

SZEGEDI TUDOMÁNYEGYETEM
Természettudományi és Informatikai Kar
Környezettudományi Doktori Iskola
Természeti Földrajzi és Geoinformatikai Tanszék

**GEOSTATISZTIKAI MEGKÖZELÍTÉSEN ALAPULÓ DIGITÁLIS
TALAJTÉRKÉPEZÉSI ÉS MINTAVÉTEL OPTIMALIZÁCIÓS
MÓDSZERTAN KIALAKÍTÁSA VALAMINT ALKALMAZÁSA
KÜLÖNBÖZŐ LÉPTÉKEKBEN**

Doktori (Ph.D.) értekezés

SZATMÁRI GÁBOR

Témavezető:
Dr. Barta Károly
egyetemi docens

Külső konzulens:
Dr. Pásztor László
tudományos főmunkatárs

**Szeged
2017**

TARTALOMJEGYZÉK

I. BEVEZETÉS ÉS CÉLKITŰZÉSEK	3
II. IRODALMI ÁTTEKINTÉS	6
II. 1. A regionalizált változók fogalmának bevezetése	6
II. 2. A geostatisztika megjelenése a talajtérképezésben	8
II. 3. A geostatisztika szerepe és trendjei a nemzetközi és a hazai talajtérképezésben	10
II. 4. A talajmintavételezés optimalizációja a nemzetközi és a hazai talajtérképezésben	14
III. ELMÉLETI MEGFONTOLÁSOK	20
III. 1. A talajtulajdonságok, mint regionalizált változók	20
III. 2. A stacionaritás kérdése	21
III. 3. A talajmintavételezések modell alapú megközelítése és optimalizációja	22
IV. MINTATERÜLETEK	24
IV. 1. Szálkai kisvízgyűjtő	25
IV. 2. Előszállási szántóföldek	26
IV. 3. Zala megye	28
V. ANYAG ÉS MÓDSZER	31
V. 1. Felhasznált talajtani adatbázisok és segédinformációk	31
V. 1. 1. Szálkai kisvízgyűjtő	31
V. 1. 2. Előszállási szántóföldek	32
V. 1. 3. Zala megye	33
V. 2. Térképezési módszertan	34
V. 2. 1. A becslési grid definiálása	34
V. 2. 2. Feltáró alapadat elemzés	35
V. 2. 3. Segédadatok előkészítése a geostatisztikai modellezésekhez	38
V. 2. 4. Regresszió krigelés	39
V. 2. 5. A regresszió krigeléssel előállított térképek hibája	43
V. 2. 6. A térbeli bizonytalanság modellezése szekvenciális sztochasztikus szimulációval ...	44
V. 3. Modell alapú talajmintavétel optimalizációs módszertan	47
V. 3. 1. Spatial simulated annealing	47
V. 3. 2. A regresszió krigelés becslési varianciája, mint az optimalizáció minőségi mértéke	49
V. 3. 3. Mintavétel optimalizációs scenáriók beállítása a módszertan tesztelésére	51
V. 3. 4. Többváltozós mintavétel optimalizációs módszertan és tesztelése	52
V. 3. 5. Az optimalizált mintavételi elrendezések értékelése	55
V. 4. A módszertanok szinergiája	55

VI. EREDMÉNYEK BEMUTATÁSA ÉS ÉRTÉKELÉSE	58
VI. 1. A definiált becslési gridek	58
VI. 2. A szervesanyag-tartalom adatok elsődleges értékelése.....	59
VI. 2. 1. Szálkai kisvízgyűjtő	59
VI. 2. 2. Előszállási szántóföldek	63
VI. 2. 3. Zala megye.....	67
VI. 3. A szervesanyag-tartalom térképek és értékelésük	72
VI. 3. 1. Szálkai kisvízgyűjtő	72
VI. 3. 2. Előszállási szántóföldek	76
VI. 3. 3. Zala megye.....	80
VI. 4. A térbeli bizonytalanság modellezése és értékelése az előszállási szántóföldeken	86
VI. 5. A mintavétel optimalizációs scenáriók és értékelésük	88
VI. 5. 1. Egyváltozós scenáriók: 1. és 2.	89
VI. 5. 2. Egyváltozós scenáriók: 3. és 4.	95
VI. 5. 3. Többváltozós scenáriók: 5. és 6.....	101
VI. 6. A módszertanok szinergiája.....	107
VII. ÖSSZEFOGLALÁS	111
VIII. SUMMARY	115
IX. KÖSZÖNETNYILVÁNÍTÁS	119
X. IRODALOMJEGYZÉK	120
XI. MELLÉKLETEK	130

„The soil varies from place to place, and many of its properties vary in time too. This is what makes the soil so fascinating”

GERARD B.M. HEUVELINK & RICHARD WEBSTER
(2001)

I. BEVEZETÉS ÉS CÉLKITŰZÉSEK

A talajok állapotára, folyamataira, funkcióira vonatkozó információk iránti igények mind hazai, mind nemzetközi szinten számottevőek és folyamatosan bővülnek (Baumgardner, 2011; Várallyay, 2012; McBratney et al., 2014; Tóth et al., 2015). A talaj multifunkcionalitásának felismerése ugyanakkor a talajfelvételezésekre fordítható erőforrások beszűkülésével egy időben következett be (Pásztor et al., 2013a). Ezért világszerte törekednek a már meglévő talajtani információk mind alaposabb felhasználására (Minasny, McBratney, 2016), illetve a terepi felvételezések minimalizálására/optimalizálására (Vašát et al., 2010). Hazánkban jelenleg a talajokkal kapcsolatos igényeket a rendelkezésünkre álló talajtani adatbázisok segítségével elégítjük ki, melyek döntő részét korábbi talajfelvételezésekből nyert tematikus térképek, illetve feltárt talajszelvények helyszíni és laboratóriumi jegyzőkönyveinek térbeli talajinformációs rendszerekbe rendezett adatbázisai jelentik (Pásztor et al., 2010a, 2015). Emellett történtek továbbá a meglévő talajtani adatbázisok kiegészítésére/reambulálására irányuló új mintavételezések is (Pásztor et al., 2002, 2013a). Az ezekben a talajinformációs rendszerekben raktározott adatok pontszerű észlelésekhez kötöttek, vagyis ezen adatponti értékek regionalizálásával elégíthetjük ki a felhasználók specifikus igényeit. A pontszerű észlelésekhez kapcsolódó talajtani információk térbeli kiterjesztése legcélravezetőbben a geostatisztika módszereivel tehető meg (Webster, 1989, 2015; Lark, 2012).

A geostatisztikai – és vele együtt, szinte elválaszthatatlanul a geomatematikai – megközelítések mára már bevált gyakorlatok a talajtérképezésben, illetve a tágabb értelemben vett talajtani kutatásokban (Webster, 2015). A geostatisztika és a geomatematika segítségével feltárhatjuk, elemezhetjük, illetve modellezhetjük a vizsgált talajtani jellemzők közötti, illetve a környezetükkel fennálló kapcsolatokat a földrajzi és/vagy attribútum térben. E modellek segítségével becsléseket tehetünk a talajjellemzők paramétereire (például: átlag, szórás), illetve ismeretlen pontbéli vagy területegység értékeire. E modellek kiemelkedő szerepet játszanak a bizonytalansági, illetve kockázati elemzésekben, ahol az adott talajtani sajátssággal, folyamattal kapcsolatos ismeretek hiányának, illetve a káros vagy nem kívánatos kimenetelű események lehetőségeinek mérlegelése történik. E modellek révén optimalizálhatjuk a talajtani

felvételezéseket az aktuális igények, megszorítások és a rendelkezésre álló talajtani, illetve segédinformációk mellett. Nem utolsó sorban e modellek segítenek hozzá, hogy jobban megismerhessük a talajban lejátszódó folyamatok térbeli és/vagy időbeli változékonyságát.

Napjaink digitális talajtérképezése és mintavétel optimalizációja nagyban épít a geostatisztika megközelítéseire (Minasny, McBratney, 2016), melyeket hazánkban is sikeresen implementáltak (például: Dobos et al., 2007; Bakacsi et al., 2012; Pásztor et al., 2014a,b). Ugyanakkor a talajtani információk iránti igények természete is megváltozott az elmúlt években, melyek a rendelkezésünkre álló talajtani információkkal és az eddig alkalmazott, „hagyományos” geostatisztikai módszerekkel egyre ritkábban fedhetők le (Pásztor et al., 2013a). Gyakori elvárás például, hogy a talajtulajdonság térképeken túl a térképek hibájáról, illetve bizonytalanságáról is álljanak rendelkezésre információk (Heuvelink, 2014; Minasny, McBratney, 2016), melyek napjainkban elengedhetetlenek egy megalapozott döntéshez. Ezeknek az elvárásoknak a „hagyományos” geostatisztikai módszerekkel kapott eredmények nem, vagy csak részben felelnek meg. Ezért kiemelten fontos, hogy hazánk digitális talajtérképezési gyakorlatában meghonosítsuk azokat a geostatisztikai módszereket, melyekkel a fenti igények kielégíthetők. E módszerek sikeres adaptálásának első lépése, hogy a geostatisztikával kapcsolatos elméleti megfontolásokat koherens és konzekvens módszertanban foglaljuk össze. Ezért doktori kutatásaim célkitűzései:

- 1) Geostatisztikai megközelítésen alapuló, koherens digitális talajtérképezési és mintavétel optimalizációs módszertan kialakítása, mely:
 - elméleti és gyakorlati keretet nyújt a mintavétel megtervezésétől a talajtérkép(ek) közreadásáig,
 - konzekvens és egymással kompatibilis lépésekből épül fel,
 - különböző léptékekben alkalmazható,
 - képes számításba venni a rendelkezésünkre álló segédinformációkat a térbeli becslés, a térbeli bizonytalanság modellezése, illetve a mintavétel optimalizáció során,
 - magában foglalja a térkép hibájának, illetve bizonytalanságának értékelését,
 - továbbá komponensei módosíthatók az adaptált valószínűségi függvény szerint.
- 2) A kialakított digitális talajtérképezési és mintavétel optimalizációs módszertan szabad-hozzáférésű, nyílt forráskódú szoftverek (*Free and Open Source Softwares, FOSSs*) segítségével legyen implementálható, mely a gyakorlat szempontjából ingyenes és bárki számára hozzáférhető felhasználást jelent.

- 3) A regresszió krigelés alkalmazása különböző léptékekben hazai mintaterületeken. A regresszió krigeléssel előállított térképek hibájának értékelése.
- 4) Regresszió krigelésre felépített szekvenciális sztochasztikus szimuláció alkalmazása hazai mintaterületen a térbeli bizonytalanság modellezésére.
- 5) A spatial simulated annealing (SSA) mintavétel optimalizációs algoritmus alkalmazása hazai mintaterületen különböző mintavétel optimalizációs scenáriók tükrében.
- 6) Az egyváltozós SSA algoritmus kiterjesztése többváltozós mintavétel optimalizációs algoritmussá és alkalmazása hazai mintaterületen.

