokat állapíthatunk meg a regresszióanalízis alkalmazásával.
- A negyedik lépcsőben a korábbi elemzés alapján a következtetések levonása és értékelése történik. Pl. a hiányzó adatok ellenére megállapítható milyen csoporthoz,

struktúrához tartoznak az adatok, ill. az, hogy a harmadik lépésben megállapított kapcsolatok helyesek-e.

A fuzzy adatanalízis módszerei közül, a klasszikus fuzzy klaszter eljárások a második lépcsőbe, a fuzzy-osztályozás, különböző szabályfelismerések, fuzzy következtető eljárások a harmadik lépcsőbe tartoznak.

## **2.2 Az adatok természete: bizonytalanság és hiba**

A bizonytalanság két jól ismert forrása a mérés és mintavétel. Jó bevezetést ad a térbeli adatok bizonytalanságába a Földrajzi Információs Rendszerrel összefüggésben Goodchild és Gopal (1989) könyve a térbeli adatbázisok pontosságáról.

Vizsgáljuk meg a hibák forrását! Mérési hiba előfordul minden mérésnél, mert a műszer korlátolt pontosságú. Ezt rendszerint meghatározzák az eszköz jellemzői, az ismételt mintavétel és a statisztikai jellemzés. Ezek a hibák rendszerint függetlenek, normális eloszlásúak és többé-kevésbé pontosan jellemezhetők. A talaj értékelésével összefüggésben ritkán szignifikánsak, ha összehasonlítjuk a mintavétellel.

Mintavételi hibát okoz, hogy szinte mindig csak a vizsgált objektum kis részét vagyunk képesek mérni. Ezek a hibák nehezebben jellemezhetők és javíthatók, mint a mérési hibák. Általában nem igazolt és nem tesztelt feltételezéseket kell tennünk a mintavételi stratégiánál. A hibák típusát és terjedelmét meghatározhatjuk ismételt és teljes körű mintavételezéssel.

Mindkét hibának terjedelmes statisztikai elmélete van.

Szokásosan a természeti erőforrásoknál és az ökonómiai vizsgálatoknál feltételezzük, hogy a várható érték és a variancia vagy normális, vagy Student eloszlású, amit használhatunk a várható értékek bizonytalanságának kifejezésére. Ha a hiba eloszlása ismert, de nem-normális alakú, megbecsülhetjük az eloszlás paramétereit (Kravchenko és Bullock, 2000).

Az adatokban (természetükénél fogva) megjelenő változékonyságot nem szabad összetévesztenünk a hiba szó hétköznapi értelemben vett jelentésével. Ezek a felismerhető vagy észlelhető hibák, amelyek esetleg nem minősíthető adatokat produkálhatnak, vagy amelyeket gondos és ellenőrzött adatgyűjtéssel elkerülhetünk.

A szisztémás hiba a megfigyelések, kísérletek során jelentkező olyan hiba, ami a folyamat során nem állandó, az egyes méréseket másképpen érinti. Jellemző szisztémás hibát okozhat a vizsgálati egységek időben, vagy térben jelentkező inhomogenitása. Például a térben egymás mellett lévő vizsgálati egyedek, földparcellák egymáshoz hasonlóak, de jelentősen eltérhetnek a távolabb elhelyezkedők. Így ha méréshez földparcellákat kell választanunk (például két műtrágya összehasonlításához) akkor biztosítanunk kell, hogy mind a két anyag egyforma eséllyel kerüljön kipróbálásra bármelyik parcellán, ugyanis nem tudhatjuk, hogy nem keverjük-e össze a vizsgálandó anyagunk hatását a parcellák csoportjai között esetleg meglévő (szisztémás) különbséggel. A mérés reprodukálhatósága gyakran fontos követelmény, hiszen a valódi értéktől való eltérés, vagy szisztémás hiba ennek során orvosolható. A mérési ingadozást a mérő eljárás javításával lehet

csökkenteni. Általában a pontosabb mérés egyben költségesebb is, és a kívánatos mérési pontosság attól is függ, hogy mekkora a mérendő tulajdonság biológiai változékonysága. Egy nagyon változékony tulajdonságot gyakran nem érdemes a változékonyságnál sokkal nagyobb pontossággal mérni.

A szántóföldi táblák értékelésében mindig szerepelnek mintapontot nem tartalmazó területek, így a mintavétel problémája kikerülhetetlen. Hogyan jellemezzük és kombináljuk a bizonytalan értékeket? Vita tárgyát képezi a változó méretű és alakú térképegység, amelyet a FIR rács celláiban használnak. Ennek akkor van nagy jelentősége, ha statisztikai, vagy dinamikus szimulációs modellt alkalmaznak, amely speciális numerikus inputok halmazát igényli, amit a rendelkezésre álló adatokból nyerhetünk (Hatfield, 2000). A klasszikus statisztika is hangsúlyozza, hogy minimális adatokból, maximális információt nyerjünk.

A térképegységet egyszerű „*tipikus*” értékkel jellemezték. Előnye, hogy megfelelő adatok hiányában az aktuális változékonyságot mutatja, és a jól választott egyszerű érték rendszerint a várható érték elfogadható becslését adja. Hátránya, hogy nem informál az egységen belüli változékonyságról, és így információvesztéssel jár.

Az „*értéktartomány*” alkalmazásánál, feltehetően átíveli a változékonyságot, és nem tudjuk megállapítani a legvalószínűbb értéket a tartományon belül. Előnye, hogy kifejezi az egészet, vagy legalább a változó tartományának lényeges részét, alkalmazható

érzékenységi vizsgálatokra. Hátránya, hogy nem egy kifejezett érték, és nincs információja az osztályon belüli értékek eloszlásáról.

A „*statisztikai eloszlás*” teljesen leírja az adatértékeket és a hozzátartozó valószínűségeket, de ez ritkán igazolt elméletileg és kísérletileg. A nem paraméteres eloszlásnál nincsenek feltevéseink az eloszlásról.

Megvizsgálták az adatok bizonytalanságát is. Ha különböző változók korrelálnak, nem függetlenek, a változó viselkedését nem megfelelően írja le az egyváltozós eloszlás. Helyette meg kell határoznunk a többváltozós eloszlásukat, ami általában magában foglalja a *variancia-kovariancia* mátrix számítását. Ez lényegesen bonyolultabb, mint az egyváltozós eloszlás. A kovariancia szerkezete fontos, mert ha kiszámítjuk a különböző változók függését (együttváltozását), az eredmény bizonytalansága nemcsak a változók egyedi varianciájától, hanem a kovarianciájuktól is függ (Podani, 1997).

Ha a rendelkezésre álló adatok hibátlanok, a változók kombinációja eredményezhet bizonytalanságot: pl. az alkalmazott függvény igazi alakja ismeretlen, vagy a szakértői döntés bizonytalan, nem egyértelmű.

A *hiba terjedése (növekedése)*: Ha az adatokat leírtuk valószínűségi eloszlásokkal, az eredményben a hibát (bizonytalanságot) meghatározhatjuk a Gauss által kifejlesztett klasszikus hibaterjedési elmélettel. Megadhatunk pontos konfidencia intervallumot. Heuvelink és Burrough (1993)

kifejlesztették a hibaterjedés elméleti és gyakorlati megközelítését GIS alkalmazásoknál.

Feltételezték, hogy különböző mennyiségeket, a független változókat  $(x_1, x_2, \dots, x_n)$  kis bizonytalansággal mértek  $(\delta x_i)$ , ez független és véletlen. Az eredmény  $z = f(x_1, x_2, \dots, x_n)$ , a függő változó bizonytalansága, ha a változók korrelálatlanok:

$$\delta z = \sqrt{\left(\frac{\partial z}{\partial x_1} \delta x_1\right)^2 + \left(\frac{\partial z}{\partial x_2} \delta x_2\right)^2 + \dots + \left(\frac{\partial z}{\partial x_n} \delta x_n\right)^2} \quad (1)$$

Ha a változók korrelálnak, a hiba nagyobb lehet (pl. pozitív korreláció esetén erősítik egymást), vagy kisebb a negatív korreláció esetén. A hiba számításánál figyelembe vesszük a kovarianciákat is.

$$\delta z = \sqrt{\sum_i \left(\frac{\partial z}{\partial x_i} \delta x_i\right)^2 + \sum_i \sum_j \frac{\partial z}{\partial x_i} \frac{\partial z}{\partial x_j} \delta x_i \delta x_j} \quad (2)$$

Leggyakoribb probléma, hogy kis távolságon belül a kovariancia kicsi, összehasonlítva a varianciával, és így általában figyelmen kívül hagyják.

A függvényekben alkalmazott műveleteket figyelembe véve, a hiba terjedése módosul, az (1) helyett az összeadás és különbség esetén (3), a szorzás és osztásnál a (4) összefüggés alapján számolhatjuk ki.

$$\delta z = \sqrt{(\delta x_1)^2 + (\delta x_2)^2 + \dots + (\delta x_n)^2} \quad (3)$$

$$\frac{\delta z}{|z|} = \sqrt{\left(\frac{\delta x_1}{x_1}\right)^2 + \left(\frac{\delta x_2}{x_2}\right)^2 + \dots + \left(\frac{\delta x_n}{x_n}\right)^2} \quad (4)$$

Hogyan mérjük a bizonytalanság terjedését?

A kis bizonytalanság ( $\delta x$ ) megbecsülhető különböző úton: Nagyon általános módszer a minta standard eltérése (5), és a négyzetes átlagos hiba (RMS).

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (5)$$

ahol  $\bar{x}$  a minta átlaga,  $n$  a minta mérete. A hiba terjedését figyelembe véve a (3) és a (4) összefüggéseket felírhatjuk az alábbi formában:

$$s_z = \sqrt{s_{x1}^2 + s_{x2}^2 + \dots + s_{xn}^2} \quad (3')$$

$$s_z = \sqrt{s_{x1}^2 (x_2 \ x_3 \ \dots \ x_n)^2 + \dots + s_{xn}^2 (x_1 \ x_2 \ \dots \ x_{n-1})^2} \quad (4')$$

Korreláló változóknál a kovarianciát is figyelve a hiba nagyobb, vagy kisebb lehet. A gyakorlatban a  $\sigma_{xy}$  értékét megbecsülhetjük a minta kovarianciával.

A bizonytalanság az adatok térképen történő ábrázolásánál vagy egységesen kezelhető, vagy cellánként különböző lehet. Így általában két térképet kezelünk: a *várt értéket* és a *bizonytalanságukat* tartalmazót.

A fentieket összefoglalva a térképező programoknál az alábbi problémákkal szembesülünk: a korreláció erősen befolyásolja a hiba terjedését, az erős pozitív kovariancia a hiba felnagyításához, az erős negatív kovariancia a hiba törléséhez vezet, ami bizonytalanná teszi az eredményt. A FIR műveleteknél



nem határozható meg zárt formában a hiba terjedése. Ha feltételezzük a térbeli függetlenséget, akkor a szomszédos mintapontokból nyert információ már valamilyen térbeli függőséget okoz.

### **2.3 Mintavételi stratégia keresése a precíziós mezőgazdaságban**

A precíziós mezőgazdaságot definiálhatjuk, mint a terméshozamhoz szükséges *inputok* meghatározását, a helyileg meghatározott követelményeknek megfelelően. Viszonylag új kutatási terület, az agrárkutatókon kívül a statisztikusok és matematikusok informatikusok is részt vesznek a munkában. A monitoring feladata térben és időben dokumentálja a változásokat, amelyek természetesen végbemennek a táblán. Megvizsgálták azokat a statisztikai módszereket, amelyek használhatók a kísérletek tervezésében. Fő kihívás azon problémák megoldása, amelyek kapcsolatban vannak azzal, ha az egyforma talajú kis területről a tágabb tábla felé mozdulunk (Tamás, 2001).

Sok probléma érintett a térbeli adatok használatával és interpretálásával, a geostatisztikai technikákkal. A másik fontos területe a döntési technikák fejlesztése. A rendszerek lehetővé teszik a bizonytalanság belefoglalását a modellekkel és az input információkkal kapcsolatban, képessé téve a felhasználót, hogy megismerje a különböző döntések kockázatait.

A táblakísérletek fő célja, hogy összehasonlítsa a különböző módszerek eredményességét. A szabatosság és a pontosság a legfőbb szempont, de a tényleges hiba becslése is nagyon fontos.

Megfelelő minta tervezése és statisztikai elemzés minimalizálja ezt a problémát.

A szántóföldön a terméshozam és a termesztési feltételek (pl. talaj termőképessége) tér és időbeli változékonyságot mutatnak. A precíziós mezőgazdaság számára fontosak azok az információk, melyek elemzik a talaj és a terméshozam térbeli változékonysága közti kapcsolatokat. A technikai gyakorlat és a gazdaságos hatásos kezelések a táblán belül kis területenként különbözhetnek, a helyspecifikus növénytermesztésnek megfelelően.

A precíziós mezőgazdaságban egyik közelítés az ún. „egységes zóna” definiálása, amelyet ugyanazzal a stratégiával kezel (ugyanaz a talajművelés, a vetéssűrűség, műtrágyaadag, növényvédelem, stb.). A kezelési zóna a tábla részterülete, amely a hozamot meghatározó tényezők viszonylag homogén kombinációja. A határai időben és térben különbözhetnek a szántóföldi műveleteknél. A precíziós mezőgazdaság helyspecifikus technológiájának alapvető kérdése, hogyan lehet meghatározni az optimális kezelési zónát (Goodwin et al. 2003). A kezelési zónák olyan földrajzi egységek az adott területen, amelyek potenciálisan külön-külön kezelhetők. A térbeli információ pontosságának nagyobb mértéke, a bemenetek optimális szintje kívánatos. A rögzített költségek zónánként, vagy gridenként kapcsolódnak a nagyobb pontossághoz. Ha megfelelően választunk ökonómiaailag optimális kezelési zónát vagy gridméretet, nagy lehetőséget kínál a profit és a kockázati kezelés kombinálásában. Mégis ez a döntés veszélyt és komplex problémát jelent. Számos

termelő egységes gridméretet kíván kialakítani (szembeállítva a változatos alakú és formájú kezelési zónák) és szemben áll a legjobb rácsméret használatával. A kutatások középpontjában kritikus és helyspecifikus technológiának megfelelő gazdaságos megvalósítás áll. Három speciális terület figyelemre méltó a kutatásokban: technológia, rácsminta és kezelési zónák.

Dillon (2002) matematikai programozási modellt alkalmazott az optimális kezelési zóna körvonalazására. Három lépésben vázolta a cél megvalósítását: 1) a komplett adatokon alapuló termelési, kezelési zóna gazdaságilag optimális körvonalazását meghatározó optimalizálási eljárás kifejlesztése; 2) a modell gyakorlati kipróbálása; 3) érzékenységi elemzés végrehajtása – milyen az optimális zónák ingadozása/változása /reagálása gazdasági döntés-előkészítő környezetben. A modell magában foglalja a kukoricatermesztés gazdasági döntéseinek keretrendszerét, használva a változó arányos inputot, és kombinálva a kevert nemlineáris egész programozási eljárással. A vizsgálatban a nagy modell méret és a nagyszámú kezelési zóna okozott nehézséget. Az eredmények érzékenyek az alapul szolgáló termelési függvényekre, a gazdasági környezetre és a rendelkezésre álló talajerőforrásokra, különös tekintettel a méretekre és térbeli szomszédokra.

A termőhely-specifikus termelés megvalósításának kulcskérdése a megfelelő pontosságú információk begyűjtése az adott területről. A precíziós gazdálkodás első szakasza azon információk összegyűjtése, amelyek leírják a növény fejlődési

feltételeit, elsősorban a talaj termőképességét meghatározó adatok. A gyakorlatban a következő módszerek terjedtek el (Gang et al, 1999):

- Talajtérképek alapján történő mintavétel,
- Távérzékelési módszerek alkalmazásával,
- Növények hozamtérképe alapján történő mintavétel,
- Szisztematikus rács talajmintavétel.

A szerzők statisztikai és geostatistikai elveket és módszereket használtak a szántó föld térbeli változatosságának tanulmányozásában. A gyakorlati precíziós mezőgazdasági technikák követelményeinek megfelelően kifejlesztették a megfelelő talajmintaszám és a mintaintervallum számítási módszereit. A tesztelt mezőt három kis területre osztották: 10\*90m, 15\*150m, 20\*70m. A talajtulajdonság térbeli változékonyságának lehetséges amplitudóját vizsgálva a talajminta rácssűrűsége 5\*5m, a minták száma 111.

A kísérleti mintánál a rácsban a talaj N, P, K és szerves anyag tartalmát mérték. A minta térbeli eloszlásának elemzése hasznos a mintavételi stratégia kialakításában.

A statisztikai elemzésben elvégezték a leíró statisztikai számításokat: a vizsgált jellemzők átlagát, szórását, varianciáját, variációs együtthatóját(CV), és az eloszlások jellemzőinek meghatározását.

Az alkalmazott geostatistika a regionális változók elméletén alapszik, a variogram függvényt használtak, mint alapvető elemzési eszközt. Segítségével tanulmányozták a változók térbeli eloszlását.

Ez a függvény bizonyos értéket vesz föl minden térbeli helyen, értéke a hellyel együtt változik. A változó bizonyos kapcsolatot mutat térbeli tartományon belül. Ha a térbeli hely a vizsgált tartományon kívül van, a kapcsolat csökken, vagy megszűnik. Ez a tulajdonság nagyon hasznosnak bizonyult a minta intervallum számításában. A változékonyság mennyiségének meghatározására korrekt modellt illesztettek a kísérleti variogramhoz, meghatározva a paramétereit. Az adatpontok, amik az ismert pontok adott hossz távolságán belül vannak, interpolálhatók. Így a hossz a maximális mintaintervallum, amit a talaj térbeli változékonyságának térképezésénél használunk.

Összefoglalva a vizsgálatok eredményei:

1. Ha a minták térben függetlenek, az *optimális mintaszám* meghatározásának szakaszai:

a) Létrehozták a talajmintarácsot a kis teszt területen és kiszámolták a hossz értékét, a változó átlagát és szórását.

b) Megválasztották a tesztfeltételnek és a műszer pontosságának megfelelően a minta pontosságát:

$$\left| \frac{\bar{Z}_N - \mu}{\bar{Z}_N} \right| \leq d, \text{ ahol } \mu \text{ a mintaátlag, } d=(5\%, 10\%, 15\%) \text{ a}$$

relatív hiba.

c) A standard normális eloszlás kritikus értékének megfelelően  $U_{95\%}= 1,96$  ill.  $U_{90\%}= 1,645$  értékek esetén a mintaszám:

$$N * 3,84 * \left( \frac{S}{Z_N * d} \right)^2 \alpha = 0,05$$

$$N * 2,71 * \left( \frac{S}{Z_N * d} \right)^2 \alpha = 0,1$$

Különböző  $d$  és  $\alpha$  esetén elemezték a három területen a kiszámított mintaszámot.

2. A variogram függvény paramétereit használták *a minta intervallum kiszámításánál*

Ennél a függvénynél teljesülnie kell a regionális változóknál az ún. belső hipotézisnek. (Steiner, 1990).

Különböző  $h$  értéknél kiszámolták  $\gamma(h)$  (variogram) értékét, modellt illesztettek az adatokhoz. A függvényben meghatározott hossz lesz a minta intervalluma. A mintában az intervallum értéke az egyes változóknál 30 és 45 m között volt.

Néhány nehézségre, problémára rámutattak a variogram paraméterek meghatározásánál:

- a) A számítási pontossághoz elegendő minta kell, azaz  $N(h) > 20$ , minden vizsgált  $h$  távolság esetén.
- b) A talajváltozó irányonként különbözően változhat, akkor a skálát a különböző irányokban is meg kell vizsgálni.
- c) Ha a belső hipotézis nem teljesül, akkor ezt adat „kilengés”-nek tekintjük.

$$E[Z(x)] = \mu(x) + \varepsilon(x)$$

Sok esetben a kilengés megszüntethető. Két módszer ismeretes: egyik módszer az előzetes felmérés, ha a térbeli változékonyság

nyilvánvaló. A vizsgált területet felosztják nagyon kis területekre, amik már hasonlóak. A másik módszer megnöveli a minta intervallumát, ha elég széles, akkor a hipotézis már teljesül.

A gyakorlatban a táblák mérete nagyon különböző néhány hektártól több száz hektárig terjed. Ha az egész táblát tesztmintának tekintjük, akkor a talajminták, a mintaszámok és intervallumok kezelése nagyon sok pénzt és időt igényel. A rácsminta sűrűségét meghatározhatjuk, ha kiszámoljuk a megfelelő mintaszámot és intervallumot és kiterjesztjük az egész táblára.

A módszer továbbfejlesztésénél a talajtulajdonságok térbeli változékonysága mellett az időbeli változékonyságot, és a különböző tényezők közötti kapcsolatot is számításba kell venni. Az idő és térbeli területi változók geostatistikai többváltozós elemzése szükséges az ilyen típusú problémák megoldásához.

Sok geostatistikai tanulmány kiindulópontja, hogy a mintaterv döntő szerepet játszik a krigelési becslés minőségében. Sacks és Schiller (1989) sztochasztikus optimalizáló algoritmust használtak a minta terv optimalizálására a becslési hiba minimalizálásakor. Az OK algoritmus egyik vonzó tulajdonsága, feltéve, hogy a geostatistika belső hipotézise (feltételezése) teljesül, a becslési pontosság kifejezhető a krigelési varianciával (Webster és Oliver, 1990). Így a krigelési variancia használható eszköz a mintaterv minőségének értékelésére.

Van Groenigen és Stein (1998) és Van Groenigen (1999) térbeli szimulációs eljárást (SSA) algoritmust használtak a mintaterv konstruálására. Folytonos megoldási térben és

valószínűbb minta számot használva olyan tervet konstruáltak, amelynek krigelési varianciája minimális. Két optimalizálási kritériumot vezettek be, és megvizsgálták a variogram paraméterek hatását az optimális minta tervre. Az első kritérium az *átlagos krigelési varianciát*, a másik a *maximális krigelési varianciát* minimalizálja. A krigelési varianciát jelentősen befolyásolták a variogram paraméterek, így a hatásukat az optimális mintatervben figyelembe kell venni. A vizsgált Gauss variogram és az exponenciális variogram különböző mintatervet eredményezett ugyanannál a röghatásnál, küszöbnél és tartománynál. Az exponenciális, a szférikus és a lineáris variogramok röghatás nélkül szabálytalan, hasonló mintatervet eredményezett. Megmutatták, hogy a variogram anizotrópia jelentős hatással van az optimalizált mintatervre, a nagyobb valószínűség irányában intenzívebb a mintasűrűség.

#### **2.4 A hozamértékek tér és időbeli stabilitása**

Az időbeli és térbeli változékonyság már régóta a mezőgazdasági termelők által jól ismert, elfogadott, de a táblán belüli változékonyság megértése és kezelése nagy kihívást jelent a jelenlegi kutatásokban (Bishop és McBratney, 2001). A termésmonitor a termelőknek közvetlen módszert adott a terméshozam térbeli változékonyságának mérésére. A terméstérképek mutatják a különböző hozamú területeket, és gyökeresen megváltoztatták a döntéshozatal útját. Arslan és Colvin (2002) összefoglalták, milyen szempontokat kell figyelembe venni



a hozamtérképek készítésénél. A változékonyságot számos tényező okozhatja, beleértve a talaj típusát, a szántó föld helyzetét, a termesztés előzményeit, a talaj fizika és kémiai tulajdonságait és a tápanyag hozzáférhetőségét. A biotikus és abiotikus tényezők közötti kölcsönhatás is vezethet a termés nagyság térbeli változékonyságához. Kimutatták, hogy a termés különbségek a talajtérkép egységek és a szántó föld helyzetének függvényei. Taylor és munkatársai (2001) megállapították, hogy a táblán belüli változékonyságot a talajban rejlő fizikai és kémiai változékonyság okozza, amely hat a tápanyagellátásra. Amióta a földrajzi információk is elérhetők a tábla paramétereiről, az adatok különböző rétegei elemezhetők, és használható alapot nyújt a hely specifikus kezelési stratégia alkalmazásához.

Blackmore (2000) által megállapított trendelemzés hasznos eszköz a tér és időbeli változékonyság jellemzésében.

Blackmore és munkatársai (2003) hat év terméshozam tér és időbeli változékonyságát, a kapcsolatukat a talaj és növény paraméterekkel vizsgálták. Az eredmények alapján javaslatot tettek a táblán a kezelési zónák kiválasztására a termésadatokat 1998-tól 2003. évig gyűjtötték dGPS-szel ellátott kombájnnal a három táblán. A normális évi csapadékmennyiség átlagos értéke 700 mm, az átlagos hőmérséklet 10 °C körül volt. Az éves napsugárzás átlagosan 11740 KJ/m<sup>2</sup>. Kukoricát termesztettek, kivéve 1999-ben őszi búzát vetettek.

A termésadatok mellett 2000. és 2003. évben megvizsgáltak különféle talajparamétereket. Az állandó adatokat 40 méterenként

gyűjtötték, így 7 adatpontot kaptak hektáronként. A talaj jellemzők változékonyságát leíró információkat a felső talajréteg 0-30 cm közötti 10 talajminta alapján kapták. A talajmintákban elemezték a talaj típusát, P, K, Mg, pH és szerves anyag tartalmát.

A termésadatokat a rácsadatok képzésével minden gyűjtő pont körül elemezték. A földrajzilag meghatározott termésadatok a megfelelő a rácsadatokkal a növény és talaj paraméterekkel korreláltak. Pearson féle korrelációs együtthatókat kiszámították az évenkénti termések között. Magas korrelációt találtak az 1999. és a 2003. évi termésadatok között ( $r= 0,9$ ). Többváltozós regresszió analízissel meghatározták a talajparaméterek és a szemtermés közötti kapcsolatot. Az eredmények szoros korrelációt mutattak, ami a vizsgált paraméterek hatását mutatja. A talaj és növényparamétereket kiválasztva megbecsülték a szemtermés hozamának időbeli és térbeli változékonyságát, javaslatot tettek, hogy ez az információ hogyan használható a hely specifikus termelési kezelési zónájának azonosításában.

Kravchenko és Bullock (2000) bemutatták, hogy a lineáris elemzés önmagában nem ad jó függvény modellt a termés változékonyság magyarázatára. Jóval összetettebb modellt használtak a kapcsolat leírására. Ez a DSSAT 4.0 termés modell az adathalmazok vizsgálatával próbálta meghatározni a térbeli változékonyságot a táblán belül. A modell térbeli eszközeivel megbecsülte a termés adatok és a limitáló tényezők közötti kapcsolatokat. A termelési modell eredményei alapján, a kezelési

zónák a magas és az alacsony termés zónák figyelembevételével világosan körülhatárolhatók és a paraméterek meghatározhatók.

## **2.5 A térbeli struktúra hatása az interpolációs módszerek pontosságára**

A talajtulajdonságok pontos térképezése kritikus eleme a sikeres hely specifikus mezőgazdaságnak. A talajminták száma és az interpolálási módszer, amellyel a diszkrét mintaadatokat folytonos térképbe konvertálják, a pontos térképezés legfontosabb tényezői közé tartoznak. A kutatás lényegi részét a megfelelő mintaszám meghatározásának módszerei alkotják, amit a vizsgált tulajdonság pontosságának a mértéke határoz meg. Bár a mintaszám szükséges a pontos térképhez, sokáig nem kapott elég figyelmet. A minta gyűjtésének és elemzésének költsége gyorsan túlszárnyalhatja a hely specifikus kezelés alkalmazásából származó előnyöket. Ezért, amikor a talajminták optimális számát választjuk, fontos, hogy a mintaszám egyensúlyban legyen a minta költségeivel. A kutatások eredményeként javasolták a 60 m-es mintarácsot, vagy a szabályos 30 m-es rácsot, de szükséges az „elfogadható” pontosságnak megfelelő talajtérkép fejlesztése (Mueller et al., 2001).

A térbeli struktúra fontosságát sok kutató felismerte, de nincs kvantitatív információ a térkép pontosságára, amikor bizonyos mintaszámmal, bizonyos térbeli struktúrára tervezik. Az adateloszlás térbeli szerkezetének leírásánál geostatistikai jellemzőt, variogramot használtak. A variogram paraméterek közül

a térbeli struktúra jellemzésére a *röghatás/küszöb* (N/S) arányt és a térbeli korrelációt alkalmazták (Kravchenko, 2003). Az N/S arány definiálja a kis tartományú változékonyságot. A térbeli korrelációs tartomány definiálja azt a távolságot, amelyen belül a tulajdonság értékei korrelációban vannak. Kicsi N/S és nagy korrelációs tartomány rendszerint jelzi, hogy nagyobb pontossággal lehet megtervezni a tulajdonság térképét.

Tamás et al. (2005) a hagyományos Pearson korreláció mintájára térbeli korrelációs eljárást a talaj pH és Cu tartalmának adatai alapján. Kereszt-variogram elemzésével lehetővé válik, hogy a térben egymással korreláló tulajdonságok mintaszámát csökkentsek, vagy azonos mintaszám megtartásával a térbeli becslés megbízhatóságát növeljék.

A másik tényező, ami hat a pontosságra az *interpolációs módszer*, amellyel a diszkrét minta adatokat folytonos térképbe konvertáljuk. Két interpolálási módszer terjedt el a mezőgazdasági gyakorlatban: az inverz távolsággal súlyozott módszer és a krigelés. Számos tanulmányban Kravchenko és Bullock (1999) összehasonlították a módszerek teljesítményét a mezőgazdaságban. A tanulmányok leginkább a kereszt-ellenőrzést, vagy a jackknife (független teszt adathalmaz) módszert használták az interpolációs módszerek összehasonlítására. A másik ok, hogy az interpolációs módszerek paraméterei (pl. az inverz módszer hatványkitevője, vagy a szomszédok száma), amelyek szükségesek az optimális működésükhöz, nagyon függenek a talajadatok változékonyságától és térbeli szerkezetétől. A leíró statisztikai mutatók: a variációs

együttható (CV), a ferdeség, az eloszlás görbe meredeksége az adathalmaz fontos jellemzői. Elméleti kiindulási pont, hogy a krigelés optimális interpolálási módszer. A korrekt alkalmazása megköveteli a variogram szerkesztésével és a modell illesztésével, a térbeli struktúra pontos meghatározását. Legalább 50-100 minta szükséges, a kívánt pontosságú variogramhoz, amely leírja a térbeli szerkezetet. Ha az elegendő mintaszám elérhető, a variogram számítása és a modell illesztése fárasztó és időigényes. A krigelési számításoknál az adateloszlás térbeli struktúráját a minta variogram segítségével vették számításba, így a másik tényező, ami megszabja a krigelés pontosságát, a variogram modell minősége.

Megbecsülték, hogy a mintasűrűség hogyan hat a különböző térbeli struktúrájú és változékonyságú talajtulajdonságok térképeinek pontosságára; és összehasonlították az inverz távolságok módszerét és az általános krigelést a talajtulajdonságok interpolálásánál

Bár a mintaszerkezet hatása szignifikánsan hat a térkép pontosságára, csak a szabályos rács mintapontjait vizsgálták.

Három különböző változékonyságú talajtulajdonságot vizsgáltak: a talaj szerves anyag tartalma kis változékonyságú CV=12%, a K tartalom képviseli a közepes (CV=40%), a P tartalom a nagy változékonyságot (CV=65%). A változékonyság szintje nagy fontosságú a helyspecifikus kezelésben, mivel a nagy változékonyságú tulajdonságok jobb jelöltek a kezelésre, mint az egyenletes eloszlásúak (Schmidt et al. 2002). Másrészt a nagy változékonyságú tulajdonságok térképei kevésbé pontosak.

Az eredeti 256 mintapontot alapul véve, mindhárom tulajdonságnál adathalmazt szimuláltak, az új adatok halmaza az eredeti adatok statisztikai és geostatistikai jellemzőinek figyelembe vételével készült. Elkészítették a kezdeti szimulált adatok variogramját, és összehasonlították a kívánt variogrammal. A szimulált adathalmazról feltételeztük, hogy reprezentálja a mintatáblát.

Mindhárom térbeli struktúra szerepelt a vizsgálatban a gyenge, a közepes és az erős. Az N/S arányt a térbeli adat struktúrájának jellemzésére használták. Pl. N/S= 0,6 megfelel a gyenge térbeli struktúrának, amikor az adatváltozékonyság 60%-a megmagyarázhatatlan, kis távolságú, véletlen ingadozású. A közepes erősségű N/S= 0,3 ill. a nagy esetén N/S=0,1. Ezek az arányok összhangban vannak az irodalomban a különböző talaj tulajdonságoknál (Mueller et al. 2001).

A térbeli korrelációs tartomány minden szimulációban 97 m volt. Átlagos korrelációs tartomány, amelyet az előzetes minta variogram alapján határoztak meg, az eredeti mérési adatokból. Két kritérium alapján ellenőrizték és hasonlították össze a térképek pontosságát: az átlagos négyzetes hiba (MSE) és a becslési jóság kritérium (G). A hibát kiszámították, mint az aktuális tesztadatok és a térkép becslések közötti különbségek négyzetösszegét. A G értékét az alábbi képlet adja:

$G = (1 - \text{MSE} / \text{MSE}_{\text{átlag}}) * 100\%$ , ahol  $\text{MSE}_{\text{átlag}}$  az átlagos hiba, a tábla átlagos értékét, mint a tesztadatok becslését használjuk. Pozitív G jelzi, a rácsminta adatok interpolálásával kapott térkép jóval pontosabb, a tábla átlagánál. Negatív vagy nullához közelítő

G értéknél a tábla átlagok pontosabb, vagy egyenletesebb becslést adnak a nem mért helyeken, mint a rács mintabecslések.

Különböző lehetőségeket mutattak be a variogram paramétereinek meghatározására, amikor az adathalmaz túl kicsi, vagy túl nagy területen megbízható variogramot produkált. Először bizonyos talajtulajdonságok, mint a talajszerkezet, topográfiai vagy termékenység információkat hasznosítanak a variogram paraméterek meghatározásában. Másodszor az átlagos variogramokat fejlesztették ki a különböző talajtulajdonságok térbeli struktúrájának közelítő becslésére. A közelítések pontosságát erősen korlátozta a térbeli struktúra változása tábláról táblára. Harmadszor, a számítógépes szimulációs eljárások fejlesztésével az optimális mintaszám és hely meghatározása vált lehetővé. További kutatások szükségesek, hogy megvizsgálják ezen lehetőségek alkalmazhatóságát különböző táblák különböző tulajdonságainál.

A pontbecslés nagyon függ az eredeti minta területétől, míg a blokkbecslés megengedi, hogy az eredetinél nagyobb területre történjen becslés. A blokkbecslés megkísérli megbecsülni a változó súlyozott átlagát bizonyos blokk fölött (pl.3x3) a becslési pont környezetében. Az átlagolás miatt a blokkhatás, a krigelési variancia gyakran kisebb, mint a pontbecslési variancia és az interpolált felület is kiegyensúlyozottabb.

Ha minden hibát nem tudunk kiszűrni az adatbázisból, Whelan et al. (2001) javasolja a blokk krigelést.

*Összefoglalva:* A talajparaméterek térképeinek tervezett pontossága erősen függ a térbeli szerkezettől. Minél erősebb a térbeli korreláció, annál pontosabb a térkép. Erős és közepes térbeli szerkezet esetén, a CV értéktől függetlenül, a talajtulajdonság viszonylag pontosan térképezhető kisszámú mintahely esetén is. Gyenge struktúránál pontos térképet csak intenzív mintázatnál határozhatunk meg. A véletlenül választott tesztadatoktól függően a térkép pontosságához használt becslési kritérium pontos értéke lényegesen változik. Ezért, legalább különböző tesztadatok szükségesek a döntő következtetések megtételéhez, az optimális mintaszerkezet és az interpolálási eljárás kiválasztásánál. Ha a variogram paraméterek ismertek előző mintákból vagy kiegészítő információkból, akkor a krigelés előnyösebb az inverz távolság módszerénél. Az inverz módszer jól használható olyan tulajdonságoknál, ahol kicsi az adathalmaz, a variogram paraméterek nem ismertek és nagy távolság van a rácspontok között.

## **2.6 A térbeli termésadatok becslése és azok bizonytalansága**

A terméstérkép egyike a helyspecifikus növénytermesztésben használt sikeres technológiáknak. A táblán belüli változékonyság ismerete segít azonosítani, milyen kezelés kell az egyedi táblarészeken. A térképező rendszerek már több különböző növényre elkészültek. A tipikus térképezési rendszer magában foglalja a termés monitort és a pozicionáló berendezést. A termés monitorok szemes terményeknél szenzorokat használnak a szemek



és a nedvesség mérésére. A nagy tömegű terményeknél, mint a burgonya, a töltött cellákat elhelyezték a gyűjtő szalag alatt és meghatározták a termés súlyát.

A térképezési rendszer hibája több különböző forrásból származik. A hibák két lényeges (jelentős) forrása a változó aratási szélesség, az aratási pozíció és a terméshozam észlelése között eltelt idő. Változatos technikát fejlesztettek ki ezeknek a hibának a csökkentésére (Blackmore és Marshall, 1996; Arslan és Colvin, 2002).

A termésadatok pontjai diszkrétnek a térben, mindenegyves adatpont a körülötte levő terület átlagtermését reprezentálja. Ezen pontok eloszlása szabálytalan a táblán, változik az aratási sebességgel. A szemléltetés és az elemzés megköveteli, hogy a szabálytalan eloszlású pontokat interpolálni kell szabályos rácsban. Ez a folyamat az adatblokkolás.

A térképező programok csak végterméket, térképet készítenek, anélkül, hogy jeleznék azok pontosságát. Világos, hogy a ritka adathalmazból valószínűleg nem készítenek olyan pontos térképet, mint a sűrűbb adatoknál. Han et al. (2004) kutatásának célja, hogy kifejlesszenek hatékony termésadat blokkoló eljárást a térkép generálására és a térkép pontosságának értékelésére.

A javasolt blokkoló eljárás magában foglalja: 1) a termés adatpontok konvertálását fix távolságú intervallumba; 2) a mozgó átlag algoritmust használva megbecsüli a cella értékét, amikor hiányzó adatok vannak; 3) a geostatisztikai algoritmussal megbecsüli a cella értékét, amikor nincs elég adatpont, de a cella

szomszédjai ismertek; 4) a bizonytalanság értékeli minden cellaérték becslésénél.

A becslések eredményeként kapott érték a cella igazi értékének csak a közelítése. Minőségileg ismerjük a különbséget a becsült és a tényleges cellaérték között. A bizonytalanság kapcsolatban van ezzel a különbséggel. Más szóval, a megbízhatóság, a konfidencia, a pontosság ugyanúgy használatos kifejezések.

A bizonytalanság értékelésénél a különféle faktorok hatását, mint a becslési hibát meg lehet határozni. Másik kézenfekvő tényező a mérési pontok száma (mérési területen az adatpontok a cellán belül). Feltételezve, hogy a mérési hiba elhanyagolható, növelve a mérések számát, csökkenteni lehet a becslési hibát. A másik tényező, amelynek talán a legjelentősebb a hatása a becslési hibára, a hozam eloszlás természete. Megbízhatóbb becslést kapunk egyenletes eloszlás esetén, mint szabálytalan viselkedésnél.

## **2.7 A talajtulajdonságok elemzése**

A területhasználat minden formájának szociális, ökonómiai, ökológiai hatása általánosságban a talaj különféle tulajdonságaitól függ, és e tulajdonságok mind térben, mind időben változékonyak. A sikeres területhasználushoz éppen ezért megfelelő előrejelzések szükségesek a kapcsolatos intézkedések hatásairól – adott helyen és időben. Így, logikusan az előrejelző modell három típusára van szükség (De Gruijter et al. 1997):

- talaj-eloszlási modell – a tervezési régió belül az adott hely vonatkozásában előrejelzi a állandó talajváltozókat

- talaj-folyamat modell - a tervezési perióduson belül, az adott időpontra vonatkozó dinamikus talajváltozók előrejelzése a permanens változók adataiból
- talaj-hatás modell – adott területhasználattal kapcsolatos intézkedések hatásainak előrejelzése azok érvényességi tartományán belül a permanens és dinamikus változók adataiból

A területhasználat sikerének titka, más tényezők közt, a használt előrejelző modellek minősége. Mivel az egyikből származó output a másik számára input, fontos szempont hogyan sikerül ezeket a modelleket egymásra hangolnunk. A folyamat és hatás modellezésében a mechanikus szimulációs modellek bevezetése és széleskörű alkalmazása volt a legnagyobb újítás. Részletesebb és specializáltabb modellek folyamatban lévő fejlesztése az előrejelzési képesség jelentős növekedéséhez vezetnek. Fontos itt megjegyezni, hogy ezek a modellek tipikusan az aggregáció alacsony szintjén jeleznek előre (gyökér zóna, talajréteg, talaj)

A talaj eloszlás modellezés területén is újítások történtek, mint a morфомetrikus osztályozási rendszerek bevezetése, a légi felvételek megfigyelése/kiértékelése, távérzékelés, továbbfejlesztett mérési technikák, statisztikai mintavételi technikák és egyes esetekben geostatisztikák, hogy egyes talajváltozók szintvonalas térképeit megkaphassuk.

Ha a folyamatos talajosztályok térbeli eloszlását oly módon ábrázolják, hogy az megőrzi a folytonosságát, az eredményt folytonos talajtérképnek nevezzük.

Mielőtt a választanánk a folytonos és a nem folytonos osztályozás között, figyelembe kell vennünk, hogy az osztályozás mindig információk elvesztéséhez vezet. Meg kell említeni, hogy vannak olyan alternatív talaj eloszlás modellezési stratégiák, amelyek teljesen kikerülik az osztályozást. Így inkább az a megválaszolendő kérdés: ha a mintapontokon/helyeken mért állandó talajváltozók értékei adottak, mi a megfelelő stratégia arra, hogy előre jelezzük a célváltozókat (dinamikus talajváltozók vagy területhasználati hatások) olyan pontokra, amelyeket nem tartalmaz a minta. Ez lefedheti az egész területet és ebben az esetben az előrejelzések egy térkép formájában jelennek meg (pl. a földhasználat egy adott fajtájának alkalmassági térképe

A célváltozók előrejelzését a folyamat – vagy hatásmodelleknél a rövidség kedvéért „*becslésnek*” nevezzük.

Egy általános helyzetben a következő körülményeket feltételezve:

Több állandó talajváltozónk van:

- A térbeli eloszlásról szóló kiegészítő információk és az elsődleges ismertek megszerezhetők, pl. egy már létező talajtérkép, látható tájképsajátosságok, távérzékelési képek, digitális terepmodellek formájában.
- Az állandó talajváltozók és a célváltozók közti kapcsolatok, mint ahogy a talajfolyamat modell és/vagy a talajhatás modell mutatja, általában nem lineárisak.

Öt különböző fő stratégiát különböztetünk meg:

1) *A becslés és interpoláció* stratégiának két szakasza van.

*Értékbecslés:* megbecsüli a célváltozók értékét a mintapontokban az állandó változók e pontokban mért értékeiből.

*Interpoláció:* előrejelzi a célváltozók értékeit az előrejelzési pontokban a mintapontokban megbecsült értékek alapján.

Ezt a stratégiát alkalmazták Bosma et al. (1994). Ezzel a stratégiával, a hibák, amelyek a becslésnél keletkeznek, a modell vagy folyamat hiányosságai miatt továbbterjednek az interpolációs lépésnél.

A stratégia hátránya, hogy kevésbé támogatja a rendelkezésre álló segéd információk és előzetes ismeretek használatát az interpolációs lépésben, mint más stratégiáknál. Ennek oka, hogy a segédinformációk és az előzetes ismeretek gyakran szorosabb kapcsolatban vannak az állandó talajváltozókkal, mint a bizonytalan dinamikus változókkal vagy többtényezős földhasználati hatásokkal. Mivel az állandó talajváltozó csak a helytől függ, kisebb erőfeszítést igényel, ha megszerezzük az információkat és ismereteket a térbeli eloszlásokról, mint a dinamikus változóknál és földhasználati hatásoknál, amely függ az időtől és más nem talajtényezőktől is. Reméljük, hogy ez a hátrány csökkenthető, fontos minimalizálni az interpolálási hibát.

A következő 4 stratégia jobb lehetőségeket kínál a segéd információk és az előzetes ismeretek használatára, mivel ezekből interpolálják az állandó talajváltozókat.

2.) *Interpoláció és becslés* is két lépésből áll:

*Interpoláció:* előrejelzi az állandó változókat a nem mért pontokban, a mintapontokban mért értékekből.

*Értékbecslés:* megbecsüli a célváltozók értékét az előrejelzési pontokban az állandó változók e pontokban előre jelzett értékei alapján.

Ezzel a stratégiával, az előzővel fordított lépésekkel, a hibák az interpolációból erednek és szaporodnak a becslési lépésben. Példát találunk Heuvelink (1993) munkáiban, aki először interpolálta szántóföldi termés tömeg és nedvesség tartalmát és kiszámította a hervadásponthoz tartozó nedvességtartalmat az interpolált értékből. A következtetések változatosak: azt találták, hogy relatív eredményesség függ a minta méretétől.

A stratégia hátránya, hogy nagyszámú változó interpolációja a térbeli változékonyság nagyon nagy terjedelmű modelljét igényli. Feltételezzük például, hogy 20 változót interpoláltak függetlenül általános krigelési módszerrel. Variogramként csak két paraméterrel, már 40 paramétert becsültek az adatokból. A legtöbb esetben ez túlparaméterezettséghez vezet.

3.) *Rendszerezés, interpoláció és becslés* a stratégia 3 lépésből áll:

*Adatcsökkentés rendszerezéssel:* az eredeti állandó változókból kisebb számú mesterséges összevont változókat definiál a rendszerezési technika segítségével (fő-komponens analízis, faktoranalízis vagy többdimenziós skálázás) helyettesíti a mintaadatokat az összevont változók értékeivel a mintapontokban.

*Előrejelzi* az összevont változókat a becslési pontokban a mintapontok meghatározott értékeiből.

*Megbecsüli* a célváltozók értékét az előre jelzési pontokban az összevont változók e pontokban mért értékeiből.

A hátránya, hogy a rendszerezési technika korrelációs mátrixon alapszik, feltételezve a linearitást, az adatstruktúra rendszerint nem lineáris a többdimenziós alakban. Valószínűnek tartjuk, hogy az adatszerkezet eredményesebben összevonható az osztályozásnál, ha az osztály középpontok a tulajdonságtér többdimenziós területén helyezkednek el. Ezt szolgálja a következő stratégia:

#### 4) *Éles osztályozás, interpoláció és becslés.*

*Éles osztályozással* csökkenti a változók számát, egymást kölcsönösen kizáró talajosztályokat szerkeszt az állandó változókkal kapcsolatban: minden mintapontnál azonosítja az osztályt, amelyikhez tartozik.

*Előrejelzi* az osztályokat a becslési pontokban a mintapontok által meghatározott osztályokból.

*Megbecsüli* a célváltozókat az előrejelzési pontokban az e pontokban becsült osztályokból

A becslést elvégezhetjük az adatkiterjesztés közbeiktatásával.

(a) kiterjesztés megbecsüli az állandó változókat az előrejelzési pontokban az osztályokból

(b) megbecsüli a célváltozókat az előrejelzési pontokban az állandó változók ezen pontokban becsült értékeiből.

Ennek a stratégiának számos alkalmazása ismert, ez

hagyományosan használt a talajvizsgálatban és táblaértékelésben. Ebben az esetben a becslési pontok lefedik az egész vizsgált területet, a talajosztályok interpolációi ábrázolhatók, mint térképek, és a változó jelentősége alkalmas rangsor az adott a földhasználati formában.

A stratégia hátránya, hogy ugyanazt a becslést készíti el az ugyanazon osztályhoz tartozó összes pont számára, figyelmen kívül hagyva az osztályokon belüli változatosságot. Ez nem megfelelő párosítás egyrészt a diszkrét talajeloszlási modell aggregációs szintje és másrészt a talaj folyamat modell vagy hatás modell között. A hibás párosítás megoldása és az adatcsökkentéshez az osztályozás következetes végrehajtása szükségessé teszi, hogy megváltoztassuk a hagyományos éles vagy nem folytonos osztályozást, fuzzy halmazokon alapuló folytonossá.

##### *5.) Folytonos osztályozás, interpoláció és becslés*

*Folytonos osztályozással* csökkenti a változók számát, folytonos osztályokból álló talajosztályokat szerkeszt az állandó változók kategóriájában, minden mintapontnak kiszámítja az osztályhoz tartozó tagsági függvényét.

*Interpolációval* előrejelzi az osztály tagsági értékét a becslési pontokban a mintapontokban számított tagsági értékekből

*Megbecsüli* a célváltozókat a becslési pontokban az ezen pontokban az osztály tagsági értékeiből

Ez a megközelítés kombinálja a rendszerezésen és az éles osztályozáson alapuló stratégiák előnyeit. Amíg a szokásos éles



osztályozásnak éles korlátai vannak a tulajdonság térben, folytonos osztályoknak nincsenek. Az osztályok tagsági értékei folytonosan változnak, innen a nevük is. Ez nem feltételezi, a földrajzi térben a folytonos osztályok tagsági értékei maguk is folytonosak. Az állandó változók „váratlan” átmenetei a táblán is fennállnak, és az ilyen átmeneteket a tagsági osztályok lépései modelleznek (McBratney és De Gruijter, 1992).

## **2.8 Becslési modellek**

A GPS bevezetésével a mezőgazdasági termelőknek lehetősége van összegyűjteni a hozamadatokat és azokat a talajadatokat, amik előidézik a benne rejlő változékonyságot. Ezen változékonyság kezelése alapkérdés, így a termelési modellek és más előrejelzési eszközök a precíziós mezőgazdasági technológia lényeges elemei.

Két fontos trend van kialakulóban a becslés végrehajtására. A hagyományos közelítés a különböző növények fiziológiájának megismerésén alapszik, és matematikai modellek illesztésével írják le a növény növekedését. Sokféle növénymodellt fejlesztettek ki a gabonaféléknél.

A másik még kevésbé elfogadott közelítés a mesterséges intelligencia alkalmazása a terméshozam előrejelzésére. Ez utóbbi közelítés megköveteli a növény fejlődésére ható tényezők specifikálását és jellemzését ugyanúgy, mint a hagyományos fiziológián alapuló modellek. A különbség a modelltípusok között

a módszerekben vannak, mellyel meghatározzák a különböző input tényezők közötti kölcsönhatásokat.

Az egyik legfontosabb kérdés a precíziós mezőgazdaságban, hogyan kezeli a szántóföldi változékonyságot. Ez a változékonyság a talaj, a klíma és a talajművelés komplex kölcsönhatásának eredménye, és minden évben különböző eredményeket ad (Moore, 1998).

Ennek következménye, hogy a termés becslése nem nagyon pontos, ha csak az előző évi terméstérképen alapszik. (Taylor et al., 2001).

A precíziós mezőgazdaságban fontos kérdés a hozamtérképek és a hozamtényezők közötti összefüggés és korreláció. Az alacsony korrelációk azt a kérdést vetik föl, hogyan készítsünk trágyázási előírásokat pl. a hozam- és a talajminta információk alapján. Számos program számol korrelációt és regressziót a tábla minden cellájánál a generált interpolációból. A másik elemzési lehetőség a hozam adatok korrelációjánál, hogy csak a rácspontoknál számol azzal az információval, amit a laboratórium szolgáltat.

A termés változékonyság komplexitása jó ismert tény. A szakirodalom sok példát mutat be, ahol felsorolnak olyan hozamtényezőket, amelyek a hozam és a talaj termőképessége közötti korreláció alapulnak. Acock és Pachepsky (1997) kifejtik, hogy nem lehet a táblán a változásokat csak kevés tényezővel magyarázni. Diszkutálták a mechanisztikus termés modell használatát, mint hathatós segítő eszközt a komplex talaj/növény/atmoszféra rendszer megértésében. Mivel a növény

viselkedése mostanáig nem teljesen ismert, ezek a modellek korlátosak és tökéletesítésre szorulnak.

A regressziós analízist és más technikákat gyakran alkalmaznak a termés változékonyságot okozó tényezőkkel kapcsolatban. Drummond et al. (1995) a tanulmányukban különféle stratégiákat használtak a regressziós analízisnél és a krigelési eljárásnál. Megbecsülték a rács távolság hatását a térbeli elemzésre és a jövedelmezőségre, a hozam és a korlátozó tényezők és a talaj termőképességének összefüggését.

## **2.9 Mesterséges intelligencia a környezeti adatok kezelésében**

### ***2.9.1 Fuzzy halmazok***

A fuzzy halmazelmélet, amit Zadeh alkotott 1965-ben ideális eszköz a bizonytalanság kezelésére a talajtulajdonságok mennyiségi információinak értelmezésében, különösen akkor, ha ez automatizált a szakértői rendszerben, vagy a földrajzi információs rendszerben (FIR) történik. A talajparaméterek fuzzy halmaza alkalmas, hogy a talajtulajdonságok lehetséges értékeit egységesen definiálhassuk. A tulajdonságok egyedi értékei a fuzzy halmazbeli tagságnak felelnek meg, amelyet a tagsági függvény meghatároz. Értéke 1 (ha a talaj besorolása egyértelmű), 0 (ha besorolása nem egyértelmű), vagy közbenső érték, amikor a folyamatok komplexitása miatt ez az értelmezés nem egészen fedi.

A gyakorlatban, amikor a FIR használója felteszi a kérdést a vizsgált helyről, vagy régióról, a talajtulajdonságok értékeiből próbálják a térbeli előrejelzési módszereivel az aktuális adatokból

megbecsülni. A becslés hibával terhelt, az előrejelzés és a kiértékelés bizonytalansága is jellemzi a FIR kimenetét (Lark és Bolam, 1997). A szerzők megvizsgálták mindkettőt a hely specifikus talajtulajdonságoknál, és javasolták, hogy az előrejelzés és értelmezés bizonytalanságát mindkét esetben prezentáljuk a fuzzy halmazban a „tagsági értékkel”. A tulajdonságok értéktartományát felosztották diszkrét intervallumokra. A tagsági értékek meghatározták, minden intervallumban összegezték a feltételes valószínűséget és a tagsági függvények értéke megadja talajtulajdonság intervallumának középértékét. Ha mind az előrejelzés, mind az értelmezés bizonytalan valamely helyen, akkor a tagsági érték 0 és 1 között lesz. Ha a tagsági érték 0,5 közelében van, ez jelzi, hogy megakadályozza az egyértelmű következtetést. A tagsági értékek használatát két adathalmazon vizsgálták. Bizonyos szituációkban kedvezőbb a két bizonytalan forrást külön kezelni, a másik lehetséges módszernél kombinálja az előrejelzés és az értékelés bizonytalanságát.

A FIR standard használata magában foglalja azokat az eljárásokat, amelyekkel az adatréteg kezelhető, bizonyos összekötő műveleteken keresztül másodlagos adatokat generál. Másodlagos adatokat létrehozhat a FIR output, vagy speciális kérdésekre adott válaszokkal az input. Rendszerint a másodlagos adatok binárisak (igen/nem). Lehet kombinálni más másodlagos adatokkal, ez általában Boole AND eljárás.

Rávilágítottak két bizonytalansági forrásra ebben az eljárásban. Először, elfogadott, hogy az adatréteg térbeli becsléssel

generálódik (osztályozás vagy numerikus interpoláció útján) a megfigyelések ritka halmazából. Az adatnak gyakran van hibatartalma, ami csak növekszik a rákövetkező műveletekkel, és hatással van az outputra. Ez a bizonytalanság lényegében statisztikai. Különböző statisztikai módszereket fejlesztettek ki a térbeli interpolációs eljárások hibájának meghatározására és minimalizálására a gyakorlatban.

Másodszor, bizonytalanság van az adatok értelmezésében a FIR-en belül. A gyakorlatban a mezőgazdászok, a farmerek döntést hoznak, az adott helyen figyelembe véve minden lehetséges tényező hatását.

Burrough (1989) felhívta a figyelmet a modellek mesterséges természetére. Például a széles körben alkalmazott küszöb-modell, amely a klasszikus éles halmazon alapszik: Legyen  $D_t$  a küszöbérték, a tagsági érték ebben a halmazban csak 0 vagy 1 lehet.

$$\mu_A(D) = \begin{cases} 0, & ha D < D_t \\ 1, & ha D \geq D_t \end{cases} \forall D \in X$$

Az  $A$  éles és az  $A'$  fuzzy halmazzal definiálhatjuk a változók összes értékeit. Az előzővel ellentétben a tagsági érték a fuzzy halmazban a változó függvénye:

$$0 \leq \mu_{A'}(D) \leq 1, \forall D \in X$$

Burrough (1989) a talajtulajdonságok tagsági függvényeit egyszerű trapézzal adta meg.

A tagsági függvény különböző alakú lehet, bizonyos megszorítások

vannak. Ross (1995) áttekintést adott arról, hogyan kell kiválasztani és definiálni a tagsági függvényeket. A fuzzy halmaz előnye különösen akkor látszik, ha több változót együtt kezelünk. Vizsgáljuk meg a termőréteg vastagságát és a talaj pH értékét. A küszöböt specifikáljuk úgy, A és B jelenti azt az értéket, amely megfelel a növénynek, így meghatározhatjuk az együttes megfeleltetést

Hagyományos halmazzal és Boole algebrával:

$$\mu_{A \cap B} = \begin{cases} 1, & \text{ha } \mu_A(X) = 1 \text{ és } \mu_B(X) = 1 \\ 0 & \text{egyébként} \end{cases}$$

Fuzzy tagsági függvénnyel:

$$\mu_{A' \cap B'} = \min(\mu_{A'}(X), \mu_{B'}(X))$$

Az  $A' \cap B'$  fuzzy metszet több információt szolgáltat, mint az  $A \cap B$ .

A fuzzy algebra lehetővé teszi, hogy a változót alacsony, közepes vagy magas szintűnek jelöljük meg. Az említett esetekben a 0 és 1 között bármilyen értéket felvehet a függvény-hozzárendelés értelmében. Zadeh (1973) egy általános matematikai keretet hozott létre a nyelvi változók „fuzzy”-ológiai leírására, ahol az általánosan használt alacsony, közepes és magas kategóriákon kívül további kategóriákat javasol felállítani. Hatékonyak tűnik az alábbi kategóriák használata: *alacsonyabb*, *alacsony*, *közepes*, *magas* és *magasabb*. A vizsgálat eredményeit alapvetően befolyásolja, hogy milyen függvénytípussal (tagsági) írjuk le az adott jelenséget. A nyelvi hármas (alacsony, közepes, magas) az alábbi fuzzy függvénytípusok alkalmazását teszi elsősorban

lehetővé.

A *szigmoid* (S-alakú), amely az A, B, C és D kontrollpontok megadása alapján különböző típusú lehet: monoton növekvő, monoton csökkenő, illetve a kettő kombinációja: monoton növekvő, majd hirtelen monotonon csökkenő; monoton növekvő, majd egy hosszabb fázis után monoton csökkenő. A másik függvénytípus a *J-alakú*, amely funkcióját tekintve hasonló az előzőhöz, azonban néhány esetben jobb eredményt ad, mint a szigmoid függvény. Ez szintén lehet monoton növekvő ill. monoton csökkenő, monoton növekvő és hirtelen csökkenő, monoton növekvő és egy platófázis után monoton csökkenő. A két utóbbi esetben a kontrollpontok helyzete az első és az utolsó az inflexiós pontnál lesz, míg a két közbülső kontrollpont a B és a C értéke a platóértékkel egyezik meg. A következő függvénytípus a *lineáris*. A lineáris funkciót is az inflexiós pont helyzete alapján határozhatjuk meg: növekvő, csökkenő, növekvő-csökkenő ill. növekvő platófázis és csökkenő formában, ezt a lineáris fuzzy összefüggést az elektronikus eszközök esetében használják széles körben

A fuzzy halmazelmélet alkalmazása a táblabecslésben gyakran csak a tagsági függvényekre korlátozódik, és gyenge pontjai vannak a tábla minőségének meghatározásában. A fuzzy relációs számítások bevezetnek a probléma megoldásába (Groenemans et al. 1997). Ez az új megközelítés a tábla minőség és a tábla egységek közötti fuzzy relációkon alapulnak. Az ilyen reláció matematikailag leírja a táblaegységek alkalmasságát az

egyres növényeknél. A reláció kalkulások lehetőséget adnak az előzőkből újabb relációk szerkesztésére, kompozíciójára. A táblaegységek kölcsönös összehasonlítása elvezet a legmegfelelőbb egységhez, a tábla indexek számításával.

Braimoh és Stein (2004) a táblabecslésben használja a fuzzy halmaz technikát három lépésben: a tábla jellemzők tagsági függvényét generálja, meghatározza a tagsági értékeket, és kombinálja a tábla megfelelési indexével. A tagsági függvényt hat talajjellemzőre (pH, szerves anyag, talajszerkezet,..) adják meg az ún. szemantikai importálással (SI), amelynek általános alakja:

$$\mu_A(Z) = \frac{1}{1 + \alpha(Z - c)^2}, \text{ ahol } 0 \leq Z \leq \alpha$$

ahol  $c$  az ideális pont, vagy standard index,  $\alpha$  a maximális érték, amit  $Z$  fölvehet.

A fuzzy logikát és az interpolációs technikát alkalmazták a kukorica termesztésénél a talaj termőképességének vizsgálatában. A megfelelési index alacsony volt, összhangban a talajjal, a magas korreláció  $R^2=0,87$  a hozam és az alkalmassági index között magyarázatot adott a csökkenő trendre. A fuzzy technika használata a becslésben elősegítette a kukoricatermesztésre ható fontosabb tényezők kiválogatását

### **2.9.2 Fuzzy osztályozás**

Az első alkalmazások a talaj vizsgálatban és a földértékelésben Burrough (1992) nevéhez fűződnek. A fuzzy logika technikája olyan osztályozáshoz vezet, ami megfelel a



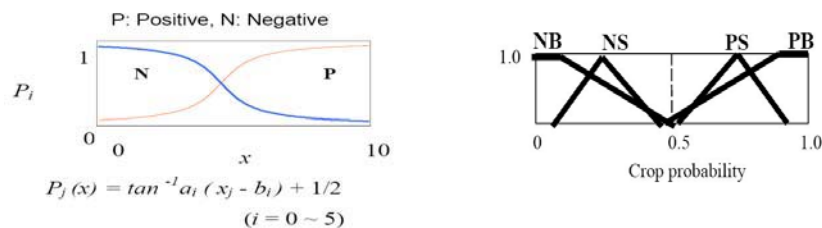
tagság folytonos skálájának. Nehézséget jelent a fuzzy halmazok összekapcsolásában a matematikai következetesség és a nyelvi koncepcióknak megfelelő kapcsolatok meghatározása. McBratney és De Gruijter (1992) kifejtették, hogy számtalan lehetőség van, egy a fuzzy halmazokon alapuló folyamatos talajtérképezés műveleti módszerének meghatározására. Ezek a választási lehetőségek megfelelően besorolhatók 4 osztályba: *talajrétegek folyamatos osztályozása, az osztálytagságok térbeli interpolációja, a célváltozók becslése és térképi megjelenítése*. De Gruijter és McBratney (1988) az elsővel kezdte egy a fuzzy klaszter analízisnek a talajadatokra való alkalmazásáról szóló tanulmányában. McBratney (1992) az osztálytagságok interpolációjának vizsgálatával folytatta. De Gruijter et al. (1997) a metodológiának egy szélesebb perspektíváját veszik, megkísérelnek mind a 4 terület felülvizsgálatát és az egymáshoz való viszonyukat értékelik. Meghatározzák és elemzik

- 1) a fuzzy halmazokon alapuló folyamatos talajeloszlási modellezés módszertanának alapjait, lehetőségeit és korlátait,
- 2) a térbeli interpoláció és a térképi megjelenítés használható módszereinek kiválasztását és ezek valóságos adatokkal történő tesztelését,
- 3) olyan kutatási eredmények azonosítását, amelyek alkalmasak arra, hogy betekintést nyerjünk a módszertan lehetőségeibe és a jövőbeli fejlődésébe.

A fuzzy halmazok alkalmazását a talaj modellezésben pl. Chang és Burrough (1987), Burrough (1992), Triantfilis és

McBratney (1993) és Davidson (1994) és Kosko (1995) vezették be.

Noguchi et al (1998) a fuzzy logikát a termés és a gyomok osztályozására használták. Az érzékelők a vizuális képesség segítségével információt szolgáltatnak a növény állapotáról, fejlettségéről és gyomfertőzöttségéről. A genetikai algoritmust optimalizálásra használták és meghatározták a fuzzy logikai tagsági függvényeket (2.9.1 ábra).



### 2. 9. 1 ábra Az input és output tagsági függvényei

A szántóföldi tanulmányok igazolták, hogy a kifejlesztett módszer képes pontosan osztályozni a termést és a gyomokat a növény egész fejlődési szakaszában

A fuzzy logika hátrányainak egyike, amikor összehasonlítjuk a hagyományos kép osztályozással (K-közép algoritmussal, vagy az egyszerű küszöb értéken alapuló szegmentációval), hogy az osztályozás pontossága függ a tagsági függvény alakjától, és a fuzzy szabályok beépülnek a tervezésnél. Többnyire próbálgatásos módszerrel dolgoztak, később Holland (1975) genetikai algoritmusát alkalmazták, mint általános célú sztochasztikus optimalizálási módszert a keresési problémában. Ezt

az algoritmust a biológiai fejlődés inspirálta és az optimalizációs problémák széles körében alkalmazható. A genetikai algoritmussal optimalizálták a fuzzy logika bemenő tagsági függvényeit, a fuzzy szabályokat és meghatározták kimeneti tagsági függvényeket. A szántóföldi vizsgálatok igazolták (6 különböző fejlődési fázisban tanulmányozták), hogy a kifejlesztett módszer képes pontosan osztályozni a termést és a gyomokat, a szójabab egész fejlődési szakaszában. Megállapítható, hogy a kifejlesztett, kombinált fuzzy logika és genetikai algoritmus megfelelő és hatékony.

A kapott tagsági függvények alakja teljesen különbözik az ember által hagyományosan kreált függvényektől. Különösen a genetikai algoritmus által optimalizált aszimmetrikus tagsági függvény számít kivételesnek a fuzzy logikában. A 2.9.2 ábrán látható a termés és a gyomok egy szegmentálási eredménye, amelyet a fuzzy logika készített. A felső ábra a nyers kép és a szürke szintek, a középsőn a termés, az alsó ábrákon a gyom meglévő valószínűségét. Teljesen világosan látható, hogy korrekten megkülönbözteti a termést és a gyomot.

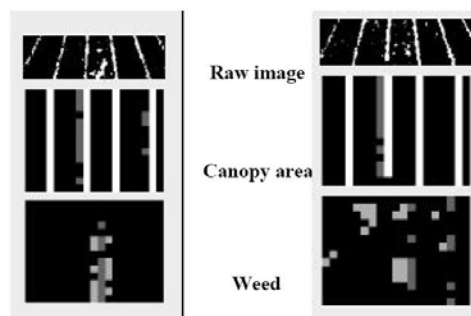


Fig.11 Results using fuzzy logic created by GA.

### **2.9.2 ábra Termés és gyomok szegmentálása**

Dobermann és Obertühr (1997) azt a problémát tárgyalták, hogyan lehet regionális szinten feltérképezni és osztályozni egyedi talajtulajdonságokat ill. komplexebb talajtulajdonságokat, tekintetbe véve azokat a bizonytalanságokat, amelyek az interpolált felületekkel és az osztályozásra használt modellekkel kapcsolatban felmerülhetnek. Vizsgálataikat az egyszerű logikai modelleken alapuló Boole (éles) és fuzzy (folyamatos) osztályozásra korlátozták.

Munkájukban feltérképezték és osztályozták a talajtermékenység kulcs jellemzőit egy Fülöp-szigeteki tipikus öntözéses rizstermesztő területen

1. azonosították azokat a területeket, amelyek a rizstermést befolyásoló talajbeli hiánytól szenvednek
2. számszerűleg meghatározták a tápanyagegyensúlyt a farmerek földjeiben
3. összehasonlították különféle Boole és fuzzy osztályozás megközelítéseket, és megbecsülték a talaj termékenységének állapotát.

### ***2.9.3 Fuzzy krigelés***

A krigelés a térbeli interpoláció legnépszerűbb módszere, de korlátozottan alkalmazható, ha az adatok száma kevés. Amikor a mérések száma nem elegendő a hagyományos krigelés számára, az adatokat kiegészítették pontatlan értékekkel. A fuzzy krigelés használ pontos (éles) mérési adatokat, és pontatlan (lágy) becsléseket (Bárdossy et al. 1989). Az értékek pontatlanságát és

bizonytalanságát kezelik fuzzy halmazokkal.

Két szakaszban van jelen a fuzziness a számításokban. Először az input fuzzy értékek okoznak változást a kísérleti variogramban, így számításba kell venni az éles elméleti variogram illesztésénél. Másodszor a fuzzy input értéket használjuk a krigelés befejező lépéseinél is, így a krigelési eredmények is fuzzy értékek. Leegyszerűsíti a fuzzy becslést és krigelést a FUZZEKS program (Bartels, 1997).

Fuzzy krigelésnél a közelítések három általános osztályát alkalmazzák az ( $\alpha$ -metszés, a kiterjesztés elve, a halmaz-érték statisztika megközelítés).

Nieto (1999) megvizsgálta a fuzzy logika és a krigelés alapkonceptióját, a fuzzy jelző krigelés egyszerű alkalmazásával. A jelző krigelés az adatértékeket ún. jelzőértékké konvertálja: az eredeti értékeket, amelyek meghaladják a választott küszöbértéket, azokat "1" kódolja, a küszöb alattiak kódja "0". Ezeket a jelzőértékeket elemezték a térbeli irányítottságukat leíró kísérleti variogramok sorozatával. A variogramok vizsgálatánál a legnagyobb és a legkisebb térbeli állandóság irányait választották. Ennek a két iránynak megfelelően modelleket illesztettek a kísérleti variogramokhoz. A jelző adatokat a szokásos krigeléssel a variogram modelleket használva a vizsgált rácshelyeken meghatározták a küszöbértékeket meghaladó valószínűségeket. Ez a meghatározás a fuzzy logikában kifejezhető, mint bizonyos küszöböt meghaladó valószínűség tagsági foka 0 és 1 között. Konvertálva a tagsági értékek eloszlását függvénybe, logikai

operátorokkal összekapcsolva más tagsági függvényekkel nagyszerű eszközt kapunk

Sunila és munkatársai (2004) bemutattak egy fuzzy talajtérképezési modellt, amely 500 véletlen mintapont alapján készült. Mind a krigeleési értéket, mind a becslési varianciát fuzzy számként kezelték és a tagsági függvényükkel jellemezték. Megállapították, hogy a fuzzy krigeleléssel létrehozott térkép kielégítő információt ad annak ellenére, hogy a fuzzy talajhatárok nem megfelelőek a részletes elemzésre. További kutatásaikban szeretnék beépíteni a geostatistikai és a környezeti döntési modellekbe

#### ***2.9.4 Fuzzy klaszterezés***

Kiindulópontként szolgálhat az elemzéshez a termés és a talaj klaszterezése, alapja lehet a kezelési zóna definiálásánk. A fuzzy klaszterezés a fokozatosan változó paramétereknél (mind a talaj, mind a hozamadatoknál) jobb besorolást ad, mint az éles csoportosítás. Odeh et al (1992) fuzzy klaszterezést használtak a zóna definiálásánál, amely a hozamot limitáló talajtényezők egyedülálló kombinációján alapszik.

Az éles klaszteranalízisek rendszeres problémáin kívül, mint a kevert objektumok, vagy kiugró értékek kezelése, további nehézséget okoz az adatok bizonytalansága. Az ökológiai adatokat gyakran használják megadott pontossággal, amikor egzakt (pontos) értéket nem tudnak rögzíteni. Ezeket szubjektíven becsülték (pl. a növény-borítottság 20%-os), vagy inter- ill. extrapolálták. Ha éles

klaszteranalízist használunk, úgy ezeket a problémákat nem tudjuk figyelembe venni. A hagyományos klaszterezési módszer, amely pontosan egy objektumot csak egy klaszterbe helyez, nem igazából használható az ökológiai adatok osztályozásánál. A fuzzy klaszterezésnél nincs pontosan definiálva az objektum helye a klaszteren belül, ennek az objektumnak a tagsági függvényértékét szétosztja a különböző klaszterek között. Összehasonlítva a hagyományos klaszterezési módszerekkel a tagsági értékek eloszlása további információkhoz vezet, az egyes objektumok tagsági értéke interpretálja a hasonlósági fokot az objektumok és a megfelelő klaszter között (Salski és Kandzia, 1996)

A fuzzy klaszteranalízis a többváltozós adatelemzés formája, az egyedek halmazának folytonos osztályozását eredményezi az adatok alapján. Megpróbál olyan osztályzást találni, amely optimálisan alkalmazható a rendelkezésre álló adatokhoz. Ez az alkalmazás megvalósítható a célfüggvény minimalizálásával vagy maximalizálásával, amelynek értéke függ az osztályozástól. Változatos célfüggvényt választhatunk, mindegyik a fuzzy klaszter analízis különböző módszerét határozza meg. Fontos megérteni, hogy a fuzzy klaszter analízis nem csak egy technika a folytonos osztályozás megvalósítására, ha ezek egyben a régió reprezentatív adatai, az eredmény az egész területen is optimális. Ez különbözik egy alternatív közelítéstől, amelynél a létező hagyományos talajosztályozást vagy talajtérképet szeretné fuzziifikálni: azaz éles osztálykorlátokat átalakítja folytonossá a távolságarány transzformációjával tagsági osztályba. (McBratney

és Whelan, 1995).

A fuzzy klaszteranalízisnek legalább kétféle típusát különböztetjük meg,

1) az egyedek közti kölcsönös távolságokon alapuló módszerek, pl. FANNY módszert fejlesztette ki Kaufman és Rousseeuw (1990),  
2) az egyedek és az osztályközpontok közti távolságon alapuló módszerek. Az a módszer, amely valószínűleg a legismertebb, a 2. típusba tartozik és fuzzy k-közép vagy c-közép módszernek nevezik. A fuzzy k-közép vonzónak tűnik, mivel fogalmilag egyszerű, széleskörűen alkalmazott módszer, a k-közép általánosítása. Összegezve, az osztályközép a „közép elmélet” numerikus analógiája, így beilleszkedik a legtöbb hagyományos talajosztályozásba.

A fuzzy k-közép alkalmazásában különféle választások vannak: az adathalmaz (egyedek és változók), az adatok lehetséges előzetes transzformációja és az osztályozási paraméterek: k -az osztályok száma, magában foglalja a kívülálló osztályokat, megadja a fuzziness fokát és a kívülálló osztály méretét. Mindezek választása szubjektív, szakértői döntés a hagyományos (éles) osztályozáshoz hasonlóan, bár ezek mind kifejezőbbek és mennyiségileg határozottak.

A fuzzy k-közép gyengéje, hogy bármely egyed, ami körülbelül egyenlő távolságra van minden osztály középpontjától a tulajdonságtérben, körülbelül egyenlő tagsági értéket kap, kivéve azokat, amelyek a középpontban vagy a külső régióban vannak. Ez okozhat nagy becslési hibát a becslési lépésben, amikor a



célváltozók a tagsági értékből származnak

Lark (1998) bemutatta a térben összefüggő osztályok generálásának módszerét több változós esetben. Az adatok fuzzy osztályozásán alapszik, az osztály tagság értékének térbeli súlyozott átlagával. Megállapították, hogy ez hasznos stratégia a térbeli összefüggő tartományok kialakításánál többváltozós esetben.

Goodwin et al. (2003) a hozam és a talaj elektromos vezetőképessége alapján csoportosított, meghatározva a talajzónákat és optimalizálta a talajmintákat a trágyázáshoz. Irányítatlan, véletlen kiindulópont a klasztereknél jelzi a természetesen előforduló csoportokat az adatokban.

### ***2.9.5 Fuzzy szakértői rendszerek***

A fuzzy következtetési mechanizmus és a szakértői rendszertechnika integrációja szolgáltató fejlesztési eszközöket a fuzzy szakértői rendszereknek (Salski, 1993). A fuzzy szabálybázison alapuló modellezés különösen hasznos, ahol nem analitikus kapcsolati modellt vizsgálnak, vagy ahol nem megfelelő az adatok mennyisége a statisztikai elemzéshez. (Salski, 1992; Salski et al. 1996). A szakértői tudás a modellezés alapja, amely gyakran bizonytalan és pontatlan, a fuzzy logikát használhatjuk a bizonytalan ismeretek reprezentálására.

Zhu (1997) egy modellt mutat be ún. hasonlósági modellt, amely arra szolgál, hogy talaj térinformációkat ábrázoljon a részletesség azon szintjén, ami megfelel a digitális terep analízisből

és a távérzékelő technikákból nyert adatoknak. A modell a fuzzy logikán és egy, a téradatak ábrázolására szolgáló raszter sémán alapszik. Tekinthejtük folyamatos térbeli modellnek. Megadott pontban lévő talajt egy, a tagsági értékek vektora ábrázol, ami leírja a hasonlóság fokait a helyi talaj és az előre megadott talajosztályok/egységek csoportja között. A vektor minden eleme a talaj és az előre megadott talajosztályok közötti hasonlóságot (tagságot) ábrázolja. Így a talajt nem kell besorolni egyetlen/kizárólagos talajkategóriába. A helyi talaj és a tipikus talajkategóriák közti eltérés megőrizhető a vektor váltakozó tagsági értékei által.

A tudásbázisú módszerekkel gyakran pontatlan szakértői adatokat kell feldolgozni, automatikusan merül fel a fuzzy-rendszerekkel való együttműködés gondolata.

Az egyik központi, de egyszersmind legkevésbé megoldott problémát -valós problémák modellezése esetén- az erőforrások korlátozottsága jelenti, azaz hogyan juthatunk közelítő eljárásokkal rövid idő alatt, korlátozott következtetési képességekkel és hiányos információk mellett használható megoldásokhoz. Az ilyen kérdésfeltevés mellett nyilvánvaló, hogy erre nagyon jól használhatók a fuzzy-módszerek. Különösen széleskörű figyelmet érdemel a Berkeleyben megalkotott "Soft Computing" kezdeményezés, melyben fuzzy-technikákat, a neurális háló módszereit, az evolúciós algoritmusokat, valamint néhány probabilitikus eljárást kombinálnak sikeresen (Mitra és Hayashi, 2000).

### ***2.9.6 Neurális háló az ökológiai modellezésben***

Az ökológiai modellezés az elmúlt három évtizedben gyorsan fejlődött. A modellek alkotásánál az ökológusok számos módszert alkalmaztak a numerikus, a matematikai és statisztikai és a mesterséges intelligencia technikáit felhasználva. A neurális hálót eredetileg biológiai függvények modellezésére fejlesztették ki. A neurális háló imitálja az agy tanulási folyamatát, és képes bonyolult nem-lineáris folyamat leírására akkor is, ha az adatok bizonytalanok és zajosak. A tapasztalatból tanulnak oly módon, mint a nem hagyományos számítógépek, és gyorsan megoldanak nehéz számítási problémákat.

A múlt század utolsó évtizedében a neurális hálók kutatása robbanásszerűen megnőtt. Az alkalmazásuk az ökológiában és a környezeti tudományokban a 90-es évek elején kezdődött. Colosanti 1991-ben hasonlóságot talált a neurális háló és az ökoszisztéma működése között, így javasolta a módszer alkalmazását az ökológiai modellezésben. Edwards és Morse (1995) kiemelték, hogy a biodiverzitás modellezésében nagy lehetősége van a neurális hálónak.

A számítástudomány eredményei: a párhuzamos számítógépek, a hierarchikus szerveződés, mint felépítésük alapelve mind közelebb vittek a komplex rendszerek megértéséhez. A neurális hálók képesek azonosítani és megtanulni a mintákat, ahol az input adathalmaz a neki megfelelő céladatokkal korrelál. Tanulás után a hálókat használhatjuk az új és független bemenő adatok alapján a kimenetének előrejelzésére. Lek és Guéan (1999)

megfogalmaztak különböző módszereket, és megvizsgálták az ökológiai adatokkal kapcsolatos megbízhatóságukat. Bemutatják a neurális háló ökológiai alkalmazásait különböző területeken: szárazföldi és vízi ökoszisztémák, távérzékelés és evolúciós ökológia. Bemutattak kétféle gyakran alkalmazott algoritmust: a felügyelt hálót a visszaterjesztéses algoritmussal, és a felügyeletlen hálót a Kohonen önszerveződő leképező algoritmussal

### ***2.9.7 Termésbecslés neurális hálóval***

A fontos nemzeti és nemzetközi gazdasági programokban a termés becslése főleg a stratégiai növényeknél, mint a búza, a gabonafélék, a rizs hosszú ideje az érdeklődés középpontjában van.

A termés nagysága fontos információ a térbeli változatosság vizsgálatánál, valamint a trágyázási döntéseknél. A traktor kezelője minőségileg vizsgálja a termés feltételeket a munka és a műveletek során a szántóföldön; a mezőgazdasági mozgó robot detektálja a termés paramétereit. Noguchi et al. (1998) a termés előrejelzési módszerre fejlesztették ki a mesterséges látást. A mesterséges neurális hálót (3 rétegű hálózat, amelyet alkot a bemenő réteg, a rejtett réteg és a kimenő réteg) használták, hogy megtalálják a függvényt a képminta és a termés paramétere között. A 9 db 20 cm négyzetes „terület” szürke szintjei voltak a bemeneten, a tanuló adatokat véletlenül választották ki a képből. Megvizsgálták a mért és a becsült értékek közti kapcsolatot az egész fejlődési periódusban, az  $r^2 = 0.92$  a tanuló adatoknál. A magas korreláció utal arra, hogy a mesterséges neurális háló - a gépi látás rendszere

használható, mint termésbecslő szenzor. Kombinálva a GPS és a képi adatokat készíthetünk térbeli térképet.

Shearer et al. (1999) bemutatták a neurális háló alkalmazását, a térben változó termés előrejelzését adják. Az adatokat 1998-ban búza betakarításnál gyűjtötték a Kentucky tartományban és tesztelték ezt a koncepciót. Az adatjellemzőket megvizsgálták, adatmodellekbe foglalták, beleértve a termékenységet, az elektromos vezetőképességet és műhold képeket. A 10 meghatározott modelltől 4 modell ígér előrejelzési eszközt az optimális szemtermésre a precíziós mezőgazdaságban. Bár az osztályozó modellek többsége előre jelezte a termésszint átlagát, de kimúlt a termés változékonyság tartományának előrejelzésénél.

A 6. modell, amely tartalmazza az összes adatforrást, ez tűnik a legjobb, átfogó munkának a termés térbeli változékonyságának előrejelzésében, de nem tettek kísérletet az időbeli változás leírására ebben a modellben.

O'Neal és munkatársai (2002) neurális hálóval modellezték a kukorica termésbecslését 1900-1998. közötti adatbázissal. A kukorica fejlődését öt fenológiai szakaszban vizsgálták, csapadék és hőmérsékleti adatokkal való kapcsolatot modellezték. Összehasonlították a regressziós és szimulációs modellezés eredményeivel. A különböző kódolási technikát alkalmazó összeállított modellek átlagos hibája 10,5 – 27,9% között van.

Az utóbbi időben az neurális hálót olyan hatásos eszközzé fejlesztették, amely képes kiszámítani a bonyolult egyenletek és

numerikus elemzések legjobb közelítését. A statisztikai modellek főleg különböző regressziót alkalmaznak a termés számításánál. Ezekhez hosszú-idősorú meteorológiai adatokra van elsősorban szükség. A regressziós modellek hátrányai közé tartozik, hogy explicite le kell írni minden input hatását a terméshozamra.

A modellek másik csoportjának szerkezete a növény fiziológiai állapotainak azonosításán alapszik, és ismerve a dinamikus mechanizmusukat, megfelelő matematikai modellben írja le. Ezek közül a CERES és WHEAT modellek a legfontosabbak és legnépszerűbbek.

Napjainkban a neurális háló alkalmazása, a fuzzy módszerek és a genetikai algoritmus mutatott eredményeket ezen a területen.

Ezen módszerek alkalmazása modellezheti a természeti folyamatokat legalkalmasabban és legnagyobb pontossággal.

Safa et al. (2004) fő célja, megbecsülni a termés mennyiségét néhány nappal, vagy hónappal a betakarítás előtt, a meteorológiai adatokat alkalmazva.

Az 1990-99 közötti években összegyűjtötték a klimatológiai adatokat a búza 11 fenológiai állapotában: a vetés, a csírázás, kelés, 3 leveles, bokrosodás, csillagalakzat, kalászolás, virágzás, tejes érés, viaszérés és teljes érés fázisában. Meghatároztak 11 meteorológiai tényezőt az adott időszakokban: az átlagos napi minimum hőmérséklet, szélsőséges napi minimum, átlagos napi maximum hőmérséklet, szélsőséges napi maximum, teljes napi esőmennyiség, a csapadékos napok száma, a napsütéses órák száma, átlagos napi szélsébség, különleges napi szélsébség, a

napi átlagos relatív nedvesség, a felhasznált napi víz mennyisége. Minden gazdálkodási évben elkülönítve gyűjtötték az adatokat két mátrixba. Összevonás után az így kapott 3D mátrix rendezett alapelemei:  $M_{ijk}$ :

A mátrix, amelynek sorai a statisztikai évek ismétlései (i), a búza minden fenológiai fázisának (j), és az oszlopok a meteorológiai faktorok (k).

A különböző hálókat elkészítették minden állapotra és a hálóparaméter optimális értéket meghatározták a fokozatos közelítés módszerével. A 8 évből 2 évet (20%-t) véletlenszerűen kiválasztottak, ennek a két évnek az adatait használták a modell pontosságának a tesztelésére. A hálóval becsült értékeket összehasonlították a valós adatokkal. Az összeállított modell a következő képességekkel rendelkezett:

1. A búzatermés becsülésének maximális hibája 45-66 kg/ha a teljes érési állapot előtt kevesebb, mint 2 hónappal
2. Minden fenológiai állapot érzékenységének meghatározása, összefüggése a meteorológiai faktorokkal. Megállapították, pl. legnagyobb érzékenység az átlagos napi minimum hőmérséklettel a csírázás, a kelés és virágzás állapotában van, ami hatást gyakorol a búza termésére.
3. A meteorológiai tényezők prioritásának és hatásának meghatározása a növény fejlődésében és terméshozamban. Négy osztályba sorolta a ható tényezőket: az I. osztályba a csapadék teljes és napi mennyisége, az átlagos maximum hőmérséklet került.

### **2.9.8 Neuro - fuzzy-rendszerek**

A mesterséges intelligencia két igen fontos területe a neurális hálózatok és a fuzzy logika, a két módszer szerencsésen ötvözhető, és együttesük kiváló eredményt szolgáltat. Egy fuzzy-rendszer kifejlesztése előtt felmerül a probléma, hogyan lehet a fuzzy mennyiségeket és a fuzzy-szabályokat megfelelő formába önteni. Általános esetben a fuzzy rendszer nem rögtön mutatja a kívánt teljesítményt, hiszen paramétereit még be kell állítani. Erre a „hangolásra” azonban még nem léteznek előírt szabályok és elvek, így a folyamat kezelése tisztán a találékonyságon múlik. Ezen a ponton kapcsolódnak be a neurális-fuzzy-rendszerek. Kihhasználva a neurális hálózatok tanulóképességét a fuzzy-rendszert a feladathoz illesztjük, illetőleg példák alapján elkészítjük. A két módszer kapcsolati formája alapján beszélünk kooperatív és hibrid modellekről. Az előbbieket egy neurális hálózatot használnak, függetlenül egy fuzzy-rendszertől, hogy a paramétereit meghatározzák. A fuzzy-rendszert később a neurális hálózat nélkül is lehet használni, üzemeltetni. A hibrid neurális fuzzy-modellek ezzel szemben egy egységes rendszert képeznek, mely neurális hálózatként és fuzzy-rendszerként is értelmezhető.

A neurális hálózatok tematikus térképezési célú alkalmazásánál hamar felmerül az a probléma, hogy a megkülönböztetni kívánt tematikák számának növelésével a neurális hálózat komplexitása rohamosan nő. A bonyolultabbá vált neurális hálózat kezelésénél felmerül:• a nehézkes és bonyolult taníthatóság,• "merev", nem megváltoztatható hálózatszerkezet,• új,



más jellegű információ figyelembevételére korlátozottan nyílik lehetőség. A problémák közül az első a legtöbb alkalmazónál kulcskérdés, hiszen komplex, sok réteget és neuront tartalmazó hálózat tanítása roppant számítás- és ebből következően erőforrás-igényes. A hatékony Levenberg-Marquardt algoritmus felhasználásával ugyan a paraméterek meghatározása gyorsítható, azonban a memóriai igény megnövekszik (Barsi, 1999).