A kialakított digitális talajtérképezési és mintavétel optimalizációs módszertan tesztelésére a talajok szervesanyag-tartalmát választottam, mely kiemelkedő jelentőséggel bír nemcsak a talajtan és az agrárium, hanem a környezetvédelem (például: éghajlatváltozás, szén biogeokémiai körforgalma) és a földtudományok számára is (Stockmann et al., 2013; Centeri et al., 2014; Hartemink, McSweeney, 2014; Jakab et al., 2016). Ugyanakkor fontos megjegyezni, hogy a kialakított módszertan alkalmas egyéb talajtulajdonságok (termőréteg vastagság, pH, mésztartalom, sótartalom, Arany-féle kötöttség stb.) digitális térképezésére, illetve mintavétel optimalizációjára is.

„A különböző talajtérképek mindenesetben a talajról szerzett információk konkrét, területre vonatkoztatott, a kor adott tudományos színvonalának, illetve a gyakorlat által megfogalmazott igényeknek megfelelő szintézisét jelentették és jelentik. Így volt ez a talajtani tudomány kialakulása óta mindig, s épp a talajtani tudomány története bizonyítja, hogy fejlődésének egyes szakaszai mindig jól jellemezhetőek az adott periódus eredményeit összegző talajtérképekkel, talajtérképezési rendszerekkel.”

VÁRALLYAY GYÖRGY & SZŰCS LÁSZLÓ
(1978)

II. IRODALMI ÁTTEKINTÉS

A talajtérképezésben, illetve a hozzá szorosan kapcsolódó mintavételezés megtervezésében meglehetősen korán, az 1980-as évek elején megjelentek a geostatisztika első alkalmazásai (Burgess, Webster, 1980a,b; Webster, Burgess, 1980; McBratney et al., 1981; McBratney, Webster, 1981, 1983). Ezek az alkalmazások és a mögöttes koncepciók azóta is meghatározzák a talajtérképezés gyakorlatát, a talajmintavételezési stratégiák és monitoring rendszerek kidolgozását, illetve a talajtérképezők talajjal kapcsolatos térbeli szemléletét és gondolkodását (Webster, 1989, 1994, 2000, 2015). Dolgozatom szakirodalmi áttekintését kronologikusan építettem fel. Az 1950-es és 1960-as évek bányászati gyakorlatából kiindulva mutatom be a regionalizált változók fogalmának bevezetését és ennek körülményeit. Erre alapozva mutatom be azon, a talajtanban analógnak tekinthető körülményeket, melyek az 1970-es és 1980-as években azt eredményezték, hogy a „hagyományos” talajtérképezést felváltották a geostatisztikai megközelítések és alkalmazások. Ennek kapcsán egy paradigmaváltásra is felhívom a figyelmet, melynek eredményeként megváltozott a talajtérképezők talajjal kapcsolatos térbeli szemlélete és gondolkodása. Ezt követően bemutatom, hogy a geostatisztikának milyen trendjeit figyelhetjük meg a talajtérképezésben, illetve a talajmintavételezések optimalizációjában.

II. 1. A regionalizált változók fogalmának bevezetése

A bányászatban dolgozó szakemberek korán felismerték, hogy az alkalmazott klasszikus statisztikai módszerek nem voltak képesek tekintetbe venni a vizsgált jelenség (például: egy érctelep fémtartalma) térbeli aspektusát (Krige, 1951, 1960; Matheron, 1963, 1971). Vegyük példaként a hisztogramot, mely az egyik leggyakrabban használt adatrepresentációs eszköz a statisztikában. A hisztogram az egyes – azonos szélességű – diszjunkt osztályokba tartozó megfigyeléseket ábrázolja grafikusán, mely által értékes információkat közöl az elemzővel.

Ugyanakkor a megfigyelések földrajzi pozícióját nem képes számításban venni azon általános okból kifolyólag, hogy a statisztika a vizsgálat alá vont jelenséget explicit módon valószínűségi változónak tekinti. A valószínűségi változónak két alapvető tulajdonsága van: (1) a kutatás alapját jelentő kísérlet elméletileg végtelen – a gyakorlatban tetszőlegesen – sokszor megismételhető és (2) a kísérletek függetlenek egymástól. Ezek a sajátosságok azonban nem voltak összeegyeztethetők az addigi tapasztalatokkal, hiszen (1) a gyakorlatban nincs lehetőség a kísérlet újbóli elvégzésére és (2) a kísérletek kimenetelei nem függetlenek egymástól, köztük – térbeli és/vagy időbeli – kapcsolat áll fent (Matheron, 1963, 1971). Ezért a korábbi statisztikai megközelítésekkel kapott eredmények, becslések nem voltak elfogadhatók az addigi tapasztalatok alapján, külön kiemelve itt a gazdaságosan kitermelhető nyersanyagra vonatkozó számításokat (Krige, 1951; Matheron, 1963). A legfőbb problémát tehát az jelentette, hogy ezek a megközelítések érzéketlenek a megfigyelési helyekre, holott ez utóbbi magával az adattal azonos értékű információ. Ezért egy olyan megközelítésre volt szükség, mellyel explicit módon hangsúlyozni lehet a vizsgált jelenség térbeli aspektusát, mely annak elidegeníthetetlen és egyik legfontosabb sajátossága. Ezzel kapcsolatosan Daniel G. Krige dél-afrikai bányamérnök végzett úttörő kutatásokat az 1950-es években a witwatersrandi aranybányákban. Georges Matheron francia matematikus és geológus felismerve a dél-afrikai kutatások jelentőségét az 1960-as években kidolgozta e megközelítés elméleti alapjait, melyet ma a regionalizált változók elméleteként (*theory of regionalized variables*) ismerünk. Georges Matheron a regionalizált változók becslésére szolgáló lineáris interpolációs eljárást Daniel G. Krige után *krigelésnek* nevezte el, így ismerve el a dél-afrikai kutató úttörő munkásságát.

Matheron (1963, 1965, 1971) munkássága nyomán a regionalizált változókat valószínűségi függvényekkel közelítjük (Journel, 1986); ahol a valószínűségi függvény olyan valószínűségi változók halmaza, melyek egymástól függését bizonyos valószínűségi mechanizmus szabályozza (Pannatier, 1996; Geiger, 2006a). *A vizsgált jelenséget kifejező regionalizált változót az ezen a jelenségen létrehozott valószínűségi függvény egyedi realizációjának tekintjük* (Matheron, 1971; Pannatier, 1996; Geiger, 2006a; Webster, Oliver, 2007). Ezért a regionalizált változó fogalmához szorosan kapcsolódik az erratikus, illetve struktúrált jellemzők, miszerint az adatponti értékek helyileg valószínűségi változók, illetve minden adatponti pár a jelenség térbeli struktúráját kifejező korrelációval van összekötve (Pannatier, 1996; Geiger, 2006a). Utóbbi a gyakorlatban – vagyis a geostatistikában – kiemelt fontosságú, hisz ez utóbbi a jelenség általános térbeli eloszlásával kapcsolatos tulajdonság, melynek egyik – de nem kizárólagos – mértéke a variogram (Matheron, 1963, 1971).

A fentiek alapján tehát elmondható, hogy *a geostatisztika és a statisztika közötti legfőbb különbség a vizsgált jelenséghez kapcsolódó megfigyelési helyek információinak rögzítésében és kiterjesztésében van* (Geiger, 2006a). Míg a statisztika a tanulmányozott jelenséget valószínűségi változónak tekinti és a mintavétel független, addig a geostatisztika regionalizált változónak tekinti és a minta adatai nem függetlenek egymástól (Matheron, 1963, 1971).

A bányászatban bekövetkezett szemléletbeli változások mellett a „társtudományokban” is születtek – habár kisebb-nagyobb időbeli eltéréssel – kísérletek a térbeliség kezelésére. Fontosnak tartom ezeknek a munkáknak a rövid számbavételét. Elsőként említendő Gandin (1965) orosz meteorológus munkája, aki Georges Matheron munkásságához hasonló elméleti megfontolásokat és gyakorlati megközelítéseket fogalmazott meg meteorológiai adatsorokra vonatkozóan. Kihangsúlyozandó Matérn (1960) svéd erdész és statisztikus munkája, akiről egy kovariancia függvényt neveztek el. A földrajztudományban kiemelendő a svéd Hägerstrand (1965), illetve az amerikai Tobler (1970) munkássága, akik elsőként helyezték előtérbe a földrajztudományban a szomszédsági kapcsolatok szerepét, melyet ma a „földrajz első törvényeként” vagy „Tobler törvényeként” ismerünk. Cressie (1990) szerint azonban Georges Matheron volt az első, aki teljes körű elméleti és gyakorlati megoldást nyújtott. Ezért az utókor Georges Matheron-nak – és gyakorlati munkássága révén Daniel G. Krige-nek – tulajdonítja ezt az eredményt (Journel, Huijbregts, 1978; Cressie, 1993; Goovaerts, 1997; Deutsch, Journel, 1998; Webster, Oliver, 2007).

II. 2. A geostatisztika megjelenése a talajtérképezésben

A geostatisztikai megközelítéseket megelőzően a talajok térképezését adott osztályozási szempontból homogénnek tekinthető területek térképi lehatárolása jelentette (Heuvelink, Webster, 2001; Webster, 2015). Ezekre számos példát találhatunk mind a hazai (Kreybig, 1934, 1937; Stefanovits, Szűcs, 1961; Szabolcs, 1966; Géczy, 1968; Várallyay, Szűcs, 1978; Várallyay et al., 1979, 1980a,b), mind a nemzetközi (Dokucsajev, 1883; Glinka, 1908; Whitney, 1909; Marbut et al., 1913; Prescott, 1933) szakirodalomban. Az általuk követett paradigma szerint a talajosztályokat ábrázoló, ún. talajfolt térképek csaknem teljes körű információval képesek szolgálni a talajokra vonatkozóan (Webster, 2015). Ez azt jelenti, hogy ha ismert, hogy az adott pont vagy terület mely osztályhoz tartozik, akkor az ottani talajtulajdonságok származtathatók, becsülhetők (Webster, 2000, 2015). Ezért a talajtérképezés legfőbb célja és kihívása az osztályozási módszerek tökéletesítése és a talajról szerzett ismeretek stabil

keretrendszerbe való integrálása volt (Webster, 2000, 2015). Ez az ún. arisztotelészi logika hosszú évtizedekig meghatározta a szakemberek szemléletét és gondolkodását a talajok térbeli változékonyságával kapcsolatosan (Webster, 2015).