A tematikus térképezés pontosságának és megbízhatóságának növelése érdekében egyre több forrásból származó adatot használnak a hálóokban: • több időpontú, többféle felbontású digitális űrfelvétel, • légi fényképek, • digitális domborzatmodellből levezethető adatok (lejtés, kitettség, stb.), • korábbi tematikus térképek, • kataszteri információk, • egyéb terepi adatok (talajtípus, földrajzi adatok). Ha felépítünk egy hálót, az újabb adatok kezelése már elkészített neurális hálózatnál csak a hálózat újraépítésével és tanításával megy. Az ismertett nehézségek megoldására a független neuronhálózatok módszerét javasolták. Az eljárás lényege, hogy annyi teljesen független neurális hálózatot hozunk létre, ahány tematikát meg kívánunk különböztetni. Megfelelő transzformációval egyetlen ekvivalens neurális hálózattá kell alakítani, a transzformált – immár egyetlen – neurális hálózat újratanítására nincs szükség, a szimulációja pedig lényegesen gyorsabb, mint a külön-külön hálózatoké.

Ebben a műveletben a fuzzy logika segíthet segítségünkre. Használata révén nincs szükség a szó igazi értelmében vett szakértői rendszerekre, hiszen a döntési szabályok kiválóan

leírhatók fuzzy jelölésekkel és fuzzy szabályokkal. A fuzzy döntéshozási mechanizmus által pedig nagy megbízhatóságú és pontosságú döntés születik.

A módszer használatának két rendkívüli előnye a pontosság növekedése és a sebesség. A fuzzy logikán alapuló döntés további erőssége abban áll, hogy a különböző típusú adat kezelésére nyújt jó megoldást. Az adaptív háló egy számítási keret, amely a fuzzy *if-then* szabályokon, a fuzzy következtetéseken és a neurális hálón alapszik.

A háló alapszerkezete 4 komponensből áll:

- Szabálybázis, amely fuzzy szabályok halmaza
- Adatbázis, amely definiálja a tagsági függvényeket, amelyeket a fuzzy szabályokban alkalmazunk
- Következtetési mechanizmus, amely végrehajtja a fuzzy szabályokon alapuló következtető eljárást
- A tanulási mechanizmus, amely tökéletesíti, vagy naprakész állapotba hozza a közelítő eljárásokat.

A fuzzy szabályoktól és az aggregálási (csoportosítási) eljárásoktól függően különböző fuzzy következtetési rendszereket alakítottak ki.

Az adaptív hálót alkalmazták különféle módokon a variogram függvény tökéletesítésére és a krigelés előállítására. A háló először bemutatja a rendszert és aztán tanulás által tökéletesíti.

Az adaptív háló lényegileg a neuro-fuzzy következtető rendszer hálós alkalmazása, ami univerzális közelítés a bizonytalanság és fuzzy rendszerek számára.

### 3. ANYAG ÉS MÓDSZER

#### 3.1 Mintaadatok gyűjtése

A szántóföldi kísérlet a Nemzeti Fejlesztési Kutatási Program keretében a Biológiai Rendszerek Műszaki Intézetének munkatársai irányításával 2001-ben indult.

A szántóföldi adatgyűjtés a NYME-MTK tangazdaságához tartozó **15,3 ha** területű 80/1. számú tábláján történt a 2001–2005. közötti időszakban.

A precíziós gazdálkodás első szakasz azon információk összegyűjtése, amelyek meghatározzák a növény fejlődési feltételeit, elsősorban a talaj termőképességét meghatározó paraméterek. A talaj jellemzésének kiindulópontja a szisztematikus rács alapján összeállított *helyspecifikus talajmintavétel* volt. A talajmintákat DGPS navigációval, 50x50 m rács mentén, 63 kezelési egységben gyűjtötték, melyek átlagosan 0,25 ha nagyságúak voltak.

A 2001. novemberében gyűjtött talajmintákból teljes körű laboratóriumi elemzést végeztek. Azokat a talajtulajdonságokat vizsgáltuk részletesen, amelyek szükségesek voltak a kijuttatandó műtrágyaadag megállapításához és terméshozammal való kapcsolat feltárásához (Mesterházi, 2004). Így a következő laboratóriumi vizsgálatokat végezték el: Arany-féle kötöttségi szám megállapítását, pHKCl érték, humusztartalom, nitrogén, a foszfor, illetve kálium tartalom meghatározását, és ezek mellett teljes körű mikroelem felmérést. A fenti analíziseket Buzás (1993) által szerkesztett Talaj- és agrokémia vizsgálati módszerkönyve alapján rendszereztük. A talaj mintavételezését 2005. novemberében

megismételték, a meghatározott kezelési egységeken. A minták értékelése a 2001. évben meghatározott előírások alapján történt.

1	2	3	4	5	6	7	8			
							9			
						10	24			
			14	13	12	11	23			
	15				21	22	25			
16		18	19		20	27	26	38		
	17				28	36	37	39		
		30	29		35	36	40	50		
	31				34	41	49	51		
		32	33		42	48	49	52		
			44		43	47	48	52	51	
		45			46	47	53	52	51	63
					55	54	53	61	62	
						59	60	61	62	
					56	58				
						57				

### 3.1.1 ábra a kísérleti tábla 80/1 kezelési egységei

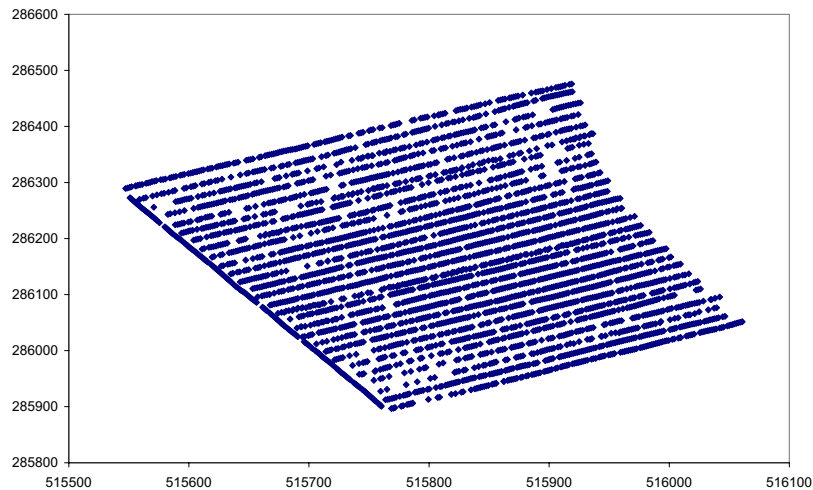
A hozam monitorozásával a 2001. és 2005. közötti években gyűjtöttük a terméshozam adatokat.

- Hozammérés: Agrocom ACT hozammérő rendszerrel, Deutz Fahr M 35.80 betakarítógépre szerelve történt.

A kísérleti táblán 2001-2002. évben kukoricát, 2003-ban tavaszi árpát 2004-ben őszi búzát és 2005. évben ismét kukoricát termesztettek. A vizsgált táblán a betakarítás során a nedves hozamot és az átlagos terméshozamot gyűjtötték.

További információt gyűjtöttek a műveléshez a talajjellenállásról, 20\*20-as rácsokban és folyamatos méréssel (3.1.2 ábra).

- Talajellenállás mérése: a NYME MÉK Környezettechnikai Tanszékén kifejlesztett folyamatos talajellenállás mérő rendszerrel végezték.



**3.1.2 ábra** A talajellenállás mérési koordinátái

A meteorológiai adatokat a Nyugat-Magyarországi Egyetem MÉK Matematika Fizika és Informatika Intézetének Meteorológiai Csoportja által létrehozott adatbank mosonmagyaróvári állomásának 1995 és 2004 közötti napi hőmérsékleti és csapadék adatai szolgálták. Ezek az Országos Meteorológiai Szolgálat által fenntartott meteorológiai megfigyelő és mérőhálózat által gyűjtött adatok

A növényi adatokat, a kukorica többéves adatsorainak elemzésénél az Agrártudományi Centrum Szigetközi Kutatási Központja által összegyűjtött adatokkal dolgoztunk.

A GPS rendszerek az ún. WGS-84 (WGS-84, World Geodetic System 1984) koordináta rendszert alkalmazzák geodéziai referencia rendszerként. A hazai EOVS (Egységes Országos Vetületi) alapú koordinátarendszer és a WGS-84 koordinátarendszer közötti átszámítás utófeldolgozással lehetséges. Az átszámítás akkor adta a legpontosabb eredményt, ha a WGS-84 koordinátarendszer EUREF (EUREF, European Reference Frame) rendszer és az EOVS közötti átszámítást végeztük el. A transzformáció alapja az ún. Bursa-Wolf modell alapján számított paraméterek (összesen 7 paraméter) meghatározása és ezek átszámítása. A modell mind 2D, mind 3D koordináták átszámítására egyaránt alkalmas.

Vizsgálatainkban és elemzéseinkben a transzformált koordinátákkal dolgoztunk. Az EEHHT (2003) transzformációs programmal megbízhatóan transzformáltuk a GPS koordinátákat, megfelelő fájlstruktúrával. A számításoknál a továbbiakban az így kapott, transzformált helykoordinátákat használtuk.

A vizsgált terület minden pontjáról nem rendelkezünk adatokkal, ezért közelítő eljárásokat kell használnunk. Az elemzésnél figyelembe kell vennünk, hogy nemcsak a mérési helyeken (mintavételezés) levő információra van szükségünk, hanem ezek alapján kell majd következtetnünk olyan helyek adataira, ahol nem volt mintavétel. A szakirodalom és a kutatások számos megoldást kínálnak: statisztikai eljárások, genetikus, evolúciós algoritmusok, fuzzy rendszerek, neurális hálók és szimulációs technikák. A felsorolt vizsgálati módszerek alkalmazását vizsgáltuk a precíziós

mezőgazdaság hozamadatainak, a talajtényezőknek a vizsgálatánál és összefüggéseik feltárásában.

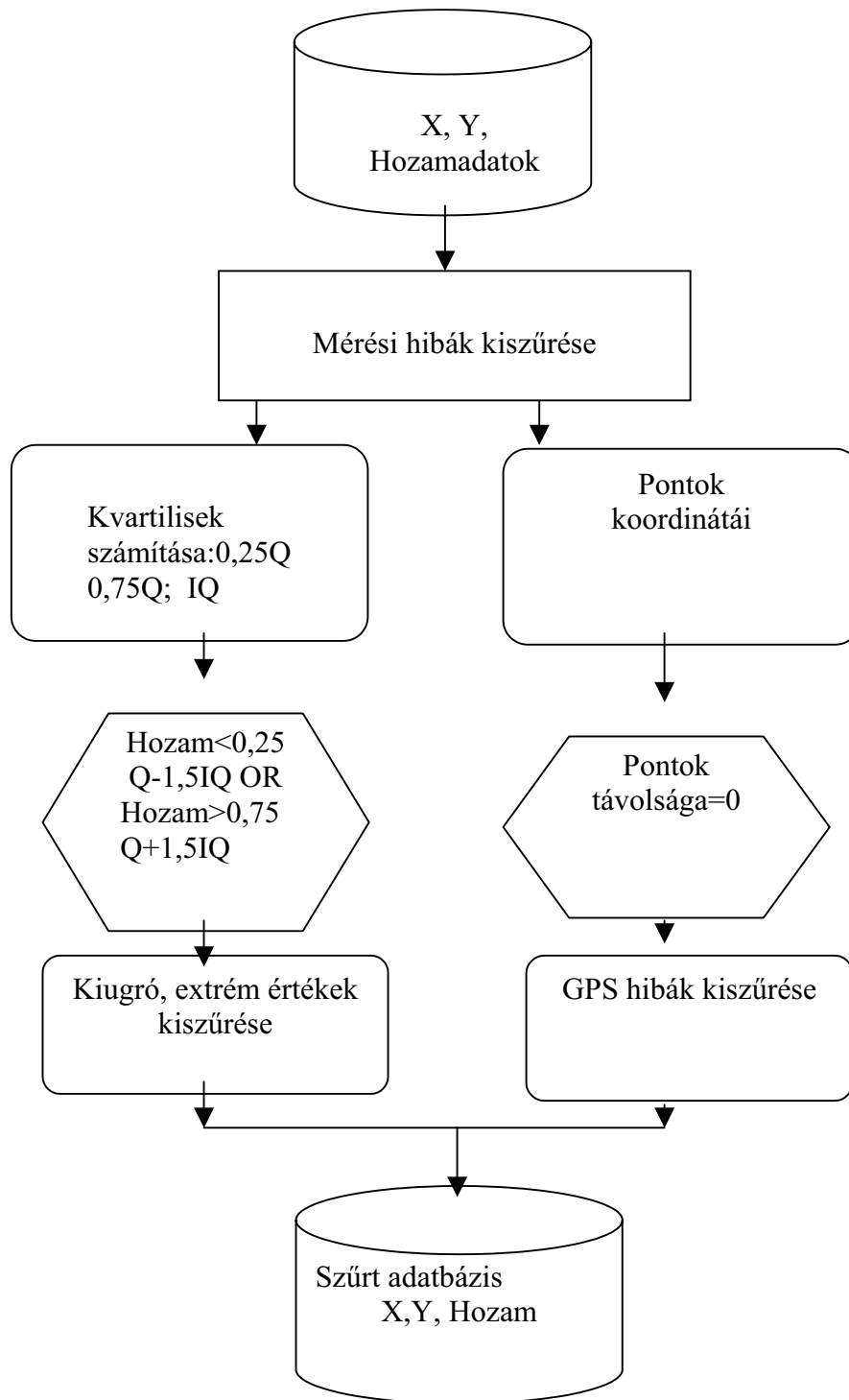
A statisztikai elemzéseknél az Excel és a Statistica 6.1 programokat, a térbeli geostatisztikai vizsgálatoknál a Surfer 8 térképező, a Variowin 2.2, az adatbányászatnál, az összefüggések feltárásánál Matlab Neural Network és Fuzzy Toolbox szolgáltatásait alkalmaztuk.

### **3.2 Statisztikai elemzés**

Gyakori igény az, hogy egy adathalmazt elemei egyenkénti felsorolása helyett néhány jellemző tulajdonságának megadásával jellemezzünk.

Leíró statisztikával meghatároztuk az elhelyezkedést, a szóródást és az eloszlást jellemző paramétereket. Ezek alapján meghatározott szűrőfeltételekkel kiválogattuk a mintavételi és mérési hibákból eredő kiugró és extrém értékeket. Kizártuk a további vizsgálatból a pozicionálás hibájából eredő azonos koordinátájú adatokat is. A további elemzés a szűrt adatbázis alapján történt. Ezek a statisztikai vizsgálatok nem adtak választ a térbeli mintahelyről, a minta térbeli változékonyságáról és eloszlásáról. A geostatisztikai vizsgálat a regionális változók elméletén alapszik. Az ún. Tobler törvény, kimondja, hogy az egymáshoz közelebb levő mintavételi pontpárok különbségeinek térbeli varianciája kisebb és autokorrelációja nagyobb, mint a távolabbi pontpároké.

A statisztikai adatelőkészítés menetét szemléltetik a 3.2.1–2. folyamatábrák.





### 3.3 Térbeli interpolálási technikák

A térbeli modellezés az adatpontok térbeli interpolálását jelenti, képezve a folytonos felületi térképet a becsült értékek (pontok) megadásával a mintaadatok között. A felület bemutatja az adatok térbeli eloszlását.

A térbeli interpoláció olyan eljárás, amely a rendelkezésre álló megfigyelések által meghatározott térség mintavétellel nem rendelkező pontjaiban *becslést* ad a vizsgált tulajdonságok értékére. Arra a kérdésre keressük a választ milyen eljárás a legalkalmasabb az egyes mintavételi pontok laborvizsgálati eredményei alapján a tulajdonság érték leírására.

Adott a vizsgált tulajdonság térbeli adatok formájában (diszkrét pontok), keressük azt a függvényt, amely legjobban modellezi, és megbecsüli a nem mért pontok értékeit.

A legtöbb esetben a tulajdonságot jellemző értéknek egy meghatározott intervallumba kell esnie, vagy egy meghatározott értékkel vett hányadosát kell tekinteni. Pontokra alapozott interpolációnál adva van bizonyos számú pont, amelyeknek helye és értéke ismert; meghatározandók bizonyos előre megadott helyen lévő más pontok értékei. A pontról pontra végzett interpoláció a leggyakrabban végzett térbeli interpolációtípus olyan adatoknál alkalmazzuk, amelyek pontszerű helyeken gyűjthetők. Az ilyen interpolációval kapott *rácspontok* gyakran szolgálnak inputként számítógépes szintvonalrajzoló algoritmusokhoz.

A rácspontok meghatározása után az *izovonalak* (pl. szintvonalak) már átvezethetők közöttük oly módon, hogy minden rácspont-pár által meghatározott egyenes szakaszon lineáris interpolációt alkalmazunk

A térbeli interpoláció alkalmazási területei a földrajzi információs rendszerekben:

- szintvonalak szolgáltatása adatok grafikus megjelenítéséhez
- a felület valamely jellemzőjének kiszámítása egy adott pontban
- az összehasonlításhoz használt mértékegység megváltoztatása, amikor különböző rétegekben különböző adatstruktúrák kerülnek alkalmazásra
- gyakran használatos segédeszköz térbeli döntéshozatali folyamatoknál

### **3.3.1 Interpolálási módszerek**

Az *inverz távolság* és a *krigelés* gyakran használt interpolálási módszerek a hozamadatoknál és talajparamétereknél. A pontosságuk is hasonló, de az utóbbi népszerűbb, mert a cellabecslések megbízhatóságának szintje jól értékelhető. A krigelés elméleti hátterét és a krigelési egyenleteket sok irodalomban megtaláljuk. (Isaaks és Srivastava, 1989; Webster és Oliver, 2001).

Az *inverz távolság súlyozási* módszer a súlyt az *1/szeparációs távolsággal* számolja:

$$W_i = \left( \frac{1}{h_i^r} \right) / \sum \left( \frac{1}{h_i^r} \right) \quad (1)$$

ahol  $h$  a szeparációs távolság az  $i$ -dik pont és a becslés pont között,  $r$  a hatványkitevő ( $r = 1, 2, 3$ ), adja a legnagyobb súlyt a legközelebbi szomszédnak.

A súlyok összege  $\sum W_i = 1$ , így biztosítja a lineáris és torzítatlan becslést

Az inverz távolsággal súlyozott becslés általános képlete:

$$Z^*(X_0) = \frac{\sum_{i=1}^n Z(X_i) \cdot h_i^{-r}}{\sum_{i=1}^n h_i^{-r}} \quad (2)$$

A *krigelés* a legjobb torzítatlan becslés. A változó ismeretlen regionális értékét és ismeretlen varianciáját nehéz meghatározni. Rendszerint jó eszköz, ha a hipotetikus kovariancia, vagy a feltételezett variogram reprezentálja az aktuális rendszert.

A krigelést megkülönbözteti más geostatistikai becsléstől, hogy a hiba varianciájának minimalizálására törekszik:

$$\sigma_r^2 = \frac{1}{k} \sum_{i=1}^k (r_i - m_r)^2, \quad (3)$$

ahol  $\sigma^2$  a becslés hibavarianciája

A képletben  $k$  jelenti a becslések számát,  $m_r$  a becslés hibája átlaga,  $r_i$  az  $i$ . becslés hibája, amely kifejezhető,  $r_i = z^*(x_i) - z(x_i)$ , ahol  $z^*(x_i)$  a változó becslés értéke,  $z(x_i)$  a tényleges érték.

A krigelésben becsljük a pont értékét, ahol nincs minta érték, a mintaátlag súlyozott lineáris kombinációjával:

$$Z^*(X_0) = \sum_{j=1}^n W_j \cdot Z(X_j), \text{ ahol } W_j \text{ a } j. \text{ súly} \quad (4)$$

Keressük a becslési hiba varianciájának minimumát úgy, hogy az egyértelműség (a súlyok összege =1) teljesüljön.

A variancia minimalizálását felírhatjuk:

$$\begin{aligned} \sum_{j=1}^n w_j \gamma(h_{ij}) - \mu &= \gamma^*(h_{i0}) \\ \sum_{i=1}^n w_i &= 1 \end{aligned} \quad (5)$$

$h_{ij}$  a távolság  $x_i$  és  $x_j$  között,  $\mu$  a Lagrange multiplikátor,  $h_{i0}$  az  $x_i$  és azon hely távolsága, amely értékét becsljük. A  $\gamma$  értéket a variogram egyenletekből kapjuk.

A krigelés célja, hogy megtaláljuk a súlyok legjobb kombinációját.

Az első lépés a kísérleti variogram felépítése, vagy a kovariancia modell meghatározása. A variogram az inercia momentum függvénye, a szeparációs távolságtól függ. Hasznos összegző statisztika, amely jelzi, hogyan *szóródtak* szét az adatok a parcellán.

$$\gamma(h) = \frac{1}{2 \cdot N(h)} \sum_{i=1}^{N(h)} [Z(X_i) - Z(X_i + h)]^2 \quad (6)$$

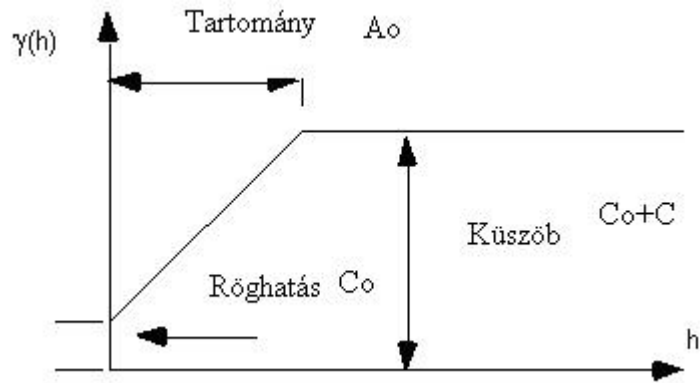
ahol  $N(h)$  a párok száma, ha  $h$  a szeparációs távolság,  $Z(X_i)$  a mért tulajdonság.

Alapfeltételezés: a variogram függvény csak  $h$ -tól függ,  $X$ -től független. A kísérleti variogram fontos jellemzői feltárják a területen jelenlévő térbeli változékonyságot. Meghatározza a krigelési súlyok szétosztását az interpoláció során, és következetesen ellenőrzi az eredmények minőségét. Minden interpolációs és kontúrozási módszer feltételezi a térbeli korreláció valamely típusát, azaz

valamely pont értékét jobban reprezentálják a szomszédos helyek, mint a távolabbi helyek. A variogram elemzésével megpróbáljuk kvantifikálni ezt a tapasztalatot. Általában a variogram növekszik, ahogy a vizsgált hosszak növekednek egy bizonyos hosszig. Végül a kritikus hosszánál, jelölje  $A_0$  azt a pontot, ahol a variancia szintje megáll és ellaposodik. Ez a távolság, az  $A_0$  tartomány, ábrázolja a térbeli függőség határát (Webster és Oliver, 1990). Ha a távolság az adatpontok és a nem mért területek között nagyobb a fenti tartománynál, akkor az adatpontok eloszlása nem használható az interpoláláshoz. Definíció szerint, ha  $\gamma(h) = 0$ , akkor  $h = 0$ . Azonban a variancia értékeket közelítő egyenletes görbe valószínűleg nem megy keresztül az origón, hanem  $\gamma(h)$  értéket közelít, ha „ $h$ ” tart nullához, ez a „*röghatás*”. Ez az  $\varepsilon$  becslése, az eltérés, a térbeli korrelálatlan zaj. A mérési hiba varianciája kombinálódik a mintaköznél kisebb távolságú térbeli változékonysággal, amelyet nem lehet meghatározni és így korlátozza az interpoláció pontosságát. Az  $A_0$  szinthez tartozó  $\gamma(h)$  érték a „*küszöb*”, amely a röghatás( $C_0$ ) +  $C$  tényező, ami ábrázolja az adatok térbeli függősége miatti változékonyság tartományát. A küszöb a variancia maximum modellezése, és ugyanaz a távolsága mint a tartományé.

A 3.3.1 ábra szemlélteti a kísérleti és elméleti variogram függvény szerkezetét.

A kísérleti variogram elemzéséből a modellt a számított varianciához illesztettük, a súlyozott legkisebb négyzetes közelítést alkalmazva.



### 3.3.1 ábra A variogramfüggvény paraméterei

A modell alapfeladata, hogy leírja, hogyan változik a variancia a távolsággal. Számításainkban a *szférikus* (1) és az *exponenciális* (2) variogram modelleket illesztettük (3.3.2 ábra)

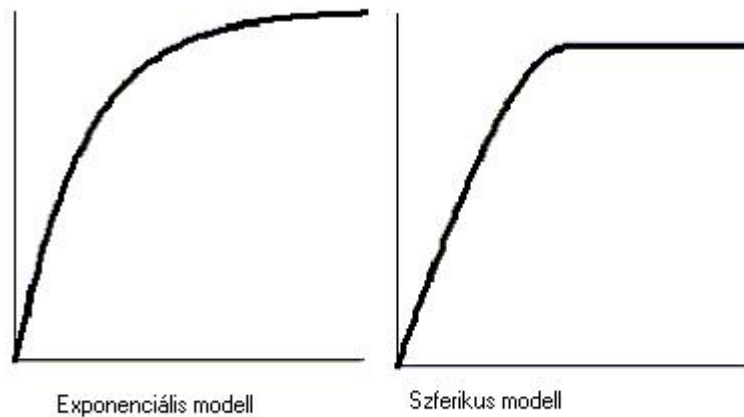
$$\gamma(h) = \begin{cases} C_0 + C \left\{ \frac{3h}{2A_0} - \frac{1}{2} \left( \frac{h}{A_0} \right)^3 \right\}, & \text{ha } 0 < h \leq A_0 \\ C_0 + C, & \text{ha } h > A_0 \end{cases} \quad (1)$$

$$\gamma(h) = C_0 + C \left\{ 1 - \exp \left( - \frac{h}{A_0} \right) \right\}, \quad \text{ha } 0 < h \leq A_0 \quad (2)$$

ahol  $C$  a szerkezeti variancia,  $C_0$  a röghatás variancia,  $A_0$  tartomány és „ $h$ ” a hosszparaméter.

Whelan et al. (2001) javasolják, ha az adatszám 500 fölé emelkedik, célszerű meghatározni a helyi variogramot a táblán belül, mert az egészre kiterjedő variogram korlátosnak bizonyul a helyi térbeli korreláció ábrázolásában. A becslés és krigelés helyi (lokális) variogrammal nyújt lehetőséget a valódi helyi térbeli

változékonyság értékelésében. Másik fontos szempont szól a lokális variogram használata mellett, ha a választott



### 3.3.2 ábra Variogram modellek

szomszédság meglehetősen kevés, a helyi variogram használata közömbösítheti a trendanalízis lehetséges előfeltételeit, és eltávolítja a trendet a variogram becslése és krigelés előtt.

Megvizsgáltuk és összehasonlítottuk a különböző interpolálási módszerek stabilitását. Megnéztük, hogyan változnak a variogram értékek és az interpolált értékek, ha kiveszünk az állományból adatokat.

## 3. 4 A mesterséges intelligencia módszerei az adatelemzésben

A térbeli statisztika a valószínűségen és a klasszikus statisztikán alapszik, az adatok rendszerint közelítők, leírók és bizonytalanok, így nem felelnek meg a valószínűségi elméletnek. A talajjellemzők folytonosságát, a térbeli változékonyság bizonytalanságát jobban leírhatjuk az ún. *mesterséges intelligencia*

eszközeivel. Tanulmányoztuk a fuzzy logika és rendszer, a neurális háló, a genetikus és evolúciós algoritmus eszköztárát és alkalmazását a precíziós mezőgazdaságban az adatbányászat, az értékelés és a döntések előkészítése területén.

### ***3.4.1 Fuzzy logika alkalmazása – fuzzy számok és tagsági függvények***

Röviden összefoglaljuk a vizsgálatainkban szereplő alapfogalmakat.

A karakterisztikus függvény jól ismert a matematikában, minden  $\{X\}$  alaphalmazbeli elemhez 0-t vagy 1-t rendel hozzá. A fogalmát úgy általánosítjuk, hogy az alaphalmaz minden eleméhez a rögzített  $[0,1]$  tartományból rendelhető érték. Az érték nagysága a halmazbeli tagság mértékével arányos. Ezt a függvényt *tagsági függvénynek* nevezzük, az általa definiált halmazt *fuzzy halmaznak*.  $A = \{X, \mu_A(X)\}$  rendezett párokkal definiáljuk a fuzzy halmazt, ahol  $\mu_A(X)$  a tagsági függvényt jelenti -  $\mu_A : X \rightarrow [0,1]$

A gyakorlatban  $X = \{x_1, x_2, \dots, x_n\}$  esetén  $A = (x_1, \mu_A(x_1); x_2, \mu_A(x_2); \dots; x_n, \mu_A(x_n))$  számpárok, ha  $\mu_A(x) = 0$ , akkor a  $(x, \mu_A(x))$  párt elhagyjuk. Így a  $\mu_A(x) = 0$  feltételezi, hogy az  $x$  nem tartozik a halmazhoz,  $\mu_A(x) = 1$  jelzi,  $x$  teljesen beletartozik,  $0 < \mu_A(x) < 1$  jelenti, hogy valamely mértékben a halmazhoz tartozik. A fuzzy szám a valós számok fuzzy részhalmaza, amely *normális* és *konvex* a valós térben, a valós számok általánosítása. A normalitás feltételezi, hogy a tagsági függvény maximuma 1, a konvexitás utal

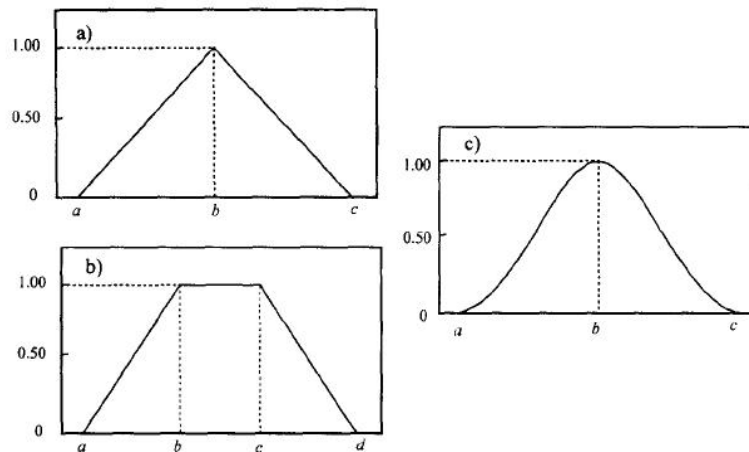


arra, hogy a fuzzy számnak van *növekvő*, *csökkenő* és néha *lapos* része.

Funkcionálisan ezt kifejezzük minden  $a, b, c \in \mathcal{R}$  esetén a következő formulával:

$$\mu_A(b) \geq \min(\mu_A(a), \mu_A(c)), \text{ ha } a < b < c$$

Az egyszerűségük miatt leginkább háromszög, trapéz vagy ehhez hasonló szakaszonként lineáris alakú tagsági függvényeket használunk (3.4.1 ábra).



### 3.4.1 ábra Háromszög, trapéz és Gauss alakú fuzzy számok

Az  $\alpha$ -vágat ( $\alpha$  szint) az egyik legfontosabb fuzzy halmazokkal kapcsolatos fogalom, az adott  $A$  fuzzy halmazon minden  $\alpha \in [0, 1]$  értékre:  $A_\alpha = \{x \mid A(x) \geq \alpha\}$  képlettel adjuk meg.

Egyes kitüntetett vágatok: az  $A$  fuzzy halmaz az alaphalmaz 0-nál nagyobb tagsági értékű pontjainak összességét a halmaz *tartójának* nevezzük, jelölése :  $supp(A) = \{x \mid A(x) > 0\} = A_{0+}$

Az  $A$  halmaz *magja* az alaphalmaz 1 tagsági értékkel rendelkező pontjainak összessége:  $core(A) = \{x \mid A(x) = 1\} = A_1$

A fuzzy halmaz *magasságát* a tagsági függvények legnagyobb értéke adja:  $h(A) = \sup_{x \in X} A(x)$ , ha  $h(A) = 1$ , a halmaz *normális*,

$h(A) < 1$ , az  $A$  halmaz *szubnormális*.

A *halmazműveletek* hasonlóak a hagyományoshoz, a klasszikus halmazműveletek általánosítását adják.

*Egyenlőség*:  $\mu_A(x) = \mu_B(x) \quad \forall x \in X,$

*Komplementálás*:  $\mu_B(x) = 1 - \mu_A(x) \quad x \in X$

Zadeh-féle metszete:  $\mu_{A \cap B}(x) = (\mu_A(x) \wedge \mu_B(x)) = \min(\mu_A(x), \mu_B(x))$

Zadeh-féle uniója:  $\mu_{A \cup B}(x) = (\mu_A(x) \vee \mu_B(x)) = \max(\mu_A(x), \mu_B(x))$

Fuzzy halmazokon értelmezett *aggregációs operátorok* több fuzzy halmaz megfelelő módon történő egyesítése által egyetlen fuzzy halmazt állítanak elő.

A  $h: [0,1]^n \rightarrow [0,1]$  függvényt fuzzy halmazokon értelmezett aggregációs operátornak nevezzük. A jól definiált aggregációs műveletnek az alábbi három axiomatikus feltételt kell kielégítenie:

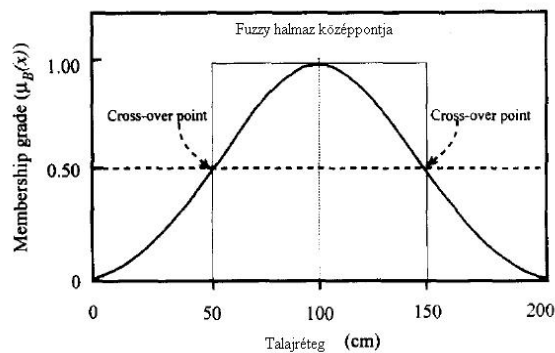
- peremfeltételekre teljesül  $h(0, \dots, 0) = 0$ ;  $h(1, \dots, 1) = 1$ ;
- *monoton növekvő* minden argumentumában  $a_i, b_i, \in [0,1]$  és  $a_i \leq b_i$ , minden  $i \in [1,n]$  esetén  $h(a_1, a_2, \dots, a_n) \leq h(b_1, b_2, \dots, b_n)$
- a  $h$  függvény folytonos.

Definiáljuk a talajréteg vastagságát leíró tagsági függvényt! Feltételezzük, hogy „ $A$ ” a vékony, „ $B$ ” a vastag és „ $C$ ” a nagyon vastag talajréteg fuzzy halmaza. Az 3.4.2 ábra mutatja, hogy 100 cm az teljesen hozzátartozik a B halmazhoz, és tanulmányozhatjuk

az átmeneteket is. A tagsági függvényt felírhatjuk exponenciális vagy Gauss alakban:

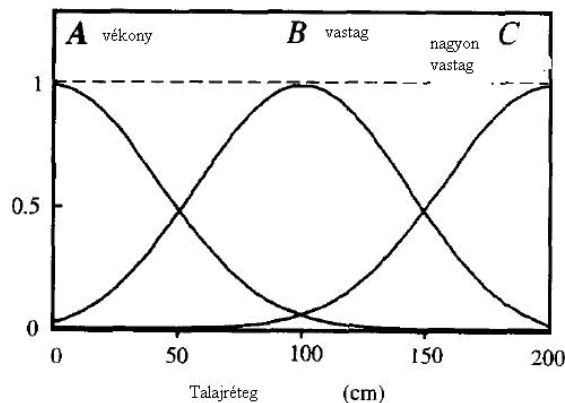
$$\mu_B(x) = e^{-\left(\frac{(x-c)}{1,44 \cdot \sigma^2}\right)}, \quad 0 < x \leq 200, c=100 \text{ és } \sigma = 50 \text{ az alsó}$$

keresztezési pont, ahol a tagsági érték 0,5.



### 3.4.2 ábra Talajjellemző tagsági függvénye

A 3.4.3 ábrán összefoglaltuk a fenti fuzzy halmazokat bemutató tagsági függvényeket ( A, C – asszimmetrikus, B szimmetrikus függvény)



### 3.4.3 ábra Talajréteg tagsági függvénye

### ***3.4.2 Fuzzy krigelés és variogram***

Az interpolációs eljárás fuzzy kiterjesztését a térbeli adatokra fuzzy krigelésnek nevezzük. A hagyományos krigelés módosítása: használ pontos, mért adatokat ugyanúgy, mint pontatlan becsléseket, amelyek szakértőtől származnak. Ha a vizsgált területen hiányoznak adatok, így magas lesz a krigelési variancia. Javíthatjuk a helyzetet, ha fuzzy adatpontokat, szakértői ismeretet fuzzy számok formájában adunk az input adathalmazhoz. Ugyanazt az elméleti variogramot használtuk mind a régi és az új adathalmazoknál, és összehasonlítottuk a két krigelési varianciát.

A fuzzy krigelési eljárás szerkezete éles és fuzzy adatokkal, az elméleti variogram meghatározásának menete látható a 3.4.4 ábrán (Bartels, 1997). A fuzzy becslési és krigelési rendszert (FUZZEKS), a kielii egyetemen kifejlesztett térbeli interpolálási programot használtuk a számításainkban. A programban alkalmazott algoritmust részletesen tanulmányoztuk Bárdossy és munkatársai (1989) összefoglaló munkájában és Bartels (1997) téziseiben és program dokumentációjában.

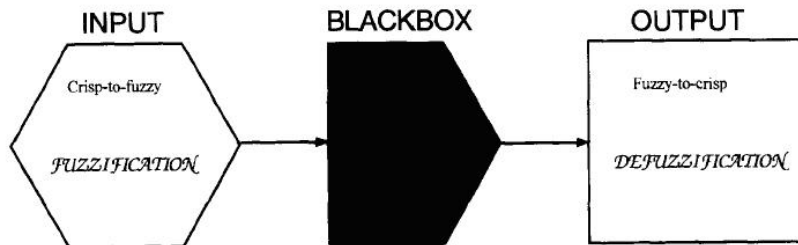
Ez a megközelítés engedélyezi az éles elméleti variogram interaktív illesztését, használva a fuzzy kísérleti variogramot. Az eredményeket térképen ábrázoltuk, a krigelt értékek izovonalaival, a krigelési varianciával. A fuzzy variogram paramétereinek meghatározásánál a választott módszer a részletes elemző modell, és a bizonytalanság együttes kezelését jelenti. Rendszerint éles szoftvereket használunk ezekhez a számításokhoz,  $\alpha$  vágatot alkalmazva. Különböző közelítéseket fejlesztettek, attól függően,

hogyan kezeli a fuzzy számokat és függvényeket. (kiterjesztés alapelvét, halmaz-érték statisztikát) Ebben lépésben a fuzzy tagsági függvényeket képeztük mind a modell, mind a variogram paraméterek választásánál. A fuzzy krigelési becslés és a megfelelő becslési variancia számításánál a kiterjesztési elvet alkalmaztuk.

### **3.4.3 Fuzzy rendszerek**

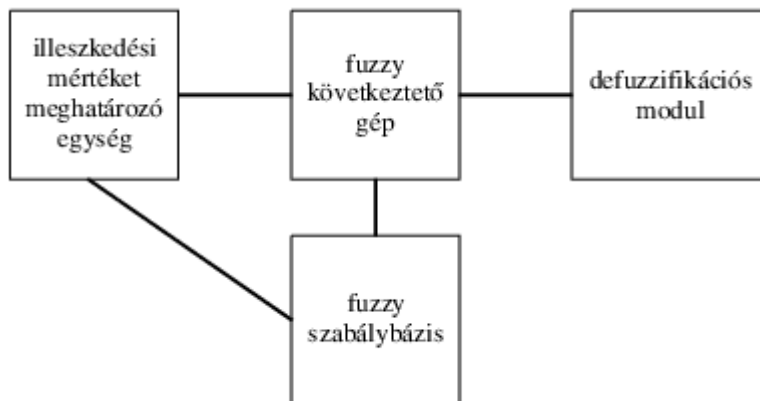
A fuzzy logikát meghatározhatjuk, mint a Boole logika általánosítását, az utóbbit kiterjesztették a részigazságok kezelésére: igaz értékek a „teljesen igaz” és „teljesen hamis” között. Formális definícióban a fuzzy logika strukturált modell-független becslés, ami megközelíti a függvényeket a nyelvi input/output sémákon keresztül. A fuzzy logika a „lágyszabály” nyelvi változókat használja, amely definiálja az igaz értékek folytonos tartományát, vagy a fuzzy tagsági függvényeket a  $[0,1]$  intervallumban, a merev bináris IGAZ-HAMIS döntések és kijelölések helyett, mint a Boole-logikában. A nyelvi input-output kapcsolatok kombinálva a *következtető* eljárásokkal a *fuzzy szabály alapú rendszereket* alkotják 3.4.5 ábra. mutatja a tipikus fuzzy szabály alapú rendszer felépítését ( McBratney és Odeh, 1997).

Három komponensből épül fel: az input a tagsági függvényekkel fuzzifikált bemeneti változókból, az ún. *fekete dobozból*, ami a következtetést végzi a fuzzy műveletek segítségével, a kimeneti komponensből, ami defuzzifikálható.



### 3.4.5 ábra Fuzzy rendszer

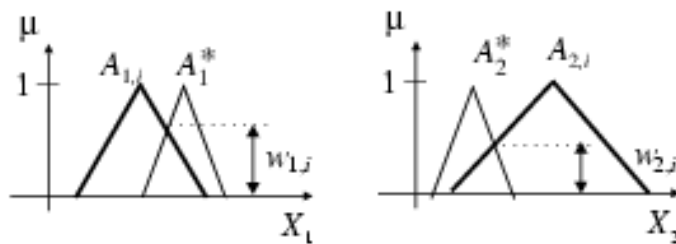
A fuzzy irányítási rendszerek (3.4.6 ábra) lényegi eleme a szabálybázis alapú modell, amely „ha  $A$  bemenet, akkor a kimenet  $B$ ” típusú szabályokból áll. A szabálybázis szerkezetileg hasonlít a szimbolikus ún. szakértői szabálybázisokra, a lényeges különbség az, hogy a szimbólumokhoz rendelt *fuzzy tagsági függvényekkel* írja le az információkat.



### 3.4.6 ábra Fuzzy irányítási rendszer szerkezete

A fuzzy irányítási rendszer további alkotórésze az illeszkedési mértéket meghatározó egység (3.4.6 ábra). Ez az egység a szabálybázis antecedens (előzmény) elemeit ( $A_{1,j}$  és  $A_{2,j}$ ) hasonlítja össze

az aktuális megfigyelés ( $A_1^*$  és  $A_2^*$ ) konkrét értékével, vagy függvényével. A tüzelő szabályoknál - ahol az antecedens és a megfigyelés metszete nem üres- meghatároz egy 0 és 1 közötti fuzzy illeszkedési mértéket ( $w_{1,j}$  és  $w_{2,j}$ ).



**3.4.7 ábra Az illeszkedés mértéke**

A harmadik egység a következtető gép, alapegysége a következtetési algoritmus, mely előállítja a megfigyelésből a következtetést.

A következtetési algoritmus első lépése az aktuális megfigyelés (bemeneti értékek) és a szabályok antecedenseinek illesztése. Minden egyes szabály antecedenshez meg kell határozni a megfigyeléssel való *illeszkedés (tüzelés vagy hasonlóság) mértékét*, melynek alapján meghatározható, hogy az egyes szabályok milyen mértékben játszanak szerepet a konzekvens (következmény) megalkotásában.

Legyen  $A^* (X_1 \times X_2, \dots \times X_n)$  n dimenziós megfigyelés vektor és r db szabály:

$$R_i : \text{Ha } x_i = A_{1,i} \text{ és } \dots x_n = A_{n,i} \text{ akkor } y = B_i$$

Az illeszkedés mértéke a *j*. dimenzióban

$w_{j,i} = \max_{x_j} \{ \min \{ A_j^*(x_j), A_{j,i}(x_j) \} \}$  súlyfaktorral számítjuk,

amely az  $A_j^*$  megfigyelés és az  $A_{j,i}$  szabályantecedens kapcsolatát jellemzi. Az adott szabály alkalmazhatóságát a súlyfaktorok minimumaként adjuk meg:

$w_i = \min_{j=1}^n w_{j,i}$ , megadja, milyen mértékben vesz részt a szabály

a végső következmény előállításában.

$B_i^* = \min(w_i, B_i(y))$ , az adott megfigyeléshez és szabályhoz tartozó következtetés.

Az egész szabálybázishoz tartozó összesített következtetés az egyes szabályokhoz tartozó  $B_i^*$  konklúziók uniójaként áll elő:

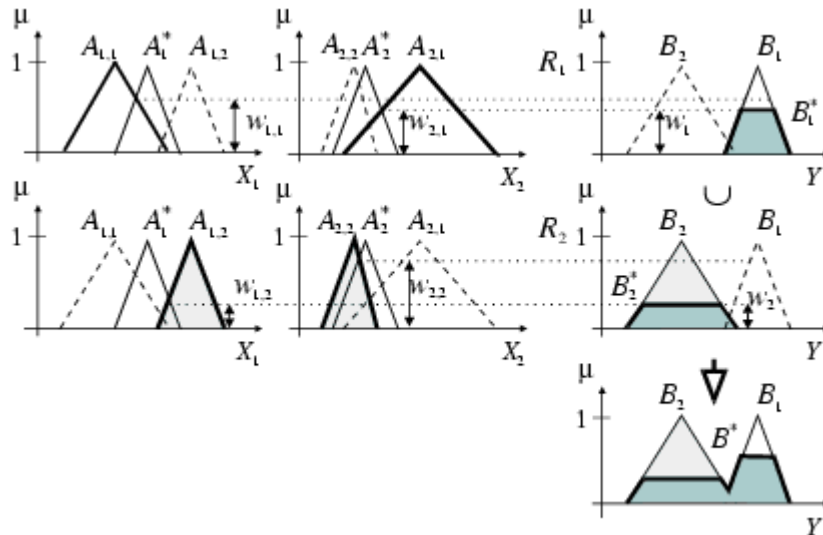
$$B^* = \bigcup_{i=1}^r B_i^* \dots B^*(y) = \max_{i=1}^r B_i^*(y)$$

A végső konklúzió meghatározása az alkalmazott MAMDANI-módszer esetén interpolatív jellegű abban az értelemben, hogy azt több szabály következtetésének egyfajta súlyozott átlagolásával kapjuk, ahol az egyes következtetéseket a bemenet és a megfigyelés illeszkedésének mértékével súlyozzuk.

Lényege, hogy az illeszkedési mérték meghatározása után a kapott súlyokat ( $w_{j,i}$ ) a tüzelő szabályok konzekvenseivel (következmény) kombinálja.

Mamdani következtető eljárást alkalmaztuk a számításainkban, amelyet röviden összefoglaltunk az alábbiakban.





### 3.4.8 ábra Mamdani következtetési algoritmus

A Mamdani-szabályozó input/output-értékei legyenek rendre az  $X_1$ ,  $X_2, \dots, X_n$  és az  $Y$  alaphalmaz elemei.

A függvényt fuzzy relációval adtuk meg, a következő típusú fuzzy szabályokkal dolgoztunk:

$$R^i: \text{IF } (x_1 \text{ is } A_{i1}) \text{ AND } (x_2 \text{ is } A_{i2}) \text{ AND } \dots \text{ AND } (x_n \text{ is } A_{in}) \text{ THEN } (y \text{ is } B_i),$$

ahol  $A_{i1}$  és  $B_i$  fuzzy halmazok,  $x_i$  és  $y$  az input-adatokat- a talajjellemzők értékeit-, valamint az outputot- a termés hozamot- jelölik. A feltételrészben az AND-művelet helyett tetszőleges T-norma művelet alkalmazható, kiértékelése hasonlóan történik, mint az általános fuzzy rendszermodellben. Következtetési eljárásként a max-min művelet alkalmazunk, és tetszőleges defuzzifikáló eljárás választható.

Megjegyzés: Bár a szabályok alakja kissé eltér az általános modell szabályaitól, ha az  $X_1, X_2, \dots, X_n, Y$  valós értékek helyett a

megfelelő nyelvi változókat használjuk, ugyanazt az eredményt kapjuk. Ennek ellenére a szabályozónak egész más a szemlélete: a szabályok nem implikációkat jelölnek, hanem egy részletekben definiált függvényt (Kruse, 1996). A k darab szabály a következő függvény definíciónak felel meg (ahol  $\approx$  jelentése: közelítőleg egyenlő):

$$y = \begin{cases} B_{i1} \text{ ha } (x_1 \approx A_{i1}^1) \text{ és } (x_2 \approx A_{i1}^2) \text{ és } \dots \text{ és } (x_n \approx A_{i1}^n) \\ B_{i2} \text{ ha } (x_1 \approx A_{i2}^1) \text{ és } (x_2 \approx A_{i2}^2) \text{ és } \dots \text{ és } (x_n \approx A_{i2}^n) \\ \dots \\ B_{ik} \text{ ha } (x_1 \approx A_{ik}^1) \text{ és } (x_2 \approx A_{ik}^2) \text{ és } \dots \text{ és } (x_n \approx A_{ik}^n) \end{cases}$$

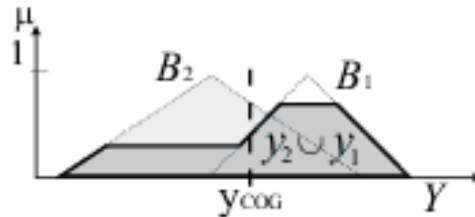
A Mamdani-szabályozó tehát egy sajátos interpolációs eljárást valósít meg.

A következtető gép kimenetén megjelenő fuzzy tagsági függvényt át kell alakítani, defuzzifikálni kell. A negyedik elem a *defuzzifikáló* egység, amely számos különböző technikát alkalmazva a függvény legjellemzőbb középértékét választja ki. A *súlypont módszerrel* (COG) dolgoztunk, amely a  $B^*$  következmény halmaz legjellemzőbb pontjául a súlypontot választja (3.4.9 ábra).

$$Y_{COG} = \frac{\sum_{i=1}^r (Y_i^* \cdot w_i^*)}{\sum_{i=1}^r w_i^*}, \text{ ahol } Y_i^* \text{ a } B_i^* \text{ részkövetkeztetés súlypontja,}$$

$w_i^*$  a súlyozási faktor.

Ez az egyik leggyakrabban alkalmazott módszer. Előnye, hogy trapéz alakú szabályoknál viszonylag egyszerűen számolható. Hátránya, hogy az eredmény szemantikusan nehezen értelmezhető.



#### **3.4.9 ábra Defuzzifikálás COG módszerrel**

A kutatás egyik fontos célja a talajadatok és a termés hozam közötti kapcsolat meghatározása, az optimális szabálybázis összeállítása. A fuzzy szabálybázis generálását és optimalizálását bakteriális algoritmussal végeztük.

#### **3.4.4 Evolúciós algoritmusok- genetikus és bakteriális algoritmusok**

A természetes szelekció "algoritmusa" alkalmas lehet matematikai, vagy programozási feladatok megoldására is. A '70-es években John Holland alkotta meg a genetikai algoritmus(GA) fogalmát (HOLLAND, 1975), melynek lényege a következő:

Adott egy feladat, melynek minden ciklus elején nagyszámú lehetséges megoldását generáljuk, a feladat különböző megoldásait egymással kombináljuk, illetve a megoldások elemein véletlenszerű változtatásokat hajtunk végre. Az így létrehozott "megoldási javaslatokat" teszteljük, majd a valamilyen szempontból legjobbnak ítélt megoldásokat újabb ciklusnak vetjük

alá. Ennek a módszernek az érdekessége, hogy a számítógéppel akkor is sikeresen megoldhatjuk a feladatot, ha eredetileg nem ismerjük pontosan a meghatározandó modellt.

Az evolúciós algoritmusok fogalmát az irodalomban sokféleképpen használják. Gutowitz (1995) például gyűjtőfogalomként, mely magában foglalja a genetikus algoritmusokat, genetikai programozást és az osztályozó/klasszifikáló rendszereket.

A számításainkban használt algoritmusok jellemző tulajdonságai:

- a) iteratívak, ahol minden iteráció (számolásmenet) egy generációt ábrázol; egy véletlenszerűen kiválasztott egyedekből álló populációval kezdődnek, és egy problémát célirányosan oldanak meg;
- b) egy generáción belül az egyes egyedek a következő generációra vonatkozó túlélési esélyei függenek az adott egyed probléma megoldás során mutatott teljesítőképességétől,
- c) az egyedeken kisebb változások (mutációk) keletkezhetnek;
- d) adaptívak, mert megfelelő környezeti feltételek között az evolúciós algoritmusok nagyobb teljesítőképességű (helyesebb problémamegoldásra képes) egyedeket generálhatnak, mint a kiindulási populáció legjobbjai.

A genetikus algoritmusok egyedek egy vagy több elemű halmazából kiindulva a szelekciós és reprodukciós műveletek felhasználásával egy sikeres egyed evolúcióját végzik el. A sikerességet a választott alkalmassági függvény méri. A genetikai algoritmusoknak is vannak gyengéi. Nem használhatók hatékonyan

olyan problémákhoz, melyek jól strukturáltak és melyeket a hagyományos matematikai módszerekkel vagy speciális algoritmusokkal is meg lehet oldani. A genetikus algoritmus szimulált evolúcióval találja meg az alkalmas egyedet, ezt mutatjuk be az alábbi eljárásban:

Alkalmazásakor a következő kérdésekre kell választ keresni: Hogyan kódoljuk a problémát, hogyan reprezentáljuk az egyedeket? Hogyan szelektáljuk az egyedeket? Hogyan reprodukálódnak az egyedek, melyek a megfelelő genetikus műveletek? Mi legyen az alkalmassági függvény, a visszacsatolás kérdése és keresés befejezése?

Az egyedek leírásában a bináris ábécét (0,1) használtuk. A szelekciós stratégia rendszerint véletlenszerű, de a kiválasztás valószínűsége az alkalmassággal arányos. A reprodukció *keresztezéssel* és *mutációval* történik. A kiválasztott egyedeket véletlenszerűen párokba sorolja, és minden párnál meghatároz egy keresztezési pontot. A szülők génjei 1-től N-ig számozottak. A keresztezési pont ( $k \in 1..N$ ) megadja, hogy az egyik utód a  $k$  db gént az első szülőtől, a többbit a második szülőtől örökl. A második utód az  $1..k$  gént a második szülőtől, a többbit az elsőől kapja. Minden egyes gén független, kis valószínűségű mutációval más értékre változhat. Az alkalmassági függvény a feladattól függ, de jellemzője, bemenete az egyed, kimenete egy valós szám, pl. a hiba mértéke

A genetikus műveletek körében három hagyományos műveletet használtunk: a *szelekciót*, *keresztezést* és a *mutációt*. A szelekcióval

több olyan egyedet állítunk elő, amelynek magasabb a túlélési esélye, és kevesebbet az alacsonyabb esélyűekből.

A keresztezéssel két létező egyed (*szülők*) két új egyed (*utódok*) hoznak létre. A keresztezés töréspontjait külön heurisztika szabályozza (egy ponton, két ponton, ciklikus, egyenletes keresztezés), a keresztezési ráta meghatározza milyen gyakran történjen. Az alábbi összeállítás az egy ponton változó *keresztezés* sémáját mutatja be.

1. szülő	[ <b>1000</b> 10011111]
2. szülő	[01101100011]
1. utód	[ <b>1000</b> 1100011]
2. utód	[01101 <b>00</b> 1111]

A mutáció során az egyedeket bitről bitre megvizsgálja az algoritmus, és véletlenszerűen a mutációs rátának megfelelően megváltoztatja. Itt elegendő egy szülő, és az eredmény egy utód. A mutáció segítségével az algoritmus elmozdulhat a lokális optimális pontokból a globális optimum felé.

Az algoritmus akkor áll le, amikor elérte a globális optimumot, célszerű szabályozni megfelelő konvergencia-kritériummal. Ez tovább kombinálható meghatározott futásidővel, vagy meghatározott ciklus idővel (generáció számmal).

A genetikus algoritmusok paramétereit a *populáció méretét*, a *keresztezési és mutációs rátát* nagyon körültekintően kell beállítani.

Ezek a paraméterek fejtik ki a legnagyobb hatást az optimum meghatározásánál, az algoritmus sebességénél és pontosságánál.

Az evolúciós algoritmus egyik legújabb típusa a bakteriális evolúciós algoritmus. Ez a módszer használja szelekciós és mutációs műveleteket. Az optimalizálásnál kiegészült gén transzfer művelettel, ami lehetővé teszi a populációban az információk átalakítását a baktériumok között.

A felsorolt problémákat részletesen tanulmányoztuk a neurális hálónál alkalmazott genetikus, és a fuzzy szimulációnál használt bakteriális algoritmusoknál.

### **3.4.5 Neurális háló**

A térbeli változók összefüggéseinek felderítésénél széles körben alkalmazzák a mesterséges intelligencia fontos modellezési eszközét a neurális hálókat.

A neurális hálók alkalmasak olyan problémáknál, ahol a háttér folyamatok nem ismertek részletesen, és az eredmény megtanítható forma. A következő szempontokat vettük figyelembe:

1. Az adott problémát hagyományos módszerrel nehéz megoldani, és a neurális háló jó alternatíva.
2. Vannak determinisztikus függvények, vagy szerkezeti tulajdonságok, amelyek hasonló kezdeti feltételek mellett hasonló eredményt adnak, általánosíthatók neurális hálóval.
3. A változók közötti *nem-lineáris kapcsolat* és az időben változó tulajdonságok is kezelhetők neurális hálóval.

4. Ha az adathalmazt nem lehet tanítani, vagy numerikusan kódolni, a probléma nem oldható meg neurális hálóval.

A mesterséges neurális háló egy *nem-lineáris leképezés, univerzális* és nagyon rugalmas *függvényközelítő* eljárás. Modellezési eszköz, különösen akkor hasznos, amikor az alapul szolgáló adatok között a függvénykapcsolat ismeretlen.

A neurális háló két objektumból áll: *csomópontokból* (neuronok) és az őket összekötő *élekből* (szinapszis). A bejövő információ (inger, impulzus) egy-egy neuronba érkezik. Az adott információ egy küszöbértéket meghalad, a neuron továbbküld egy jelet a következő, vele összekötött neuronba. Az idegsejtek bonyolult hálózaton továbbítják, míg el nem érik a kérdéses ingerre adandó válaszáért felelős egységet. A tanulás során az intenzíven használt szinapszisok megerősödnek. Az emberi agy tanulásának ezt a mechanizmusát utánozzák le a számítógépes modellek.

Az alkalmazott mesterséges neurális háló jellemzői:

- 1.) az egyes neuronok több más neuronnak is adhatnak impulzust,
- 2.) minden élnek meghatározott az impulzus-átadási határfoka, azaz meghatározott *súlyal* továbbítja az információt,
- 3.) az ingerek különböző mértékűek, és továbbításra nem egy egyszerű küszöbértéket meghatározó kritériumot használ, hanem bonyolult *aktivitási függvények* segítenek megadni a kimenő impulzus mértékét,



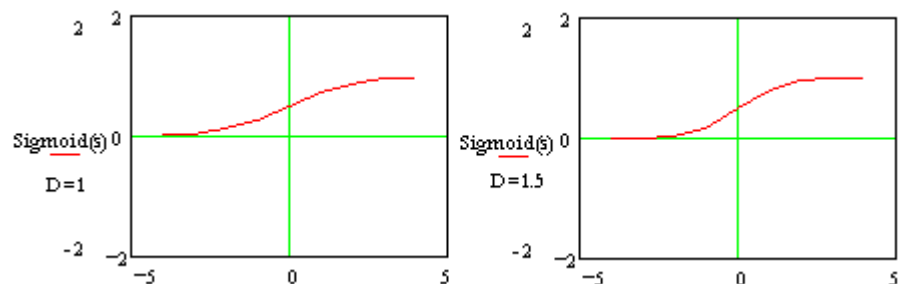
4.) a tanulás mechanizmusát matematikai-statisztikai módszerek szabályozzák (pl. *hiba visszaterjesztéses eljárás* – *back-propagation*)

A neurális háló működését úgy vizsgáltuk, hogy a struktúrát fekete doboznak tekintettük, csak a bemenő és a kimenő jelek kapcsolatát írtuk le. A hálózati topológia, a kapcsolatok iránya és az alkalmazott tanulási módszer megválasztásával eltérő viselkedésű hálókat képeztünk.

A neurális háló alapegysége, a feldolgozó egység – a *neuron*. Ebben az elemben zajlanak a nemlineáris számítási folyamatok. A neuron a bemeneti jeleket összegzi valamilyen módon (additív, multiplikatív), majd egy függvény segítségével transzformálja ezt a jelet.

A különböző típusú neurális háló-architektúrák más és más aktivitási állapotokat engednek meg, megkülönböztetünk *folytonos*, illetve *diszkrét* értékű állapotokat. Ennek megfelelően az aktivitási függvények lehetnek: lineáris, lépcsős, szigmoid vagy tangens hiperbolikus függvény. Vizsgálatainkban a szigmoid függvényeket alkalmaztuk:

A logsigmoid függvény alakja:  $y = \frac{1}{1 + e^{-Dx}}$ ;  $D \geq 1$

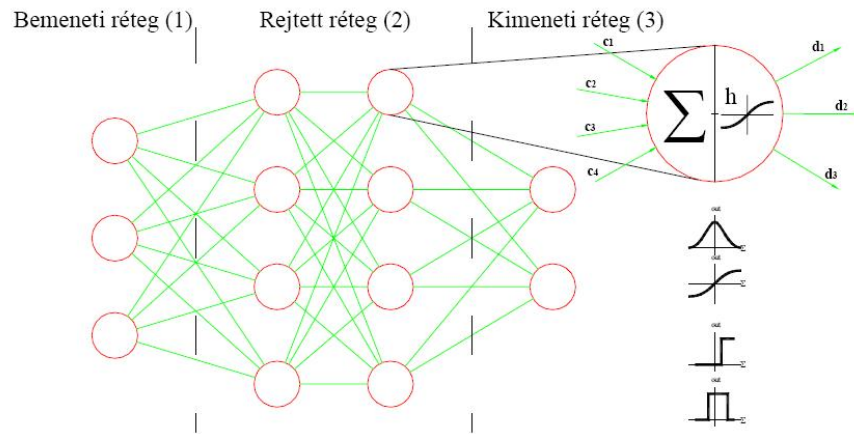


**3.4.10 ábra A háló aktivitási függvénye**

A neurális hálóban a neuronok funkciójuk szerint csoportokat, rétegeket alkotnak. Külön réteg tartozik az adatbevitelhez – *beviteli réteg*-, az eredmények megjelenítéséhez – *kiviteli réteg*-, és további különböző rétegek szervezhetők – *rejtett rétegek*. A rétegekbe szervezett hálókat perceptronoknak nevezzük. A rétegek közti kapcsolatnál a két szomszédos réteg minden elemét összekötjük egymással, teljes kapcsolatot létesítünk. Ez a kapcsolat lehetővé teszi a szétbontva tárolt tudás általánosítását. A 3.4.11 ábrán a számítási modellben alkalmazott hálószerkezetet foglaltuk össze. A szerkezetben jelzett nemlineáris aktivációs függvény (*logsig*) a súlyozott értéket a végső értéké transzformálja.

Az információfeldolgozás iránya alapján - visszacsatolás nélküli – *feed-forward* - hálóval dolgoztunk. Ebben az esetben az információ csak egy irányba, a beviteli réteg felől a kiviteli réteg felé halad.

A visszacsatolósos – *feedback* hálónál a továbbított információ közvetlenül, vagy közvetve visszakerülhet a küldő elemhez



### 3.4.11 ábra A neurális háló egyszerűsített ábrája

A neurális hálóstruktúra különböző kapcsolatait, rétegeit, visszacsatolásait a *súlyokkal* fejeztük ki.

$$\text{Összefoglalva mátrix táblázatban } W(n \times n) = \begin{pmatrix} W_{11}, W_{12} \dots W_{1n} \\ \vdots \\ W_{n1}, W_{n2} \dots W_{nn} \end{pmatrix},$$

amely minden lehetséges kapcsolatot ki tud fejezni.

Ha  $W_{ij} = 0$  – nincs kapcsolat, és  $W_{ij}, W_{ji} \neq 0$  – közvetlen visszacsatolás.

A kutatások szempontjából lényeges, hogy ez a struktúra *adaptív*, azaz válaszával képes a bemeneti adatokhoz tanulással alkalmazkodni.

A neurális háló *tanulása* az a folyamat (algoritmus), amely során a mintapéldák hatására a háló egyes összetevői megváltoznak. Változtathatjuk a kapcsolat erősségét – a *súlyok módosításával*, a feldolgozóelem jellemző *függvényeinek* változtatásával. A megoldás során a közbenső réteget

véletlenszerűen elrendezi, paraméterezi, majd az eredményt az előre megadott értékekhez hasonlítja. Addig finomítja a közbenső rétegeket, míg az általa generált eredmény az előre megadott "minta eredményt" adott pontossággal meg nem közelíti. Az sem jó, ha a hálót "túltanítottuk", mert a mintában jelenlevő hibákat is megtanulja és előállítja.

Hibafüggvényként a háló kimenete ( $O_i$ ) és céleredmény ( $T_i$ ) értékek négyzetes különbségével számoltunk:

$$E = \frac{1}{2} \sum_{i=1}^m (T_i - O_i)^2, \text{ ahol } m \text{ az adatok száma.}$$

A kívánt kimenet és háló által kiszámított aktuális kimenet közti hibát minimalizáltuk. Két fő kérdés volt, a hiba megállapítása és hibát létrehozó súlyok szétosztása. A választ az alábbi megfontolás alapján kerestük:

Jelölje a súlyok függvényében az eredmény hibáját  $E(W) = E(W_{11}, W_{12}, \dots, W_{mn})$  ez  $(n+1)$  dimenziós hiperfelületet képez, és a felület globális minimumpontja a legjobb megoldás. A hibafüggvény minimalizálására a *gradiens menti csökkentés* módszerét alkalmaztuk.

Általános kétrétegű hálózatra felírjuk a hibát („g” jelöli az aktivációs függvényt, esetünkben szigmoid),  $a_j$  a  $j$ -edik csomópont kimenete  $W_{j,i}$  a súly:

$$E(W) = \frac{1}{2} \sum_i (T_i - g(\sum_j W_{j,i} \cdot a_j))^2 = \frac{1}{2} \sum_i (T_i - g(\sum_j W_{j,i} g(\sum_k W_{k,j} I_k)))^2$$

A tanulási szabály a két tetszőleges kapcsolat erősségét kifejező súly változtatási módját leíró algoritmus. Összetett asszociációs feladatoknál az általánosított delta szabály, más néven *back-propagation* szabályt alkalmaztuk.

$\Delta W_{ij} = \eta \cdot O_i \cdot \delta_j$ , ahol  $\Delta W_{ij}$  a  $W_{ij}$  súly módosítása,  $\eta$  a *tanulási ráta*,  $O_i$  az output és  $\delta_j$  a j-dik hiba.

A tanulási folyamatot két fázisban vizsgáltuk – a tanulási fázisban a súlyok módosítása, az eredmények ellenőrzése történt, a tesztelési szakaszban a háló becslési képességét vizsgáltuk. A két fázisban más-más adatokkal dolgoztunk, így a rendelkezésre álló mintahalmazt – *tanuló*, *ellenőrző*- és *teszthalmazokra* bontottuk. A vizsgált halmaz nagysága, az input minta és a véletlen módon kiválasztott részhalmazok is nagymértékben befolyásolták az eredményt.

A tanulási módok kiválasztását, a tanulás célját, a tréningmintákat és a tanulási alapeladatokat külön-külön megvizsgáltuk.

A többrétegű neurális hálózatokra Cybenko (1989) által megfogalmazott és bebizonyított tétele alapján: „*Ha egy neurális háló legalább egy rejtett réteget tartalmaz, akkor tetszőleges folytonos függvény reprezentálására képes*” – A tanulás nem más, mint a paraméterek hangolása annak érdekében, hogy a mintahalmaz adataihoz illeszkedjen, azaz alkalmas legyen a nem-lineáris kapcsolat, regresszió meghatározására:

$$\text{Terméshozam} = f(\text{talajtényezők, időjárási tényezők, kezelési tényezők, véletlen faktorok})$$

## 4. EREDMÉNYEK

### 4.1 Adatok elemzése

A vizsgált 5 évben összegyűjtött hozamadatokat és a mért talajadatokat a technikai előkészítés után mind statisztikai, mind geostatisztikai módszerekkel jellemeztük. Ha a ferdeségi mutató értéke nagyobb, mint 1, vagy kisebb, mint -1, jelzi, hogy a változó eloszlása eltér a normálistól.

#### 4.1.1 táblázat Terméshozamok statisztikai jellemzői

	2001	2002	2003	2004	2005
	Kukorica (t/ha)	Kukorica (t/ha)	Tavaszi árpa(t/ha)	Őszi búza (t/ha)	Kukorica (t/ha)
Minimum:	0,1	0,1	0	0	0
Alsó kvartilis	<b>3,3</b>	<b>4,6</b>	<b>3,5</b>	<b>6,12</b>	<b>9,4</b>
Median:	4,6	6,2	4,1	7,2	10,5
Felső kvartilis	<b>5,9</b>	<b>7,6</b>	<b>4,6</b>	<b>7,92</b>	<b>11,3</b>
Maximum:	14,1	17,4	8,5	10,44	16,9
Terjedelem	14	17,3	8,5	10,44	16,9
Interkvartilis	<b>2,6</b>	<b>3</b>	<b>1,1</b>	<b>1,8</b>	<b>1,9</b>
Átlag	<b>4,591</b>	<b>6,089</b>	<b>3,919</b>	<b>6,829</b>	<b>9,879</b>
Szórás	1,953	2,139	1,078	1,788	2,656
Variancia	3,813	4,577	1,162	3,199	7,056
<b>Variációs együttható</b>	<b>43%</b>	<b>35%</b>	<b>28%</b>	<b>26%</b>	<b>27%</b>
Ferdeségi együttható	<b>-0,05</b>	<b>-0,19</b>	<b>-1,58*</b>	<b>-1,60*</b>	<b>-2,52*</b>

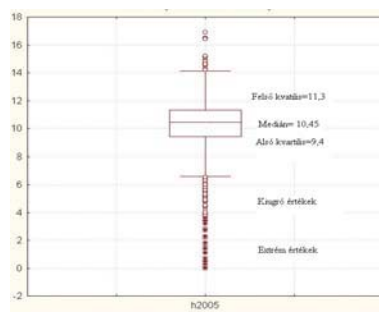
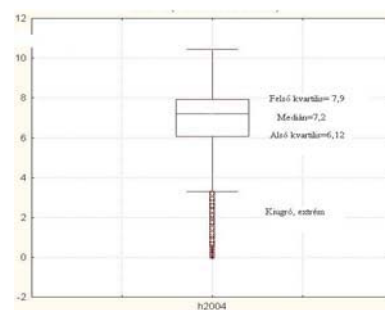
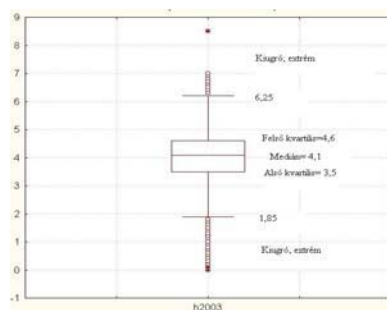
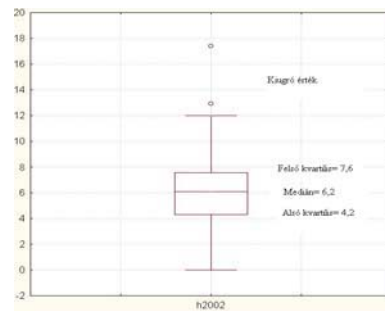
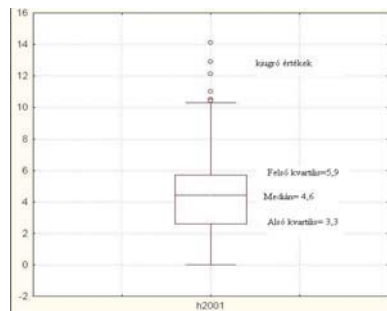
A 4.1.1 táblázatban (\*)-gal jelölt értékek esetében megvizsgáltuk, mi okozta ezt ferdeséget az egyes állományokban. A Box &

Whisker ábrát alkalmaztuk az adatokkal való ismerkedés fázisában. Ez az ábrázolás hasonló, mint az igen széles körben használt *átlag* +/- *szórás*, de annál sokkal informatívabb. Az ábrázolás lényege, a mediántól felfelé és lefelé vizsgáltuk a felső és alsó kvartilisig. Az ezen kívül eső értékeket egy- egy ponttal jelöltük, ezeket kiugró (outlier), ill. extrém értékeknek tekintettük (4.1.1 ábra). Az irodalomban többféle szűrési kritériumot javasoltak a szóródási jellemzők figyelembe vételével: *átlag* ± 2 *szórás*, *átlag* ± *átlag*/2: (Kerry és Oliver, 2003) tartományt vették be a további elemzésbe. A hozam adatok elemzésénél a szűrési feltételt a kvartilisek(Q) vizsgálatával (Robinson, 2004) állapítottuk meg, és a 4.1.2 táblázatban adtuk meg a tartományokat. Az így meghatározott alsó és felső korláttal sikerült kiszűrni a különböző hibákból eredő kiugró és extrém értékeket.

#### 4.1.2 táblázat Szűrési feltételek

<b>Szűrési feltétel</b>	<b>2001.</b>	<b>2002.</b>	<b>2003.</b>	<b>2004.</b>	<b>2005.</b>
<i>Alsó korlát:</i> $0,25Q - 1,5 IQ$	0	0,1	1,85	3,42	6,55
<i>Felső korlát</i> $0,75 Q + 1,5I Q$	9,8	12,1	6,25	10,62	14,15

A kiugró alsó és felső értékek kiszűrése után a ferdeségi együttható is megjavult, mindegyik évben negatív, ez átlag alatti értékek felé mozdult el, de a mutató értéke [0,5;-0,5] intervallumban van, így az adatokat nem kell transzformálni. A hozamértékek gyakorisági eloszlásait mutatják a 4.1.2-3. ábra hisztogramjai.



4.1.1 ábra Hozamاداتok szóródása box & whisker grafikonon



**4.1.3 táblázat A szűrt hozam adatok statisztikai jellemzői**

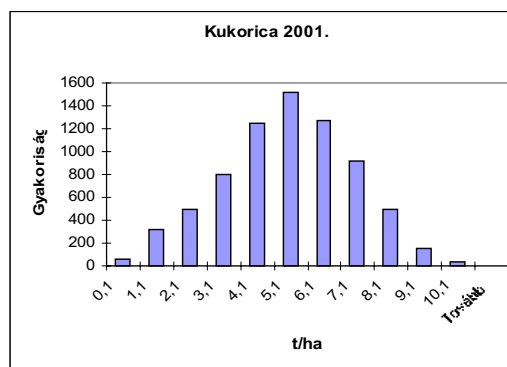
	2001	2002	2003	2004	2005
	<b>Hozam (t/ha)</b>				
Várható érték	<b>4,579</b>	<b>6,088</b>	<b>4,079</b>	<b>7,097</b>	<b>10,463</b>
Standard hiba	0,023	0,026	0,009	0,011	0,014
Medián	4,6	6,2	4,2	7,26	10,6
Módusz	4,6	6,2	4,3	7,5	10,9
Szórás	1,935	2,132	0,754	1,291	1,267
Minta varianciája	3,743	4,544	0,569	1,667	1,606
Csúcsosság	-0,358	-0,403	-0,018	-0,003	-0,085
Ferdeség	<b>-0,10</b>	<b>-0,21</b>	<b>-0,38</b>	<b>-0,44</b>	<b>-0,32</b>
Tartomány	9,7	11,8	4,3	7,02	7,5
Minimum	0,1	0,2	1,9	3,42	6,6
Maximum	9,8	12	6,2	10,44	14,1
Darabszám	7297	6804	7253	13987	8013
CV	<b>42%</b>	<b>35%</b>	<b>18%</b>	<b>18%</b>	<b>12%</b>

Mindegyik adathalmaz normális eloszlásúnak tekinthető.

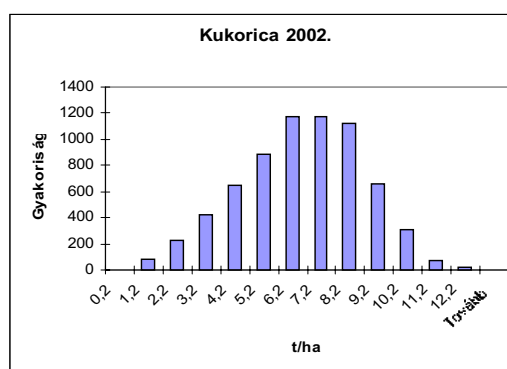
A kukorica esetében 3 év adatsorát tanulmányoztuk, így összehasonlításra, további részletes vizsgálatra is alkalmas, ezért csoportosítottuk az ábrákat.

Három különböző változékonyságú a kukorica hozama a vizsgált években: 2005-ben kis változékonyságú  $CV=12\%$ , a 2002. év képviseli a közepes  $CV=35\%$ , a 2001. évi termés a nagy változékonyságú  $CV=42\%$ . Az árpa és az őszi búza esetében a változékonyság egyenlő  $CV=18\%-18\%$ . A hozamparaméter értékeinek variációs együtthatója is fontos információval szolgál a térbeli vizsgálatoknál.

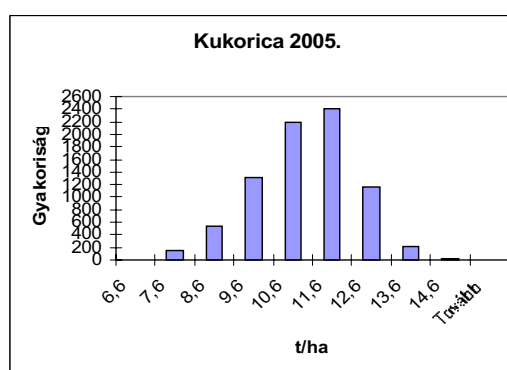
Közök	Gyakoriság
0,1	56
1,1	313
2,1	489
3,1	805
4,1	1250
5,1	1517
6,1	1272
7,1	912
8,1	493
9,1	151
10,1	39



Közök	Gyakoriság
0,2	4
1,2	83
2,2	223
3,2	427
4,2	651
5,2	885
6,2	1170
7,2	1172
8,2	1126
9,2	663
10,2	305
11,2	76
12,2	19

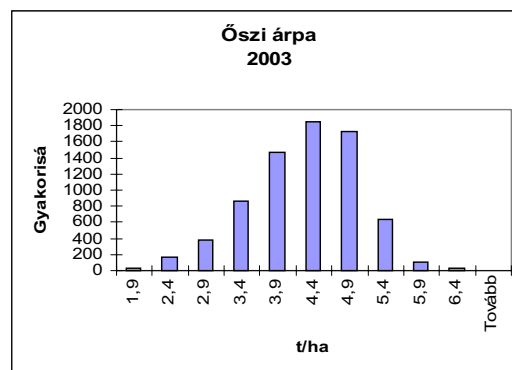


Közök	Gyakoriság
6,6	9
7,6	157
8,6	547
9,6	1311
10,6	2195
11,6	2409
12,6	1151
13,6	216
14,6	18

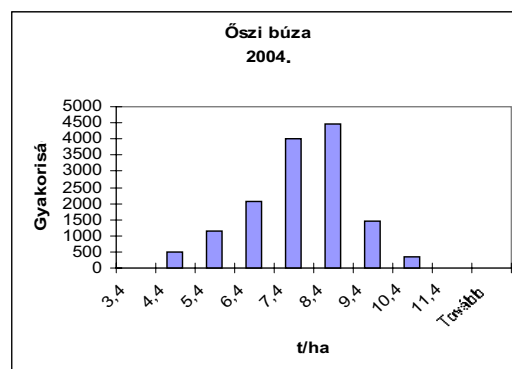


4.1.2 ábra Kukorica termés hozamának hisztogramjai

Közök	Gyakoriság
1,9	28
2,4	173
2,9	378
3,4	860
3,9	1468
4,4	1853
4,9	1722
5,4	638
5,9	109
6,4	24
Tovább	0



Közök	Gyakoriság
3,4	0
4,4	499
5,4	1145
6,4	2054
7,4	3999
8,4	4458
9,4	1469
10,4	357
11,4	6
Tovább	0



#### 4.1.3 A 2003. és 2004. évi hozam adatok hisztogramja

A kukorica hozam adatok varianciánálízise mutatja, hogy a termésátlagok a vizsgált években szignifikánsan különböznek  $p=0,1\%$ -os szinten. Az átlagérték a 2001. és 2002. évi hozamnál  $1,51t/ha$  -ral növekedett, a 2002. és 2005. évi hozamnál a növekedés  $4,37t/ha$ . A 2001. és a 2005. évet összevetve ez a növekedés az átlagtermésben már  $5,88 t/ha$ .

Megvizsgáltuk a talajmintavétel értékelése során kapott fontosabb jellemzőket a 2001. és 2005. év mintavétele alapján. Az 50\*x50 m-es rácson gyűjtött talajparaméterek összegző statisztika jellemzését 4.1.4 és 4.1.5 táblázatokban adtuk meg. A táblázatból kiolvashatók az egyes változók eloszlását leíró mutatók.

A kvartilisek vizsgálatával a szűrőfeltételnek megfelelően a kiugró értékeket a \* -gal jelzett ferdeségű paraméterekből kizártuk, és az eloszlásnak megfelelő adattal pótoltuk. A talaj *K* és *Na* tartalmának eloszlása megváltozott, ezt jelzi a ferdeség értéke: 1,65-ről 0,45-re, ill. 2,18-ről 0,17-re változott.

A talajparaméterek mindkét évben hasonló változékonyságot mutattak. Alacsony CV= 1% és 8% között a talaj pH és kötöttség, közepes CV<30% az összes többi tulajdonságnál.

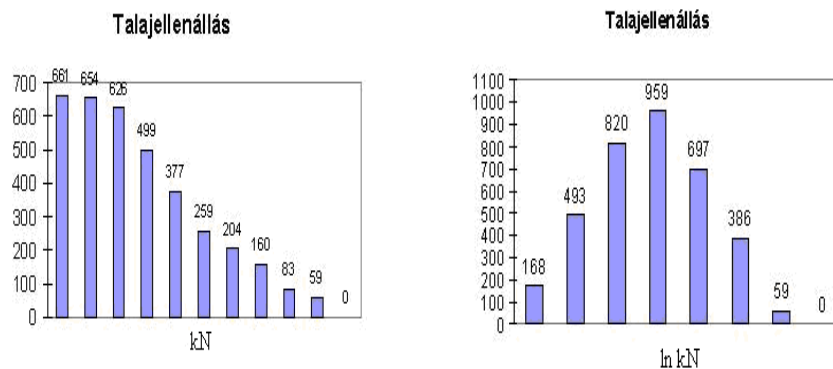
A vizsgált talajtulajdonságoknál megnéztük van-e különbség a két év talajmintái között. Szignifikáns különbséget kaptunk a kötöttség ( $p=1\%$ ), a humusztartalom ( $p=5\%$ ) értékei között, mindkét paraméter értéke növekedett. Az össz N%, a P<sub>2</sub>O<sub>5</sub> és K<sub>2</sub>O értékei is szignifikánsan különböznek ( $p= 1\%$ -os) a két évben, mindegyik értéke csökkent. A Nitrát értéke és a Nitrát N összetevő  $p=0,1\%$ -os szinten emelkedett. A mikroelemek vizsgálatánál a Na kivételével szignifikáns csökkenést mutatnak a 2005. évi adatok (*Mg* és *Cu* -  $p=5\%$ ; *Zn*, *Mn* és a *Fe*  $p= 0,1\%$  szinten).

További vizsgálatot igényelnek ezekben a talajtulajdonságokban bekövetkezett pozitív és negatív irányú változások.

A talajellenállás adatai a szűrés után sem voltak normális eloszlásúak, így ezeket átalakítottuk a szokásos logaritmus transzformációval, és a további elemzéseket ezen az adathalmazon végeztük (4.1.4 ábra). A vizsgált fizikai talajparaméter nagy változékonyságú CV=72%.

#### 4.1.3 táblázat A talajellenállás statisztikai paramétereit

	Talajellenállás (kN)	
	Szűrés előtt	Szűrés után
Várható érték	3,459	3,208
Standard hiba	0,046	0,039
Medián	2,786	2,711
Módusz	1,165	1,165
Szórás	2,765	2,310
Minta varianciája	7,644	5,334
Csúcsosság	3,762	-0,144
<b>Ferdeség</b>	<b>1,481*</b>	0,760
Tartomány	25,475	9,992
Minimum	0,000	0,000
Maximum	25,475	9,992
Darabszám	3682	3582
CV	<b>80%</b>	<b>72%</b>



4.1.4 ábra Talajellenállás adatok – eredeti és logtranszformált

Kiszámoltuk a talajtulajdonságok kapcsolatát leíró és jellemző paramétereiket. A korrelációs együttható (Pearson-féle) az összetartozó érték párok lineáris kapcsolatát jellemző, dimenzió nélküli szám.

A talajváltozók korrelációs mátrixában, az *1. mellékletben* az együtthatók jelzik a változók függőségét, ill. függetlenségét. A *kötöttség* ( $K_A$ ) szignifikáns pozitív korrelációban van a talaj *humusz*, *Na*, *Mg* tartalmával  $p=0,1\%$  szinten. Erős negatív korreláció jellemzi a *kötöttség* és *P2O5*, a *Zn* a *Cu* kapcsolatot ( $p=0,1\%$ ). Pozitív korrelációs kapcsolatban van a *humusz* a talaj *Na*, *Mg*, *Cu* tartalmával ( $p=0,1\%$ ), és az összes *Nitrogén%*- kal  $p=1\%$ -os szinten. A *pHKCl* a erős negatív korrelációs kapcsolatban van a *kötöttség* mellett a talaj *humusz*, a *Mg* és *Cu* tartalmával.

Kiszámítottuk a talajparaméterekre a kovariancia becslését (*2. melléklet*), az együtváltozás mértékét, az alábbi módon értelmeztük:

$$COV(\xi_i, \xi_k) = \frac{1}{n} \sum_{j=1}^n (X_{ij} - \bar{X}_i) \cdot (X_{kj} - \bar{X}_k),$$

ahol az *i.* és *k.* változók

átlagától való eltéréssel számoltunk. Ha létezik a kovariancia, akkor ennek nagyobb, vagy kisebb mértékétől függően létezik egy olyan tendencia, hogy a két vizsgált változó esetén az azonos *j* sorszámhoz a számtani középértől azonos értelemben térjenek el az adatok (az  $X_{ij}$  és  $X_{kj}$ ). Így belátható, hogy a kovariancia a valószínűségi függés egy lehetséges értéke. A pozitív értékek jelzik, hogy a két változó esetén az átlagtól azonos értelemben térnek el. A kovariancia negatív előjele, arra figyelmeztet, hogy

ellentett értelmű együttváltozásról van szó, az abszolút értéke ugyanúgy a függés mértékéről tájékoztat. Megjegyezzük, hogy a függés kis mértéke mellett is adódhat nagy kovarianciaérték, ha nagyok a vizsgált változók ingadozásai. Egyértelmű a kapcsolat a *kötöttség* és a *foszfor* változása között, a nagy negatív korreláció és negatív kovariancia jelzi. Pozitív korreláció és kovariancia jellemzi a *kötöttség* és *Mg*, a *Fe* és *K<sub>2</sub>O* kapcsolatát.