Elsőként az építő- és hadmérnökök ismerték fel a statisztikában rejlő potenciált a talajtulajdonságok (például: szervesanyag-tartalom, szénsavas mésztartalom, termőréteg vastagság) adott pontra vagy területre vonatkozó becslése, illetve térképezése kapcsán (Morse, Thornburn, 1961; Kantey, Williams, 1962). A statisztikai mintát rétegzett véletlen mintázásból nyerték, ahol is a mintavételezés alapját jelentő „rétegzést” a talajfolt térképek szolgáltatták. A vizsgált talajtulajdonságra vonatkozó becslések az ún. térbeli változékonyság diszkrét modelljét (Heuvelink, Webster, 2001) követték. E modell szerint a vizsgált talajtulajdonság értéke becsülhető az adott talajfoltra – illetve a magába foglalt összes pontra vonatkozóan – az ugyanazon talajfoltból származó statisztikai minta segítségével, mely becslés „érvényességét” a fentebbi paradigmának kell(ene) igazolnia. Hazánkban is történtek ezen az elven alapuló, ún. céltérképezések (Várallyay et al., 1979, 1980a,b; Kummert et al., 1989; Várallyay, 1989a,b); például ezen az elven alapulva készült el 1980-ban Magyarország talajainak vízgazdálkodási tulajdonság térképe 1:100.000 méretarányban (Várallyay et al., 1980b). Mindazonáltal a térbeli változékonyság diszkrét modellje sok esetben kétes eredményeket szolgáltatott. Az ezzel kapcsolatos megfigyeléseket Webster és Beckett (1970) munkája foglalja össze, melyből két fontos megállapítást kell kiemelni: (1) több talajtulajdonság esetén is megfigyelhető volt, hogy az osztályon belüli változékonyság nagyobb, mint az osztályok közötti, továbbá (2) megfigyelhető volt, hogy a visszamaradó reziduumok változékonysága struktúrált, azaz nem ún. fehérzaj. Mindemellett az angol talajtérképező Richard Webster 1950-es évek végén Afrikában töltött térképezési munkálatai során is „meglepő” jelenséggel találkozott: az ott található talajok fokozatos átmenetei miatt nem lehetett az osztályokat precízen definiálni, továbbá az adott munkaléptékben egyértelműen lehatárolni. Ezek a megfigyelések és tapasztalatok ellentmondtak az addigi talajtérképezési gyakorlatnak, illetve a talajokkal kapcsolatos térbeli szemléletnek és gondolkodásnak (Webster, 2015).

A fenti ellentmondások miatt Webster és Cuanalo (1975) egészen új szemszögből kezdték vizsgálni a talajokat: úgy tekintették egy keresztmetszvény mentén mért talajtulajdonságok változékonyságát, mintha az egy véletlenszerű folyamat realizációja lenne. Ezt az új megközelítést az idősor elemzésekkel foglalkozó szakemberektől vették át, akik az idősorok fluktuációinak vizsgálata során analógnak tekinthető nehézségekkel szembesültek, mint a talajtanos szakemberek a térképezések során. A származtatott korrelogramok alapján

megfigyelhető volt, hogy bizonyos talajtulajdonságok hosszabb, míg mások rövidebb távolságban mutatnak korrelációt. Ugyanakkor nem volt számukra világos, hogy ezeket az információkat miként is lehetne alkalmazni a talajtérképezésben.

Néhány évvel később Richard Webster-nek ausztráliai útja során Daniel Sampey ausztrál geológus említette meg Daniel G. Krige és Georges Matheron eredményeit, vagyis a krigelést, illetve a regionalizált változók elméletét. Webster (2015) visszaemlékezése alapján a Daniel Sampey által elmondottak alapjaiban oldották meg a talajok térképezését övező elméleti ellentmondásokat és gyakorlati problémákat. Hazatérve Angliába, Trevor M. Burgess angol matematikus közreműködésével „átültette” a bányászati gyakorlatban sikeresen alkalmazott geostatisztikai megközelítéseket a talajtérképezésbe. Ennek eredményeként 1980-ban három tudományos dolgozatot publikáltak (Burgess, Webster, 1980a,b; Webster, Burgess, 1980), melyekben explicit módon tekintették a talajtulajdonságok térbeli változékonyságát regionalizált változónak, és használtak hagyományos krigelést (*ordinary kriging*), egyszerű krigelést (*simple kriging*), blokk krigelést (*block kriging*), illetve univerzális krigelést (*universal kriging*) a talajjellemzők térképezéséhez.

II. 3. A geostatisztika szerepe és trendjei a nemzetközi és a hazai talajtérképezésben

Az 1980-as évet követően a geostatisztika elfogadott megközelítés lett a talajtérképezésben, illetve a tágabb értelemben vett talajtani kutatásokban (Webster, 1989, 1994; Goovaerts, 1998a, 1999a; McBratney et al., 2000, 2003; Rajkai, 2001; Lark, 2012; Oliver, Webster, 2014, 2015), mely azóta is meghatározza a talajtérképezők talajjal kapcsolatos térbeli szemléletét és gondolkodását. A hazai gyakorlatban – igen korán – az 1980-as évek második felében jelentek meg a geostatisztika első alkalmazásai (Oertli, Rajkai, 1988; Rajkai et al., 1988; Csillag, Kertész, 1989).

Webster (2015) szerint a talaj ideális médiuma a geostatisztikának; hiszen (1) a talajtakaró egy folytonos közeget képez a földfelszínen, (2) könnyen hozzáférhető és relatíve olcsón mintázható, (3) a talajjellemzők eloszlása közel normális vagy lognormális, illetve (4) a talajban érvényre jutó folyamatok – mind tér- és időbeli – megismerhetőségének korlátai miatt a talaj egy sztochasztikus folyamatként közelíthető és modellezhető.

Fontos kiemelni, hogy a talajtanos szakemberek nem csak átvették a kidolgozott geostatisztikai módszereket, hanem lépést is tartottak, sőt aktívan hozzá is járultak a geostatisztika eszköztárának fejlesztéséhez, bizonyos részeinek újragondolásához és

megújításához. A talajtérképező, illetve talajtanos szakemberek érdeklődése és hozzájárulása főként a nem stacionárius valószínűségi modellek, a nem normál eloszlások, illetve a lokális és térbeli bizonytalanság értékelése és modellezése felé irányult (Heuvelink, Webster, 2001). Ezen irányú kutatásaikat és eredményeiket számos szakkönyv (például: Goovaerts, 1997; Webster, Oliver, 2007; Webster, Lark, 2013; Oliver, Webster, 2015), doktori disszertáció (például: Goovaerts, 1992; van Groenigen, 1999; Hengl, 2003; Kempen, 2011) és publikáció (például: Webster, Oliver, 1989; Goovaerts, 1993, 1996, 1998b, 1998c, 2000, 2001, 2009; McBratney et al., 2000, 2003; Hengl et al., 2004, 2007; Papritz, Schwierz, 2016) fémjelzi, megannyi nemzetközi konferencia (például: Pedometrics; International Geostatistics Congress; Digital Soil Mapping Workshop; IAMG Meeting & Conference) szóbeli közlései mellett. Magyarországon is számos geostatisztikai megközelítésen alapuló talajtani munka született (például: Csillag et al., 1992, 1996; Rajkai, Rydén, 1992; Tóth, Kertész, 1996; Tóth, Várallyay, 2001; Douaik et al., 2004, 2005; Bakacsi et al., 2012). Továbbá az elmúlt tíz évben – a nemzetközi trendhez hasonlóan – hazánkban is számos geostatisztikai alapú digitális talajtérképezési munka látott napvilágot (Dobos et al., 2006, 2007; Pásztor et al., 2010a, 2013a, 2014a,b, 2015, 2016a,b; Illés et al., 2011; Szatmári, Barta, 2013; Szatmári et al., 2013, 2015a,b; Waltner, 2013; Tóth et al., 2016).

A hagyományos krigelés volt az első módszer, melyet talajtulajdonságok (például: nátriumtartalom, kövesség, termőréteg vastagság) térképezésére alkalmaztak (Burgess, Webster, 1980a); ugyanakkor mind a mai napig igen gyakran és széles körben használt módszer, amit „egyszerűségének”, szofisztikált algoritmusának és az elérhető jó minőségű eredményeinek köszönhet (Webster, 2015). Mind a hazai, mind a nemzetközi szakirodalom széles körű alkalmazhatóságát bizonyítja (Webster, 1989; Webster, Oliver, 2007; Farkas et al., 2008; Szalóki, Somodi, 2014; Oliver, Webster, 2015; Szalóki, Szabó, 2016).

A talajtani felmérések gyakori jellemzője, hogy a mintáknak csak egy bizonyos hányadára áll rendelkezésre az adott talajtulajdonságra vonatkozó konkrét mért információ, melynek oka lehet például a laboratóriumi vizsgálat magas költsége. Ugyanakkor ezt a talajjellemzőt is gyakran térképezni kívánjuk. Már 1983-ban megfogalmazódtak ezek az igények (McBratney, Webster, 1983); ezért olyan módszerre volt szükség, mely szimultán képes figyelembe venni a térképezendő talajtulajdonság adatponti értékeit, illetve a rendelkezésre álló egyéb, az előzőnél „sűrűbben” mintázott talajjellemzőket a térbeli becslés során. A bányászati gyakorlatban az ilyen jellegű problémákra kidolgozott kokrigeléses (*cokriging*) megközelítést McBratney és Webster (1983) alkalmazta elsőként a

talajtulajdonságok térképezésében. A kokrigelés mögött azon megfontolás húzódik, hogy a becsülni kívánt regionalizált változó mellett egy (vagy akár több) regionalizált változó áll rendelkezésre, melyek térben korrelálnak egymással. E kapcsolat felhasználható az adott talajtulajdonság becslésére és pontosabb térkép származtatható (McBratney, Webster, 1983). A térképezési igények és az elérhető talajtani adatbázisok révén a kokrigelés igen elterjedt módszer lett a talajtérképezésben. De nemcsak talajtani, hanem a talajtulajdonságok becslésében alkalmazható egyéb környezeti adatok is betölthetik a segédinformáció szerepét. Az elmúlt két évtizedben a digitális domborzatmodellek előállítása, illetve a távérzékelés rohamos fejlődésnek indult, mely egyértelműen kedvezett a kokrigelés talajtani alkalmazásának. E fejlődésnek köszönhetően egyre olcsóbban, egyre nagyobb mennyiségben és nem utolsósorban növekvő geometriai, időbeli, illetve spektrális felbontásban állnak rendelkezésre segédinformációk, melyeket aktívan fel lehet használni a talajtulajdonságok kokrigelésen alapuló térbeli becslésére. Erre számos hazai és nemzetközi példát találhatunk (Odeh et al., 1994, 1995; Meul, Van Meirvenne, 2003; Bourennane et al., 2006; Dobai et al., 2012; Rakonczai, Fehér, 2015). Kiemelendő, hogy ezek a segédinformációk a vizsgálat alá vont területre egységes térbeli lefedettséggel állnak rendelkezésre, mely számos új perspektívát is eredményezett.

Tipikusan ilyen megközelítés a krigelés külső drifttel (*kriging with an external drift*), mely megkívánja a segédinformációk teljes térbeli lefedettségét, ugyanakkor a segédinformációk számával kapcsolatban rugalmasabb, mint a kokrigelés. Ez utóbbi abból a tényből fakad, hogy a nagyszámú segédinformáció „túlbonyolítja” a direkt és keresztvariogramok modellezését (Goovaerts, 1997). Továbbá a krigelés külső drifttel módszer egészen új szemszögből közelítette meg az adott talajtulajdonság térbeli változékonyságát. Ahogy az a nevéből is következik, képes kezelni az adatok átlagában megjelenő szisztematikus változást – vagyis a trendet – mely a rendelkezésre álló segédinformációk révén modellezhető. A térbeli becslés feltétele a lineáris kapcsolat a becsülni kívánt talajtulajdonság és a segédinformációk között, melynek együtthatóit implicit becsli a krigelési rendszeren keresztül (Goovaerts, 1997). Habár a gyakorlatban jelentősen több segédinformációt képes szimultán kezelni, mint a kokrigelés, azonban a nagyszámú segédinformáció ellehetetlenítheti a mátrixszámításokat (Hengl, 2009). Továbbá a segédinformációk térbeli értékeiben megjelenő ugrásszerű változások a krigelési rendszer instabilitását okozzák (Goovaerts, 1997). Ugyanakkor ezt az új megközelítést az 1990-es évek második fele óta igen széles körben

alkalmazzák a talajtulajdonságok digitális térképezésében (Goovaerts, 1997, 1999b; Bourennane, King, 2003; Baxter, Oliver, 2005; Simbahan et al., 2006).

A digitális talajtérképezésben a krigelés külső drifftel megközelítés mellett – szinte vele egy időben – további perspektívaként jelent meg a regresszió krigelés (*regression kriging*), mely az előzőhöz hasonlóan képes kezelni a trendet, ugyanakkor a trendmentesítést apriori végzi el (Odeh et al., 1994, 1995; Goovaerts, 1997). A trendmentesítés szintén az adott talajtulajdonság és a rendelkezésre álló segédinformációk közötti kapcsolat számszerűsítésén alapszik, de nem feltétel a lineáris kapcsolat, továbbá a segédinformációk térbeli értékeiben megjelenő ugrásszerű változások is lehetségesek (Hengl, 2009). Ezek tették lehetővé a regresszió krigelés elfogadását és egyre szélesebb körű alkalmazását a talajtulajdonságok digitális térképezésében. A korábban felhasznált segédinformációk köre – mint a digitális domborzatmodell és a belőle származtatott morfológiai paraméterek, illetve a távérzékeléssel nyert képinformációk – tovább bővült kategória típusú segédinformációkkal, mint például földtani, területhasználati, illetve archív talajtérképekkel. A kezdetben alkalmazott többszörös lineáris regressziók mellett megjelentek a GLM (*generalized linear model*) és GAM (*generalized additive model*) módszerek a trend leválasztásra (Hengl, 2009; Poggio et al., 2010, 2013; Kempen, 2011). A regresszió krigelés gyakorlatias volta miatt mind a nemzetközi, mind a hazai digitális talajtérképezésben elfogadott és az utóbbi években gyakran alkalmazott geostatistikai módszer a talajtulajdonságok térbeli becslésére és térképezésére (Hengl et al., 2004, 2007; Simbahan et al., 2006; Minasny, McBratney, 2007a; Dobos et al., 2007; Szatmári, Barta, 2013; Szatmári et al., 2013, 2015a,b; Heuvelink, 2014; Pásztor et al., 2014a,b, 2015, 2016a; Tóth et al., 2015, 2016).

Az 1980-as évek végén kialakult precíziós mezőgazdaság, illetve a talaj- és környezetvédelmi tervezések és beavatkozások részéről új igények fogalmazódtak meg (Webster, Oliver, 1989). Az újonnan megjelenő igények olyan megközelítést kívántak, mellyel nem az adott pontra vagy területegységre vonatkozó legjobb – egyértelmű – becslés, hanem adott, például műtrágyázási szaktanácsadás által meghatározott hatóanyag koncentráció alatti vagy bizonyos szennyezéssel kapcsolatos beavatkozási határérték feletti „valószínűségi térkép” előállítás volt a cél (Webster, Oliver, 1989; Goovaerts, 1994, 1997; Goovaerts et al., 1997). Ezeket az igényeket a bányászati gyakorlatba már bevezetett diszjunktív krigelés (*disjunctive kriging*), indikátor krigelés (*indicator kriging*), illetve sztochasztikus szimulációk képesek kielégíteni (Webster, Oliver, 1989; Goovaerts, 1994, 1997). E megközelítések segítségével adott pontra vagy területrészre vonatkozó bizonytalanság modellezésére és értékelésére nyílik

lehetőség, melyek széles körű alkalmazásra leltek a talajtani kutatásokban (Goovaerts, 1994, 1997, 2000, 2001, 2009; Goovaerts et al., 1997; Lark, Ferguson, 2004; Kerry, Oliver, 2007; Webster, Oliver, 2007; Poggio et al., 2013; Poggio, Gimona, 2014).

Az utóbbi években a digitális talajtérfépezésben is erőteljesen megnövekedett az igény az előállított talajtulajdonság térképek térbeli bizonytalanságának modellezésére, illetve értékelésére (Heuvelink, 2014). A szakirodalomban két – geostatistikai alapokon nyugvó – megközelítést találunk: (1) a krigelési (vagy becslési) variancia alkalmazása konfidencia intervallum származtatásához (Heuvelink, 2014; Kempen et al., 2014; Pásztor et al., 2016a), illetve (2) a korábban említett sztochasztikus szimulációk alkalmazása a térbeli bizonytalanság modellezésére (Goovaerts, 1997, 2000, 2001; Poggio et al., 2013; Poggio, Gimona, 2014; Szatmári et al., 2015a; Geiger et al., 2016). A krigelési variancia számítása a „krigelés” szerves része (Webster, Oliver, 2007); ismeretében igen egyszerűen megadható például a krigeléssel adott becslés körüli 95%-os konfidencia intervallum (Goovaerts, 1997, 1999a; Heuvelink, 2014). Ugyanakkor csakis szigorú feltételek (lásd: V.3.2. fejezet) teljesülése esetén alkalmazható a lokális bizonytalanság jellemzésére, mely feltételek a gyakorlatban többnyire nem adóttak (Journel, 1986; Isaaks, Srivastava, 1989; Journel, Rossi, 1989; Goovaerts, 1997, 1999a). A krigelési eljárásokkal szemben a sztochasztikus szimulációk célja olyan alternatív, de egyenlően valószínű realizációk generálása az adott regionalizált változóra, melyek visszaadják a modell statisztikákat (Deutsch, Journel, 1998; Goovaerts, 1997; Geiger, 2012). A szimuláció során kapott realizációk összessége a térbeli bizonytalanság modellje.

II. 4. A talajmintavételezés optimalizációja a nemzetközi és a hazai talajtérfépezésben

A talajok felvételezése szorosan összefügg a talajok térképezésével, hiszen mindig is a térképezendő területen feltárt talajszelvények, illetve az ott begyűjtött minták szolgáltatták azokat az alapinformációkat, melyek elengedhetetlenek az adott talajtulajdonság(ok) térképezéséhez (Kreybig, 1937; Szabolcs, 1966; Géczy, 1968; McBratney, Webster, 1981; Carter, Gregorich, 2007; Webster, Lark, 2013). Ugyanakkor a terepen végzett munka erőforrás- és időigénye, illetve a laboratóriumi vizsgálatok költségei mindig is korlátozó tényezők voltak (Carter, Gregorich, 2007; Webster, Lark, 2013). Ezért már az 1980-as évek elején megjelentek azon irányú igények, hogy miként is lehetne optimalizálni a mintavételi stratégiákat az adott talajtulajdonság geostatistikai módszereken alapuló térképezéséhez (Burgess et al., 1981; McBratney et al., 1981; McBratney, Webster, 1981). Ezeknek a törekvéseknek a célja volt,

hogy adott költségkeret és/vagy mintaméretbeli korlátozás mellett a mintavételi helyek azon optimális térbeli elrendezését találják meg, mely mellett a térkép várható pontossága – valamilyen alkalmasan megválasztott minőségi mérték tekintetében – maximális (Burgess et al., 1981; McBratney et al., 1981; McBratney, Webster, 1981, 1983).

Napjainkban a nemzetközi szakirodalom a mintavételezésekre két megközelítési módot különböztet meg (Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006; Delmelle, Goovaerts, 2009; Wang et al., 2012; Webster, Lark, 2013): a terv alapút (*design-based*), illetve a modell alapút (*model-based*). A terv alapú megközelítés a klasszikus mintázási elmélethez kapcsolódik, míg a modell alapú mintavételezések kifejezetten a geostatisztikához, illetve az idősor elemzésekhez kapcsolódnak (de Gruijter, ter Braak, 1990, 1992; Brus, de Gruijter, 1993, 1997; de Gruijter et al., 2006; Wang et al., 2012; Webster, Lark, 2013). A két megközelítési mód tulajdonképpen a mintavételezésbe bevezetett „véletlenszerűség” természetében különbözik. Särndal (1978), illetve Hansen és munkatársai (1983) szerint a „véletlenszerűségnek” kitüntetett szerepe van a mintavételben, hiszen a „randomizálás” bevezetésével elkerülhető, hogy a populációra és a mintára szükségtelen apriori feltételezéseket tegyünk. A terv alapú mintavételezések esetén a minta kiválasztása egy előre definiált véletlenszerű kiválasztási metodika alapján történik (Cochran, 1977; Särndal, 1978; Hansen et al., 1983). E módszer lényeges tulajdonsága, hogy a kiválasztás valószínűsége a mintázott populáció minden kiválasztott mintájára ismert (Geiger, 2007a,b); vagyis a „véletlenszerűséget” a minták kiválasztási szintjén vezetjük be, mely a statisztikai alapú következtetések alapját fogja jelenteni (Cochran, 1977; Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006). E megközelítés szerint a vizsgálat alapját képező populáció értékei rögzítettek (Brus, de Gruijter, 1993, 1997; Wang et al., 2012; Webster, Lark, 2013). A terv alapú megközelítésekkel szemben a modell alapú megközelítések szerint a vizsgálat alapját képező populációt egy sztochasztikus folyamat realizációjának tekintjük, melyet egy valószínűségi modellen keresztül közelítünk (Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006; Webster, Lark, 2013). E szerint a mintákon mért értékek nem mások, mint a mérési helyek végtelen kicsiny sugarú környezetében helyet foglaló valószínűségi változók egy-egy véletlenszerű kimenetelei, mely valószínűségi változók egymástól függését bizonyos valószínűségi mechanizmus szabályozza. Ebből az következik, hogy a vizsgálat alapját jelentő „populáció” értékei nem rögzítettek, mint a terv alapú megközelítések esetén (Brus, de Gruijter, 1993, 1997). Ha apriori ismerjük a valószínűségi modellt vagy annak bizonyos komponenseit, akkor ezek az előzetes ismeretek felhasználhatók a mintavételi helyek kiválasztásában, mely a

modell alapú mintavételezések alapját jelenti (Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006; Webster, Lark, 2013). E megközelítések esetén is megkülönböztetett jelentősége van a „véletlenszerűségnek”, mely a statisztikai következtetések alapját fogja jelenteni; ugyanakkor e „véletlenszerűséget” nem kell „saját kezűleg” bevezetnünk a mintavételezésbe (például: véletlenszerű minta kiválasztással), hiszen ez a vizsgálatunk középpontjában elhelyezkedő sztochasztikus folyamat sajátosága (Brus, de Gruijter, 1993, 1997; de Gruijter et al., 2006; Webster, Lark, 2013). A fentiek alapján a két megközelítés közötti legfőbb különbség, hogy míg a modell alapú megközelítés esetén a „véletlenszerűség” a sztochasztikus folyamathoz kapcsolódik, addig a terv alapú megközelítések esetén a „véletlenszerűség” a mintavételi helyek kiválasztásához fűződik (Särndal, 1978; Hansen et al., 1983; de Gruijter, ter Braak, 1990, 1992; de Gruijter et al., 2006).