#### 4.1.4 táblázat A talajtulajdonságok statisztikai jellemzői (2001. év)

	<i>pH</i> <i>H<sub>2</sub>O</i>	<i>pH</i> <i>KCl</i>	<i>kötöttség</i>	<i>CaCO<sub>3</sub></i> %	<i>humusz</i> %	<i>összN</i> %	<i>Nitrát</i> ppm	<i>Nitrát N</i>
Várható érték	7,76	7,3	42,6	20,56	2,86	0,14	22	4,84
Standard hiba	0,006	0,009	0,427	0,316	0,039	0,002	0,772	0,173
Medián	7,77	7,3	42	21	2,85	0,15	22	4,95
Módusz	7,78	7,26	42	22	2,5	0,15	23	5,17
Szórás	0,05	0,069	3,392	2,507	0,31	0,018	6,125	1,373
Variancia	0,002	0,005	11,504	6,283	0,096	0	37,512	1,884
Csúcsosság	1,948	-0,82	-0,672	0,469	-0,734	0,463	1,55	1,585
<b>Ferdeség</b>	<b>-0,08</b>	<b>-0,08</b>	<b>0,5</b>	<b>-0,6</b>	<b>-0,18</b>	<b>0,13</b>	<b>0,89</b>	<b>0,90</b>
Tartomány	0,31	0,27	12	12	1,3	0,09	33	7,42
Minimum	7,62	7,15	38	14	2,15	0,1	10	2,25
Maximum	7,93	7,42	50	26	3,45	0,19	43	9,67
CV	<b>1%</b>	<b>1%</b>	<b>8%</b>	<b>12%</b>	<b>11%</b>	<b>12%</b>	<b>28%</b>	<b>28%</b>

	<i>Al P<sub>2</sub>O<sub>5</sub></i>	<i>Al K<sub>2</sub>O</i>	<i>Al Na</i>	<i>Mg(KCl)</i>	<i>Zn</i>	<i>Cu</i>	<i>Mn</i>	<i>Fe</i>
	<i>ppm</i>							
Várható érték	271	91	18	121	3,6	3,48	30,2	24
Standard hiba	8,58	2,021	0,387	3,276	0,055	0,075	1,055	0,576
Medián	293	88	17	115	3,7	3,55	31	23
Módusz	207	85	16	107	3,4	3,55	21	23
Szórás	68,101	16,038	3,075	26,002	0,434	0,599	8,37	4,575
Variancia	4637,81	257,211	9,458	676,107	0,189	0,359	70,063	20,931
Csúcsosság	-1,3	5,16	6,63	3,44	-0,29	0,14	-1,22	-1
<b>Ferdeség</b>	<b>-0,23</b>	<b>1,65*</b>	<b>2,18*</b>	<b>-0,44</b>	<b>-0,1</b>	<b>0,34</b>	<b>0,27</b>	<b>0,1</b>
Tartomány	265	97	18	177	2	2,85	31,5	15
Minimum	123	65	14	18	2,7	2,15	18	17
Maximum	388	162	32	195	4,7	5	49,5	32
CV	<b>25%</b>	<b>18%</b>	<b>17%</b>	<b>21%</b>	<b>12%</b>	<b>17%</b>	<b>28%</b>	<b>19%</b>
Darabszám	63	63	63	63	63	63	63	63

4.1.5 táblázat A talajtulajdonságok statisztikai jellemzői (2005.év)

	pH H <sub>2</sub> O	pH KCl	kötöttség	CaCO <sub>3</sub> %	humusz %	összN %	Nitrát	Nitrát N
Várható érték	7,76	7,33	44,44	20,63	2,73	0,13	44,84	10,1
Standard hiba	0,014	0,013	0,378	0,424	0,034	0,004	1,004	0,228
Medián	7,73	7,32	44	20	2,72	0,13	43,6	9,81
Módusz	7,72	7,32	44	20	2,72	0,12	44,8	10,08
Szórás	0,107	0,105	2,999	3,364	0,271	0,029	7,966	1,807
Variancia	0,012	0,011	8,993	11,317	0,074	0,001	63,456	3,266
Csúcsosság	-0,421	-0,942	-0,059	-0,46	0,527	17,653	1,128	1,18
<b>Ferdeség</b>	<b>0,167</b>	<b>0,02</b>	<b>0,211</b>	<b>0,043</b>	<b>0,362</b>	<b>3,206*</b>	<b>1,044*</b>	<b>1,069*</b>
Tartomány	0,47	0,4	13	14	1,43	0,21	39,2	8,82
Minimum	7,5	7,11	39	14	2,13	0,09	32,4	7,29
Maximum	7,97	7,51	52	28	3,56	0,3	71,6	16,11
Darabszám	63	63	63	63	63	63	63	63
CV	<b>1%</b>	<b>1%</b>	<b>7%</b>	<b>16%</b>	<b>10%</b>	<b>22%</b>	<b>18%</b>	<b>18%</b>

	Al P <sub>2</sub> O <sub>5</sub>	Al K <sub>2</sub> O	Al Na	Mg(KCl)	Zn	Cu	Mn	Fe
	ppm							
Várható érték	244	79	20	117	3,06	3,29	22,73	19
Standard hiba	5,444	1,735	0,52	3,153	0,048	0,076	0,72	0,407
Medián	237	77	17	115	3	3,15	23,5	18
Módusz	212	77	17	101	3	3	15	18
Szórás	43,207	13,773	4,126	25,025	0,378	0,601	5,716	3,232
Variancia	1866,853	189,709	17,02	626,272	0,143	0,361	32,668	10,444
Csúcsosság	3,896	3,653	-0,576	-0,196	-0,065	0,296	-0,315	4,731
<b>Ferdeség</b>	<b>1,564*</b>	<b>1,445*</b>	<b>0,847</b>	<b>0,43</b>	<b>-0,141</b>	<b>0,882</b>	<b>0,314</b>	<b>1,558*</b>
Tartomány	234	77	15	114	1,8	2,65	24,5	20
Minimum	181	58	15	76	2,2	2,35	13	12
Maximum	415	135	30	190	4	5	37,5	32
CV	<b>18%</b>	<b>17%</b>	<b>20%</b>	<b>21%</b>	<b>12%</b>	<b>18%</b>	<b>25%</b>	<b>17%</b>
Darabszám	63	63	63	63	63	63	63	63

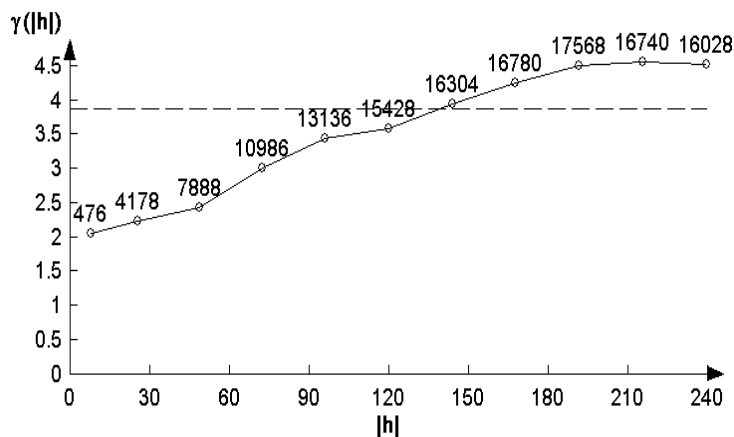


## 4.2 A hozam és talajadatok térbeli becslése

Az adatok elemzése, szűrése után a további számításokhoz szöveges adatállományokba szerveztük (.TXT ill. .DAT típusú állományokat hoztunk létre, fejléccel, tabulátorral, szóközzel vagy vesszővel tagolva). Az első két oszlopban a helykoordináták értéke, a következőkben a vizsgált változók. pl adat2002.txt fájl tartalma:

```
x,y,hozam2002
515579.664, 286280.796, 6.0
515584.948, 286282.898, 6.3
515593.297, 286288.267, 7.7
515599.328, 286290.353, 7.2
```

.

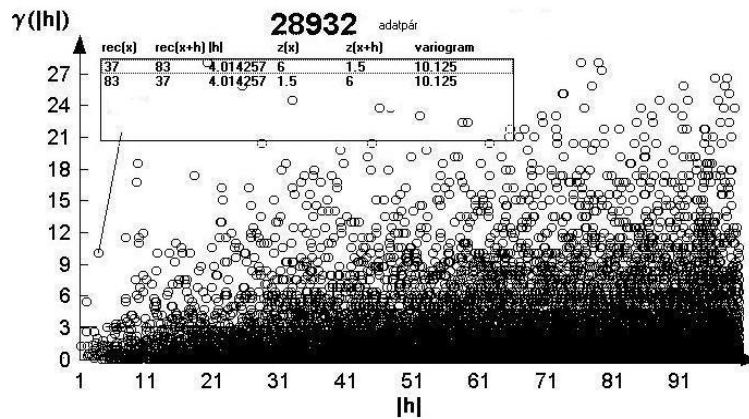


### 4.2.1 ábra A kísérleti variogram felépítése

A variogram modellel meghatároztuk a különböző változók adathalmazának térbeli korrelációs strukturáját.

A kísérleti variogrammal ábrázoltuk a mérési pontokból képzett párok átlagos különbségét (aktuálisan a négyzetes különbség felét: szemi-variogram, vagy másnéven varianciát). Megvizsgáltuk a

párok gyakoriság eloszlását néhány jellemző hossz (lag) távolságnál, és optimalizáltuk a méretét és számát. Azonosítottuk a fontos adatjellemzőket, a heterogenitást és a kiugró párokat (a variogram-felhő és szóródási térkép alapján), vagy azokat a hatásokat, amelyek a torzítást/zajt okoztak a variogram becslésben. Az 4.2.2 ábrán jeleztük, hogy különböző távolságok esetén pl.  $\approx 4$  m távolságban levő sorszámokkal azonosítható pontok értékei nagyon különböznek. Azt is leolvashatjuk, hogy 100 m-en 28932 adatpárt vizsgált meg 10 m-s hosszakat vizsgálva.



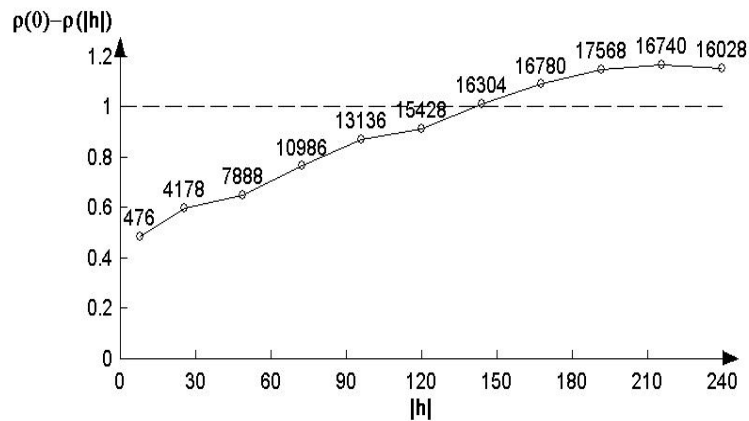
#### 4.2.2 Variogram felhő – pont-párok szóródása

A variogramfelhő segít a térbeli folytonosság jellemzését zavaró adatpárok azonosításában, adatanomáliák kiszűrésében. Azokat a pontokat vizsgáltuk, amelyek távolsága alapján besorolt közös intervallumban vannak, de a hozam értéke nagyon eltérő. Ezek a pontok kitűnnek a felhőből: pl 4 m távolságban levő termésadatok:

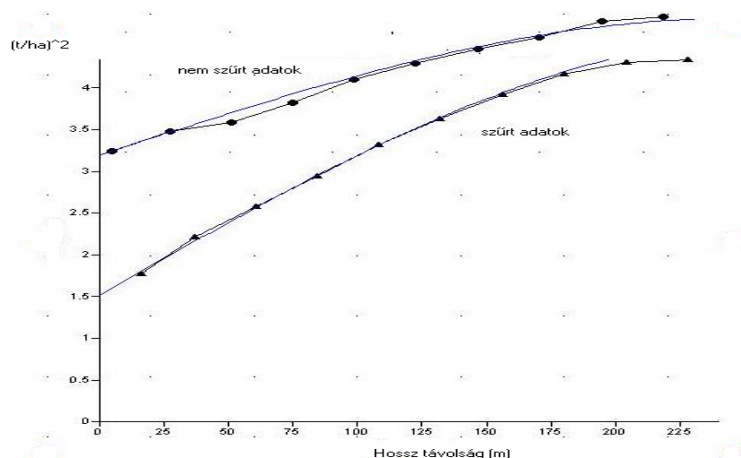
6 t/ha és 1,5 t/ha. Minden kritikus pontpárt összekapcsoltunk a pontok térképével.

Ha olyan térbeli folyamatok vannak, amelyek létrehoznak regionalizált változókat, elemeztük a potenciális hatásukat a térbeli korrelációban. Összehasonlítottuk a térbeli korreláció különféle becsléseit (szemivariogramot, a kovarianciát és korrelogramot). Ezek kevésbé érzékenyek a csoportosulásokra és a kiugró adatokra, és így a térbeli szerkezet tisztább becslését adják. Ha a struktúra reprezentálása hasonló, mi döntjük el, hogy melyiket választjuk. Megjegyezzük, hogy a variogram számolásához az ún. belső hipotézis elegendő, nem szükséges a regionalizált változók átlaga, míg a kovariancia meghatározásánál a stacionaritás feltétele és a regionalizált változók átlaga is szükséges. A 4.2.3 ábrán a fenti kísérleti variogramban szereplő adatpontok korrelogramját ábráztuk.

A térbeli változások nyomon követésére a variogramok legalkalmasabbak, és nagyon jó elemző programok állnak rendelkezésre. Az adatállományokban bekövetkezett változásokat, korrekciókat is jól tükrözik. Megvizsgáltuk a szűretlen és szűrt adatok térbeli korrelációjának alakulását a kísérleti variogram segítségével. A 4.2.4 ábrán a röghatás jelentős csökkenése mutatja, hogy az adatállományban levő hiba (mérési, gyűjtési) is csökkent.



4.2.3 ábra Térbeli pontok korrelogramja



4.2.4 ábra Variogram illesztése eredeti és szűrt adatokhoz

A termésadatok szűrésének elsődleges célja az interpolálás minőségének a javítása, a krigelésnél az adatminőség tökéletesítése, amit a röghatás varianciájának csökkenése jelez. Egyértelműen látszik, hogy a bizonytalanság csökkent a hozamban

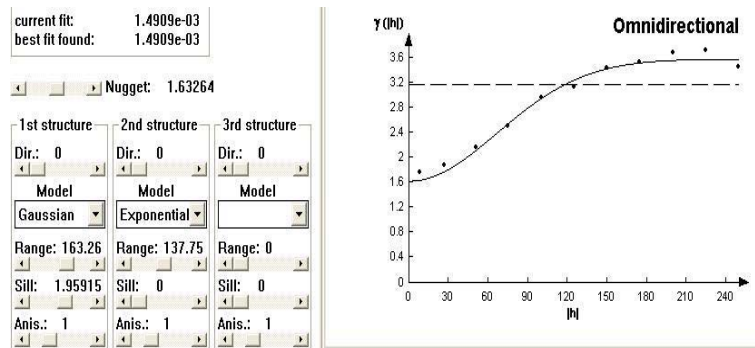
a szűrt adatoknál, mint ez látható a variogram röghatásának jelentős csökkenésében. A szűretlen adatoknál az értéke 3,35 a szűrt adatoknál lecsökkent 1,52-re.

A zajos variogramok esetén transzformáltuk az aszimmetrikus adateloszlásokat, csillapítva az extrém értékek hatását.

A variogramok összeállításánál az alábbi általános megfontolásokat vettük figyelembe:

- a) a maximális hossz nem lehet nagyobb a maximális pártávolság felénél, és legalább 20-50 pár legyen minden átmeneti zónában.
- b) megvizsgáltuk a variogram területét az izotrópia és irány szempontjából; övezetek esetén, ellenőriztük a hossz beállításokat;
- c) megvizsgáltuk az autokorrelációs struktúrát az alternatív variogram mértékek (standard variogram, kovariancia, korrelogram) között
- d) a becsült variogram struktúrát kipróbáltuk a hossz beállítások különböző kombinációjával (4.2.5 ábra).

Minden vizsgált változónál ellenőriztük a variogram becslés szórásértékét az átmeneti mezőn belül, ha szükséges kizártuk a kiugrókat, ezzel javítottuk a becslést. Megpróbáltunk nagyobb hosszal simább variogramot szerkeszteni, de minimális lag távolságot választottuk, így a legtöbb hossznál elfogadható volt a párok száma. Ha túl nagy volt a variogram-zaj átlapolt hosszakat használtunk.



#### 4.2.5 ábra Variogram struktúra illesztése

A variogram interpretációja szubjektív, előzetes tapasztalaton és szakértői véleményen alapszik. Pl. kevés adat rossz hatással van; így ha a folyamat, ami a térbeli változékonyságot generálta jól ismert, használjuk ezt az információt a legjobb variogram kiválasztásában.

A valódi variogram a mérési párok varianciája, mely kielégíti a távolság és az irány minden kombinációját. A gyakorlatban a meghatározott adatokkal kiszámoltuk a mérési párok csoportjainak varianciáját a hasonló távolságú és irányú intervallumban. Megrajzoltuk a variancia grafikonját a távolság függvényében bizonyos irányoknál, és modell görbét illesztünk a grafikonhoz. A modell a valódi variogram feltételezett közelítése.

A munka menete minden hozam és talajadatnál a következő volt:

1. Elkészítettük az adatpárok közbülső állományát,
2. Kiszámítottuk, megrajzoltuk a kísérleti variogramot,

### 3. Szferikus és/vagy exponenciális modell variogramot illesztettünk.

Először meghatároztuk az intervallum osztályok távolságát (lag = hossz, különbség) és az iránytoleranciát a variogram számításánál. Próbálgatással megkerestük a jó kombinációt, de a szisztematikus közelítés is jó segítség volt. Kezdetben az alapértelmezett irányt használtuk, amely az ún. ” minden irányú”(omnidirectional) variogram.

A 90 fokos szögtoleranciát (figyelembe vett szögtartományt) megengedtünk egy tetszőleges irányvonal bármely oldalán, így lehetővé tettük, hogy az iránytól függetlenül minden párt figyelembe vegyünk. Ez maximalizálta az egyes távolságosztályokba eső párok számát, amely rendszerint a legjobb, vagy a legsimább variogramot adta.

Következő lépésben különböző hossz intervallumokkal megrajzoltuk a kísérleti variogramot. Megbízhatóbb a modell, ha a kísérleti variogram számításánál több különböző hosszúságú intervallumot használunk.

Az alapértelmezett intervallumokat a hozzávetőleges számítások alapján határoztuk meg, amely variogramok általában nem lépik túl a minták közötti maximális távolság felét.

A maximális pártávolságot ezért felosztottuk kettő, és továbbosztottuk 10 egyenlő távolságosztályra: pl. 240 m volt a maximális távolság, így a vizsgált kiindulási hossz: 12 m.

A variogram jól definiált szerkezetet mutat, különböző modelleket illesztettünk hozzá.

A szferikus elméleti modell – a kezdeti lineáris növekedés a Y értéknél viszonylag élesen növekszik, majd vízszintesen konstans értéket ér el. A szferikus modell illesztésénél meg becsültük a *rögghatást* vagy Y eltérést 0-tól, a *küszöböt* vagy a maximum érték és a rögghatás közötti távolságot, és a *tartományt*, vagyis azt a távolságot, amely alatt a maximumot eléri.

Az illesztés jóságát a négyzetes hibaösszeg (SSE), vagy az Akaike információs kritérium (AIC) (Webster és McBratney, 1989) mutatók jelzik. Minél kisebb az értékük annál jobb a függvény illeszkedése a kísérleti variogramhoz. A 2001. évi vizsgálatnál feltüntettük az exponenciális függvény paramétereit is, a többi évben csak a szferikus függvény adatait közöltük, mivel ez adta a jobb megoldást. A variogram paramétereit alapján ábrázoltuk a függvényeket az 4.2.6 ábrákon. A térbeli korrelációs tartomány definiálja azt a távolságot, amelyen belül a tulajdonság értékei korrelációban vannak.

#### 4. 2.1 táblázat A hozam adatok variogram modelljei

Paraméter	2001.		2002	2003	2004	2005
	Szferikus	Exponenc.				
Rögghatás	1,52	1,37	1,02	0,37	1,18	1,04
Küszöb	4,39	5,64	5,90	0,57	1,64	1,75
Tartomány	241	176	229	188	116	218
SSE súlyozott	0,015	0,051	0,970	0,002	0,040	0,009
RMSE	0,027	0,049	0,215	0,010	0,040	0,020
AIC	-81,92	-56,45	-26,51	-125,00	-61,39	-93,39
<i>N/S</i>	<b>0,35</b>	<b>0,24</b>	<b>0,17</b>	<b>0,65</b>	<b>0,72</b>	<b>0,59</b>



A variogram paraméterek közül a térbeli struktúra jellemzésére a táblázatban megadott röghatás/küszöb (N/S) hányados definiálja a kis tartományú változékonyságot. Kicsi N/S és nagy korrelációs tartomány rendszerint jelzi, hogy nagyobb pontossággal lehet megtervezni a tulajdonság térképét. Az irodalomban elfogadottterminológia alapján (Kravchenko, 2003) a 2003. 2004. és a 2005. évi  $N/S \geq 0,6$  megfelel a gyenge térbeli struktúrának, ami azt jelenti, hogy az adatváltozékonyság 60%-a megmagyarázhatatlan, kis távolságú, véletlen ingadozású. A 2001. évben közepes erősségű  $N/S \approx 0,3$  volt jellemző, míg 2002-ben nagy erősségű térbeli struktúrát jelent  $N/S \approx 0,1$ .

Az első szempont volt az elméleti variogram illesztésénél természetesen az, hogy hasonlítson a kísérleti variogramhoz, és a becslésnél értelmes krigelési eredményt adjon.

A második fontos szempont, hogy számításba veszi a becsült krigelési variancia eloszlását és maximális abszolút értékét. Nagyon elővigyázatosnak kell lennünk a becsült krigelési varianciával, mivel ez az érték fejezi ki a krigelés minőségét, ami az elméleti variogramon alapszik, ebből számítottuk.

A harmadik szempont az illesztésnél, az izovonalak ellenőrzése, ha legvalószínűbb értékek nem tipikusak, akkor a variogram paramétereit meg kell változtatni.

A keresztértékelés alkalmazásával a becslés jóságát határoztuk meg: a véletlenszerűen kiválasztott minták értékeit a becsült adatokkal hasonlítottuk össze. A következő kérdésekre kerestünk választ:

Melyik modell a legjobb? Hogyan határozzuk meg, hogy melyiket használjuk? Mi történik, ha hibásan döntünk?

A legjobb modellnek azt fogadtuk el, amelyik legszorosabban illeszkedik az igazi variogramhoz. Annak ellenére, hogy a programok a legkisebb négyzetes eltérést használják az illesztéseknél, a modell kiválasztása szubjektív, egyszerűen amelyik a legjobb illeszkedést mutatja a különböző hosszak tartományán.

Megvizsgáltuk, hogy a szferikus és az exponenciális modell közötti különbségek csak csekély eltéréseket okoztak a krigelt értékekben, így elfogadható bármelyik választás. Megállapítható volt, hogy az igazi variogram definiálásának képessége korlátozott a felhasználható adatok egyedi halmaza miatt.

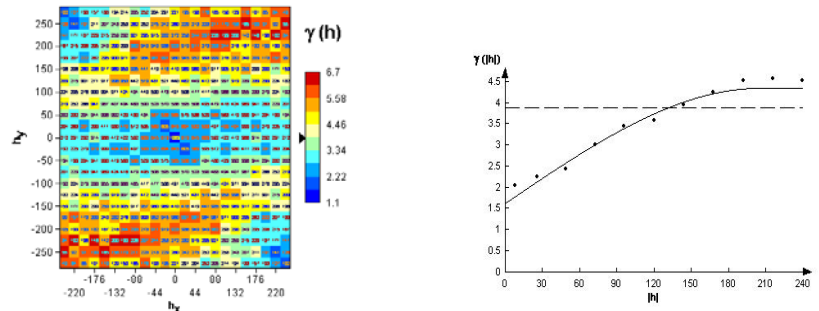
Ezen a ponton a szerkezeti elemzésben az anizotrópia a legnagyobb megválaszolendő kérdés. Az anizotrópia azt jelenti, hogy a variogram nem azonos minden lehetséges irányban: eltér az alakja, küszöbértéke és hatástávolsága.

Ha az irányított variogramot modelleztük az anizotrópia derékszögű fő komponenseivel (a maximum és a minimum tartomány irányai), kombináltuk a térbeli korrelációval.

A röghatás variancia mindig izotropikus paraméter, ez még további vizsgálatokat igényel.

A variogram felszín (4.2.6 ábra) mutatja valamennyi keresési távolságra a mért hozam átlagos variabilitását, az egyes távolságokon belül a pontpárok számát, valamint a keresés irányát.

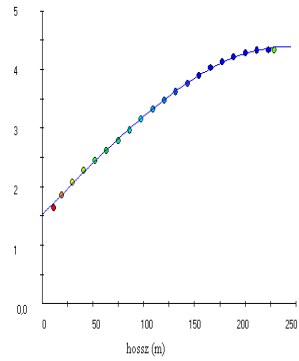
Az ábrán jól lehatárolható a legalacsonyabb variabilitási irány ( $\gamma(h)= 1,1$ ), és a térbeli autokorreláció határa.



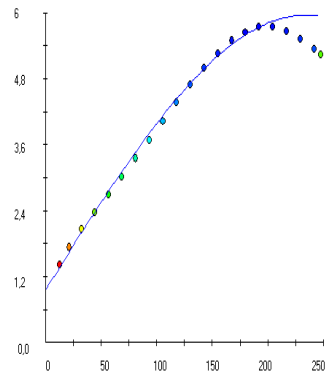
**4.2.6 ábra Terméshozam mintapontjainak variogram felszíne és variogram modellje**

A 4.2.7 ábrán a variogram modellek jól mutatják az évjáratok hatását, a tartományok változását. A kukoricatermés három különböző évben más –más tartományban van, de a térbeli eloszlás heterogenitása hasonló képet mutat.

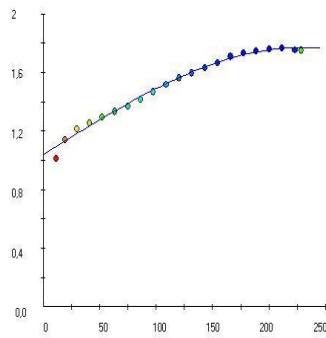
A másik tényező, ami hat a pontosságra az interpolációs módszer, amellyel a diszkrét minta adatokat folytonos térképbe konvertáltuk. Két interpolálási módszerrel próbálkoztunk, az inverz távolsággal súlyozott módszert és a krigelést hajtottuk végre mind a hozamértékeknél, mind a talajadatoknál. Számos tanulmányban összehasonlították a módszerek teljesítményét a mezőgazdaságban. A tanulmányok leginkább a kereszt-ellenőrzést, vagy a jackknife (független teszt adathalmaz) módszert használták az interpolációs módszerek összehasonlítására.



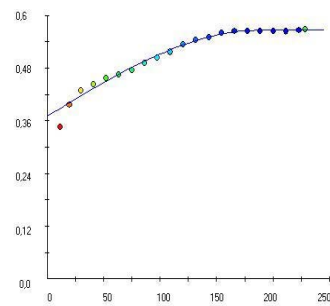
**Kukorica 2001.**



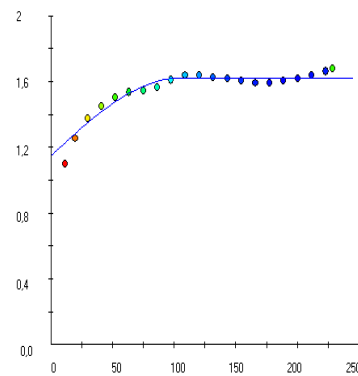
**Kukorica 2002.**



**Kukorica 2005.**



**Árpa 2003.**



**Búza 2004.**

**4.2.7 A hozamértékekvariogram függvényei**

Különböző szerzők javasolják a kereszt-értékelési technikát az interpolálás pontosságának értékelésére (Kravchenko és Bullock, 1999; Webster és Oliver, 2001). Ez a módszer széles körben használatos, mert nincs szükség külön teszt adathalmazra. Az interpolációs technika valódi pontosságának becslését a jackknife közelítéssel kapjuk meg Gooverts, P. (2000). Ez az eljárás független tesztadatokat igényel, amelyek nem vesznek részt az interpolálásban. A halmaz készítésénél a mintákból kiválogattunk véletlenszerűen, vagy szisztematikusan adatokat. Minél több a rendelkezésre álló adat, annál kisebb a valószínűsége, hogy az eredmények torzítanak. A hozamadatoknál minden 10. adatot kiválasztottuk, ebből képeztük a tesztadathalmazt. Ezek a halmazok 650-1200 eleműek voltak az eredeti adatszámoknak megfelelően.

Statisztikailag definiáltuk az interpolálás pontosságát, meghatároztuk a tesztadathalmazban, a valódi és az interpolálással nyert értékek különbségét. Kiszámoltuk az egész vizsgált területen, az átlagos négyzetes hiba gyökét (RMSE) képezve (Webster és Oliver, 2001) nyomán.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \{Z(X_i) - Z^*(X_i)\}^2}$$

ahol  $Z^*(X_i)$  a becsült érték,  $Z(X_i)$  a mért érték és  $N$  a vizsgált pontok száma ( $i= 1, 2, \dots, n$ ).

Kívánatos, hogy a becslések a mért értékekhez tartsanak, a hiba minél kisebb legyen.

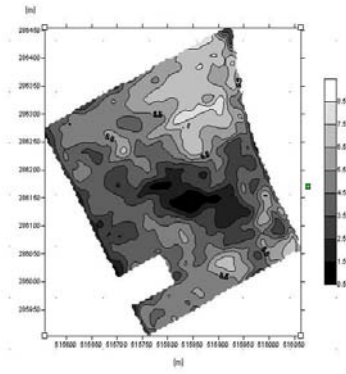
A szűrt adatbázison elvégeztük a következő interpolációs eljárásokat: krigeléssel az egész területi és a helyi variogrammal, az inverz távolság súlyozásával különböző kitevőkkel (1, 2, 3). Az interpolált értékekkel elkészítettük a térképeket. (4.2.8 ábra). Megállapítható a térképek alapján, hogy jól nyomon követhető a hozamérték változásai, és hasonló mintázatot adnak az egyes években.

Mindegyik interpolálásnál megvizsgáltuk a becsült és a mért adatok eltérését. Az eltérések négyzetösszege alapján számított RMSE a kiválasztott tesztmintánál azt mutatták, hogy a krigelés és az inverz távolság módszerek közel azonos értéket adtak.  $RMSE_{Krigelt}=0,449$ ;  $RMSE_{inverz}=0,491$ .

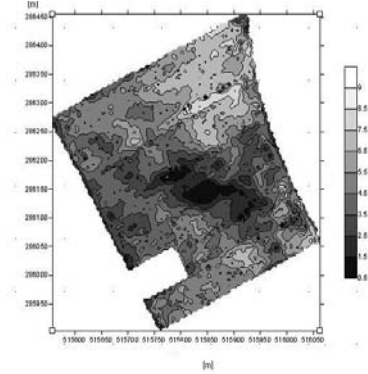
A krigelt értékek jóságát meghatároztuk az egyes években:  $RMSE_{2001}=0,449$ ;  $RMSE_{2002}=0,458$ ;  $RMSE_{2003}=0,214$ ;  $RMSE_{2004}=0,219$ ;  $RMSE_{2005}=0,329$ .

Fontos kérdés az interpolációnál hogyan változik a terület egyes részein a becslés pontossága.

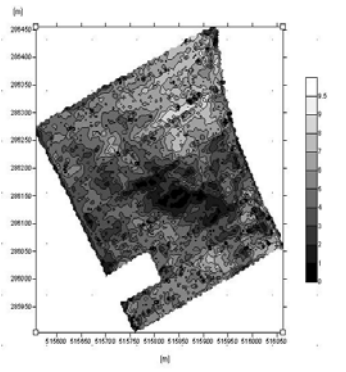
Megvizsgáltuk, hogyan változik az adott táblán a becslési hiba. A hiba térképe mutatja, hogy alacsony szinten mozog az interpolálási területen (4.2.10 ábra).



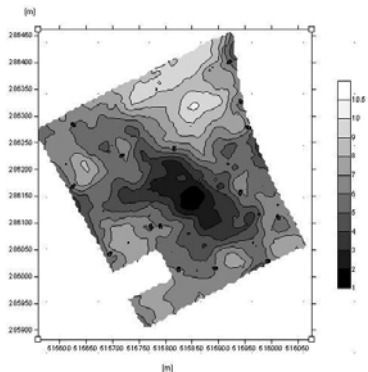
*Inverz távolság (kitevő=1)*



*Inverz távolság (kitevő=3)*

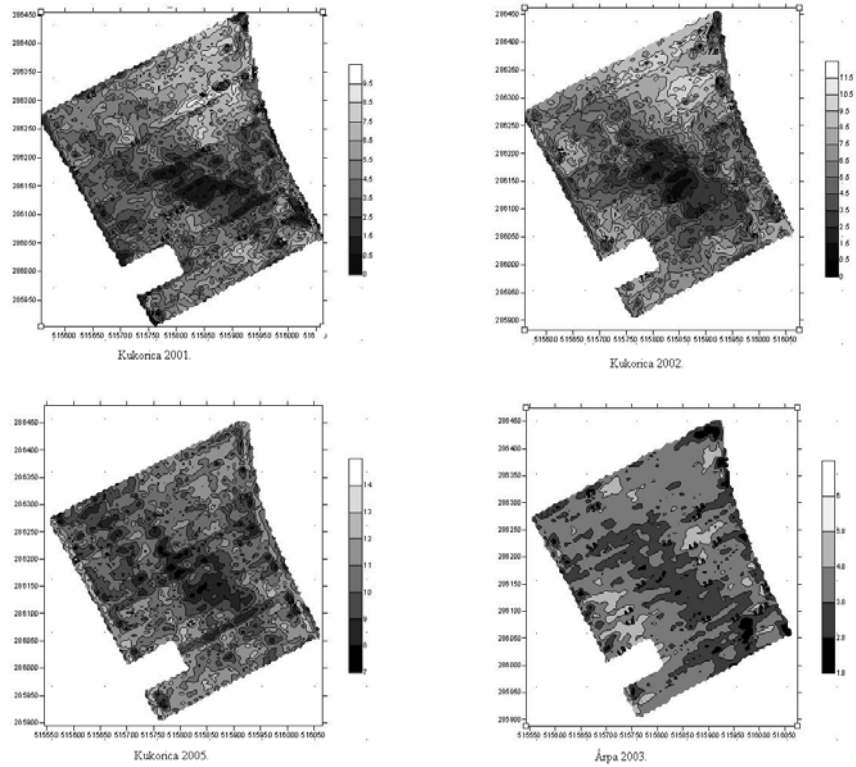


*Inverz távolság (kitevő=2)*

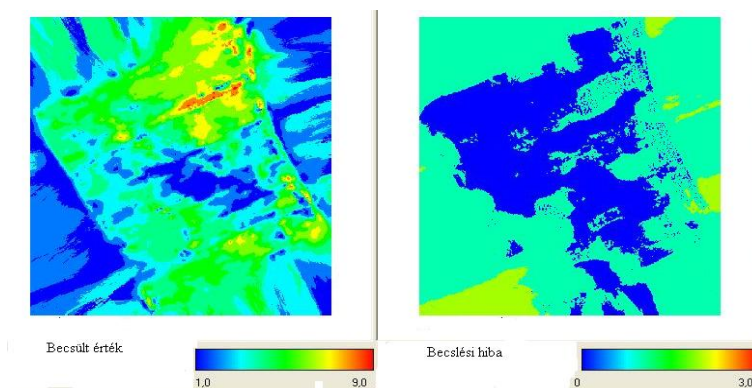


*Krigelés variogrammal*

**4.2.8 ábra Interpolálási módszerek**



**4.2.9 ábra Krigelt hozamtérképek**



**4.2.10 ábra Hozam becslése és hibaterképe**



Elemeztük, hogyan változik a térbeli szerkezet, ha különböző nagyságú adathalmazzal végezzük vizsgálatokat (4.2.2 táblázat). Kiválasztottuk minden *második, negyedik, nyolcadik és tizenhatodik* adatot, és az így kapott adatokra külön-külön elvégeztük a térbeli becsléseket. Az elkészített térképeket értékeltük, meghatároztuk a pontosságukat, és kiválasztottuk a kezelési döntések előkészítésére legalkalmasabbat.

Az elemzések azt mutatták, hogy a térbeli szerkezet nem változik lényegesen az adatok számának csökkenésével, jelentőst változást a hibában a 2001\_16 adatállománynál figyelhetünk meg.

#### **4.2.2 A 2001. évi hozam adatok variogram statisztikája**

	C0	C1	A1	SSE súlyozott	RMSE	AIC	N/S
2001	1,52	2,88	241	0,015	0,027	-81,92	0,35
2001_2	1,53	2,89	242	0,012	0,024	-86,19	0,35
2001_4	1,52	3,01	242	0,010	0,022	-91,13	0,33
2001_8	1,48	2,98	228	0,018	0,029	-78,84	0,33
2001_16	1,69	2,96	220	0,210	0,120	-26,76	0,36

Az interpolált térképek hozambecslést adtak a vizsgált táblán, a következő lépésben megvizsgáltuk milyen kapcsolata van a talaj adatok térképeivel.

Jelenleg az egyik legnagyobb nehézség a helyspecifikus táblakezelésnél, hogy annyi információhoz jussunk a talaj-tulajdonságokról, ami alapján megbízható becslést tudunk készíteni. A talajinformációkat gyűjtöttünk a táblán a mintahelyekről. Ezek a minta adatok térbeli autokorrelációban

lennének az interpolált értékekkel, bármilyen módszer esetén. Ennek megvalósításához a minta intervallumának elég kicsinek kell lennie, hogy figyelembe vegye a szántóföld térbeli változékonyságát. Ezért a mintapontok közötti távolság kapcsolatban van e változékonyság skálájával. Széleskörű kutatások mutatták, hogy a hibás, túl széles intervallumok gyenge minőségű becslést eredményeztek. A térbeli skálához viszonyított túl kicsi intervallum fölösleges erőfeszítés, és erőforrás igényes.

A variogram leírja a változók korrelációs struktúráját, így ha vannak a szántóföld előző vizsgálataiból a talajtulajdonságokat leíró függvények, ezeket használhatjuk a jövőbeli minták tervezésénél. A variogram tartománya alapvetően megadja a térbeli szerkezetek vagy minták változékonyságának átlagos kiterjedését. Első közelítésben a *minta intervallumnak*. Az átlagos variogram tartomány felénél kisebbnek kell lennie. A megbízható variogram kezdetben legalább 100 mintapontot igényel (Kerry és Oliver, 2003).

A precíziós mezőgazdaság irodalomban több szerző is javasolta az ún. *kiegészítő* adatok- *hozamterképek, légi fotók, talaj vezetőképessége, talaj ellenállása*, stb.- használatának lehetőségét a mintaterv előkészítésében (Mulla et al. 1997; Yang és Anderson, 1999).

Megvizsgáltuk néhány állandó talajtulajdonság térbeli struktúráját, a hozamok, a talajtulajdonságok és a talajellenállás közti kapcsolatokat, felhasználva a geostatistikai és korrelációs

elemzéseket. *Kiegészítő* adatokat az 5 éven keresztül gyűjtött hozam adatok és a talajjellenállás mérései adták.

A variogram modellek egyenleteit Webster és Oliver (2001) összefoglaló művében megadottak alapján alkalmaztuk. Az 4.2.3 táblázatban összpontosítsunk a *variogram* tartományokra, amelyek megmutatják a jelenlevő térbeli struktúra átlagos kiterjedését. A hatástartomány nagyon hasonló az egyes tényezőknél. A nagy N/S arány a talajjellenállásnál (84%) felveti, hogy jelentős helyi zaj, nagy változékonyság, vagy hiba van ebben az adatban.

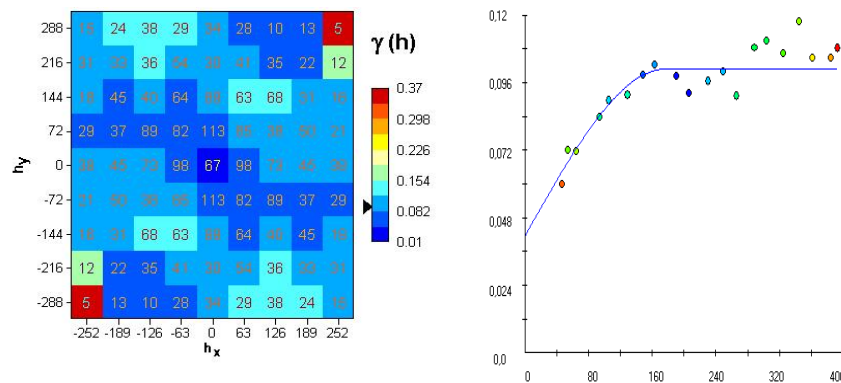
Ha a kiegészítő adatok variogramjai által azonosított térbeli struktúra hasonlóságot mutat az állandó talajtulajdonságok struktúrájával, felveti azt a lehetőséget, hogy alapul szolgálhat a talajtulajdonságok alkalmas mintaszerkezetének kialakítására. A mintasűrűséget igazítani kell a térbeli függőség tartományához, hogy elkerüljük mind az alul és felülmintázást.

#### **4.2.3 táblázat Variogram-paraméterek**

<i>Tényezők</i>	<i>Röghatás</i>	<i>Küszöb</i>	<i>Tartomány</i>	<i>N/S</i>
Humusz	0,0385	0,064	210	0,6
AlP2O5	585	4333,8	230	0,1
AlpK2O5	78	174	227	0,45
pHKCl	0,0007	0,0047	195	0,15
Kötöttség	0,5	12	250	0,4
Talajjellenállás	0,421	0,5	100	0,84
Kukorica 2001	0,95	2,97	241	0,32
Kukorica 2002	1,1	3,23	229	0,34
Árpa 2003	0,28	0,75	188	0,37
Őszi búza 2004	0,6	1,8	116	0,33
Kukorica 2005	1,04	1,75	218	0,6

Ha a mintasűrűség kiválasztásánál kiegészítő adatok variogramjai szolgálnak alapul, akkor az átlagos variogramtartomány harmada és fele között a javasolt minta intervalluma, ez a vizsgált táblán 55-65m mintatávolságot jelent.

A variogram felszínének vizsgálatával megnéztük talajminták térbeli variabilitásának irányultságát és kiterjedését. Az ábrázolt felszín ÉD irányú tájolású, a középpontjában jelzi a nulla távolságot, kiterjedése a vizsgált variogramok maximális tartománya

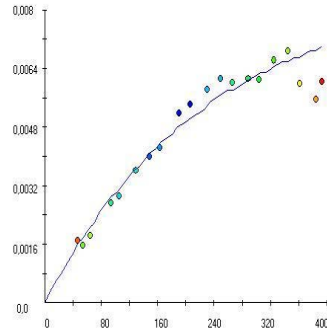
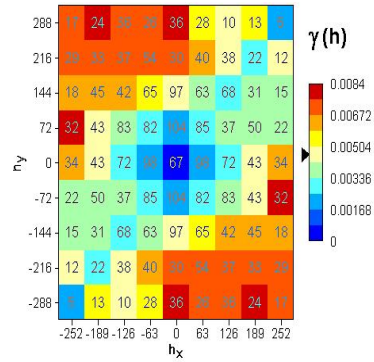


#### 4.2.11 Humusz mintapontjainak variogram területe és variogram modellje

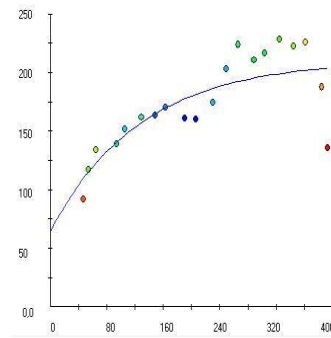
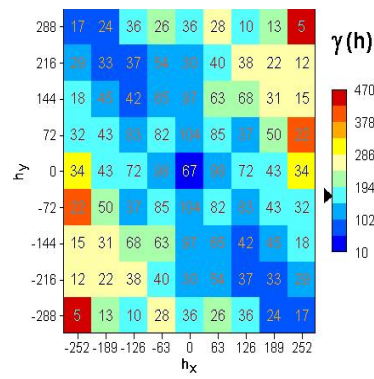
A humusz variogramja jól mutatja térbeli homogenitást, és a legkisebb variabilitás irányát. A pH értékeknél nem jellemző a homogenitás és a legkisebb variancia iránya is eltérő a humusztól.

A két tulajdonság korrelációs együtthatója =-0,575.

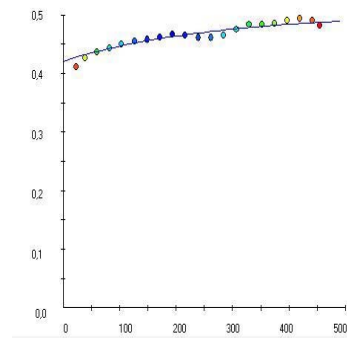
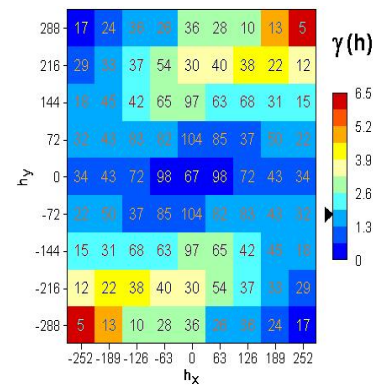
A talaj K tartalma is meghatározott irányban mutatja a legkisebb variációt.