A hagyományos, illetve az egyszerű krigelés talajtérképezésben történő alkalmazásaival szinte egy időben megjelentek az ezeknek a módszereknek a becslési varianciáján alapuló mintavétel optimalizációk (Burgess et al., 1981; McBratney et al., 1981; McBratney, Webster, 1981). A krigelési variancia számítása a „krigelés” szerves része (Webster, Oliver, 2007). Mivel független az adatponti értékektől, igen idő- és költséghatékony (Burgess et al., 1981; McBratney et al., 1981; Flatman, Yfantis, 1984); továbbá alkalmas a geometriai adatelrendezések (konfigurációk) rangsorolására (Journel, Rossi, 1989; Geiger, 2006a). Yfantis és munkatársai (1987) kimutatták, hogy izotróp térbeli változékonyság mellett optimális mintavételezés szabályos háromszög alapú grid hálóval érhető el. McBratney és Webster (1981) szerint anizotróp térbeli változékonyság esetén a mintavételi helyek az anizotrópia ellipszis kis- és nagytengelyeinek megfelelően sűrítendőek, illetve ritkítandók. Mindemellett rámutattak arra, hogy a mintavételi távolságnak rövidebbnek kell lennie, mint a variogram hatástávolságának. Fontos megjegyezni, hogy a hagyományos és az egyszerű krigelés becslési varianciáján alapuló mintavétel optimalizációk feltételezik a variogram apriori ismeretét (McBratney, Webster, 1981, 1983). A fenti megállapításokat széles körben alkalmazzák a hazai és nemzetközi mintavétel tervezésekben, illetve monitoring rendszerek kidolgozásában (például: Csillag, Kertész, 1989; Cressie et al., 1990; Hagyó, Rajkai, 2004; Farkas et al., 2008). Hazánkban Füst (1982, 1984, 1997, 2011, 2012), illetve Füst és Geiger (2010) fogalmazott meg hasonló – általános – alapelveket, melyek számos hazai kutatás alapját képezték (például: Hatvani et al., 2012; Kovács et al., 2012; Kohán, 2014; Kohán, Szalai, 2014; Szalóki, Somodi, 2014; Szatmári et al., 2015b, 2016).

A nagy volumenű talajfelmérések általában számos talajtulajdonság szimultán történő térképezését célozzák meg (Kreybig, 1937; Szabolcs, 1966; Géczy, 1968). A mintavételezések, illetve a laboratóriumi vizsgálatok idő- és költségigénye sikeresen csökkenthető, ha a drágább vagy bizonyos szempontból körülményesebb mérési metodikával meghatározható talajtulajdonságo(ka)t ritkábban, míg az olcsóbb vagy relatíve egyszerűbb mérési metodikával meghatározható tulajdonságo(ka)t sűrűbben mintázzuk (McBratney, Webster, 1983). Ebben az esetben az optimalizáció alapját a mintázott talajtulajdonságok közötti direkt- és kereszt-variogramok, illetve az ezekre felépített koregionalizáció lineáris modellje jelenti (McBratney, Webster, 1983; Vašát et al., 2010). Ezért a mintavétel optimalizáció feltételezi a direkt- és kereszt-variogramok apriori ismeretét.

Az 1990-es évek végén van Groenigen és Stein (1998) kidolgozta az SSA (*spatial simulated annealing*) algoritmust, mellyel hatékonyabban, számítógépes környezetben lehet végrehajtani a talajmintavételezés optimalizációját. Továbbá számos célfüggvény kidolgozására is sor került, mely függvények minimalizálása vagy maximalizálása révén határozhatjuk meg az optimális mintavételi elrendezést (van Groenigen, 1999). Az SSA algoritmus nagy forduló pontot jelentett a talajmintavételezések optimalizációjában (Brus, Heuvelink, 2007; Heuvelink et al., 2007; Webster, Lark, 2013). Ugyanakkor a módszer egyik legnagyobb hátránya, hogy a mintavétel optimalizációt csakis egy talajtulajdonságra képes elvégezni (Vašát et al., 2010; Szatmári, 2014).

A trendet explicit módon kezelő krigelési módszerek (például: krigelés külső drifftel, regresszió krigelés) esetében előtérbe került azon igény, hogy a becslés során felhasználható segédinformációkra (például: domborzatmodellek, műholdképek) is optimalizált legyen a mintavételezés (Hengl et al., 2003; Heuvelink et al., 2007). Ez utóbbi fontossága abban rejlik, hogy a mintavételi pontok felhasználási köre a „krigelésen túl” tovább bővült a trend leválasztásával (Hengl et al., 2003; Heuvelink et al., 2007). Ezért Brus és Heuvelink (2007) a regresszió krigelés becslési varianciáját alkalmazta a mintavételezés optimalizációja során, melyet az SSA algoritmussal kötöttek össze. E krigelési varianciából képzett célfüggvény minimalizálása szimultán keresi a földrajzi tér, illetve a segédinformációk által meghatározott attribútum tér optimális „lefedettségét”. A regresszió krigelés becslési varianciája alkalmasnak bizonyult mintavételezési tervek optimalizációjához (például: Baume et al., 2011; Melles et al., 2008, 2011; Romary et al., 2011; Bechler et al., 2013; Heuvelink et al., 2013). E megközelítés feltételezi a regressziós modell struktúrájának, illetve a regressziós reziduumok variogramjának

apriori ismeretét (Brus, Heuvelink, 2007). A módszer hátrányaként továbbra is fennmaradt, hogy az optimalizáció csakis egy talajtulajdonságra végezhető el (Szatmári, 2014).

A segédinformációk által definiált attribútum tér optimális „lefedésére” az LHC (*latin hypercube*) is gyakran alkalmazott módszer a talajtanban (például: Minasny, McBratney, 2006, 2007b, 2010; Chu et al., 2010; Lin et al., 2014; Pahlavan Rad et al., 2014; Brus, 2015), melyet gyakran az SSA módszerrel kötnék össze. Az attribútum tér mellett a földrajzi tér optimális „lefedettsége” a földrajzi koordináták (Minasny, McBratney, 2006, 2007b) vagy a segédinformációk direkt- és kereszt-variogramjainak (Lin et al., 2014) segítségével érhető el.

Habár a hazai talajtani kutatásokban kevés mintavétel optimalizációs munkát találunk, mindazonáltal a társtudományokban, mint a hidrológiában és a hidrogeológiában számos ilyen témájú dolgozat született. Fontos ezekre kitérni a hazai vonatkozásuk és a talajtannal való szoros kapcsolódásuk miatt. A talajvízállások modellezése nagyban támaszkodik a geostatistika módszereire (például: Bárdossy, Molnár, 2003; Mucsi et al., 2013; Kohán, Szalai, 2014; Fehér, 2015a,b; Geiger, 2015), mely arra vezethető vissza, hogy a talajvízszint regionalizált változónak tekinthető (Geiger, Mucsi, 2005). Bárdossy és Bogárdi (1983), illetve Bogárdi és munkatársai (1985) ezekre az alapokra építették fel kutatásaikat, mely során egy többváltozós optimalizációs módszert dolgoztak ki. A kutatásuk négy vizsgált paraméterre (két földtani réteg vastagságára és porozitására) vonatkozott. Az optimalizáció során figyelembe vették a pénzügyi ráfordításokat, mely által egy olyan döntéstámogatási rendszert építettek fel, mely számos kritérium alapján tudja meghatározni az optimális megfigyelések térbeli elrendezését. Szidarovszky (1983a,b) többfunkciójú megfigyelési hálózatokra vonatkozóan dolgozott ki általános modelleket, mely a hagyományos krigelés becslési varianciájának területi minimalizálásán alapult. Dolgozatában három optimalizációs modellt dolgozott ki: (1) egy új megfigyelési pont optimális elhelyezése a meglévők figyelembe vételével, (2) adott számú új mintavételi pont optimális elhelyezése pont vagy blokk krigeléshez, illetve (3) adott minőségi érték eléréséhez szükséges mintavételi pontok számának meghatározása és optimális elhelyezése pont vagy blokk krigeléshez. Az optimalizáció során szintén figyelembe vette a költséghatékonyságot, mint fontos gazdasági tényezőt. Fontos kiemelni Bárdossy és Molnár (2004), illetve Molnár (2009) kutatásait, akik talajvízszint észlelő hálózatok tér- és időbeli optimalizálását célozták meg. Munkájuk szintén a hagyományos krigelés becslési varianciájának minimalizálásán alapult. Az optimalizáció során használt összetett célfüggvény három komponensből épült fel: (1) legkisebb információvesztéssel elhagyható kutak, (2) a krigelés becslési varianciájának összterületi minimalizálása az optimális térbeli elrendezéshez,

illetve (3) az észlelő hálózat részét képező „szelvények” optimalizálása. A célfüggvény egyes komponenseit súlytényezőkkel látták el, mely segítségével az adott igényekhez igazítható a megfigyelési hálózat kialakítása. Optimalizációs módszerként az SA (*simulated annealing*) algoritmust alkalmazták. Napjaink mintavételezési gyakorlatában az SA algoritmust felváltotta a van Groenigen és Stein (1998) által kidolgozott SSA algoritmus, mely a folytonos földrajzi térre kidolgozott verziója az alapjaiban diszkrét SA módszernek (van Groenigen, 1999).

A korábban bemutatott megközelítések feltételezik a variogramok – esetenként kereszt-variogramok – apriori ismeretét, melyek nem feltétlenül állnak a rendelkezésünkre a kutatások kezdeti szakaszán (Burgess et al., 1981; Warrick, Myers, 1987; van Groenigen et al., 1999; Füst, Geiger, 2010; Webster, Lark, 2013). A talajmintavételezések optimalizációjához használatos variogramok, illetve kereszt-variogramok származtathatók (van Groenigen, 1999): (1) direkt erre irányuló új mintavételezésből, illetve (2) archív talajtani információkból. A variogram becslésére irányuló új mintavételezések esetén a talajtanban gyakran alkalmazzák Warrick és Myers (1987) módszerét (például: van Groenigen, 1999; van Groenigen et al., 1999), az összetett mintavételezést (például: Oliver, Webster, 1986; Csillag, Kertész, 1989; Webster, Oliver, 2007; Lark, Webster, 2013), illetve a rendelkezésre álló segédinformációk (például: domborzatmodellek, légifotók) tapasztalati variogramjait, melyek analógiát mutathatnak a térképezendő talajtulajdonság variogramjával (például: Kerry, Oliver, 2003). Lark (2000a,b, 2002) a variogram becslését „maximum likelihood” módszerrel végezte el, melyet SSA alapú mintavétel optimalizációval kötött össze. Eredményei azt mutatták, hogy a „maximum likelihood” módszeren alapuló mintavétel optimalizáció jó alapot nyújt a variogram becsléséhez, illetve jelentősen csökkenti az ehhez megkívánt mintaméretet. A variogramok becslésének alapját jelenthetik archív talajtani információk is, melyek származhatnak (van Groenigen, 1999): (1) az adott területen folytatott korábbi talajfelmérésekből, illetve (2) az adott területtel talajtani analógiát mutató, egyéb területről származó talajtani információkból. Míg az előbbire számos példát találhatunk a szakirodalomban (például: Burgess et al., 1981; Brus, Heuvelink, 2007; Vašát et al., 2010; Szatmári et al., 2015b, 2016), addig az utóbbira csak igen keveset (például: van Groenigen et al., 1999).

„I was joined from Mexico by H. E. Cuñanalo in 1968. He pointed out that time-series analysts have similar problems, and they treat actuality as realizations of stochastic processes to describe quantitatively fluctuations in time. Could we not do the same for soil? So we switched our thinking from the classical mode and took a leap of imagination; we should treat the soil as if it were random – against all the tenets of the day!”

RICHARD WEBSTER
(2015)

III. ELMÉLETI MEGFONTOLÁSOK

E fejezetben a célkitűzéseimben megfogalmazott, geostatisztikai megközelítésen alapuló digitális talajtérképezési és mintavétel optimalizációs módszertan elméleti alapjainak a bemutatására kerül sor. Ezen elméleti megfontolások sarokköve, hogy a talajtulajdonságok térbeli változékonysága egy sztochasztikus folyamatként közelíthető és modellezhető.

III. 1. A talajtulajdonságok, mint regionalizált változók

A geostatisztika kiindulási alapja azon explicit feltételezés, hogy a vizsgálat tárgyát jelentő jelenség térbeli és/vagy időbeli sztochasztikus folyamatként közelíthető és modellezhető (Journel, 1986; Isaaks, Srivastava, 1989; Pannatier, 1996; Goovaerts, 1997; Deutsch, Journel, 1998; Webster, Oliver 2007). A talajtulajdonságok és a talajban megfigyelhető jelenségek bonyolult, komplex, egymásra kölcsönösen pozitív vagy negatív irányú hatást gyakorló fizikai, kémiai, illetve biológiai folyamatok eredői, melyből következik, hogy ezeknek a tulajdonságoknak, jelenségeknek – fizikai értelemben véve – determinisztikusnak kell lenniük. Jenny (1941) munkája ezen a tézisen alapszik, miszerint ha ismerjük, hogy a talaj miként képződött és formálódott a talajképző közetben az idők során, akkor elméletileg ismernünk kell jellemzőit és ezek múlt-, jelen- és jövőbeli viselkedését is. Fontos megjegyezni, hogy e szemléletmód megkívánja a talajtulajdonságok és a talajban lejátszódó folyamatok, ha nem is teljes körű, de kielégítő ismeretét, illetve ezeknek a tényezőknek a pontos, de legalábbis megbízható mérésének lehetőségét is. Ugyanakkor e fizikai, kémiai, illetve biológiai folyamatok közötti interakciók – mind irányukban és erősségükben – oly komplexek és változékonyságúak a talajképződés előrehaladtával, hogy a talajokkal kapcsolatos ismereteink közel sem mondhatók teljesnek (Webster, 2000; Oliver, Webster, 2014). Továbbá a talajon végzett méréseinken keresztül is csak adott pont(ok)ban, illetve adott időpillanat(ok)ban vagyunk képesek „szemlélni” a talajban lejátszódó folyamatokat, jelenségeket. Ezért ezek térbeli csakúgy, mint az ezeket irányító folyamatok időbeli változékonyságáról töredékesek az ismereteink. Ennél fogva a talaj sok esetben úgy jelenik meg számunkra, mintha az egy

véletlenszerű folyamat realizációja lenne (Webster, 2000, 2015). *Természetesen a talaj nem egy véletlenszerű folyamat „eredménye”* – egy determinisztikus rendszer, mely alá van rendelve a fizika törvényeinek –, *ugyanakkor a talajokat érintő ismereteink hiánya, illetve a megismerhetőségének korlátai miatt a talaj egy sztochasztikus folyamatként közelíthető és modellezhető* (Webster, 2000, 2015; Oliver, Webster, 2014, 2015). E megközelítés alapján *a talajtulajdonságok olyan jellemzői a talajnak, melyek a mérési hely végtelen kicsiny sugarú környezetében valószínűségi változónak tekinthetők, ugyanakkor megadható – nem feltétlenül egy – olyan térben értelmezett függvény, mely ezeket a valószínűségi változókat összekapcsolja. Az ilyen tulajdonságokat a geostatisztika regionalizált változóknak nevezi* (Geiger, 2006b).

III. 2. A stacionaritás kérdése

A geostatisztika további feltételezése a stacionaritás, mely egyben számos félreértés okozója is (Journel, 1986; Myers, 1989; Webster, 2000). A vizsgált jelenségről egy-egy adatponti érték áll a rendelkezésünkre, azonban a számításokhoz több ismételt mérésre lenne szükségünk, melyekre nincs lehetőségünk (II.1. fejezet). Mindazonáltal, ha a vizsgált jelenség homogénnek feltételezhető az adott területen, akkor ennek adatpontonkénti értékét úgy tekinthetjük, mint a regionalizált változó önmaga ismétlését a térben. Ez az ismétlődés ugyanazon valószínűségi függvény sok realizációjának azonosságát adja, és statisztikai számításokat tesz lehetővé (Geiger, 2006a). A stacionaritás hipotézisét pedig a vizsgált jelenség térbeli homogenitásának különböző fokaihoz kapcsoljuk (Geiger, 2006a). A geostatistikában a következő stacionaritási típusok a leginkább használatosak (Myers, 1989; Webster, 2000; Geiger, 2006a; Füst, Geiger, 2010):

- *Szigorú stacionaritás*: Egy valószínűségi függvény szigorúan stacionárius, ha térbeli törvénye invariáns a translációra.
- *Gyenge stacionaritás*: Egy valószínűségi függvény gyengén stacionárius, ha a várható érték létezik és független a hely megválasztásától, továbbá a kovariancia létezik és csakis a pontokat elválasztó szeparációs vektortól függ.
- *Belső hipotézis*: Egy valószínűségi függvény belső hipotézissel bír, ha a várható érték helytől független, továbbá a félvariogram létezik és csakis a pontokat elválasztó szeparációs vektortól függ.

A feltételezett stacionaritási típus határozza meg, hogy a valószínűségi modellen belül milyen következtetéseket tehetünk. Fontos hangsúlyozni, hogy *a stacionaritás a valószínűségi függvény attribútuma, s nem jellemzője sem az adatoknak, sem pedig a valóságnak* (Journel,

1986; Myers, 1989; Goovaerts, 1997; Webster, 2000). Ezért arra kérdésekre, hogy „a talajban megfigyelt jelenség stacionárius-e?”, nem létezik válasz. Sokkal inkább helyénvaló az a kérdés, hogy „az adott jelenségre felépített stacionárius valószínűségi modell reális-e?” (Webster, 2000). Az utóbbi kérdés a modellezés céljainak a tükrében válaszolható meg. Fontos kiemelni, hogy *a stacionaritás típusának kiválasztására nincs teszt, ezért ezeket a „hipotéziseket” a minta alapján általában nem lehet igazolni* (Füst, Geiger, 2010). Ezért a stacionaritás „hipotézise” szabatosan nem igazi hipotézis – mint inkább egy döntés a szakértő részéről – hisz az adatok alapján sem elfogadni, sem visszautasítani nem lehet azt (Journel, 1986; Goovaerts, 1997; Füst, Geiger, 2010). *Ez azt is jelenti, hogy a mintavételi stratégia kidolgozása során olyan „helyzet” kialakítása fontos, mely ezeket a stacionaritási feltételezéseket nyilvánvalóvá teszi* (Geiger, 2007b; Füst, Geiger, 2010).

III. 3. A talajmintavételezések modell alapú megközelítése és optimalizációja

A mintavételezés fogalmába tartozik a *minta kiválasztása azon célból, hogy a rajta végzett mérések, megfigyelések alapján következtetéseket tegyünk a (cél) populációra* (de Gruijter et al., 2006; Geiger, 2007b; Wang et al., 2012). A következtetéseink vonatkozhatnak (Wang et al., 2012): (1) a populáció valamely paraméterére (például: átlag, szórás), (2) a populáció értékeire ismeretlen pontokban vagy területegységekre (blokkokra), illetve (3) a populációhoz kapcsolódó entitás(ok) térbeli helyzetére.

De Gruijter és munkatársai (2006), illetve Webster és Lark (2013) szerint a mintavételezés kardinális kérdése, hogy milyen következtetés(ek)e)t kívánunk levonni. Ez a mintavételi stratégia kidolgozásának az alapját fogja jelenteni, ahol *mintavételi stratégián a mintavételi helyek kiválasztásának mikéntjét és a keresett paraméter becslésére szolgáló módszer együttesét értjük* (Särndal, 1978; Hansen et al., 1983; Brus, de Gruijter, 1993; 1997; de Gruijter et al., 2006). Talajtérképezés esetén a cél, hogy becsléseket tegyünk az adott talajtulajdonság értékeire ismeretlen pontokban vagy blokkokra. A vizsgált talajtulajdonságot regionalizált változónak tekintve ezeket a becsléseket egy valószínűségi modellen keresztül tehetjük meg, ahol a térképezendő talajtulajdonságot egy sztochasztikus folyamat egyedi realizációjának tekintjük. Ha ez a valószínűségi modell – vagy annak bizonyos komponense(i) – apriori ismert, akkor e modell felhasználható a mintavételi helyek kiválasztásában, mely a modell alapú mintavételezés sarokkövét jelenti (Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006; Webster, Lark, 2013). E modell alapú megközelítést alkalmazva nagyban növelhetők a begyűjtött mintákból levonható statisztikai következtetések hatásfoka (van

Groenigen, 1999, 2000; de Gruijter et al., 2006; Heuvelink et al., 2007; Lark, Lapworth, 2012; Webster, Lark, 2013). Ez mindaddig igaz, amíg a célunk, hogy becsléseket tegyünk a vizsgált talajtulajdonság értékeire ismeretlen pontokban vagy blokkokra. Ahogy a terv alapú megközelítések esetén, úgy a modell alapú megközelítések esetén is megkülönböztetett szerepe van a „véletlenszerűségnek”, mely a statisztikai alapú következtetések alapját fogja jelenteni (Särndal, 1978; Hansen et al., 1983; de Gruijter et al., 2006). Ugyanakkor ez a „véletlenszerűség” természetében más, hiszen ezt nem kell „saját kezűleg” bevezetnünk a mintavételezés folyamatába (például: véletlenszerű minta kiválasztással), mert ez a vizsgálatunk középpontjában elhelyezkedő sztochasztikus folyamat sajátja (Särndal, 1978; Hansen et al., 1983; Brus, de Gruijter, 1993, 1997). Ezért *a modell alapú megközelítések esetén a minták kiválasztása nem egy előre definiált véletlenszerű kiválasztási metodika alapján történik, hanem az apriori valószínűségi modell alapján* (de Gruijter et al., 2006; Brus, de Gruijter, 1993, 1997; Webster, Lark, 2013).