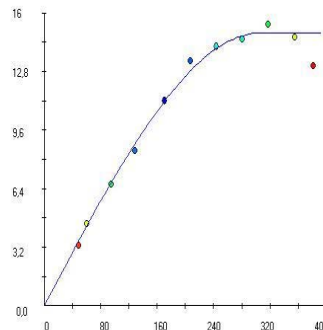
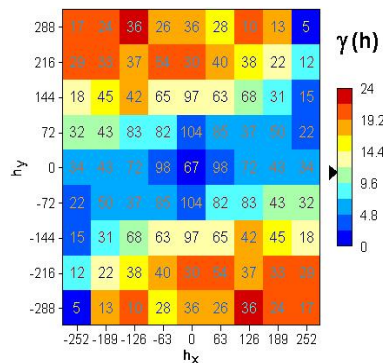
*pH mintapontjainak variogram területe és variogram modellje*



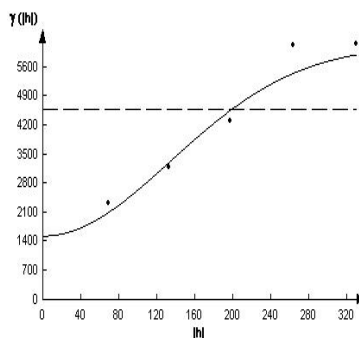
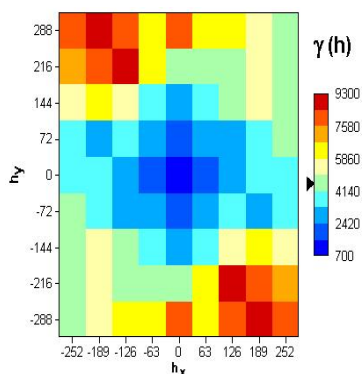
*K tartalom variogramja*



*Talajellenállás*



*Talaj kötöttsége*



*P tartalom variogramja*

**4.2.13 ábra Talajjellemzők térbeli szóródása**

Ha a kísérleti variogramnak volt konkáv emelkedési szakasza, vagy gyorsan növekedett, közben a küszöbértéke, a tulajdonság változékonyságánál trenddel kell számolnunk. Mivel a geostatistika feltételezi, hogy az alapjául szolgáló folyamat véletlen, a trendeket ki kell zárnunk, és a következő elemzéseket a

függvényből visszamaradón kell elvégezni (P tartalom variogramja).

*Összegezve:* Az adatok szűrni és javítani kell az interpolálás előtt azért, hogy bemutassuk az igazi változékonyságot a vizsgált területen, szisztematikus és műveleti hibák nélkül. A vizsgálati módszerrel végigjártuk a térbeli minta kiértékelésének problémáját. A vizsgált táblán a hozamok stuktúrája hasonló képet mutat a különböző években a talajtulajdonságokkal, de nyilvánvaló sok más tényezőt is meg kell még vizsgálni. A kukorica terméshozamának változása is jelzi, hogy az időjárási tényezők is módosítják.

Az eredeti mérési adatokból származó variogramok nem tekinthetők teljesen stabilnak, mert túl kevés az alapul szolgáló mérési pont. A szakirodalom legalább 50-100 közötti adatpont esetén tekinti megbízhatónak, vizsgálatainkban 63 mintaponthoz tartozó adatokat vizsgáltunk.

### **4.3 A talajadatok fuzzy variogramja és krigelése**

A másik megközelítés a hiányos és bizonytalan adatok kezelésére a fuzzy logika elemeinek alkalmazása. Speciális adatelemző és modellező módszer a környezeti adatok bizonytalanságának és heterogenitásának kezelésére (Lark, 2000).

A pontosság növelésére és a bizonytalanság csökkentésére a talajadatokat kiegészítettük az egyes talajparaméterek eloszlása alapján szimulált „mérésekkel”. Az eredeti 63 adatpontot bővítettük 146-ra, ill. 242 mintapontra. Az így összeállított adatok

„pontatlanságát” és bizonytalanságát fuzzy halmazokkal kezeltük. Megadtuk az egyes változók értéktartományát leíró tagsági függvényeket. Az adatokat fuzzy számmal írtuk le, *háromszög* alakú tagsági függvénnyel:  $T(x/a, b, c) \in [0, 1]$ , ahol  $a \leq b \leq c$  a változók intervallumát jelenti a vizsgált helyen. A paraméterekhez rendelt tagsági függvények lehetővé teszik a talajtulajdonságok és hozamérték folytonosságának és az egyes értéktartományok átlapolásának kezelését. A mért (éles) adatokat beágyaztuk a fuzzy számok halmazába, a fuzzy számok speciális esetei, amelyek tagsági függvénye  $\mu = 1$

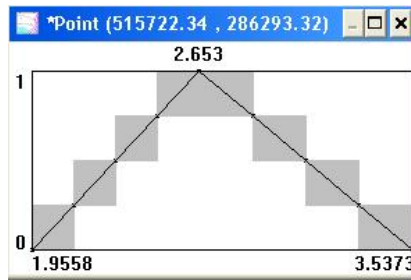
A feldolgozásnál az input adatállomány előkészítését egyszerűsíti, hogy speciális ASCII fájl formátumot hoztunk létre. Kombináltuk az éles és fuzzy adatokat egységes formában. Az input adatokat, ha szükséges volt, transzformáltuk (pl. logaritmus függvénnyel), mind a kísérleti variogram, mind a krigelés számítása előtt.

Az input állomány felépítése: Megadtuk a tartomány koordinátáit, ahol a krigelést végrehajtottuk, előkészítettük a koordinátarendszert. Fontos lépés a fuzzy számok vágatainak megadása, a „fuziness” felbontása:  $\alpha = 1; 0,75; 0,5; 0,25; 0$  értékekkel adtuk meg ( $A_\alpha = \{x \mid A(x) \geq \alpha\}$ ). (legalább két értéket meg kell adni:  $\alpha = 1, 0$ )

Összeállítottuk a paramétereket, néhány fontosabb talajjellemezővel dolgoztunk. Az 4.3.1 ábrán leírtuk a talaj humusztartalmának lehetséges értékét az adott helyen: Az értéke a  $[2, 3.5]$  intervallumba esik, a leginkább lehetséges értéke 2,65; Az ábráról



leolvashatjuk a különböző vágatokhoz  $(0; 0,25; 0,5; 0,75;1)$  tartozó lehetséges értékeket.



#### 4.3.1 ábra Az adatok fuzzy alakja

A fent definiált fuzzy alakot használva összeállítottuk a további számítások alapjául szolgáló adatállományokat, a fuzzy krigelő programnak megfelelő formátumban. Ez tartalmazza a helykoordinátákat és az egyes jellemzők mintaadatait, az alábbi formában:

x	y	humusz
515571.161	286268.757	2.5/0:2-3.5
515581.861	286246.266	2.5

A fuzzy számot leírja a „2.5/0:2-3.5” kifejezés értelmezése: 2.5 a leginkább lehetséges érték  $\alpha=1$ ;

A lehetséges értékek a  $[2, 3.5]$  tartományban vannak,  $\alpha=0$ ; A szint-vágatok a  $(0,1)$  intervallumban lineáris interpoláltak, így ezek háromszög alakú fuzzy számok.

A hagyományos éles értéket beágyaztuk a fuzzy számok halmazába, a „2.5” a fuzzy szám speciális esetének tekintjük. A szint-vágatok  $[0,1)$  intervallumban  $[2.5,2.5]$ , ha fuzzy számnak tekintjük.

Sok előnye van a fuzzy szám használatának. Egyik lehetőség, hogy azokon a helyeken, ahol kevés a pontos mérés, a szakértői ismeret bevonásával értékeket adjunk meg. Így csökken a krigelési variancia és a fuziness megjelenik az eredményben. Az első pillanatban ez nem látszik előnynek, csak később. Az eredmény most több információt jelent, mert a határozatlan bizonytalan információt is figyelembe veszi, amit a hagyományos módszerek nem használnak. A másik lehetőség, hogy az adatok gyakran természetüknél fogva magukban foglalják a „lágyságot”, amit kifejezhetünk fuzzy számmal- pl a mérési toleranciát. Nagy előnyt jelent, ha ismerjük az eredmény toleranciát a becsléseknél. Ha az „átmenetet” nem vesszük figyelembe, az eredményt nem tudjuk pontosan becsülni.

A kísérleti variogram kiszámításánál az adatpárokat a hasonló távolság alapján besoroltuk. A csoportok számát meghatároztuk és megadtuk a lehetséges átlapolást. Ha az átlapolás = 0, akkor a maximális távolságot elosztjuk a csoportok számával. Ha nullánál nagyobb értéket választunk, a csoportokat kiterjeszti az átlapolásnak megfelelően. Kezdő lépésként a variogram tartományát megbecsültük a két pont közötti legnagyobb távolsággal, a küszöb a két adatpont érték legnagyobb különbsége négyzetének a fele. A paraméterek változtatásával meghatároztuk a variogramot. A legjobban illeszkedő variogram modell kiválasztását és illesztését végeztük: az *exponenciális*, *szferikus*, *gauss*, *lineáris típusú* alapfüggvények, és ezek kombinációjával írtuk le a modellt. A tapasztalati és az elméleti variogram ugyanazt

a koordinátarendszert használja. A kísérleti variogramhoz illesztettünk modellt, meghatároztuk a variogram paramétereit és megkomponáltuk azt a típust, amellyel a krigelési becslést végrehajtottuk. A program a Bárdossy (1989) 1. típusú fuzzy krigelést hajtotta végre, amely éles és fuzzy adatokat, éles variogramot, és a kijelölt  $\alpha$ -vágatot alkalmazta.

A 4.3.2 ábrán a mért mintapontokat  $\square$ , a fuzzy pontokat  $\nabla$  jelzi. A mintasűrűséget a csoportok számával határoztuk meg (10, 12, 16) csoportot vizsgáltunk meg. Az illesztett éles variogram függvény paramétereit a hagyományos krigeléssel közel egyezőek voltak.

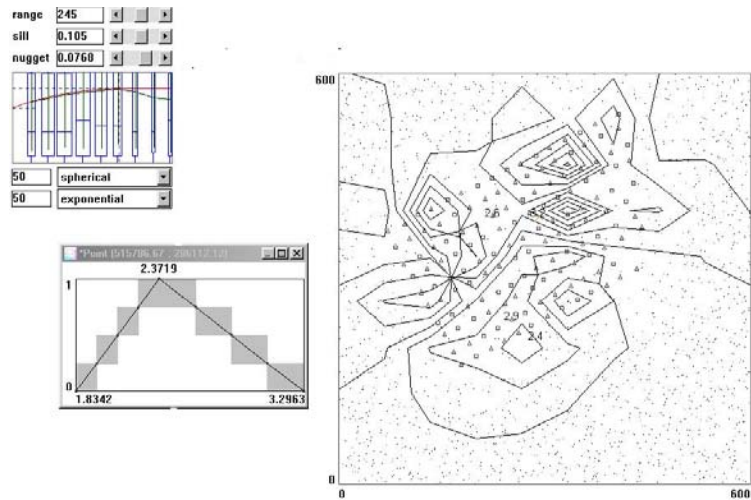
A fuzzy krigelés outputja a *becsült érték* a specifikált helyen. A krigelés az elméleti variogramon alapszik, ami a geostatistikai elemzés eredménye. Beszámol az értékek hasonlóságáról a távolságuk függvényében.

Az eredményt izovonalakkal szemléltettük, ezek átvihetők ASCII állományba (4.3.2 ábra). A krigelés eredményét exportáltuk SURFER térképező programba, és elkészítettük a szokásos kontúrvonalas térképet (4.3.3 ábra).

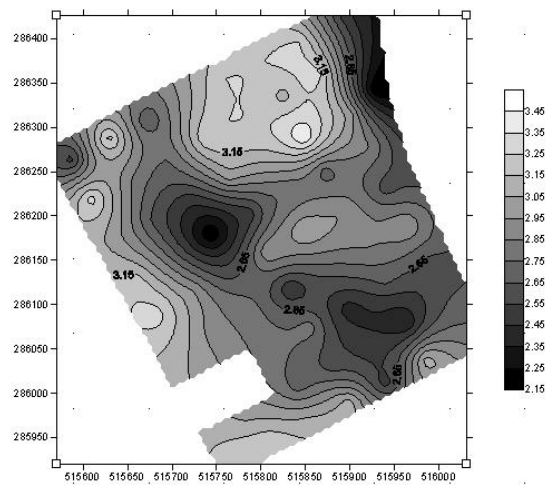
Megvizsgáltuk a mező egyedi pontjait, mint fuzzy számot, vagy a metszetet két adott pont között, amit a tagsági függvény megoszlása jellemez a metszetvonal mentén.

A fuzzy krigeléssel előállított térképen az értékek átmenete jobban követi a természetben előfordulókat, jobban tükrözi a talajadatok folytonosságát.

A fuzzy variogram meghatározásánál 4 fontos szempontot vettünk figyelembe:



4. 3.2 ábra A humusztartalom fuzzy krigelése



4. 3.3 ábra Humusztartalom izovonalas térképe

1. Az összeállított modell legjobban illeszkedjen az adatainkhoz, azzal a céllal, hogy segítségével megfelelő interpolálási eredményt kapjunk.

2. A becsült krigelési variancia abszolút értékét és eloszlását is számításba veszi. Elővigyázatosnak kell lennünk a krigelési varianciával, mert ez fejezi ki az interpolált értékek minőségét, és ez az elméleti modellen alapszik.

3. Az illesztésnél a modell ellenőrzi, átvizsgálja az aktuális krigelési eredményeket. Ha az izovonalak nem a vizsgált paraméter lehetséges tipikus értékeit adják (a szakértő döntése szerint), a modell meg tudja változtatni.

4. Az illesztést ellenőrizhetjük kereszt-értékeléssel, ha elegendő adatunk van.

A fuzzy variogram modell és krigelés biztosítja a többféle feltétel kompromisszumát.

A krigelés előtt lehetőség van a változók aggregálására, több változó együttes kezelésére az AND (minimum), OR (maximum) és SUM (súlyozott összeg) függvényekkel. Aggregálás közben a paramétereket a tagsági függvényeik segítségével közös skálára transzformáljuk és kombinálhatjuk a különböző műveleteket. Az aggregáció eredménye fuzzy szám lesz, így alkalmas az együttes krigelésre.

#### 4.4 A termés hozam becslése fuzzy szimulációval

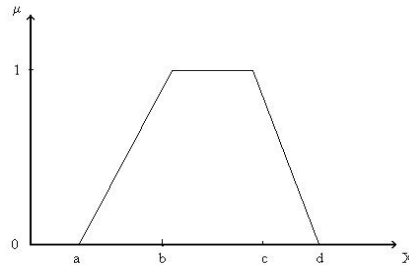
Matematikailag a többváltozós térbeli interpoláció a valós térben kifejezhető: a  $Z = f(x, y, v_1, \dots, v_n)$  függvénnyel, ahol a  $(x, y)$  a helykoordináták és  $v_1 \dots v_n$  a vizsgált változókat jelölik. Különböző interpolációs modellek vannak a fenti probléma megoldására. A technikák közül a co-krigelés a direkt és kereszt-variogram strukturális modelljét igénylik. A módszerek hiányosságai: nehéz készíteni a tapasztalati variogramhoz megfelelő modellt, az adatvektor magasabb dimenziói több variogramot igényelnek.

A fuzzy ismeretalapú modellezés különösen hasznos azon a területen, ahol a komponensek között a kapcsolatok nem ismertek pontosan, vagy hiányosak az adatok a statisztikai elemzéshez.

A talajparaméterek és a hozam adatok közötti összefüggés meghatározását fuzzy szabály alapú rendszerrel 6 és 8 változóval vizsgáltuk meg. Az újabb változók bevonása már hierarchikus szabálybázis felállítását feltételezi, ez még további vizsgálatokat és fejlesztést igényel.

A szabályok leírásánál trapéz alakú tagsági függvényeket használtunk, melyek a trapéz 4 töréspontjával adtuk meg.

Célunk az adott feladat szempontjából az optimális szabálybázis megtalálása. Az optimális fuzzy szabálybázist bakteriális algoritmusmal kerestük.



#### 4.4.1 ábra A szabálybázis tagsági függvénye

Esetünkben egy baktérium egy fuzzy szabálybázist jelent. Így a baktériumok fejlődése tulajdonképpen a fuzzy szabálybázisok fejlődését szimulálja.

A szabálybázis a következőképpen van belekódolva egy baktériumba:

Szabály1	Szabály2	Szabály3	Szabály4	.....	SzabályN										
a <sub>31</sub> = 7.13	b <sub>31</sub> = 7.32	c <sub>31</sub> = 7.28	d <sub>31</sub> = 7.36	a <sub>32</sub> = 43.4	b <sub>32</sub> = 50.9	c <sub>32</sub> = 51.5	d <sub>32</sub> = 51.9	a <sub>33</sub> = 2.19	b <sub>33</sub> = 2.6	c <sub>33</sub> = 2.86	d <sub>33</sub> = 3.37	a <sub>3</sub> = 6.5	b <sub>3</sub> = 7.15	c <sub>3</sub> = 7.61	d <sub>3</sub> = 8.34
A31				A32				A33				B3			

A fenti összeállítás egy 3 bemenetű és 1 kimenetű  $N$  darab szabályból álló szabálybázis elrendezését mutatja be. A 3. szabály az összeállítás alapján a következő :

$$Ha x_1 = A_{31}(7.13, 7.32, 7.28, 7.36) \text{ és } x_2 = A_{32}(43.4, 50.9, 51.5, 51.9) \text{ és } x_3 = A_{33}(2.19, 2.6, 2.86, 3.37) \text{ akkor } y = B_3(6.5, 7.15, 7.61, 8.34)$$

Az egyed összesen tehát  $4 \cdot N \cdot (k+1)$  darab valós számot tartalmaz, ahol  $N$  a szabályok száma,  $k$  a változók számát jelenti, és egy tagsági függvényen belül megtartja a rendezettséget ( $a < b < c < d$ ). Mindegyik valós szám benne van a hozzá tartozó változó

(dimenzió) korlátos intervallumban. A feladat során ugyanis a változók korlátos tartományba eső mintapontjaihoz kell megtalálnunk a megfelelő szabálybázist, és az intervallumok ismertek.

A programban megfogalmazott bakteriális algoritmus paraméterei:

$N_{ind}$ : az egyedek száma – véletlenszerűen létrehozott populáció

$N_{clones}$ : klónok száma a bakteriális mutációban – a lemásolt egyedek száma

$N_{inf}$ : az infekciók száma a géntranszfernél-

$N_{gen}$ : a generációk száma

$N$ : a szabályok száma (ez az algoritmusban rögzített volt, tehát a szabálybázis méretét nem változtattuk generációról-generációra)

Az egyed kiértékelése azt jelenti, hogy megmondjuk, hogy az egyed által reprezentált fuzzy szabálybázis mennyire illeszkedik jól a tanítómintákra. Többféle hiba definíció van, melyeknek a lényege, hogy a minták alapján kívánt kimenet és a fuzzy rendszer által számított kimenet közötti különbséget számítják ki.

Az egyedek kiértékelésénél a következő hibadefiníciót használtuk:

$$\varepsilon_{norm} = \frac{1}{n} \sum_{i=1}^n \frac{|y_{(i)} - \hat{y}_{(i)}|}{O_{max} - O_{min}},$$

ahol  $n$  a minták száma,  $O_{max}$  a kimeneti változó intervallumának felső korlátja,  $O_{min}$  pedig az alsó korlátja,  $y_{(i)}$  az  $i$ . mintához tartozó kívánt kimenet,  $\hat{y}_{(i)}$  pedig ugyanahhoz a mintához a fuzzy rendszer által számított érték.

A bakteriális evolúciós algoritmus folyamata a következő (Botzheim, 2001).



- (1) Létrehoztunk véletlenszerűen egy  $N_{\text{ind}}$  egyedből álló populációt
- (2) Minden egyedre alkalmaztuk a bakteriális mutációt.

A mutáció-kiértékelés-kiválasztás-behelyettesítés lépéssorozatban kiválasztottuk a legjobb egyedet.

- (3) Ezután következik a géntranszfer művelet:

- (3.1) A populációt rendeztük az egyedek jósági értéke szerint, majd felosztjuk két fél populációra. Az egyik alpopulációba a jobb egyedek kerülnek, a másikba pedig a rosszabbak

- (3.2) Kiválasztunk véletlenszerűen egy egyedet a „jó” közül, ez lesz a *forrás-baktérium*, és egy másikat a „rosszak” közül, ez a *cél-baktérium*

- (3.3) Kiválasztunk egy „jó” részt a forrás-baktérium kromoszómájából, és átadjuk a cél-baktériumnak. A cél-baktérium ezzel felülírja a kromoszómájának egy „rossz” részét

- (3.4) A (3)-as pont lépéseit ismételjük  $N_{\text{inf}}$ -szer, ahol  $N_{\text{inf}}$  egy paramétere az algoritmusnak, és az „infekciók” számát jelöli

- (4) Ha a populáció megfelelő a megadott hibadefiníció alapján, vagy elértünk egy maximális generációs számot ( $N_{\text{gen}}$ ), akkor az algoritmus megáll, különben folytatjuk a (2) lépéstől

Először összeállítottuk a mintafájlt, ami alapján a tanítást végeztük.

A mintafájl egyszerű szerkezetű szövegfájl, a következőket tartalmazza: a *változók számát* (kijelöli a dimenziót), a *minták számát* (tanuló adatok), a változók *alsó és felső határait* (a

bemeneti és a kimeneti változók érvényes tartományai), az *adatsorokat* (mérési helyek és mintapontok tulajdonság adatai). Mindegyik minta külön sorban van. Az első kísérleteknél 6 és 8 bemenő változóval és 1 kimenettel, 63 mintaponttal, az alábbi adatokkal dolgoztunk:

Az input adatok: a *helykoordináták*(x, y), a talaj *humusz%*, *pHKCl* érték, *kötöttség*, *P*, *K*, *talajjellenállás*, az output változó a 2001. évi kukorica termés hozama.

A 8 változós modell adatállománya:

8	63						
515557.63		516063.01					
285902.88		286455.18					
2	4						
7.2	8.5						
30	50						
100	400						
50	180						
0	9						
0.7	8.3						
515587.33		286263.22	2.5	7.29	46	216	75
	8.78	3.57					
515629.42		286286.97	3.3	7.24	48	320	85
	0.03	4.4.....					

Minden egyes szimulációs futtatás során keletkezik egy *szabálybázis-fájl* (rulebase.txt). A szabálybázist szövegfájl tárolja, szerkezete a következő:

Az első sorban a létrehozott szabályok száma (10), és a bemeneti változók száma (8) található. Ezután következnek - a mintafájlhoz hasonlóan - a változók tartományai. Nálunk ez most 9 sorból áll, az első 8 a bemeneti változóhoz tartozik, a 9. pedig a kimeneti változóhoz. Utána következnek a szabályok. Egy szabály (bemeneti változók száma +1) sorból áll, azaz most 9 sort jelent. Minden sor

egy tagsági függvényt tartalmaz, azaz a trapéz 4 töréspontját. Ilyen módon minden szabálybázisfájl egyértelműen azonosít egy rendszert, mert tartalmazza a változók határait, és az összes tagsági függvényt.

Number of rules : 10 Inputnumbers : 8

515557.625000	516063.031250		
285902.875000	286455.187500		
2.000000	4.000000		
7.000000	7.700000		
30.000000	53.000000		
100.000000	400.000000		
50.000000	180.000000		
0.000000	9.200000		
4.000000	10.000000		
515575.843750	515601.000000	515876.375000	515946.156250
286010.843750	286296.000000	286375.968750	286428.000000
2.145573	2.316904	3.516953	3.800104
7.138154	7.219056	7.374386	7.546870
36.463333	44.240669	49.487564	50.637318
104.239021	199.786377	312.665192	391.897339
68.079468	84.397415	92.732933	118.398087
0.854945	4.939030	6.590802	9.088534
4.184210	4.396618	5.264565	6.010560

.....

A szimulációs futtatások során keletkezik egy másik fájl is, ami szintén a szabálybázist tartalmazza, csak olvashatóbb formában. Ebben a fájlban a változók intervallumai nem találhatók meg, csak a szabályok. A fájl szerkezete:

Rule 1 :

*If input1 is MF1-1(515575.843750, 515601.000000, 515876.375000, 515946.156250) and input2 is MF1-2(286010.843750, 286296.000000, 286375.968750, 286428.000000) and input3 is MF1-3(2.145573, 2.316904, 3.516953, 3.800104) and input4 is MF1-4(7.138154, 7.219056, 7.374386, 7.546870) and input5 is MF1-5(36.463333, 44.240669, 49.487564, 50.637318) and input6 is MF1-6(104.239021, 199.786377, 312.665192, 391.897339) and*

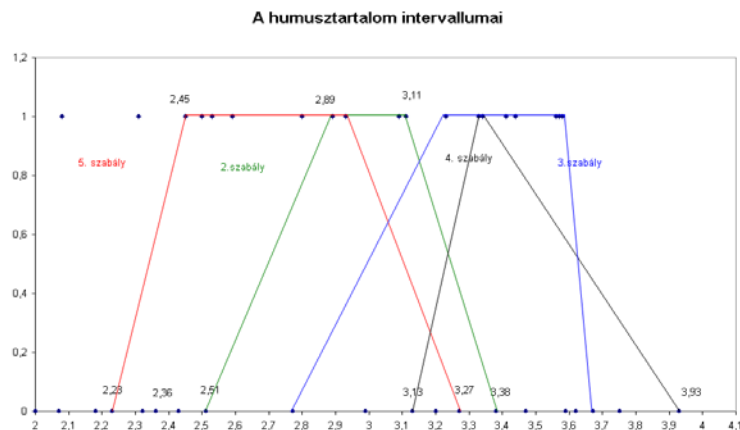
*input7 is MF1-7(68.079468, 84.397415, 92.732933, 118.398087) and input8 is MF1-8(0.854945, 4.939030, 6.590802, 9.088534) then output is MF1-9(4.184210, 4.396618, 5.264565, 6.010560)*

A kapott szabályok megfelelnek egy speciális függvénydefiníciónak, fuzzy jellegű függvénydefinícióról van szó. A szimulációs eredményben az "y=Bi1 ha az 1. szabály, Bi2, ha a 2.szabály..." nem éles határokat tartalmaz, hanem fuzzy átfedéseket. Vagyis azt jelenti, hogy a kimenet mekkora része jön az 1. szabályból, mekkora része jön a 2. szabályból stb...

A szabálybázis tulajdonképpen egy fuzzy reláció, de egyváltozós kimenet esetén a reláció függvényt jelent. A szabályok kimenete független egymástól. A szabályokat, mint egy relációt fejeznek ki, így egy szabály, egy "X" fuzzy vektorhoz egy "Y" fuzzy számot rendel. Tehát egy fuzzy pontot ír le egy szabály, és a pontok között nincs kapcsolat.

A döntéshozatalban a szabálybázis minden szabálya részt vesz, felhasználva a tagsági függvényeket és a bemenetekből nyert igazság értékeket. A szabályok értékeit leképezzük a kimeneti értéket meghatározó tagsági függvényre és az igazság értékre. Majd ezek az értékek kombinációja adja a kimenet (crisp = éles, konkrét) értékét. Ez az eljárás a defuzzyfikálás

A szabályokban levő tagsági függvények lefedik a változók teljes intervallumát. Az 4.4.4 ábrán a pontok jelentik a humusztartalom tagsági függvényeinek töréspontjait. Az 5. szabály 1. változójának értelmezése: ha a humusztartalom  $x < 2,23 \rightarrow \mu(x) = 0$ ,  $x \in [2,23; 2,45] \rightarrow \mu(x) \in (0,1)$ ,  $x \in [2,45; 2,94] \rightarrow \mu(x) = 1$ ,  $x > 2,94 \rightarrow \mu(x) = 0$



#### 4. 4.4 ábra A szabályok tagsági függvényei

Vizsgáljunk meg egy 6 input változós modell 5. szabályának kiértékelését:

If input1 is MF5-1(2,230781; 2,457167; 2,938444; 3,270547) and input2 is MF5-2(7,117176; 7,302157; 7,302734; 7,472100) and input3 is MF5-3(33,845852; 44,673054; 45,487289; 52,863827) and input5 is MF5-4(121,927551; 208,484146; 296,569733; 347,749268) and input5 is MF5-5(55,46087; 87,242043; 99,842522; 100,453506) and input6 is MF5-6(1,389813; 3,38,3001; 4,596771; 6,279708) then output is MF5-7(8,182257; 8,716941; 8,990875; 9,657582)

Ha a feltételben X1 változó a [2,457167; 2,938444] intervallumba esik, az X2 változó [7,302157; 7,302734] intervallumba ...és X6 változó [3,38,3001; 4,596771] intervallumba – „saját trapézának középső részébe esik”, akkor a kimeneti Y értéke a [8,716941; 8,990875] intervallumba kerül ( itt a tagsági érték = 1). Mindegyik szabály leír egy intervallumot.

Ha X1 változó értéke (humusz)  $\approx 2,37$  (tehát a trapéz baloldali szárának közepe,  $\mu(x)=0,5$ ) a többi változó marad a trapéz középső

részében  $\mu(x)=1$ , akkor a kimenet a [8,716941; 8,990875] intervallumba  $\mu(x)=0,5$  tagsági értékkel tartozik („félíg igaz”). Az illeszkedési mértéket a minimum operátorral határoztuk meg.

A 10 szabály együttesen választ adott arra, hogy a bemenetek alapján a kimenetek hibája a szimulációnál 12,3%, az együttes közelítés eredménye.

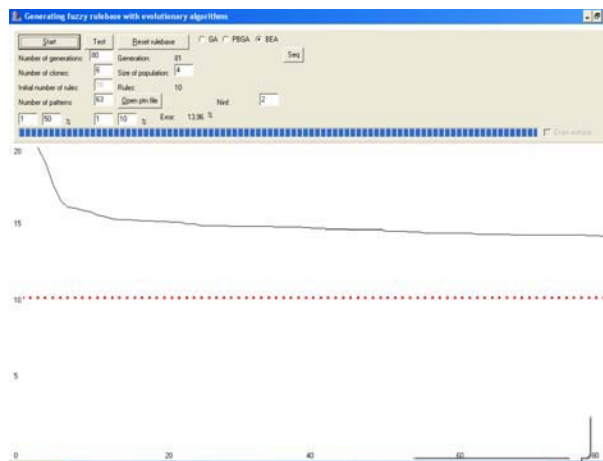
Az egyes futtatások eredményeit (paraméter beállítások és hibaérték) a következő táblázatban összefoglaltuk:

	$N_d$	$N_{clones}$	$N_{inf}$	$N_{gen}$	$N$	Hiba [%]
1. szimuláció	4	6	2	40	10	13.64
2. szimuláció	4	6	2	80	10	13.96
3. szimuláció	4	8	2	80	15	13.34
4. szimuláció	6	6	2	80	15	13.38
5. szimuláció	8	8	3	80	20	11.80

A létrehozott szabályokat tovább tanítottuk, ezzel pontosabb szabálybázist hoztunk létre. Csak ugyanolyan szerkezetű adatokkal folytatható a tanítás, azaz meg kell egyezni a dimenziószámnak (pl.6), a változók jelentésének és a változók intervallumának is.

A szimulációs futtatások mutatják, hogy legjobb modell hibája is meghaladja a 10%-ot. Ebben a vizsgálati sorozatban 5 szimulációt hajtottunk végre: az induláshoz képest (ld. 1.szimuláció) kétszeresére növekedett az induló egyedek száma, a generációk száma és a szabályok száma az 5. szimulációban. Ezek a változtatások  $\approx 2\%$  javulást eredményeztek a hibában. Ez tovább már ebben az összeállításban nem javítható. Ha növeljük vagy az

egyedszámot, vagy a generációs számot, vagy a szabályok számát, akkor jobb eredményt kaphatunk (ilyenkor a számítási igény is nagyobb lesz). A klónok számának megtalálása nem egyértelmű. Több klónnal látszólag jobb eredményt érhetünk el, de egy idő után a klónok számának növelése már nem eredményez számottevő javulást (a globális minimum helyett a lokális minimumba kerülés veszélye). Nem könnyű az optimális infekciós szám megtalálása sem. A szimulációk alapján a 2 tűnik megfelelő értéknek.



#### 4. 4.3 ábra A hiba alakulása a szimulációban

Fontos kérdés a változók terjedelme. Érdekes a lehető legszűkebb intervallumokat megadni, ami a változók tartományát jelenti, de érvényesnek kell lennie az egész vizsgálandó területre. A szabályokban szereplő trapézokat ezen intervallumok alapján hozzuk létre. A cél, hogy a trapézok a lehető legjobban fedjék az intervallumokat.

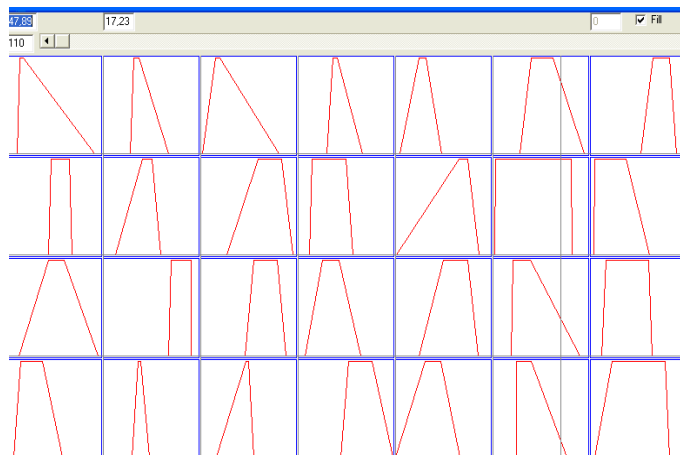
A másik gond a helykoordinátákkal volt. Kezdetben a GPS hálózat rendszerében megadott formátumú koordinátákkal dolgoztunk. Ebben a formában a számok első jegyei nem hordoznak információt. (mindegyike 47.90-nel, ill. 17.25- tel) kezdődtek, és csak utána jönnek az "értékes" számjegyek. Ezért itt is a másik fajta kódolást az EOV rendszert használtuk, ezek már több jegyben különböznek, a nagyságrendjük is egészen más (515587.33, 286263.22).

Az adatok alsó és felső korlátjának (a tartományának) kis mértékű módosításával újabb szimuláció vizsgálatokat végeztünk a 2002. évi kukoricaadatok elemzésével. Ez a szabálybázis jobb közelítést adott, mindegyik szimuláció hibája 10% alatt van.

	$N_{ind}$	$N_{clones}$	$N_{inf}$	$N_{gen}$	$N$	Hiba [%]
1. szimul.	4	6	2	40	10	6.425
2. szimul.	4	6	2	20	5	9.198
3. szimul.	4	6	2	60	5	8.386
4. szimul.	4	6	2	40	7	8.139
5. szimul.	2	6	1	40	7	9.088
6. szimul.	2	6	1	40	10	4.869
7. szimul.	4	6	1	40	10	4.352
8. szimul.	4	6	1	40	7	6.212
9. szimul.	4	10	1	40	7	6.322
10. szimul.	4	4	1	40	7	7.396

A 4.4.4 ábrán az algoritmus által szolgáltatott szabálybázis egy részlete látható, szintén a 4. szimulációs futtatás során.





**4.4.4 ábra A 6 változós szabálybázis részlete**

A következő szimulációban a 2004. évi őszi búza adatok alapján állítottuk össze a szabálybázist. Két változatot vizsgáltunk meg: A térbeli interpolálásnak megfelelően a helykoordinátákat is bevettük az input változók közé, és ezáltal 8 dimenziós problémánk volt. Ez a modell alkalmas arra, hogy a tábla egész területét a helykoordinátákkal folytonosan lefedjük Így a létrehozott szabálybázissal egy többváltozós interpolálást valósítottunk meg. A másik összeállítás nem tartalmazza a helykoordinátákat, 6 változós problémát elemeztünk. Ennek a modellnek a felhasználása kiterjeszhető más táblákra is, ha a vizsgálandó táblán a változók száma és tartománya az összeállított szabálybázis követelményeinek megfelel. Az adatállományok szerkezete a fentebb leírtak alapján elemezhető:

*8 dimenziós:*

8	63	
515557.631		516063.019
285902.888		286455.181
2	4	
7	7.7	

30	53						
100	400						
50	180						
0	9.2						
4	10						
515587.33	286263.22	2.5	7.29	46	216	75	
	8.78	6.48					
515629.42	286286.97	3.3	7.24	48	320	85	
	0.03	7.14					

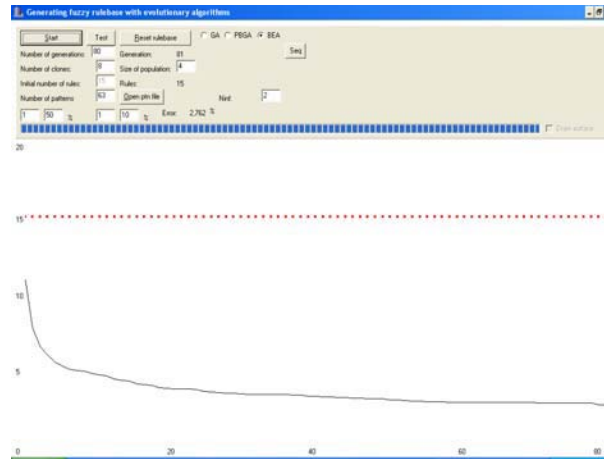
A vizsgált 5 szimulációs futtatásban szereplő paraméterek beállításait, és a hibaértékeket az alábbiakban foglaltuk össze:

	$N_{ind}$	$N_{clones}$	$N_{inf}$	$N_{gen}$	$N$	8 dimenzió	6 dimenzió
1.	4	6	2	40	10	5.312	4.277
2.	4	6	2	80	10	5.800	4.859
<b>3.</b>	<b>4</b>	<b>8</b>	<b>2</b>	<b>80</b>	<b>15</b>	<b>2.762</b>	<b>2.378</b>
4.	6	6	2	80	15	2.805	2.515
5.	8	8	3	80	20	1.367	2.630

Ezek a futtatások már nagyon jó közelítést adtak mindkét dimenzióban, ebből is kiemelkedik a 3. szimuláció, amelyben 4 egyeddel, 8 klón, 2 infekció és 80 generáció 15 szabállyal leírja a búza termés hozamának tartományát, a megfigyelt érték és szimulációval becsült értékek hibája 2,762% ill. 2,378% (4.4.5 ábra).

Hasonlóan jó eredményt kaptunk a 2003. évi árpa adatainak összeállításánál, a legjobb szabálybázis hibája 4.869% volt.

A vizsgálatok azt mutatták, hogy a modell jóságát befolyásolják az input adatállomány jellemző tulajdonságai.



#### 4.4.5 ábra A szimuláció hibája

A tavaszi árpa és a az őszi búza termésének változékonysága lényegesen kisebb (18%), mint a 2001. és 2002. évi kukorica hozamoknak (42% és 35%).

A szimuláció eredményeként létrehozott szabályokkal további elemzéseket végeztünk

A vizsgálatunk célja egy alkalmas függvény meghatározása, ami a megadott változók (talajtulajdonságok és terméshozamok) közötti kapcsolatot leírja. Az alkalmazott Mamdani következtetés felfogható úgy, mint egy sajátos interpolációs eljárás,  $A \rightarrow B$  ( $A$  implikálja  $B$ -t) kifejezést nézzük, ez Mamdani esetben egy  $(A, B)$  fuzzy pontpárnak felel meg.

A szimulációval összeállított szabálybázist újabb minták kiértékelésére alkalmaztuk.

A már megtanított szabálybázissal egy beérkező mintára (melynek nem ismerjük a kimenetét), kiszámítottuk a kimenetét a fuzzy következtető algoritmussal (pl. Mamdani). A 2001. évi kukorica

adatok alapján összeállított szabálybázist kipróbáltuk a mintatábláról választott tesztadatokkal, a becslés eredménye 13%-os hibát tartalmaz.

A másik fajta kiértékelés, amikor a megtanított szabálybázist egy másik mintafájlon, a tesztfájlon vizsgáltuk. Ennek során a tesztfájlból szereplő minden egyes bemenetre, a 6 input változó értékeivel számítottunk egy kimenetet. Ezt műveletet egy mintafájtra, a használt szabálybázis szerkezetének megfelelő adathalmazra elvégeztük, ezt összehasonlítottuk a tesztminta kívánt kimenetével, és ezek alapján számoltuk a szimulációs algoritmusban használt hibaátlagot.

A vizsgálatok azt mutatták, hogy a modell jóságát befolyásolják az input adatállomány jellemző tulajdonságai. A tavaszi árpa és az őszi búza termésének változékonysága lényegesen kisebb (18%), mint a 2001. és 2002. évi kukorica hozamoknak (42% és 35%).

A szimulációval összeállított szabálybázist újabb minták kiértékelésére alkalmaztuk. A már *megtanított* szabálybázissal a beérkező mintára elvégeztük az interpolálást, 87-93%-os pontossággal megbecsültük a hozamértékeket.

Ha kiértékelésre alkalmazzuk a szabálybázist, a változók tartományára –a minimum és maximum korlátokra ugyanolyan értékeket kell megadni. A szabálybázis felépítése során minden változónak van egy intervalluma, és a szabályokban szereplő tagsági függvények ezekre a tartományokra vonatkoznak.

Mennyi lehet az input változók maximális száma modellben, ez egyike azon kérdéseknek, amelyekre keressük a választ. A következő általános vélekedések vannak:

1. Minél nagyobb a változók száma, annál rosszabb lesz a közelítés pontossága.
2. Az algoritmus paramétereinek növelésével, pl. sokkal több szabály alkalmazásával a közelítés pontossága javítható, ami a hiba csökkentését jelenti.
3. Kompromisszumot kell kötni a modell komplexitása - jelen esetben a *szabályok száma*, vagy az alkalmazott tagsági függvények száma és a közelítés pontossága között. A modell komplexitása a számítási idővel is összefüggésben van, hiszen minél nagyobb a komplexitás, annál tovább tart a megoldás kiszámítása. Azaz ha növeljük a modell méretét, akkor hasonló pontosság eléréséhez több szabályra van szükségünk, de ez maga után vonja a számítási idő növekedését.

Ezen törvényszerűségek értelmében, ezzel a módszerrel 8-9 input változóval tervezhetünk. Ez függ a minták változékonyságától is (eloszlás, szórás), tehát problémafüggő.

Ha jelentősen szeretnénk megnövelni a változók számát, akkor már mindenképpen a hierarchikus fuzzy szabálybázisokat célszerű alkalmazni (Kóczy, 2000).

## 4.5 Terméshozam becslése neurális hálóval

A termelés sikere függ, hogy milyen részletes ismeretekkel rendelkezünk az adott tábláról. Nagy jelentőségű annak ismerete, hogy milyen kapcsolat van a táblán belül a terméshozam és a talajtulajdonságok között. Az első lépés ebben a folyamatban olyan technika keresése, amely képes a függvénykapcsolatot azonosítani a mért talajtulajdonságok, a helyi jellemzők és a terméshozam között. Öt év hozamait és a megfelelő helyi és talajtulajdonságokat tanulmányoztuk. A többrétegű, előrecsatolt hálóval a megadott minták alapján a háló tanításával feltártuk a bemenő adatok és a kimenő értékek közötti kapcsolatot. Segítségükkel az adott minták alapján elkészítettük a terméshozam és a meghatározó tulajdonságok kapcsolati modelljét.

### 4.5.1 Az adatok előkészítése és elemzése

Az alkalmazás legfontosabb szakasza, mivel az eredmény pontossága a legtöbb hálónál függ a tanuló adatok minőségétől és mennyiségétől. Bár a neurális háló elfogadja az input adatok széles választékát, csak numerikus formában kódolt adatokkal dolgoztunk. Két fő vizsgálati szempont az adat előkészítésben: a *vizsgált változók száma*, amely meghatározza a dimenziót, a határozottság vagy *adatfelbontás*, a hálóban szereplő adatok terjedelme és nagysága. Az adat-előkészítő folyamatban az alábbi műveleteket végeztük el:

- a) Az adatok **típusának** meghatározása (diszkrét vagy folytonos)

- b) *Adatok generálása*: a tanuló adatokat lehet generálni méréssel, lényeges modellek szimulálásával vagy virtuális példák származtatásával.
- c) *Egyszerű statisztikák számítása*: átlag, szórás a folytonos változóknál, különböző események száma a diszkrét adatoknál.
- d) *Kiugró adatok kizárása*: azokat az adatpontokat vettük figyelembe, amelyek az átlagtól kétszórásnyi távolságon belüliek, ez a normális eloszlású adatok 95%-át fedi le.
- e) *Adatok minőségi és mennyiségi vizsgálata*: az összegyűjtött adatok mennyiségét főleg a háló mérete (a változók száma) és a kívánt felbontás határozza meg (Sárközi, 2001). A minőség ellenőrzés céljából végrehajtott statisztikai tesztek biztosítják, hogy olyan adathalmazt készítettünk, amelyek a kívánt információt tartalmazzák.
- f) *A dimenzió csökkentése*: ha a változók száma túl nagy, megnöveli a tanulási időt. Kívánatos csökkenteni az input változók számát, ameddig a háló bonyolultság megfelelő szintű. Kiválasztottuk azokat a változókat, amik legfontosabbak és legjobban befolyásolják a kimenetet. A változók kovarianciája és korrelációja segít meghatározni a leghasznosabb változókat (1-2. melléklet).
- g) *Elkészítettük az adatok skálázását*. Ha a változóhalmazok terjedelme túlságosan különböző, kívánatos az adatokat a skála ugyanazon a tartományába transzformálni, hogy a

normált bemenő adatok az aktiváló függvény hatásos szakaszára esznek.