A gyakorlatban rendszeresen felmerülő kérdéskör, hogy a mintavételezés céljaiban megfogalmazott kérdések milyen mintavételi elrendezés mellett válaszolhatók meg a leghatékonyabban. Ezért a mintavételi helyek kijelölése optimalizációs problémának tekinthető. Dolgozatomban *optimalizációnak tekintem azt a folyamatot, mely során a lehetséges, alternatív mintavételi elrendezések (vagy konfigurációk) közül azon konfiguráció kerül kiválasztásra, mely mintavételi elrendezés az előre deklarált szempont vagy szempontrendszer tekintetében a legjobb*. A kiválasztás egy előre definiált célfüggvény (*objective function*) globális minimumának (vagy maximumának) a meghatározásán alapszik. A célfüggvény definiálása egy alkalmasan megválasztott minőségi mérték alapján történik, mely igazodik a mintavételezés céljaihoz (de Gruijter et al., 2006). A minőségi mérték segítségével objektíven rangsorolhatók a lehetséges alternatív mintavételi elrendezések, mely alapján kiválasztható a legjobb konfiguráció. A fentiek alapján *adott valószínűségi modell és előírányzott mintavételi megszorítások mellett optimálisnak tekintem azt a mintavételi elrendezést, mely a – mintavételezés céljaihoz alkalmasan megválasztott minőségi mértékből képzett – célfüggvény globális minimumát (vagy maximumát) adja*.

IV. MINTATERÜLETEK

A módszertani kutatásaimhoz szükséges mintaterületek kiválasztása az alábbi szempontok alapján történt:

- Elsődleges szempont volt, hogy a mintaterületek területi kiterjedésükben, vagyis a térképezés léptékében különbözzenek egymástól.
- Az előzővel azonos súlyú szempont volt, hogy a mintaterületeken érvényre jutó talajképződési folyamatok (például: humuszosodás, kilúgozás) lényeges különbségeket mutassanak.
- További elvárás volt, hogy a választott területeken a talajok szervesanyag-tartalmának térbeli változékonysága eltérő legyen; vagyis a viszonylagos homogenitástól a komplex heterogenitásig reprezentálják a kutatási területek a szervesanyag-tartalom lehetséges térbeli változékonyságát.
- További szempont volt a kiválasztásnál, hogy kellő mennyiségű talajtani észlelés (minimum 30 adatpont) álljon a rendelkezésemre.
- Az utóbbival azonos súlyú elvárás volt, hogy kellő mennyiségben és minőségben álljanak rendelkezésemre a területekre vonatkozóan segédinformációk (például: digitális domborzatmodellek, légifotók, archív talajtérképek).

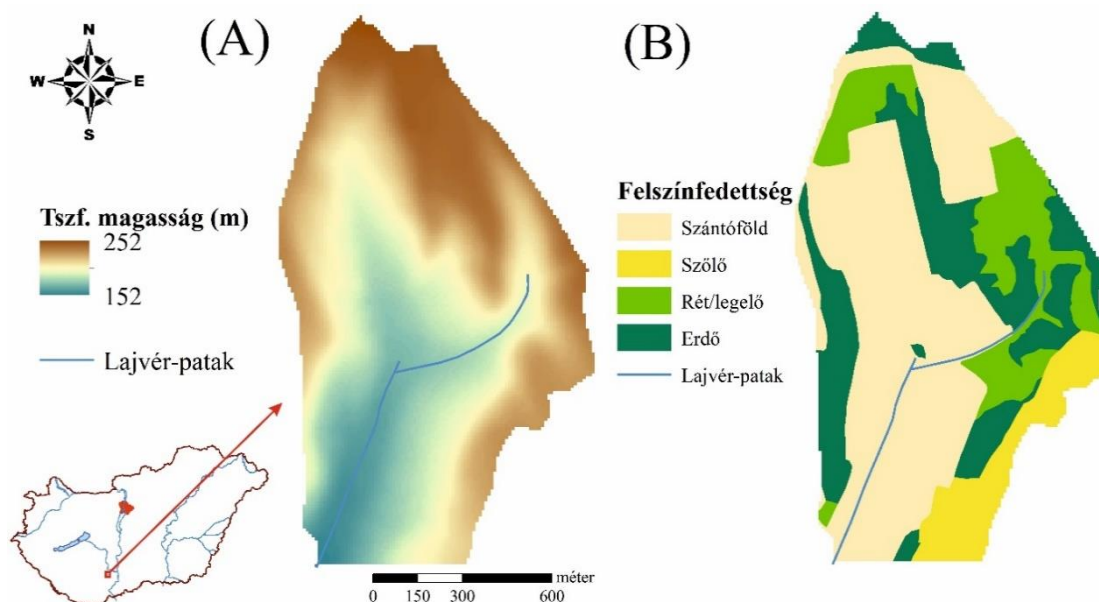
A felsorolt szempontok alapján az alábbi három mintaterületre esett a választásom:

- Szálkai kisvízgyűjtő (1,3 km²),
- Előszállási szántóföldek (16,6 km²),
- Zala megye (3784 km²).

A mintaterületek nem csak területi kiterjedésükben, hanem az ott érvényre jutó talajképződési folyamatok tekintetében is lényeges különbségeket mutatnak. Ennek következtében a talajban raktározott szerves anyag mind mennyiségében, mind minőségében rendkívüli változatosságot mutat a mintaterületek között. A szervesanyag-tartalom térbeli változékonysága is igen eltérő az egyes területeken. Míg az Alföldön található előszállási mintaterület viszonylag egyveretűnek mondható, addig a Szekszárdi-dombságban található szálkai mintaterület a kisvízgyűjtőkre jellemző anyag- és energia-forgalmának köszönhetően változatosabb képet fest. Zala megye térbeli változékonyságában kiemelkedik az előzőkhöz képest, melyet rendkívül tagolt domborzatának, illetve a láptalajok kiugróan magas szervesanyag-tartalmának köszönhet.

IV. 1. Szálkai kisvízgyűjtő

A közel 132 hektáros mintaterület a Szekszárdi-dombságban található, Szálka községtől északkeletre (IV.1. ábra). A mintaterület éghajlata mérsékelt meleg – mérsékelt száraz, az évi középhőmérséklet 10,2–10,5 °C, az évi csapadékösszeg 650–670 mm között alakul (Dövényi, 2010).



IV.1. ábra

A szálkai mintaterület (A) digitális domborzatmodellje, illetve (B) felszínfedettségi térképe (2005-ös légifotó alapján)

A kisvízgyűjtő 152–252 m tengerszint feletti magasságban található. Lejtőinek többsége az 5–12%-os, illetve a 17–25%-os lejtőkategóriákba sorolható, melynek következtében a mintaterület vízerózióval erősen veszélyeztetett (Borcsik et al., 2011). Ehhez társul még a terület közel felére jellemző intenzív mezőgazdasági művelés, mely következtében a veszélyeztetettség konkrét formát is ölt: a felületi rétegerózió, illetve a vonalas erózió okozta anyag-áthalmazódás jelei mind a domboldalakon, mind a völgytalpakon jól megfigyelhetők. Míg az előbbi esetben számos helyen felszínre bukkan a talajképző laza kőzet, addig az utóbbinál a kisvízgyűjtőt elhagyó időszakos vízfolyás fokozatos feliszapolódása követhető nyomon. A mintaterület délkeleti részén a szőlőtelepítés miatt teraszok találhatók, melyek igen gyakori műszaki, talajvédelmi beavatkozások a Szekszárdi-dombságban (Benyhe, 2013).

A mintaterület talajképző kőzetét löszös üledékek, illetve harmadkori üledékek képezik. A kisvízgyűjtőn fellelhető főbb talajtípusok: a csernozjom barna erdőtalaj, illetve a Ramann-féle barna erdőtalaj (Borcsik et al., 2011). A területet sújtó vízerózió eredményeként a talajtípusok gyengén, közepesen, illetve erősen erodált fokozatai jelennek meg. A lejtők alján

felhalmozódó talajanyag lejtőhordalék-talajokat képez, míg az időszakos vízfolyás mentén a talajképződést vízhatás is kíséri. A csernozjom barna erdőtalajokat, illetve a Ramann-féle barna erdőtalajokat meghatározó talajképződési folyamatok: a humuszosodás, a kilúgozás, illetve a vasas agyagosodás (Stefanovits et al., 2010). Csernozjom barna erdőtalajok esetén a humuszanyag felhalmozódás kifejezettebb. Ezért ezek a talajtípusok eredendően humuszanyaggal jól, illetve közepesen ellátottak. Mindazonáltal a területet sújtó vízerózió, illetve az intenzív mezőgazdálkodás következtében igen szélsőséges értékek (0,7–3,5%) között változik szervesanyag-tartalmuk, mely térbeli változékonyságukban is drasztikusan megmutatkozik. Ezek az erőteljes változások jól kapcsolódnak a területhasználathoz; míg az erdők alatt a talajok humuszanyaggal jól ellátottak, addig a szántók talajai humuszban igen szegények (Borcsik et al., 2011).

IV.1. táblázat

A szálkai mintaterület felszínfedettségi kategóriáinak területi megoszlása (2005-ös légifotó alapján)

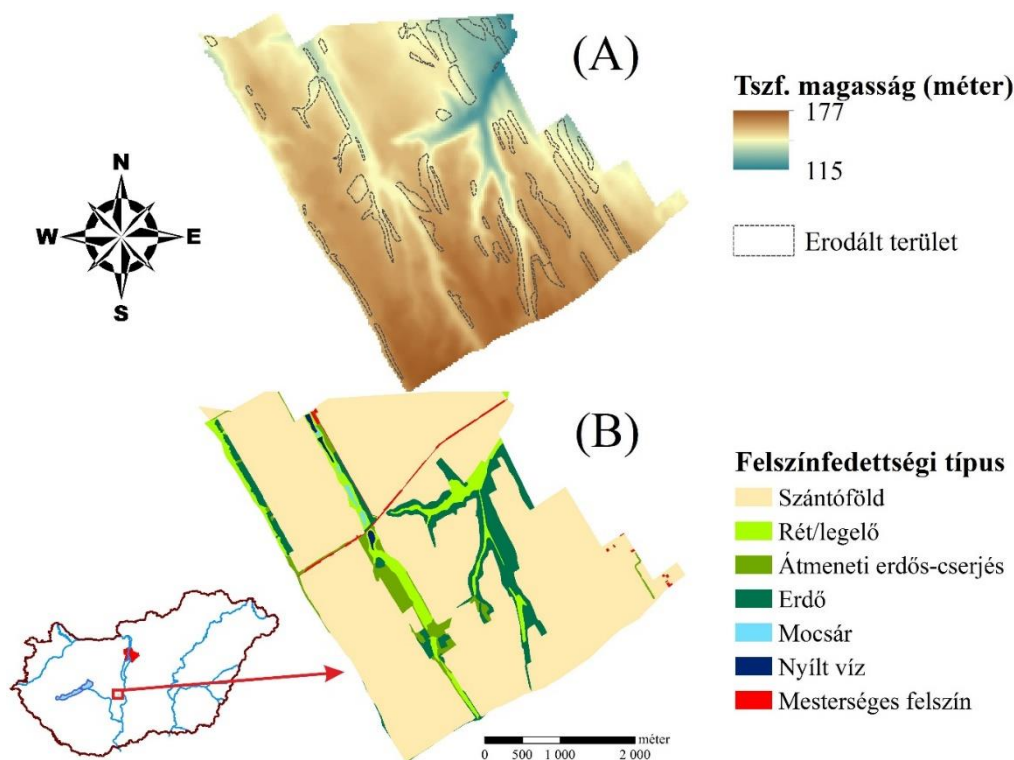
Felszínfedettség típus	Terület [ha]	Terület [%]
Szántóföld	65,2	49,5
Szőlő	12,6	9,5
Rét/legelő	21,2	16,1
Erdő	32,7	24,9
<i>Összesen</i>	<i>131,7 hektár</i>	<i>100%</i>

A kisvízgyűjtő területhasználata igen változatos. Szántóföldi növénytermesztés folyik mintegy 65 hektáron (IV.1. ábra és IV.1. táblázat), ahol többnyire monokultúras kukoricatermesztés folyik (Borcsik et al., 2011). A kukorica rossz talajvédő hatása miatt a szántók közepesen, illetve erősen erodáltak. A terület meredekebb lejtőit rétek, akác-, illetve tölgyerdők fedik (IV.1. ábra), melyek kellő védelmet nyújtanak az vízerózió káros hatásaival szemben. A mintaterület délkeleti részén kialakított teraszokon szőlőültetvények találhatók (IV.1. ábra).