- h)** *Az adatok kódolása* az adat előkészítésének utolsó fázisa, a szükséges esetekben. pl. kategória adatokat mindig kódolni kell.

#### ***4.5.2 Terméshozam és a talajadatok kapcsolata***

A vizsgálat első fázisában a hálók tartalmazták a talajtulajdonságokat – pH, szerves anyag, kötöttség, N, P, K, Mg, talajjellenállás – mint bemeneteket, és a hozamot (kimenet) becsültük meg a mintavételi pontokban. Az így összeállított háló ígéretes segítség a termés változékonyság megértésében, de ezeket a modelleket is tovább kell tökéletesíteni a pontosság növelésére.

##### *1. lépés: Az input tényezők elemzése*

A neurális háló szerkesztését az inputok meghatározásával kezdtük. Az input tényezők kiválasztása agronómiai ismereteken alapul, és fokozatos megközelítéssel a háló futtatásán keresztül pontosítottuk. Az adatelemzéssel kiszűrtük az adatanomáliákat – a hiányzó, a kiugró és a hibás adatok kiválogatását jelentette.

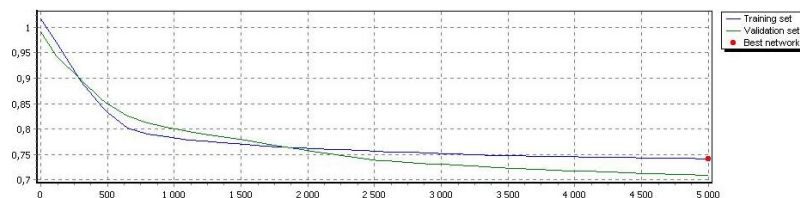
A kiugró értékek gátolják a neurális háló tanulását, és szignifikánsan csökkentik a háló hatékonyságát. A vizsgálat eredményeként a talaj K tartalma (K<sub>2</sub>O) változónál a 162 ppm értékű adatot, és a Mg tartalomnál 18 ppm értéket találtuk szignifikánsan különbözőnek, kizártuk a további értékelésből.

##### *2. lépés: A tanuló, értékelő és tesztadatok előkészítése*



Véletlen particionálással három részre osztottuk az adathalmazt. A tanuló adathalmaz 70%-t, az értékelő (validációs) és a teszt adathalmaz 15-15%-t képvisel.

A validációs halmazt a háló hangolására használtuk, a topológia vagy a hálóparaméterek, súlyok beállításánál. Kiprobáltuk a rejtett egységek számának meghatározásánál. A tanuló mintától függően elindultunk a bemenő változók számának feléről és a változók számának kétszereséig vizsgáltuk. 10 input változó esetén 5 rejtett egységgel indulva 20 egységig vizsgáltuk. Az értékelő halmazon a hálóhiba vizsgálatával megkerestük azt az állapotot, amikor a neurális háló teljesítménye romlott, így segített meghatározni a legjobb hálót. A 4.5.1. ábra mutatja, hogyan változik iterációról-iterációra a hiba értéke a tanuló és a validációs halmazon. Ebben az esetben az iterációs szám a legjobb hálónál 5000 volt.



#### 4.5.1 A tanulási folyamat hibagrafikonja

A teszhalmaz az input adatoknak azon részhalmaza, amelyet a háló kiprobálására használhatunk. Segítségükkel teszteltük, milyen hibák fordulhatnak elő a háló későbbi alkalmazásainál.

##### 3. lépés Az input változók kiválogatása, rangsorolása

A változók közül kiválasztottuk azokat az input jellemzőket, amelyek szignifikánsan hozzájárulnak a neurális háló hatékonyságához. Kizártuk a nem determináns inputokat, így

javítottuk a hálót, annak ellenére, hogy néhány input információt elveszítünk. Négy módszert próbáltunk ki. Az *előrelépegetésnél* (forward) egy-egy új változót von be, a *visszalépegetés* (backward) minden lépésben kizár egy inputot. Lehetőség van az inputok minden lehetséges kombinációjának tesztelésére – ez nagyszámú input változó és nagy adathalmaz esetén nagyon időigényes.

Az inputok kiválasztásánál a *genetikus algoritmust* alkalmaztuk. Ez az input konfiguráció véletlen populációjával indul. A módszer időigényes, de különösen jó, „*kölcsönösen*” szükséges inputok meghatározására, az egymásra utaltság felderítésére.

Az algoritmusban lényeges paraméterek: a *populáció mérete*: ha nagyobb a populáció, nagyobb a jó megoldások száma, de több idő kell.

Számításainkban a populáció mérete: 50

A *generációk száma*: a genetikus algoritmus lépéseit jelenti. Minden generáció azonos időintervallummal bír, amelyben előállítja az új populációt, a keresztezési és a mutációs ráta különböző kombinációjánál. Minden generáció az előző módosítása, megkísérelve a legjobb inputok kiválasztását. A generációk számának növelése megjavítja az optimális konfiguráció meghatározásának valószínűségét. Vizsgálatainkban 50 és 100 közötti generációszámmal dolgoztunk.

A *keresztezési ráta* a keresztezés valószínűsége minden generációban, amely során a szülői konfigurációkból új egyed (utódot) hoz létre. A nagyobb rátánál nagyobb a konvergencia

valószínűsége a hasonló utódhalmaznak, kisebb rátánál nagyobb generációs szám szükséges.

A mutáció a változtatási folyamat, az egyedeket bitről bitre megvizsgálja és megváltoztatja. A mutációs ráta a változtatás valószínűségét adja minden konfigurációnál és minden új generációban. Nagyobb mutációs rátánál jobbak az algoritmus keresési lehetőségei, de nagyobb a jó konfiguráció „összeomlási” lehetősége is.

A változókat binárisan kódolva a genetikus algoritmussal összeállítottuk a bemenő változók körét. *A populáció mérete = 50, a generációk száma = 100, a keresztezési ráta = 0,7 és a mutációs ráta = 0,4 esetén pl. a 2003. évi változók válogatásának eredménye: 1011100111-* ami azt jelenti, hogy a hozamot legjobban befolyásoló tényezőket jelöli 1, a kevésbé hatókat 0.

*pHKCl, Kötöttség, humusz, CaCO<sub>3</sub>%, AlP<sub>2</sub>O<sub>5</sub>, AlK<sub>2</sub>O, Mg, Zn, Mn, Talajjellen*

Ez a konfiguráció az öt legjobb összeállítást figyelembe véve átlagosan 63%-os megfelelést mutatott.

*4. lépés* Az adatok előkészítését és transzformációját jelenti a hálóba kapcsolás előtt. A neurális hálóban a neuronok műveleti értékeinek *korlátos* tartománya van, így a változók értékeit arányosítani kell, megfelelő skálázási tényezővel transzformáltuk

Az input változók a [-1, 1] tartományba kerülnek, a kimeneti skálatartomány függ a kimeneti réteg *aktivációs* függvényétől: *lineáris [-1, 1], logisztikus [0, 1], hiperbolikus tangens [0, 1],*

Megvizsgáltuk a lineáris és logisztikus függvény alkalmazása esetén, hogyan változott az eredmény jósága.

Az egyes változókat a skálázási tényezőkkel transzformáltuk. A szakirodalomban a regressziós neurális háló input adatainál az adatok kódolásánál alkalmazták a *maximum*, a *maximum-minimum*, a *logaritmus*, „*thermometer*” és a *bináris sémát*. O’Neal és munkatársai (2002) tanulmánya és megállapításai alapján a *maximum-minimum* transzformációt választottuk, ez jól használható a folytonos értékeknel:

$$SF = (ST_{\max} - ST_{\min}) / (X_{\max} - X_{\min})$$

$$X_{\text{transzformált}} = ST_{\min} + (X - X_{\min}) SF$$

$X$  – az aktuális változó értéke,  $X_{\min}$ ,  $X_{\max}$  – a változó minimuma, maximuma,  $ST_{\min}$  – a skálatartomány minimuma (-1 vagy 0),  $ST_{\max}$  – a skálatartomány maximuma (1),  $X_{\text{transzformált}}$  – az előkészített értéket jelöli.

Fontos szempont, ha előrejelzésre használjuk a hálót, akkor a maximális és minimális értéket úgy állítsuk be, hogy a tartományba beleférjenek az értékek:  $X_{\min}$ ,  $X_{\max}$  ( $\pm 10\text{-}30\%$ ).

##### 5. lépés A háló tervezése és a háló architektúra specifikálása

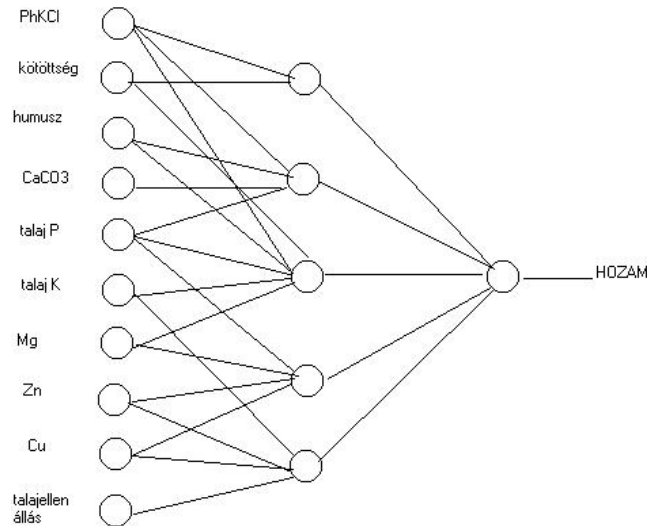
A háló konstruálása az input változókból kiindulva a rejtett rétegek számát, a rétegben levő egységek számát és a hálótulajdonságok (a hiba és az aktivációs függvények) meghatározását jelentette.

A tanuló minta létrehozásával már megterveztük a bemenő és kimenő rétegben található neuronok számát. Ezután már 'csak' azt kellett megterveznünk, hogy hány rejtett rétegünk lesz, hány neuron lesz az egyes rétegekben, és milyen *aktiváló függvényeket* alkalmazunk a rejtett rétegek neuronjaiban

Az elemek számának megbecsülésére eléggé ellentmondásos irodalmi adatokkal találkozunk. Ismert az állítás, hogy legalább annyi sornak (összetartozó ismert be és kimenő adatnak) kell lenni a tanuló adatfájlban, ahány súly van a hálóban. Egy másik javaslat alapján bármely folytonos függvény közelíthető olyan hálóval, amelynek egy rejtett rétege pedig  $N-1$  neuront tartalmaz, ahol  $N$  a tréning adatok száma. Ismeretlen  $f(x_i)$  függvény közelítő definíciójánál, nem az a cél, hogy megtanulja a tanuló adatok pontos ábrázolását, hanem az adatok generálási folyamatának statisztikai modelljét alkotjuk meg.

Nagyon kevés rejtett egység esetén csak nagyjából tárja fel az adatainkban a rejtett összefüggéseket. Túl sok rejtett egység esetén, a kapcsolatok megtalálása helyett memorizálja az adatainkat, zajt tartalmaz, így nagy hálóhibához vezet. Az induló hálóstruktúra kiválasztásánál a változók számának felével indultunk. Lépésenként kiszámoltuk a teszt hibát, a korrelációt, az illesztési (alkalmassági) értéket.

Az indulási konfiguráció: 1 rejtett réteg, ahol az egységek száma – *input változók száma* /2. Az induló háló konfiguráció: [10–5–1] (4.5.2 ábra).

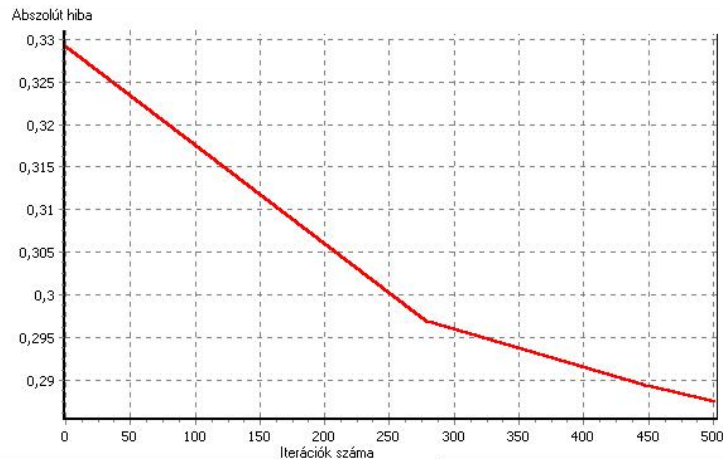


#### 4.5.2 ábra A tanuló háló a termésbecslésnél

A 10-5-1 szerkezetből kiindulva megvizsgáltuk, hogyan változik a háló struktúrája, az egyes összeállításoknak megfelelő hiba, az Akaike információs kritérium (AIC) és a korrelációs tényező. Az AIC az alkalmassági kritérium a háló építése során, a változásával teszteltük, a rejtett rétegek számának változása hogyan hat hálóra. Segítségével meghatároztuk az optimális hálószerkezetet. 4.5.1 táblázatban összefoglaltuk az egyes évek mintái alapján összeállított struktúrákat.

#### 4.5.1 táblázat Optimális hálóstruktúrák

Évek	Struktúra	Súlyok	Tanulási hiba	Validációs hiba	Teszt hiba	AIC	Korreláció
2001.	[10-15-1]	181	0,815	0,876	1,197	191,50	<b>0,67</b>
2002.	[10-13-1]	157	0,860	0,619	0,913	145,78	<b>0,69</b>
2003.	[10-15-1]	181	0,322	0,428	0,226	151,65	<b>0,53</b>
2004.	[10-14-1]	169	0,7	0,788	0,938	160,99	<b>0,78</b>
2005.	[10-16-1]	193	0,402	0,335	0,013	156,49	<b>0,73</b>



#### 4.5.3 ábra A kiválasztott háló (10-15-1) hibája

A leállítási feltételeket meghatároztuk a háló tanulásánál. Az első feltétel a *hiba értéke*. Ha a háló hiba a tanuló halmazon kisebb, mint az előírt hiba a tanulás leáll. A regressziós problémáknál specifikáltuk az átlagos abszolút hibát, vagy átlagos négyzetes hibát, az osztályozásnál definiáltuk a korrekt klasszifikációs rátát. A második feltétel a *hibaváltozás*. Amikor a háló hiba változása kisebb, mint az utolsó iterációban volt, a tanulás befejeződik. A hiba finomítása mutatja, hogyan viselkedik a neurális háló, hogyan javul az előrejelzési képessége. Ajánlatos volt megfigyelni azt a helyzetet, amikor a háló nem javul, a további tanulás haszontalan. Mintánként becslési hibát számoltunk (Liu et al., 2001) nyomán:

$$\text{Becslési\_hiba} = \frac{\text{Becsült\_hozam} - \text{Cél\_hozam}}{\text{Cél\_hozam}} * 100\%$$

Voltak alul és felül becslült értékek, hogy megkapjunk egy általános értéket az összes adatra és külön-külön az egyes részhalmazokra, meghatároztuk a négyzetes átlagos hibát:

$$RMS\_hiba = \sqrt{\frac{\sum_{i=1}^N becslési\_hiba^2}{N}}, \text{ ahol } N \text{ az adatok számát}$$

jelenti.

A harmadik feltétel az *iterációk száma*. Ez a legegyszerűbb és a legáltalánosabban használt feltétel - 500-12000 közötti értékekkel dolgoztunk.

A túltanulás azonosítása a validációs halmaz használatával történt. A túltanulás kezdőpontját úgy azonosítottuk, a háló hiba növekedni kezd a validációs halmazon, ugyanakkor a tanulási halmazon csökken.

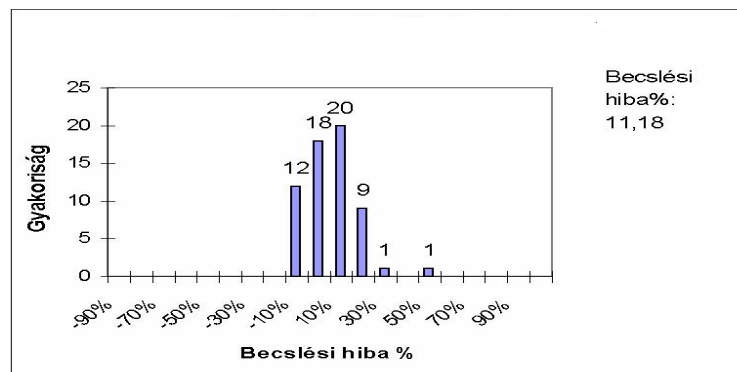
Megvizsgáltuk az egyes években a mintahalmazokra hogyan sikerült a háló tanítása. A 4.5.2 táblázatban összefoglaltuk a 2003. évi tavaszi árpa mintái alapján tanított háló statisztikai jellemzőit. A tanítás kimenetét a mintavételi helyeken becsült termés hozamokat, és azok hibáját a 3. melléklet részletesen tartalmazza. A 4.5.3 ábrán a hiba hisztogram mutatja az alul és felülbecsült értékek eloszlását.

A becslések jósága 88-94% között van. A tanított hálóval elvégeztük az érzékenységi elemzést, az input tényezők tartományainak  $\pm 10\%$ -os változtatásával. A számításokat elvégeztük külön-külön minden input tényezőre. A háló összekapcsolja a termés hozamot a mintapontonként ható tényezőkkel, így segíti a táblán belüli változékonyság megismerését.



**4. 5.2 táblázat A 2003. évi háló kimeneti statisztikája**

	CÉL	Háló Output	Abszolút hiba	Relatív hiba
<i>Tanuló</i>				
Átlag	3,913	3,911	0,320	8,3%
Szórás:	0,443	0,145	0,220	0,059
Min:	3,020	3,644	0,009	0,002
Max:	4,760	4,174	0,866	0,287
<i>Teszt</i>				
Átlag:	4,051	3,974	0,253	6%
Szórás:	0,303	0,164	0,235	0,053
Min:	3,560	3,745	0,003	0,001
Max:	4,630	4,210	0,705	0,158
<i>Validációs</i>				
Átlag	4,070	3,945	0,424	11,4%
Szórás	0,536	0,143	0,364	0,124
Min:	2,850	3,762	0,082	0,022
Max:	4,830	4,175	1,265	0,444
<i>Összes adat</i>				
Átlag	3,957	3,925	0,326	8,4%
Szórás	0,446	0,149	0,253	0,073
Min:	2,850	3,644	0,003	0,001
Max:	4,830	4,210	1,265	0,444



**4.5.3 ábra Termésbecslési hiba hisztogramja**

A következő lépésben a tanított hálóval megbecsültük az új inputok alapján a várható terméshozamot (4.5.3 táblázat). Ez a minta a 2003. évi hozam adatok alapján készült, az input adatokat behelyettesítettük a megtanított mintába és a becslőfüggvény alapján kiszámította a mintapontokhoz tartozó becsült hozamot. A mintában a várható hozam hibája közelítőleg 13% volt.

**4.5.3 táblázat A minta értékelése neurális hálóval**

hozam2003	Becsült	Hiba	RMSE
4,76	4,15	0,129	0,017
4,15	4,21	-0,014	0,000
3,96	3,98	-0,004	0,000
3,98	4,10	-0,030	0,001
3,02	3,89	-0,287	0,082
4,40	3,87	0,120	0,014
4,36	4,15	0,049	0,002
4,11	4,17	-0,016	0,000
4,31	4,05	0,061	0,004
4,37	4,08	0,066	0,004
3,97	3,88	0,023	0,001
4,15	3,94	0,050	0,003
2,85	4,11	-0,444	0,197
3,46	4,05	-0,170	0,029
3,81	3,88	-0,019	0,000
3,45	3,78	-0,095	0,009
3,53	3,91	-0,108	0,012
4,00	4,13	-0,034	0,001
4,00	4,21	-0,053	0,003
3,93	4,07	-0,036	0,001

A tanított háló végső súlyait elmentettük és ezekkel a súlyokkal a korábban összeállított háló topológia alapján kiszámítottuk a bemeneti értékekhez tartozó kimeneteket. Ez a számítás gyakorlatilag pillanatok alatt kész van. Következésképpen, ha a

jelenség, amit modelleztünk állandó és korlátos is, úgy a súlyokat bármikor újabb tanítás nélkül felhasználhatjuk az eredmények meghatározására.

A háló újratanításával elvégeztük a számításokat a többi növénynél is a vizsgált években. A 4.5.4 táblázatban a cél a mintaadatok átlagát, a háló output a becsült értékek átlagát jelzi a tanulási, a validációs és a teszt adatok halmazán. A háló átlagos teljesítménye évről-évre javult. A három év kukorica termését vizsgálva a konstruált „regressziós” függvénnyel 2001-ben 80%, 2002-ben 85 % és 2005-ben 97 % pontosságú volt a közelítés.

#### 4.5.4 táblázat A hálóbecslések statisztikája

Évek		Cél (kg)	Háló output (kg)	Standard hiba	Relatív hiba
2001	TRN	4259	4262	744	19%
	VLD	4342	4364	1005	25%
	TST	4532	4392	652	17%
	MIND	4233	4296	769	20%
2002	TRN	5314	5339	711	15%
	VLD	5446	5635	950	22%
	TST	6029	6121	707	12%
	MIND	5479	5498	746	15%
2003	TRN	3913	3911	320	8.3%
	VLD	4051	3974	253	6%
	TST	4070	3945	424	11.4%
	MIND	3957	3925	326	8.4%
2004	TRN	7041	7045	653	10%
	VLD	6713	7066	708	12%
	TST	7160	6900	920	12.5%
	MIND	7010	7027	700	10.5%
2005	TRN	10338	10502	330	3%
	VLD	10262	10256	268	3%
	TST	10343	10557	374	4%
	MIND	10329	10474	328	3%

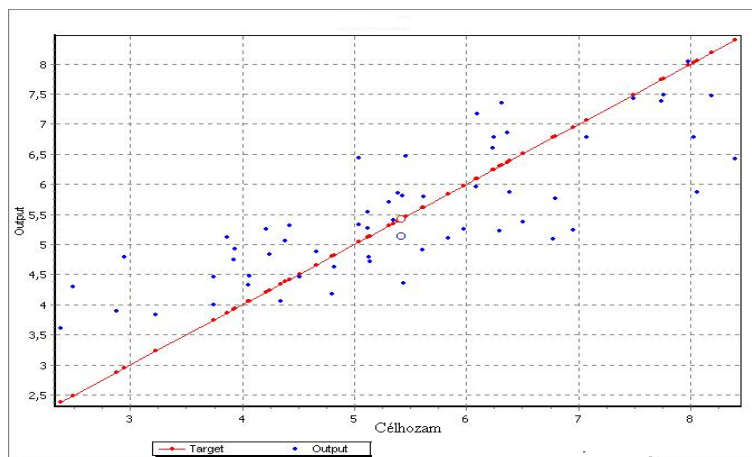
A kukorica hozamértékeinek becslésénél mindhárom vizsgált évben ugyanazokat talajváltozókat vonta be a tanításba, de a kifejtett hatások változtak (4.5.4 táblázat).

**4.5.5 táblázat A talajváltozók hatása a kukorica termés hozamára**

Input Változók	Hatás % 2001.	Hatás % 2002.	Hatás% 2005.
pHKCl	12%	21%	5%
kötöttség	25%	15%	22%
humusz%	8%	5%	3%
CaCO <sub>3</sub> %	7%	2%	10%
P <sub>2</sub> O <sub>5</sub>	2%	2%	6%
K <sub>2</sub> O <sub>5</sub>	5%	11%	3%
Mg(KCl)	7%	22%	31%
Mn ppm	33%	16%	7%
Zn (EDTA)	1%	5%	13%
Talajellenállás	0%	0%	0%

A hatás % vizsgálata segít eldönteni, hogy melyik változókat lehet kivonni a vizsgálatból. Ezt nagy körültekintéssel kell végrehajtani, mert a változók nem függetlenek, így a kölcsönhatás miatt csak együtt szolgáltatják ugyanazt az információt.

A háló összes elemére vonatkozó hibajellemző a 2001. évi adatoknál  $\approx 20\%$ , az adatállomány változékonysága ( $CV=42\%$ ) magyarázza. A 2002. évi adatoknál is hasonló kapcsolatot találtunk, de a hálót tovább lehetett tanítani. Elvégeztük a számításokat 6000, 12000 iterációra és a háló hibája tovább csökkent. Az összes adat hibája az első tovább tanításnál 15%-ra, majd 12%-ra csökkent. A mért és a hálóval becsült hozamértékek szóródását határoztuk meg a 4.5.4. ábrán.



**4. 5.4 ábra A 2002. évi becsült és mért hozam adatok szóródása**

Az adatok szóródása azt mutatja, hogy főleg a kisebb és a nagyobb terméshozam értékek tartományában vannak eltérések. A legjobb közelítést a 2005. év mintája alapján kaptuk, 97%-os pontossággal sikerült megbecsülni a hozamokat. A terméshozamok becslésének pontosságát befolyásolta a tanítás alapjául szolgáló minták heterogenitásának különbözősége.

A 2004. évi őszi búzánál a tanulóhalmazon a hiba 10%, a validációs és a teszt adatokon 12% az összes adaton 10,5%-os hibát számoltunk.

A terméshozam alakulásában a néhány vizsgált talajparaméter mellett számos tényező fontos szerepet játszik.

#### **4.5.3 A kukorica terméshozamának becslése – talajadatokkal és meteorológiai tényezőkkel**

A reális hozam becslése a tábla minden egyes részén a precíziós mezőgazdaság egyik kritikus problémája. A kukorica hozamára

ható tényezők, mint a talaj, az időjárás és technológia nagyon bonyolultak, így a hagyományos statisztikai elemzés nem ad pontos eredményeket. A neurális háló előnyös választás a nagy tömegű adathalmaz feldolgozásánál a precíziós mezőgazdasági termelésnél és kutatásnál.

A további kutatásban újabb adatforrások bevonásával célul tűztük ki átfogó neurális hálós modell felépítését– a hozam, a talaj, az időjárás és művelési tényezők, - nemlineáris függvénykapcsolatának közelítését. A hálós modell alkalmazását megkönnyíti, ha a kezdeti tanulást széles adatbázison sikerül végrehajtani, akkor az újratanulásnál, pl. egy másik táblán már lényegesen kisebb adatállományra van szükség.

A hiba visszaterjesztés (back-propagation) neurális háló univerzális közelítés, ha adott elegendő rejtett réteg, a többrétegű előrecsatolt hálóarchitektúra képes a bármely, kívánt pontosságú függvény közelítésére.

A változékonyság elemzésénél, egy adott termőhely értékelésénél segít, ha ismerjük az ún. évjáratok hatását.

Összegyűjtöttük és csoportosítottuk azokat az input tényezőket, amelyek legjobban befolyásolták a kukorica terméshozamát. 1995-2004. között 10 év adataival dolgoztunk.

A fentiekben vizsgált talajtényezők (pH, humusz, P, K, kötöttség) mellett megvizsgáltuk az időjárási tényezőket és a termesztési technológiát is (Varga-Haszonits et al. 1994).

*Időjárás tényezők: GDD:* a fejlődési szakaszban összesített hőegység, amelyet alábbi képlet szerint számoltuk- a kukoricánál a

vegetációs időszakban  $\approx$  április 20- október 20 közötti napokat jelentette, a konkrét vetési és betakarítási időpontokat ismerve.

$$GDD = \frac{Napi\ max.^{\circ}C + Napi\ min.^{\circ}C}{2} - 10^{\circ}C$$

*Az eső mennyisége:* A kukorica vízigénye függ a fejlődési fázisától, vízfogyasztása csúcsot ér el, a beporzás idején. Két héttel előtte és utána, érzékeny az aszályra, a szárazság ebben a periódusban komoly termés csökkenést okoz. Ennek megfelelően szakértői konzultáció alapján a különösen kritikus szakaszokban vettük figyelembe a csapadékmennyiségeket.

Májusi, júniusi, korai júliusi, késői júliusi, augusztusi, előző évi – a csapadék időszakainak szegmentálása a növény fejlődési periódusában a vetés időpontján alapult. A több éves tapasztalatot figyelembe véve a májusi eső a vetést követő 30 napot, a júniusi 30 napos, a júliusi két 15-15 napos periódust jelent.

Kiegészíthetők még a májusi hőmérsékleti adatokkal, és kritikus napok számával- amikor a beporzás és a szem kitöltési szakaszban a hőmérséklet  $35^{\circ}C$  fölött van.

Nagy szerepe van az alkalmazott termesztési technológiának, ezek vizsgálata és adatok gyűjtése újabb kihívást jelent a szakembereknek..

*Technológiai tényezők:* vetőmag szelekció, a kukorica genetikai potenciálja, N műtrágya, elővetemény, növényvédelem (gyomok, rovarok, kártevők), vetési sűrűség, vetésforgó, öntözés

Ha megfigyelések rendelkezésre állnak, bevonhatjuk az ún. *véletlen tényezőket* (rendkívüli, szokatlan események): jégverés, fagy,..stb.

Az első lépés volt a talaj, az időjárás és művelési tényezők közötti kapcsolati modell tervezése.

A háló tanításánál a kukoricatermesztésre vonatkozó minta adatait a szigetközi adatbázis és a részletesen elemzett tábla adatai szolgáltatták.

94 mintát vizsgáltunk az adott területről, az alábbi tényezőket vontuk be az elemzésbe:

*talaj pH, humusz, kötöttség, hőösszeg(GDD), májusi eső, júniusi eső, júliusieső1, júlieső2, összes kritikus időszaki eső, összes eső, elővetemény.*

Az adatok statisztikai vizsgálata és elemzése után, megfelelő skálázással és kódolással előkészítettük az inputokat a rangsorolásra. Az *elővetemény* tényezőt kategória változónak választottuk: a kukorica szempontjából: *rossz, közepes, jó* csoportokat képeztünk. A három elemű bináris kódolást választottuk: *001, 010, 100*.

A kísérleti ún. előzetes számításoknál a fentebb felsorolt tényezők közül egyértelmű kapcsolatot az inputok rangsorában a *talajtulajdonságok*, a *hőösszeg(GDD)*, a *májusi eső*, és az *összes kritikus időszaki eső* mutatott. Így ezek a tényezők adták a tanuló mintát. Az irodalom és szakértői tapasztalat alapján, a háló építésénél a mintahelyenkénti hozambecslésnél megismert módszert követtük. Ennél a vizsgálatnál a kimenő aktivációs függvénynél a lineáris és a logisztikus kapcsolatot is megvizsgáltuk. A mintahalmaz particionálása a számításához és ellenőrzéshez szükséges különböző részhalmazokra mindig



véletlen kiválasztással történt. A háló struktúra a tervezés eredményeként: [8-12-1] szerkezetű, a korrelációja = 0,98. Az aktivációs függvény változtatásával a háló illesztési hibája az összes adat esteén logisztikus függvénynél 11%, a lineáris aktivációs függvénynél 21% volt.

A kimeneten megjelenő hatás% alakulását az egyes hálóstruktúrákban a 4.5.6 táblázatban foglaltuk össze.:

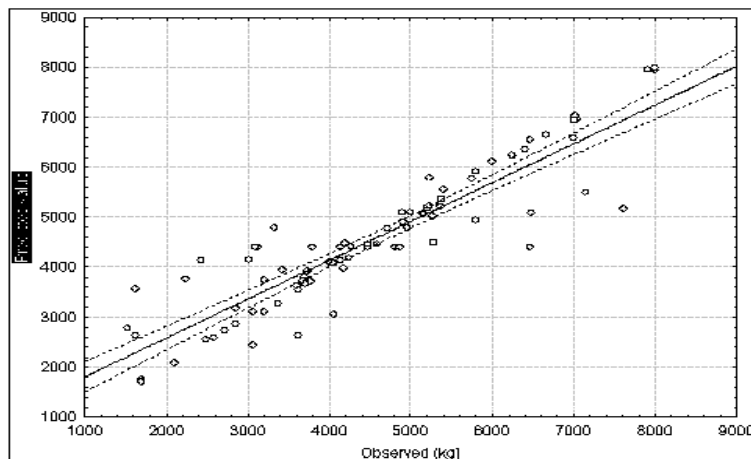
A tanított hálóban a talajtényezők  $\approx 40\%$ -t, a vizsgált időjárási tényezők  $\approx 60\%$ -t képviselnek

**4.5.6 táblázat Az input változók hatása a kukorica terméshozamára**

	Logisztikus függvény	Lineáris függvény
Input tényező	Hatás	
pH	16%	21%
Humusz	10%	7%
Kötöttség	14%	7%
Hőösszeg	5%	9%
Májusi eső	20%	24%
Összes kritikus eső	35%	32%

A 4.5.7 táblázatban összefoglaltuk a logisztikus illesztéssel megvizsgált adathalmazok becslési hibáit. A tanulóhalmaz becslésnek pontossága 92% és az egész mintahalmazon eléri a 89%-os pontosságot.

A becslt és a mért adatok szóródási görbéje (4.5.5 ábra) jó illeszkedést mutat.



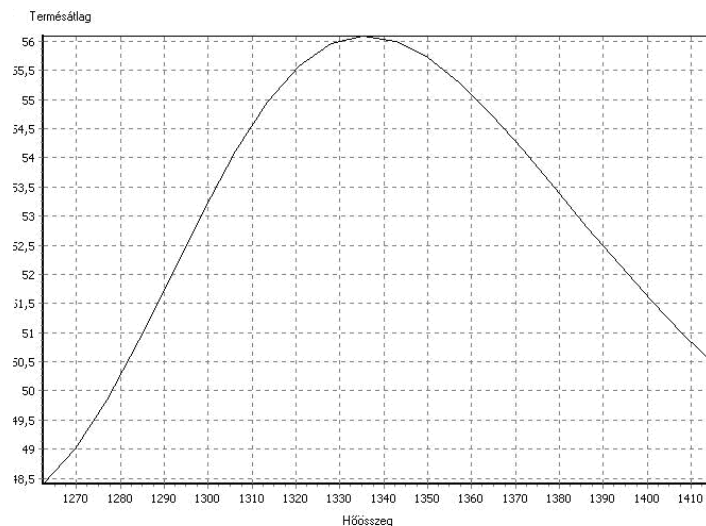
4. 5.5 ábra A becsült és a mért hozam szóródása (log)

4.5.7 táblázat A kukorica termésbecslése (1994-2005)

	CÉL	Háló Output	Abszolút hiba	Relatív hiba
<i>Tanuló</i>				
Átlag	5,52	5,49	0,436	<b>8%</b>
Szórás:	1,63	1,42	0,402	0,080
Min:	1,52	1,80	0,017	0,003
Max:	8,72	8,21	1,468	0,394
<i>Teszt</i>				
Átlag:	5,47	5,23	0,972	<b>21%</b>
Szórás:	1,44	0,93	1,072	0,303
Min:	3,01	3,91	0,074	0,015
Max:	7,61	6,34	3,024	1,005
<i>Validációs</i>				
Átlag	4,67	4,44	0,644	<b>14%</b>
Szórás	1,43	0,97	0,705	0,118
Min:	2,10	2,67	0,043	0,012
Max:	7,15	6,48	2,471	0,371
<i>Összes adat</i>				
Átlag	5,38	5,29	0,549	<b>11%</b>
Szórás	1,60	1,35	0,633	0,150
Min:	1,52	1,80	0,017	0,003
Max:	8,72	8,21	3,024	1,005

A többváltozós függvény illesztése mellett megvizsgáltuk az egyes bevont változók és a hálóval becsült hozam kapcsolatát.

A jól definiált és megfelelően tanított hálóval leírtuk az input tényező „hatásfüggvényét”. A hőösszeg és a termés hozam kapcsolatát jellemeztük hálóval becsült értékek segítségével. Ezt a függvénykapcsolatot mutatja a 4. 5.6. ábra.



#### **4.5.6 ábra A becsült termés és a hőösszeg kapcsolata**

Az ábra jól mutatja a telítődés menetét, jól leírja a hőösszeg és a termésátlag közötti kapcsolat jellegét.

Kipróbáltuk az összeállított háló csoportosítási képességét is: a termés hozamot három csoportba soroltuk: *alacsony* (4t alatt), *közepes* (4-6t) és *magas* (6t fölött) termésosztályba. A vizsgált és rangsorolt tényezők alapján megpróbáltuk megbecsülni a besorolást. A talaj és meteorológiai adatok alapján összeállított

input változókkal, a [6-4-3] háló topológiát kreáltuk, ezzel tanítottuk. A kimenetet a három terméskategóriába történő besorolás jelentette. Míg a függvénykapcsolat meghatározásánál a regressziós, a csoportosításnál a valószínűségi neurális hálóval dolgoztunk.

A kiválasztásnál a *legnagyobb aktivációs szint*, az ún. nyerő szint stratégiát választottuk, a hibát a kereszt-entrópia értéke mutatta. A csoportosítás eredményeként meghatároztuk a besorolás jóságát. A mintaadatainknál 86 %-os volt, ami azt jelenti, hogy az összes adat közül ennyit sorolt jól az ismert hozam alapján megállapított osztályba.

#### 4.5.8 táblázat Hozamkategóriák megfelelési mátrixa

	HÁLÓ OUTPUT		
CÉL	ALACSONY	KÖZEPES	MAGAS
ALACSONY	4	2	0
KÖZEPES	3	24	3
MAGAS	0	1	22

A besorolások jóságának eloszlását összefoglaltuk az ún. *megfelelési* mátrixban (4.5.8 táblázat), aminek a sorai és oszlopai tartalmazzák a célkategóriákat. A mátrix (i,j) cellájában azon mintapontok számát tárolja, melyek a cél *i. (sor)* és az output *j. (oszlop)* kategóriájába tartozik, ahol a főátlóba kerültek azon minták száma, ahol jó volt a besorolás. A besorolás jóságát újabb tényezők vizsgálatával lehet javítani.

A részletes besorolás eredményét a 4.5.9 táblázatban foglaltuk össze, feltüntetve a minta és a becsült hozam alapján várható besorolásokat.

**4.5.9 táblázat. A minta csoportok besorolása**

	ID	CÉL	Output	Megfelel?		ID	CÉL	Output	Megfelel?
TRN	0	közepes	közepes	<b>OK</b>	TRN	30	magas	magas	<b>OK</b>
TST	1	közepes	közepes	<b>OK</b>	TRN	31	közepes	közepes	<b>OK</b>
TRN	2	magas	magas	<b>OK</b>	TRN	32	közepes	közepes	<b>OK</b>
TRN	3	magas	magas	<b>OK</b>	TRN	33	magas	magas	<b>OK</b>
VLD	4	magas	magas	<b>OK</b>	TST	34	közepes	közepes	<b>OK</b>
TRN	5	közepes	közepes	<b>OK</b>	TRN	35	magas	magas	<b>OK</b>
TRN	6	közepes	közepes	<b>OK</b>	VLD	36	közepes	magas	<b>Rossz</b>
TRN	7	magas	magas	<b>OK</b>	TST	37	magas	magas	<b>OK</b>
TST	8	magas	közepes	<b>Rossz</b>	TRN	38	magas	magas	<b>OK</b>
TRN	9	magas	magas	<b>OK</b>	TRN	39	közepes	közepes	<b>OK</b>
TRN	10	alacsony	alacsony	<b>OK</b>	TRN	40	magas	magas	<b>OK</b>
TRN	11	közepes	magas	<b>Rossz</b>	TST	41	magas	magas	<b>OK</b>
VLD	12	alacsony	alacsony	<b>OK</b>	TRN	42	közepes	közepes	<b>OK</b>
TRN	13	közepes	közepes	<b>OK</b>	VLD	43	magas	magas	<b>OK</b>
TRN	14	közepes	alacsony	<b>Rossz</b>	TRN	44	közepes	közepes	<b>OK</b>
TRN	15	alacsony	közepes	<b>Rossz</b>	TRN	45	magas	magas	<b>OK</b>
TRN	16	közepes	közepes	<b>OK</b>	TRN	46	közepes	magas	<b>Rossz</b>
TRN	17	közepes	közepes	<b>OK</b>	TRN	47	magas	magas	<b>OK</b>
TST	18	közepes	alacsony	<b>Rossz</b>	TRN	48	magas	magas	<b>OK</b>
VLD	19	közepes	közepes	<b>OK</b>	VLD	49	közepes	közepes	<b>OK</b>
TRN	20	alacsony	alacsony	<b>OK</b>	TRN	50	magas	magas	<b>OK</b>
TRN	21	közepes	alacsony	<b>Rossz</b>	TST	51	magas	magas	<b>OK</b>
TST	22	alacsony	közepes	<b>Rossz</b>	TRN	52	közepes	közepes	<b>OK</b>
TRN	23	magas	magas	<b>OK</b>	TRN	53	közepes	közepes	<b>OK</b>
TRN	24	közepes	közepes	<b>OK</b>	TST	54	magas	magas	<b>OK</b>
TRN	25	közepes	közepes	<b>OK</b>	TRN	55	közepes	közepes	<b>OK</b>
VLD	26	közepes	közepes	<b>OK</b>	VLD	56	magas	magas	<b>OK</b>
TRN	27	magas	magas	<b>OK</b>	TRN	57	alacsony	alacsony	<b>OK</b>
VLD	28	közepes	közepes	<b>OK</b>	TRN	58	közepes	közepes	<b>OK</b>
TRN	29	közepes	közepes	<b>OK</b>					

## 5. KÖVETKEZTETÉSEK, JAVASLATOK

A kutatás középpontjában egy elemzési rendszer létrehozása volt, amely támogatja a termelők és a szakértők munkáját. Megvizsgáltuk, hogyan tudjuk komplex módon kezelni a helyspecifikus termelésben összegyűjtött adatok jellemzőit, a térbeli változékonyságot, a kapcsolatok nem-linearitását, a geostatistika és a mesterséges intelligencia módszereivel. Az agronómiai kutatás igényli a precíziós mezőgazdasági termelésben az adatok és információk új elemzési technikáinak bevezetését.

A térbeli változások nyomon követésére a variogramok alkalmasak, és nagyon jó elemző programok állnak rendelkezésre. Az adatállományokban bekövetkezett változásokat, korrekciókat is jól tükrözik

A krigelés pontosságára ható tényező a variogram modell jósága, célszerű a függvény tökéletesítése.

A fuzzy logika, a genetikus algoritmus és a neurális hálók elemeinek együttese nélkülözhetetlen intelligens eszközöket definiál a precíziós mezőgazdaságban.

A fuzzy logika kiterjeszti a talajtulajdonságok folytonos változékonyságát a földrajzi térről a tulajdonságtérre. A részleges átlapolást a fuzzy logika tagsági függvényeinek bevezetésével fejeztük ki. Ez a megközelítés mind az adatok becslési bizonytalanságának, mind a megjelenítésében alkalmas eszközt szolgáltat.

Az alkalmazott fuzzy-rendszer stratégiája nem más, mint a részben speciális hiányzó adatokat, vagyis a egyfajta “ködösséget”, bizonytalanságot a modellben képes tolerálni.

A kifejlesztett fuzzy ismeretalapú modellezés különösen hasznos azon a területen, ahol a komponensek között a kapcsolatok nem ismertek pontosan, vagy hiányosak, bizonytalanok az adatok a hagyományos statisztikai elemzéshez.

Felmértük a neurális háló becslési képességét a többéves adathalmazon, azonosítottuk azt a technikát, amely a mérési adataink alapján a legjobb függvényközelítést adta. A talajadatokat összekapcsoltuk a megfelelő meteorológiai adatokkal. A neurális hálók segítségével adott minták alapján jó pontossággal modelleztük a termés és a meghatározó tulajdonságok bonyolult kapcsolatát.

Az alkalmazott módszerek közelebb visznek a természeti jelenségek megértéséhez, jobban követik a természet folytonos jellegét. A dolgozat eredményei elősegítik a térbeli változékonyság jobb megértését, a térbeli becslés és a bizonytalanság kezelését. Megfelelő előkészítés és szakértői ismertek segítségével a fenti módszerek alkalmazásával, nagyobb pontossággal modellezhetjük a természeti folyamatokat.

## 6. ÖSSZEFOGLALÁS

A szántóföldön a termés hozam és a termesztési feltételek (pl. talaj termőképessége) tér és időbeli változékonyságot mutatnak. A precíziós mezőgazdaság számára fontosak azok az információk, melyek elemzik a talaj és a termés hozam térbeli változékonysága közti kapcsolatokat. A technikai gyakorlat és a gazdaságos hatásos kezelések a táblán belül kis területenként különbözhetnek, a helyspecifikus növénytermesztésnek megfelelően. A termőhely-specifikus termelés megvalósításának kulcskérdése az adott területről a megfelelő pontosságú információk begyűjtése és komplex elemzése. A precíziós gazdálkodás első szakasza azon információk összegyűjtése, amelyek leírják a termőhely tulajdonságait, a növény fejlődési feltételeit, elsősorban a talaj termőképességét meghatározó adatok.

Hagyományos statisztikai és geostatistikai módszerekkel megvizsgáltunk különböző adatállományokat: az 5 éven keresztül összegyűjtött hozam adatokat, az állandó talaj tulajdonságokat. Meghatároztuk azokat a jellemzőket, amelyek leírják a térbeli változékonyságot és jellemzik a különféle hibákat, torzításokat. Az elemzésnél figyelembe vettük, hogy nemcsak a mérési helyeken (mintavételezés) levő információra van szükségünk, hanem ezek alapján kell majd következtetnünk olyan helyek adataira, ahol nem volt mintavétel.

Azonosítottuk a fontos adatjellemzőket, a heterogenitást és a kiugró párokat (a variogram-felhő és szóródási térkép alapján),



vagy azokat a hatásokat, amelyek a torzítást/zajt okoztak a variogram becslésben. Az 5 éves hozam adatok térbeli változékonyságának elemzésére elkészítettük a térbeli korrelációt leíró kísérleti variogramokat, és az adatok szóródását leíró függvényt illesztettünk hozzá. Megvizsgáltuk és variogramokkal leírtuk az összegyűjtött talaj adatok jellemzőinek térbeli változékonyságát. A variogram paraméterek közül a térbeli struktúra jellemzésére a *röghatás/küszöb* (N/S) arányt és a térbeli korrelációt alkalmaztuk. Az N/S arány definiálja a kis tartományú változékonyságot. A térbeli korrelációs tartomány definiálja azt a távolságot, amelyen belül a tulajdonság értékei korrelációban vannak. Kicsi N/S és nagy korrelációs tartomány rendszerint jelzi, hogy nagyobb pontossággal lehet megtervezni a tulajdonság térképét.

A vizsgált terület minden pontjáról nem rendelkezünk adatokkal, ezért közelítő eljárásokat használtunk. A térbeli változékonyság mellett a másik fontos tényező, ami hat a pontosságra az *interpolációs módszer*, amellyel a diszkrét minta adatokat folytonos térképbe konvertáljuk. Két interpolálási módszer terjedt el a mezőgazdasági gyakorlatban: az inverz távolsággal súlyozott módszer és a krigelés. A módszer megválasztása nagyon függ a vizsgált adatok változékonyságától és térbeli szerkezetétől. Összehasonlítottuk, kereszt-értékeléssel és jackknife független teszteléssel a két interpolálási módszert, különböző adatállományokon. Az elemzés azt mutatta, hogy nagyságrendben nincs eltérés a két módszer által szolgáltatott

becslésben. Ha megfelelően előkészített adatállomány van, és a kiegészítő információk, egy jól szerkesztett, megbízható variogram (helyi és területi) rendelkezésre áll, akkor a krigelés jól alkalmazható becslés a nem mért területek előrejelzésében.

A becslés hibával terhelt, az előrejelzés és a kiértékelés bizonytalansága is jellemzi a kimenetet, amikor a terület jellemzését a talajtulajdonságok értékeiből, a térbeli előrejelzés módszereivel, az aktuális adatokból próbáltuk megbecsülni. A helyspecifikus talajtulajdonságoknál megvizsgáltuk, hogyan lehet az előrejelzés és értelmezés bizonytalanságát a fuzzy halmazokkal kezelni. A pontatlan, bizonytalan adatok használatánál, a fuzzy halmazban alkalmazott „tagsági értékek”, a tulajdonságok vizsgálatánál és a térbeli becslésnél alkalmazott fuzzy módszerek jól illeszkednek a talajértékelési módszerek csoportjába. A talajadatoknál a mérések száma nem elegendő a hagyományos becslés számára, így az adatokat pótoltuk „pontatlan” értékekkel. Ezen értékek pontatlanságát és bizonytalanságát kezeljük fuzzy halmazokkal, kombináltuk az éles és fuzzy adatokat egységes formában.

A fontos nemzeti és nemzetközi gazdasági programokban a termés becslése főleg a stratégiai növényeknél, mint a búza, a gabonafélék, a rizs hosszú ideje az érdeklődés középpontjában van. A statisztikai modellek főleg különböző regressziót alkalmaznak a termés számításánál. Ez bonyolult többváltozós nem-lineáris függvények illesztését jelenti, és megköveteli a kapcsolatok részletes ismeretét és leírását.

A fuzzy ismeretalapú modellezés különösen hasznos azon a területen, ahol a komponensek között a kapcsolatok nem ismertek pontosan, vagy hiányosak az adatok a statisztikai elemzéshez. A talajparaméterek és a hozamok közötti összefüggés meghatározását fuzzy szabály alapú rendszerrel 6 és 8 változóval vizsgáltuk meg. Összeállítottuk a mintaállományokat, ami alapján a tanítást végeztük. Az input adatok: a *helyi koordináták*( $x, y$ ), a talaj *humusz%*, *pHKCl* érték, *kötöttség*, *P*, *K*, *talajellenállás*, az output változó az évenkénti termés hozam volt. A kapcsolatot meghatározó optimális fuzzy szabálybázist kiválasztottuk az evolúciós algoritmussal. A szabályokban levő tagsági függvények lefedték a változók teljes intervallumát, és megfelelnek egy speciális fuzzy interpolációnak..

Az utóbbi években az neurális hálót olyan hatásos eszközzé fejlesztették, amely képes kiszámítani a bonyolult egyenletek és numerikus elemzések legjobb közelítését.

A talaj és hozamok térbeli változékonyságának és kapcsolatának elemzése után kísérletet tettünk a függvénykapcsolat meghatározására, a neurális hálók eszköztárát használtuk.

A háló összeállítása után évente elvégeztük az összeállított minták tanítását, így meghatároztuk a becsült termés hozamokat.

Újabb adatforrások bevonásával célul tűztük ki átfogó neurális hálós modell felépítését– a hozam, a talaj, az időjárás és művelési tényezők, - nemlineáris függvénykapcsolatának közelítését. A kísérleti ún. előzetes számításoknál a fentebb felsorolt tényezők közül egyértelmű kapcsolatot az inputok

rangsorában a *talajtulajdonságok*, a *hőösszeg(GDD)*, a *májusi eső*, és az *összes kritikus időszaki eső* mutatott. A mért és a hálóval becsült hozamértékek szóródását határoztuk meg. Megvizsgáltuk a háló csoportosítási képességét: a vizsgált jellemzők alapján 3 hozamkategóriába soroltuk.

A dolgozat eredményei elősegítik a térbeli változékonyság jobb megértését, becslését és a bizonytalanság kezelését. Megfelelő előkészítés és szakértői ismertek segítségével a fenti módszerek alkalmazásával, nagyobb pontossággal modellezhetjük a helyspecifikus folyamatokat.

A mesterséges látás, a vizuális technika fejlesztésével újabb adatforrások kapcsolódnak be a vizsgálatba. A növények állapotáról készített képek kiegészítik az összegyűjtött adatokat, és kiértékelésükkel a vizsgált területről újabb információkhoz jutunk.

Tervezzük a mesterséges intelligencia módszereinek összekapcsolásával a témához kapcsolódó neuro-fuzzy módszereken alapuló szakértői rendszer kidolgozását.

## 7. ÚJ KUTATÁSI EREDMÉNYEK (TÉZISEK)

1. *A hozam adatok táblán belüli térbeli változásainak nyomon követésére az összeállított variogramok alkalmasak, segítségükkel meghatározta a térbeli változékonyságot.*

2. *A szerző megállapította, hogy a termés hozamok a kukorica hároméves adatsorában szignifikánsan különböznek, de a térbeli korrelációjuk hasonló kiterjedésű. A hozamtérképek jó alapul szolgálnak a talaj-mintavételi stratégia kialakításánál*

3. *A vizsgált változókhoz (talajparaméterek, hozamértékek) egységesen használható elemző és modellező eszköz a fuzzy halmaz. A bevezetett tagsági függvénnyel jól kezelhető az adatok „pontatlansága” és bizonytalansága, a talajtulajdonságok és hozamérték folytonossága.*

4. *A különböző mértékegységű és nagyságrendű változókat a tagsági függvényeik segítségével közös skálára transzformálta, így lehetőség van a változók speciális elemzésére és összevonására.*

5. *Meghatározta a termés hozam térbeli 6 és 8 változós interpolációját a fuzzy szabálybázis alapú következtető rendszerrel. A hozam közelítésének pontossága 88% és 97% között volt a vizsgált táblán.*

6. *Az összeállított regressziós neurális hálókkal a minta adatok alapján történő tanulással közel 90%-os pontossággal megbecsülte a mért termés hozamokat.*

## 9. IRODALOMJEGYZÉK

*Acock, B.- Pachepsky, Y.* (1997) Holes in precision farming: mechanistic crop models. *Precision Agriculture* 1, 397-404.

*Amaral, J.- Tanscheit, R. – Pacheco, M. – Vellasco, M.* (2002) Evolutionary fuzzy system design and implementation (ICONIP'02), Vol.4.

*Arslan, S. - Colvin, T.* (2002) Grain yield mapping: yield sensing, yield reconstruction and error. *Precision Agriculture* 3, 135-154.

*Baker, W.* (1991) An application of neural networks to forecasting corn yields. MS Thesis, Dep.of Agricultural Economics, Purdue University

*Banai, R.* (1993) Fuzziness in geographical informations systems: contributions from the analytical hierarchical process. *Int. Jour. Geogr. Inf. System* 7. 315-329.

*Bárdossy A.-Bogárdi I.- Kelly, W.* (1990) Kriging with imprecise (fuzzy) variograms , *Math. Geology* 22, 63-79.

*Barsi Á.* (1999) Felszínborítottság térképezése neuro-fuzzy módszerrel GIS-környezetben.

*Bartels, F.* (1997) Ein Fuzzy-Auswertungs- und Krigingsystem für raumbezogene Daten, *Inst. Für Informatik und Praktische Mathematik, Universität, Kiel*, 92 pp.

*Bishop, T.–McBratney A.B* (2001) A comparison of prediction metods for the creation of field-extent property maps. *Geoderma* 103, 149-160.

*Blackmore, B.-Marshall, C.* (1996) Yield mapping; errors and algorithms. 3rd International Conference on Precision Agriculturae,Cranfield University

*Blackmore, B.* (2000) The interpretation of trends from multiplie yield maps. *Computers and Electronics in Agriculture* 26, 37-51.

*Blackmore, B.-Godwin, R.-S., F.* (2003) The analysis of spatial and temporal trends in. yield map data over six years, *Biosystems Engineering* 84, 455-466

*Blackmore, B.-Moore, M.* (1999) Remedial correction of yield map data. *Precision Agriculture* 1, 53-56.

*Borgulya I.* (1998) Neurális hálók és fuzzy rendszerek. *Dialog Campus*

- Borgulya I.* (2004) *Evolúciós algoritmusok.* Dialóg Campus, 303 p.
- Bosma, W.–Marinussen, M.–Van der Zee, S.* (1994) Simulation and areal interpolation of reactive solute transport. *Geoderma* 62, 217-231.
- Botzheim J. - Hámori B.- Kóczy T.L.* (2001) Extracting Trapezoidal Membership Functions of a Fuzzy Rule System by Bacterial Algorithm. *Fuzzy Days* : 218-227.
- Braimoh, A. - Stein, A.* (2004) Land evaluation for maize on fuzzy setand interpolation. *Environmental Management* 33. 2. 226-238.
- Burrough, P.A. - MacMillan, R.A. - van Deursen, W.* (1992) Fuzzy classification methods for determining land suitability from soil profile observations and topography. *J. Soil Sci.* 4,: 193-210.
- Burrough, P.A.*(1989) Fuzzy mathematical methods for soil survey and land evaluation. *J. Soil Sci.* 40, pp. 477–496.
- Buzás, I.,* (1993) *Talaj- és Agrokémiai vizsgálati módszerkönyv 1.* INDA kiadó, Budapest.
- Chang, L - Burrough, P.* (1987) "Fuzzy reasoning: a new quantitative aid for land evaluation." *Soil Survey and Land Evaluation*, 7, 69-80.
- Colosanti , R.* (1991) Discussions of the possible use of neural network algorithms in ecological modelling. *Binary* 3, 13-15.
- Cybenko, G.* (1989) Approximation by superpositions of a sigmoid function. *Mathematics of Control, Signals and Systems* 2 303-314.
- Csathó P. - Árendás T. és Németh T* (1998): New, environmentally friendly fertiliser advisory system, based on the data set of the Hungarian long-term field trials set up between 1960 and 1995. *Communications in Soil Science and Plant Analysis.* 29, 2161-2174.
- De Gruijter, J. J. - Walvoort, D. J. - van Gaans, P. F. M.* (1997) Continuous soil maps — a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models *Geoderma* 77, 169-195.
- Dillon, C.* (2002) A mathematical programming model for optimal management zone delineation in precision agriculture. American Agricultural Economics Ass.annual meeting, Long Beach
- Dimensions of precision agriculture.* (1997) National Academy Press, Washington, D.C.

- Dobermann, A. - Oberthur, T. (1997) Fuzzy mapping of soil fertility – a case study on irrigated riceland in the Philippines. Geoderma 77, 317-339.*
- Drummond, S. T. - Sudduth, K. A. - Birell, S. J. (1995) Analysis and correlation methods from spatial data ASAE Paper 95-1335.*
- Edwards, M.-Morse, D. (1995) The potential for computer-aided identification in biodiversity research. Trends Ecol. Evol. 10, 153-158.*
- Farkas B.- Kohán B. (2003) EEHHTT program (EUREF EO V Hivatalos helyi térbeli transzformáció )*
- Friderichs, M. - Fränzle, O. - Salski, A. (1996) Fuzzy clustering of existing chemicals according to their ecotoxicological properties. Ecological Modelling, 85. 1, 67-74.*
- Gang, L. - Jishuang, K. - Maohua, W. (1999) Research on Soil Sampling Strategy for Precision Agriculture Proceed of 99 Int. Conf on Agricultural Engineering*
- Goodchild, M.F. - Gopal, S. (ed). 1989. The accuracy of spatial databases. London: Taylor & Francis. xviii, 290 pp.*
- Goodwin, B. – Knight, T. – Coble, K.- Rejesus, R. – Shaik, S (2003) „ Current reference yield analysis report”*
- Goovers, P. (2000) Estimation or simulation of soil properties? An optimisation problem with conflicting criteria. Geoderma 97, 165-187.*
- Groenemans, R.-Van Ranst, E.-Kerre, E.(1997) Fuzzy relational calculus in land evaluation. Geoderma 77, 283-298.*
- Gutowicz, H.-Howard A. (1995) Artificial-Life Simulators and Their Applications  
(<http://alife.santafe.edu/alife/topics/simulators/dret/dret.html>)*
- Han S.-Schneider S.-Evans R.-Davenport J. (2004) Block Estimating of Spatial Yield Data and its Uncertainty Precision Agriculture 5, 73-84*
- Hatfield J. (2000) Precision Agriculture and Environmental Quality: Challenges for Research and Education. U.S. Department of Agriculture’s Natural Resources Conservation Service, [www.ars.gov](http://www.ars.gov).*
- Hellendoorn, H.-Thomas, C. (1993) Defuzzification in fuzzy controllers. Journal of Intelligent and Fuzzy Systems, 1(2):109–123,.*



- Heuvelink, G.* (1993) Error propagation in quantitative spatial modelling. Applications in geographical information systems. Doctoral thesis, Utrecht
- Heuvelink, G.- Burrough, P.A.* (1993) Error propagation in cartographic modelling using Boolean logic and continuous classification. *Int. J. of GIS* 7(3): 231-246.
- Heuvelink, G.- Webster R.* (2001) Modelling soil variation: past, present, and future. *Geoderma* 100, 269-301.
- Holland, J.* (1975) Adaptation in natural and artificial systems. Michigan Press, MI
- Höppner, F.-Klawonn F.- Kruse, R.* (1997) Fuzzy-Clusteranalyse Vieweg, Braunschweig, Wiesbaden
- Hufnagel L. – Gaál M.* (2002) Többváltozós állapot síkrendszerek alkalmazása valós és szimulált adatsorok kezelésében. In. *Proc. Agrárinformatika 2002*, Debrecen 315-325.
- Isaaks, E. H. - Srivastava, R. M.* (1989) An Introduction to Applied Geostatistics , Oxford University Press
- Kalmár S.-Pecze Zs.* (2000) Hozamtérkép készítése AgroMap Basic 3.0 programmal. *Növényvédelmi Tanácsok* 09, 1: 16-18
- Kaufman, L. - Rousseeuw, P.* (1990) Finding groups in data: An introduction to cluster analysis. Wiley, New York
- Kerry R. – Oliver M.* (2003) Variograms of ancillary data to aid sampling for soil surveys. *Precision Agriculture* 4, 261-278.
- Kirschfink, H. - Lieven K.* (1998) Basic tools for fuzzy modelling. Aachen
- Kóczy, L. - Sugeno, M.* (1996) Explicit functions of fuzzy control systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 4, 515–535.
- Kóczy, L. - Tikk, D.* 1998) Approximation of transfer functions by various fuzzy controllers. In L. Reznik, V. Dimitrov, and J. Kacprzyk, editors, *Fuzzy System Design: Social and Engineering Applications*, number 17 in *Studies in Soft Computing and Fuzziness*, pages 202–224. Physica-Verlag, Heidelberg–New York
- Kóczy, T. L. - Tikk, D.* (2000) Fuzzy rendszerek. Typotex Kiadó

*Koltai G. –Mike-Hegedűs F. –Palkovits G.- Schummel P. (2002): A kukorica terméseredményei a talajvízszint és a tápanyagellátás függvényében a Szigetközben, 2002. Növénytermelés 51, 581-593.*

*Koltai G. –Mike-Hegedűs F. –Palkovits G.-Schummel P. (2002): Az őszi búza terméseredményei a talajvízszint és a tápanyagellátás függvényében a Szigetközben, 2002. Növénytermelés 51, 61-69.*

*Kosko, B. (1995) Function approximation with additive fuzzy systems. In H. Nguyen, M. Sugeno, R. Tong, and R. R. Yager, editors, Theoretical Aspects of Fuzzy Control, pages 313–347. Wiley, New York*

*Kravchenko A. - Bullock D. (1999) A comparative study of interpolation methods for mapping soil properties. Journal of Agronomy 91, 393-400.*

*Kravchenko A. - Bullock D. (2000) Correlation of corn and soybean grain yield with topography and soil properties. Agronomy Journal 92, 75-83.*

*Kravchenko, A. (2003) Influence of Spatial Structure on Accuracy of Interpolation Methods Soil Sci. Soc. Am. J. 67, 1564-1571*

*Kruse, R. (1996) Fuzzy-Systeme Positive Aspekte der Unvollkommenheit . Informatik Spektrum, 19. 4-11EE'92), pages 1153–1162, San Diego, 1992.*

*Lark, R.M. - Bolam, H.C. (1997) Uncertainty in prediction and interpretation of spatially variable data on soils. Geoderma 77, 263–282.*

*Lark, R.M.(1998) Forming spatially coherent regions by classification of multivariate data: an example from analysis of maps of crop yield. Geographical information science, 12, 83-98*

*Lark, R.M. - Catt, J.A.; Stafford, J.V. (1998) Towards the explanation of within-field variability of yield of winter barley: soil series differences. J. Agric. Sci. 131, 409–416.*

*Lark, R.M. (1999) Use of ancillary data sources to aid spatial prediction of soil properties at within-field scales. In: Stafford, J.V., Editor, , 1999. Precision Agriculture '99. Proceedings of the 2nd European Conference on Precision Agriculture, SCI/Sheffield Academic Press, Sheffield, pp. 407–417*

*Lark, R. M. (2000) Designing sampling grids from imprecise information on soil variability, an approach based on the fuzzy kriging variance, Geoderma 98, 35-59*

*Lek, S. & Guéan J. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. Ecological Modelling 120, 65-73.*

- Lloyd, C. - Atkinson, P. (1998) Scale and the spatial structure of landform: optimising sampling strategies with geostatistics.*
- Marsili-Libeli S. (1989) Fuzzy clustering of ecological data, Coenoses, 4(2) 95-106.*
- McBratney A.B. - Webster, R. - Burgess, T.M. (1981) The design of optimal sampling schemes for local estimation and mapping of regionalised variables: I. Theory and method. Computer. Geoscience. 7, 331–334*
- McBratney, A. - De Gruijter, J. (1992) A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal Soil Sci. 43, 159-175.*
- McBratney A. B. – Whelan, B. M. (1995) Continuous models of soil variation for continuous soil management. In. Site-Specific Management for Agricultural Systems. Madison, 325-338.*
- McBratney A.B. - Odeh, I.O.A. (1997) Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma 77, 85–113.*
- McBratney, A. B. - Pringle, M.J. (1999) Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precis. Agriculture 1, 125–152*
- Mesterházi P. (2004) Development of measurement technique for GPS-aided plant production. PhD. Dissertation*
- Mesterházi P. Á. – Neményi M. – Kacz K. –Stépán Zs. (2002) Data transfer among precision farming system. ASAE 2002, 021047*
- Mitra, S.–Hayasi, Y.(2000) Neuro-fuzzy rule generation: Survey in soft computing framework. IEE Transactions on neural network 11. N. 3.*
- Moore, M. (1998) A strategy for better crop management based on yield mapping. AGCO Ltd. Coventry*
- Mueller, T.G. - Pierce F.J. - Schabenberger O. - Warneke D.D. (2001) Map quality fo site-specific fertility manegement, Soil Sci. Soc. Am. J., 65, 1545-1558.*
- Neményi M. – Mesterházi P. Á. – Pecze Zs. (2001) A preciziós helyspecifikus növénytermesztés műszaki és térinformatikai feltételrendszere. Növénytermelés, 50. 4.*

*Neményi M. – Mesterházi P. Á.* (2002) Hozam és talajellenállás térképezés és helyspecifikus tápanyag-visszapótlás a precíziós növénytermesztésben. *Agrárinformatika* 2002, Debrecen, 164-172.

*Neményi, M. - Mesterházi, P. - Pecze, Zs. – Stepán, Zs.* (2003) The role of GIS and GPS in precision farming. *Computers and Electronics in Agriculture* **40**, 45-55.

*Nieto, A.* (1999) Potential applicability of fuzzy logic in geostatistics.

*Noguchi, N. - Reid, J. - Zhang, Q. -Tian, L.* (1998) Vision intelligence for precision farming using fuzzy logic optimized genetic algorithm and artificial neural network. ASAE N. 983034

*O’Neal, M.R.- Engel, B. A.- Ess, D. R- Frankenberger, J. R.* (2002) Neural network prediction of maize yield using alternative data coding algorithms. *Biosystems Engineering* 83(1), 31-45.

*Odeh, I. - McBratney. A. - Chittleborough, D.* (1990) Design of optimal sample spacings for mapping soil using fuzzy-k-means and regionalized variable theory. *Geoderma* 47, 93-122.

*Odeh I. - McBratney A. B. - Chittleborough, D.* (1992) Soil pattern recognition with fuzzy c-means: application to classification and soil-landform interrelationship. *Soil Sci. Soc. Am.* 56, 505-516.

*Papritz, A. - Webster, R.* (1995) Estimating temporal change in soil monitoring: I. Statistical theory. *Eur. J. Soil Sci.* 46, 1–12.

*Pardo-Igúzquiza, E. - Dowd, P.A.* (1998) Maximum-likelihood inference of spatial covariance parameters of soil properties. *Soil Sci.* 163, pp. 212–219.

*Podani, J.*(1997) Bevezetés a többváltozós biológiai adatfeltárás rejtelmeibe. Scientia

*Recknagel, F.* (ed.) (2003) *Ecological Informatics: Understanding ecology by biologically – inspired computation.* Springer-Verlag

*Ross, T.J.* (1995) *Fuzzy Logic with Engineering Applications,* McGraw Hill, New York.

*Rossel, R. A. - McBratney A.B.*(1998) Soil chemical analytical accuracy and costs: Implications from precision agriculture. *Aust. Journ. Exp. Agric.* 38. 765-775.

*Rossiter, D. G.* (1996) A theoretical framework for land evaluation. *Geoderma* 72, 165-190.

- Russell, S. - Norvig, P. (2000) Mesterséges intelligencia modern megközelítésben. Panme-Pretice-Hall*
- Sacks, J. - Schiller S. B. - Welch, W. J. (1989) Designs for Computer Experiments. Technometrics, 31(1):41--47,.*
- Safa, B. - Khalili, A. - Tehnehlab, M. (2004) Artificial neural networks Application to predict wheat yield using climatic data*
- Salski A. - Kandzia, P. (1991): Einsatz der Fuzzy-Set-Theorie in der Ökosystemforschung. In: Hälker & Jaeschke (Hrsg.): Informatik für den Umweltschutz. Springer-Verlag, Informatik-Fachberichte 296: 303-310*
- Salski, A.-Sperlbaum C. (1991): Fuzzy logic approach to modelling in ecosystem research. In: B. Bouchon-Meunier, R. Yager & L.A. Zadeh (Eds.): Uncertainty in Knowledge Bases. Springer-Verlag, LNCS 521: 520-527*
- Salski, A. (1996): Fuzzy approach to ecological modelling and data analysis. Proc. of FUZZY'96 , Zittau, 316-325*
- Salski A. - Kandzia, P. (1996): Fuzzy sets and fuzzy logic in ecological modelling. EcoSys., 4, 85-97.*
- Salski, A. - Fränzle, O. - Kandzia, P. (Eds) (1996): Fuzzy Logic in Ecological Modelling. Ecological Modelling special issue, 85, No. 1.*
- Salski, A. (1999): Fuzzy logic approach to data analysis and ecological modelling. Proc. of European Symposium on Intelligent Techniques (ESIT'99)*
- Salski, A. (2002) Ecological Applications of Fuzzy Logic. In: F. Recknagel (ed): Ecological Informatics. Springer*
- Salski A. - Kandzia, P.-Bartels, F. (2004) A fuzzy approach to land evaluation. The 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'2004), Perugia, July, 4-9,*
- Salski, A. - Bartels, F.(2005): A fuzzy approach to land evaluation. IASME Transactions, Issue 5, V.2, p.774-780.*
- Samra, J.S. - Rajput, R.K. - Katyal, V. (1992) Structured heterogeneity of soil-pH and grain-yield of rice and wheat grown in a sodic soil. Agron. J. 84, 877–881.*
- Schmidt J.P.-Taylor R. K.-Milliken G.A. (2002) Evaluating the potential for site-specific phosphorus applications without high-density soil sampling. Soil Sci. Soc. Am. J. 66, 276-283.*

*Shearer T. - Burks, T. - Fulton, J. - Higgins, S. (1999)* Yield prediction using a neural network classified tained soil landscape features and soil fertility data. ASAE 993042.

*Steiner F. (1990)* A geostatistika alapjai. Tankönyvkiadó Budapest

*Sudduth K. Drummond S. - Birrel S. - Kitchen N. (1996)* Analysis of spatial factors influencing crop yield. Proc. of 3<sup>rd</sup> International Precision Agriculture Conference p. 129-140

*Sunila R.-Laine E.-Kremenova O. (2004)* Fuzzy model and kriging for imprecise soil polygon boundaries. Geoinformatics 2004. 489-495.

*Tamás J. (2001)* Precíziós mezőgazdaság. Mezőgazdasági Szaktudás Kiadó

*Tamás J. (2003)* Problems and solutions of field scale agr-ecological data acquisition and data interpraetations in agroinformatival domain. Applied Ecology and Environmental Research 1, 143-157.

*Tamás, J. - Buzás I. - Nagy, I. (2005)* A mintapontok folytonos GIS térbeli elemzése a culorrépa termésének és minőségének vizsgálata során. Agrártudományi Közlemények, 2005/18.

*Taylor, R. - Kluitenberg, M.D. - Schrock, N. - Zhang, J.P. -Schmidt, J.-Havlin, J.L. (2001)* Using yield monitor data to determine spatial crop production potential. Transactions of the ASAE 44(6).

*Triantafilis, -McBratney, A. (1993)* Applications of continuous methods of soil classification and land suitability assessment in the lower Namoi valley. CSIRO Australia Division of Soils, Divisional Rep. 121.

*Van Groenigen, J.W.-Stein, A. (1998)* Constrained optimisation of spatial sampling using continuous simulated annealing. Journal of Environmental Quality 27, 1078–1086.

*Van Groenigen, J.W. (1999)* Constrained optimisation of spatial sampling, a geostatistical approach. Ph.D. diss. ITC Publ. Ser. 65. ITC, Wageningen, the Netherlands

*Van Groenigen, J. W. (2000)* The influence of variogram parameters on optimal sampling schemes for mapping by kriging Geoderma 97, 223-236

*Várallyay, Gy. (1998)* Multifunctional soil management for sustainable development in Hungary. Agrokémia és Talajtan, 47. 7-22.

*Varga-Haszonits Z.–Mike-Hegedűs F.*(1993) Az éghajlati változékonyság és a növénytermesztés. *Növénytermelés* 42, 361-373.

*Varga-Haszonits Z. – Schmidt R. - Mike-Hegedűs F.*(1994) Az éghajlati változékonyság és a gazdasági növények. *Növénytermelés* 43, 485-497.

*Wang L. X.*( 1992) Fuzzy systems are universal approximators. In Proc. of the 1st IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'92), 1163–1169, San Diego.

*Van Groenigen, J.W.* (1999) Constrained optimisation of spatial sampling, a geostatistical approach. Ph.D. diss. ITC Publ. Ser. 65. ITC, Wageningen, the Netherlands

*Webster, R., - McBratney, A.,* (1989) On the Akaike Information Criterion for choosing models for variograms of soil properties. *Journal of Soil Science* 40, 493-496

*Webster, R. - Oliver, M.* (1990) *Statistical Methods in Soil and Land Resource Survey*, Oxford Univ. Press, Oxford.

*Webster, R. - Oliver, M.* (1992) Sample adequately to estimate variograms of soil properties. *J. Soil Sci.* 43, 177–192.

*Webster, R. - Oliver, M.* (2001) *Geostatistics for environmental Scientists*. Wiley, Brisbane

*Whelan, B. -; McBratney, A. - Minasny, B.* (2001) VESPER- spatial prediction software for precision agriculture. In: *Proceedings of the 3rd European Conference on Precision Agriculture*, 139-144.

*Zadeh L.* (1965) Fuzzy Sets. *Information and Control* 8, 338-353.

*Zhu A.X.* (1997) A similarity for representing soil spatial information. *Geoderma* 77, 217-242.

*Zhu, A. - Hudson, B. - Burt, J. - Lubich, K. - Simonson, D.*(2001) Soil mapping using GIS, expert knowledge and fuzzy logic. *Soil. Sci. Soc. Am. J.* 65, 1463-1472.

*Zimmerman, H.-J.* 1991. *Fuzzy set theory and its applications*. 2nd ed. Boston: Kluwer Academic. 399 pp.

## 10. SUMMARY

Within a field, crop yield and the conditions of cultivation (e.g. soil fertility) vary in space and time. Precision agriculture necessitates information about the relationship between the spatial variability of soil properties and the spatial variability of crop yield. The key question of the realisation of site specific production is whether sufficiently accurate information is gathered from the field. Precision agriculture must begin by gathering information about the factors that determine the conditions of plant growth. Parameters influencing the productiveness of soil are most important among these.

In our field experiment, we have analysed crop yield and site and soil properties based on data collected in a period of five years.

The data on yield and soil properties for five years were analysed with both statistical and geostatistical methods after technical preparations. Descriptive statistics were used to determine location, variance and the characteristic parameters of distribution.

These statistical examinations did not provide answers to questions on the spatial location or the spatial variability and distribution of the sample. Geostatistical analysis is based on the theory of regional variables. The spatial distribution of variables was described by a variogram function. It is a useful aggregating statistic, which shows how data are *dispersed* in the field. In order to analyse the spatial variability of yield, we prepared experimental variograms, and fitted (exponential and spherical) functions to the variograms.



Among the parameters of variograms, the nugget effect/sill (N/S) ratio was used to define small scale variability in order to characterise spatial structure. The spatial correlation interval shows the distance within which the values of the soil property correlate with one another. A small N/S and a large correlation interval usually mean that a soil property can be mapped with relatively great precision.

The accurate mapping of soil properties is critical for the success of location specific cultivation. The number of soil samples is also decisive. Moreover, the method of interpolation that is used for converting discrete sample data into continuous maps is one of the most important factors for the success of soil mapping.

Two interpolating techniques were used to draw maps: the method of inverse distance weighting and kriging. Their correct application demands a precise definition of the spatial structure, which was given by geostatistical characteristics: experimental variograms and functions. The goodness of predictions was checked by cross-checking and the method of independent testing.

The number of soil samples is insufficient for drawing a reliable variogram. Therefore, data were supplemented with numbers drawn from random distributions of the variables.

The tools of artificial intelligence are better suited to the description of the continuous character of soil properties and the uncertainty of spatial variability. First, imprecise (no 'sharp') data were represented and treated as fuzzy numbers. Second, uncertain information was represented and treated as a set of fuzzy rules.

Traditional kriging was extended with the help of fuzzy kriging, which enabled treating together precisely measured data and imprecise predictions defined as fuzzy numbers in the spatial interpolation. The fuzzy set of soil parameters enables us to define the possible values of soil properties in a unified way. The individual values of soil properties correspond to their membership in the fuzzy set, as defined by the membership function.

Fuzzy modelling is especially useful in a field of research where the relationships between components are not precisely known or the incompleteness or uncertainty of data renders statistical analysis unreliable.

The relationships between variables were determined with the aid of rule based systems. The function was given by the fuzzy relationship. The system was constructed from factual data and rules, and processed based on inference mechanisms.

An important goal of our research was to establish a relationship between soil data and crop yield, and thus construct an optimal rule base. A bacterial algorithm optimised the fuzzy rule base.

The relationship between soil parameters and yield data was examined by a fuzzy rule based system with six and eight variables. The membership functions contained in the rules cover the variables' entire intervals and thus ensure continuity. The obtained rules correspond to a special fuzzy-like function definition. The simulation result: "predicted yield value" on the site does not contain sharp boundaries but fuzzy overlaps. The rule base constructed through simulation was employed to evaluate

additional, new, samples. The existing, trained, rule base for a new sample carried out the interpolation.

The factors affecting crop yields, such as soil, weather and management are so complex that traditional statistics cannot give accurate results. Therefore, a more complete understanding of relationships between yield and soil-site properties is of critical importance to precision farming. As an automatic learning tool, an artificial neural network is a useful alternative for processing the massive data set generated by precision farming production and research. Using neural networks and the available samples, we constructed a complex model of the relationships between crop yield and important soil properties. Neural networks models may be used for estimating the parameters of nonlinear ecological models and implemented in other fields of agricultural research. The framework for neural network development presented for crop-soil-weather modelling could be extended.

Our results contribute to the understanding of spatial variability, the management of spatial estimation and the treatment of uncertainty. After ensuring an adequate preparation of data and drawing on expert knowledge, the methods developed in our PhD thesis may facilitate a more precise modelling of site-specific plant production.

1. melléklet A talajparaméterek korrelációs együtthatói 2001. év

	pH KCl	kötöttség	CaCO <sub>3</sub> %	humusz%	N%	Nitrát	Al P <sub>2</sub> O <sub>5</sub>	Al K <sub>2</sub> O	Al(Na)	Mg(KCl)	Zn	Cu	Mn
pHKCl	1												
kötöttség	-0,6377 <i>p=,000</i>	1											
CaCO <sub>3</sub> %	-0,1472 <i>p=,099</i>	0,2095	1										
humusz%	-0,5752 <i>p=,000</i>	0,5553	0,1404	1									
N%	-0,16 <i>p=,000</i>	0,1115	-0,0561	0,3684	1								
Nitrát	0,194102	-0,24098	0,0858	-0,03573	-0,27896	1							
Al P <sub>2</sub> O <sub>5</sub>	0,3908 <i>p=,002</i>	-0,4417 <i>p=,000</i>	-0,136 <i>p=,288</i>	-0,1982 <i>p=,119</i>	0,0462	0,13031	1						
Al K <sub>2</sub> O	-0,0314	0,0488	-0,0926	0,0296	-0,0539	-0,1832	0,2237	1					
Al Na	-0,3192 <i>p=,011</i>	0,5473 <i>p=,000</i>	0,2248 <i>p=,077</i>	0,4212 <i>p=,001</i>	0,3473 <i>p=,003</i>	-0,1603	-0,0785	0,0415	1				
Mg(KCl)	-0,4333 <i>p=,000</i>	0,5537 <i>p=,000</i>	0,4485 <i>p=,000</i>	0,4595 <i>p=,000</i>	0,233 <i>p=,097</i>	-0,2096	-0,1565	0,0118	0,4702 <i>p=,000</i>	1			
Zn	0,2994 <i>p=,017</i>	-0,5511 <i>p=,000</i>	-0,0181 <i>p=,888</i>	-0,1511 <i>p=,237</i>	-0,0712	0,21093	0,5142 <i>p=,000</i>	0,268 <i>p=,034</i>	-0,2728 <i>p=,031</i>	-0,0728 <i>p=,571</i>	1		
Cu ppm	-0,5126 <i>p=,000</i>	0,1695 <i>p=,184</i>	-0,0977 <i>p=,446</i>	0,5184 <i>p=,000</i>	0,3243 <i>p=,014</i>	-0,0907	-0,1348	0,1695	0,1174 <i>p=,360</i>	0,3104 <i>p=,013</i>	0,2531 <i>p=,045</i>	1	
Mn ppm	0,222 <i>p=,080</i>	-0,5665 <i>p=,000</i>	-0,0447 <i>p=,728</i>	-0,0711 <i>p=,580</i>	0,0149	0,22664	0,0756 <i>p=,556</i>	0,0433 <i>p=,736</i>	-0,399 <i>p=,001</i>	-0,2017 <i>p=,113</i>	0,6994 <i>p=,000</i>	0,3982 <i>p=,001</i>	1
Fe ppm	0,0565 <i>p=,660</i>	-0,3209 <i>p=,010</i>	-0,1575 <i>p=,218</i>	0,0065 <i>p=,960</i>	-0,1251	0,15468	-0,0509 <i>p=,692</i>	0,3232 <i>p=,010</i>	-0,3719 <i>p=,003</i>	-0,2671 <i>p=,034</i>	0,5175 <i>p=,000</i>	0,4644 <i>p=,000</i>	7730 <i>p=,000</i>

2. melléklet A talajparaméterek kovariancia együtthatói (2001.év)

	pH KCl	kötöttség	CaCO <sub>3</sub> %	hu%	összN %	összN %	összN %	P <sub>2</sub> O <sub>5</sub>	K <sub>2</sub> O <sub>5</sub>	Al Na	Mg(KCl)	Zn	Cu ppm	Mn ppm
pH KCl	0,005													
kötöttség	-0,147	11,322												
CaCO <sub>3</sub> %	-0,025	1,753	6,183											
hu%	-0,012	0,575	0,107	0,095										
összN%	0	0,007	-0,002	0,002	0									
Nitrát	0,081	-4,927	1,296	-0,067	-0,03	36,916								
Nitrát N	0,018	-1,111	0,29	-0,016	-0,01	8,27	1,854							
Al P <sub>2</sub> O <sub>5</sub>	1,805	<b>-100,411</b>	<b>-22,85</b>	-4,123	0,055	<b>53,488</b>	12,454	4564,2						
Al K <sub>2</sub> O <sub>5</sub>	-0,034	2,61	-3,665	0,145	0,042	-14,4	-3,258	<b>240,49</b>	253,13					
Al Na	-0,067	5,619	1,705	0,396	0,018	-2,482	-0,56	-16,179	2,016	9,308				
Mg(KCl)	-0,764	<b>48,055</b>	<b>28,765</b>	3,65	0,106	<b>-32,84</b>	<b>-7,443</b>	<b>-272,71</b>	4,826	<b>37,003</b>	665,375			
Zn	0,009	-0,799	-0,019	-0,02	-0	0,552	0,127	14,969	1,838	-0,359	-0,809	0,186		
Cu	-0,021	0,339	-0,144	0,095	0,003	-0,327	-0,075	-5,412	1,602	0,213	4,756	0,065	0,353	
Mn	0,126	<b>-15,827</b>	-0,922	-0,182	0,002	11,435	2,549	<b>42,421</b>	5,718	-9,43	<b>-43,209</b>	2,502	1,965	68,951
Fe	0,018	-4,9	-1,778	0,009	-0,01	4,265	0,974	-15,596	23,34	-5,15	-31,265	1,012	1,252	29,13

3. melléklet Háló becslése a mintapontokban

	Minta száma	CÉL	Háló Output	AE	ARE		Minta száma	CÉL	Háló Output	AE	ARE		Minta száma	CÉL	Háló Output	AE	ARE	
TRN	0	3,300	3,779	0,479	15%	TST	22	3,930	4,072	0,142	4%	TST	43	4,630	4,043	0,587	13%	
TRN	1	4,030	4,101	0,071	2%	VLD	23	4,000	3,880	0,120	3%	TRN	44	4,450	4,136	0,314	7%	
VLD	2	3,990	3,795	0,195	5%	TRN	24	4,400	3,967	0,433	10%	VLD	45	4,570	3,856	0,714	16%	
TRN	3	4,760	4,146	0,614	13%	TRN	25	4,760	3,915	0,845	18%	TRN	46	4,000	3,822	0,178	4%	
TST	4	4,150	4,209	0,059	1%	TRN	26	3,800	3,915	0,115	3%	TRN	47	4,230	3,767	0,463	11%	
TRN	6	3,980	4,101	0,121	3%	TRN	27	3,640	3,862	0,222	6%	TRN	48	3,630	3,809	0,179	5%	
TRN	7	3,020	3,886	0,866	29%	TRN	28	3,820	3,738	0,082	2%	TRN	50	4,210	3,766	0,444	11%	
TRN	8	4,400	3,870	0,530	12%	TRN	29	4,370	3,902	0,468	11%	TRN	51	3,110	3,644	0,534	17%	
TRN	9	4,360	4,147	0,213	5%	TRN	30	4,030	4,039	0,009	0%	TRN	52	3,560	3,695	0,135	4%	
TRN	10	4,110	4,174	0,064	2%	TRN	31	3,980	4,094	0,114	3%	TRN	53	3,860	3,771	0,089	2%	
TRN	11	4,310	4,047	0,263	6%	TRN	32	4,570	3,972	0,598	13%	TST	54	4,450	3,745	0,705	16%	
TRN	12	4,370	4,083	0,287	7%	VLD	33	4,220	3,868	0,352	8%	TRN	55	3,600	4,039	0,439	12%	
TST	13	3,970	3,878	0,092	2%	TRN	34	3,370	3,776	0,406	12%	VLD	56	4,830	4,175	0,655	14%	
VLD	14	4,150	3,942	0,208	5%	TST	35	3,560	3,937	0,377	11%	VLD	57	4,340	4,111	0,229	5%	
VLD	15	2,850	4,115	1,265	44%	TRN	36	4,620	3,879	0,741	16%	TRN	58	4,030	3,866	0,164	4%	
TRN	16	3,460	4,049	0,589	17%	TRN	37	3,650	3,788	0,138	4%	TRN	59	3,760	3,967	0,207	5%	
TRN	17	3,810	3,882	0,072	2%	TST	38	3,920	3,917	0,003	0%	TRN	60	3,410	3,703	0,293	9%	
TRN	18	3,450	3,776	0,326	9%	TRN	39	3,910	3,923	0,013	0%	TRN	61	3,340	3,744	0,404	12%	
TRN	19	3,530	3,911	0,381	11%	TRN	40	3,240	3,767	0,527	16%	TST	62	3,850	3,752	0,098	3%	
TRN	20	4,000	4,134	0,134	3%	VLD	41	3,680	3,762	0,082	2%							
TST	21	4,000	4,210	0,210	5%	TRN	42	4,040	3,832	0,208	5%							