IV. 2. Előszállási szántóföldek

A mintegy 1661 hektáros mintaterület a Közép-Mezőföld déli részén, a Sárobozvárdi löszplatón található, Előszállás községtől délnyugatra (IV.2. ábra). A terület éghajlata mérsékelt meleg, száraz vidék. Az évi középhőmérséklet 10,2–10,4 °C, az évi csapadékösszeg 540–580 mm között ingadozik. A csapadék gyakran heves zivatarok formájában jelentkezik a késő tavaszi időszakban (Dövényi, 2010).

A mintaterület 115–177 m tengerszint feletti magasságban fekszik a löszplatón. A terület a Közép-Mezőföld kistájnak megfelelően síkvidéki, ezért lejtőinek döntő többsége a 0–5%-os lejtőkategóriába tartozik. Ugyanakkor az alsó-pleisztocénben bekövetkezett mozaikos feltöredeзések és kiemelkedések révén a terület – alföldi jellegéhez képest – igen tagolt (Dövényi, 2010). Így a nagyobb hullámokkal tarkított felszínrészletei és a platóba vágódó löszvölgyek meredek oldalai potenciálisan erózió-veszélyeztetetté teszik a területet (Szatmári, Barta, 2012).



IV.2. ábra

Az előszállási mintaterület (A) digitális domborzatmodellje és az erózióval sújtott térszínnek, illetve (B) felszínfedettségi térképe (2005-ös légifotó alapján)

A kutatási területet befedő, néhányszor tíz méteres vastagságú eolikus löszön típusos mészlepedékes csernozjom talajokat találunk (Stefanovits et al., 2010; Dövényi, 2010). Ezek a talajok igen kedvező tulajdonságokkal rendelkeznek, mint például: a humuszanyag felhalmozódása, a morzsalékos talajszerkezet, illetve a kalciummal telített talajoldat kétirányú mozgása (Stefanovits et al., 2010). A mészlepedékes csernozjomok mellett a vízerózió bizonyítékeként jelennek meg a lejtőhordalék és földes kopár talajok. A löszvölgyekben – ahol a talajvíz közelsége, illetve az összegyülekező felszíni vizek miatt a talajt vízhatás éri – réti csernozjomokat találunk. A csernozjom talajokat rendszerint a humuszanyagok felhalmozódása jellemzi, mely az ősi mezősségi növénytakaró alatt bekövetkezett talajképződés eredménye.

Ennél fogva ezek a talajok eredendően humuszanyaggal gazdagon ellátottak (Stefanovits et al., 2010), ugyanakkor a területen folytatott intenzív és monokultúras mezőgazdálkodás, illetve a területet érintő vízerózió révén a szervesanyag-tartalom igen tág határok (1,5–4,4%) között változik.

IV.2. táblázat

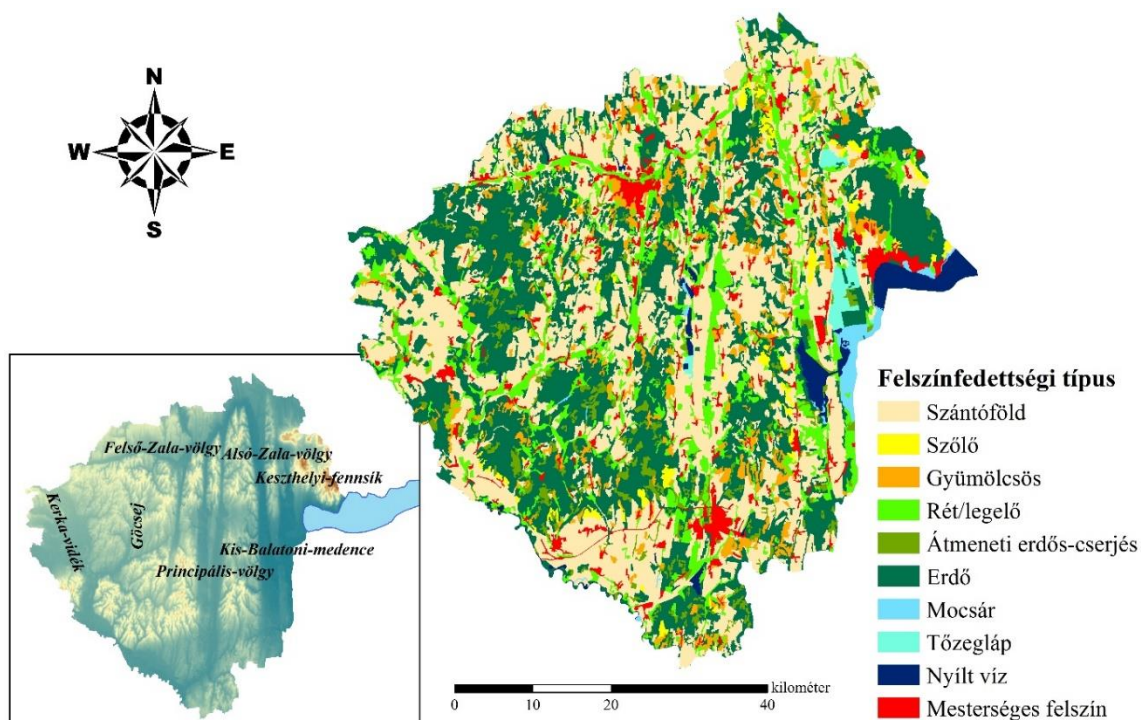
Az előszállási mintaterület felszínfedettségi kategóriáinak területi megoszlása (2005-ös légifotó alapján)

Felszínfedettség típus	Terület [ha]	Terület [%]
Szántóföld	1401,2	84,3
Rét/legelő	79,9	4,8
Átmeneti erdős-cserjés	46,0	2,8
Erdő	124,8	7,5
Mocsár	3,0	0,2
Nyílt víz	4,1	0,3
Mesterséges felszín	2,0	0,1
<i>Összesen</i>	<i>1661,0 hektár</i>	<i>100%</i>

A kutatási terület földhasználata igen „egyhangú”, lévén hogy kiemelkedő termőhelyi adottságainak köszönhetően főként szántóföldi használat alatt áll. A szántók több mint 80%-át teszik ki a területnek (IV.2. táblázat), ahol őszi búzát, napraforgót, illetve kukoricát termesztene, gyakran monokultúras termesztésben. A mintaterületen húzódó löszvölgyek meredekebb oldalain akácerdőket találunk (IV.2. ábra), melyek a korai vonalas erózió által kikényszerített felhagyás eredményei (Szatmári, Barta, 2012). A völgyek lankásabb részeit rétek, legelők borítják (IV.2. ábra).

IV. 3. Zala megye

Zala megye az ország nyugat-délnyugati határán található (IV.3. ábra). A megye magába foglalja: a Zalai-dombságot, a Keszthelyi-riviérát, a Kis-Balaton-medence északnyugati részét, a Keszthelyi-fennsík és a Tátika-csoport nyugati részeit, illetve a Felső-Kemeneshát déli részét. Zala megye éghajlata mérsékelt hűvös – mérsékelt nedves (a megye keleti területei közelítik a nedves éghajlati típust), az évi középhőmérséklet 9,2–9,8 °C, az évi csapadékösszeg 700–780 mm között változik (Dövényi, 2010).



IV.3. ábra
Zala megye felszínfedettségi térképe (CORINE LC 2012 alapján)

A megye változatos természetföldrajzi adottságokkal rendelkezik, melynek kialakításában hosszú időn át tartó komplex geológiai folyamatok játszottak közre (Dövényi, 2010; Stefanovits et al., 2010). A megye területét észak-déli irányú széles völgyek, illetve azokra közel merőleges elhelyezkedésű kisebb völgyhálózatok szabdalják fel. A viszonylag hűvös és csapadékos éghajlat, illetve a változatos térszínek következtében a megye gazdag vízfolyásokban. A széles völgytalpak lejtése igen alacsony (0–5%-os lejtőkategória), továbbá a talajképződést vízhatás kíséri. A domboldalak lejtői az 5–12%-os, illetve a 17–25%-os lejtőkategóriába sorolhatók, ezért az erősen tagolt dombsági tájon a vízerózió meghatározó talajdegradációs folyamat (Stefanovits et al., 2010).

A megye fő talajképző kőzetei nyugaton szénsavas meszet nem tartalmazó periglaciális vályog, illetve pannon agyagos üledék, míg keleten vékony lösztakaró (Stefanovits et al., 2010). Zala megye nyugati területei a pangóvízes (vagy pseudoglejes) barna erdőtalajok hazája, melyet az agyagos, rossz vízgazdálkodási tulajdonságokkal rendelkező talajképző kőzetnek, illetve a hűvös és csapadékos éghajlatnak köszönhet. A keleti területeken agyagbemosódásos barna erdőtalajokat, illetve löszön képződött Ramann-féle barna erdőtalajokat találunk. A domborzat és a vízfolyások nyújtotta lehetőségek kedveztek a réti és lápos talajok képződésének, melyek nem elhanyagolható területi kiterjedéssel jelennek meg, például a Kis-Balaton-medencében és a Principális-völgyben (Dövényi, 2010). A Keszthelyi-fennsíkon kevés

– mészkövön, illetve dolomiton képződött – rendzinát találunk, továbbá szintén kicsiny területi kiterjedéssel fekete nyiroktalajok is előfordulnak a Tátika-csoport bazaltjain (Dövényi, 2010). Az agyagbemosódásos, illetve a pangóvizes barna erdőtalajokra jellemző talajképződési folyamatok: a humuszosodás, a kilúgozás, az agyagosodás, illetve az agyagvándorlás (Stefanovits et al., 2010). Míg az agyagbemosódásos barna erdőtalajokban közepes mértékű savanyodás figyelhető meg, addig a pszeudoglejes barna erdőtalajok esetén a savanyodás igen erőteljes, továbbá az agyagszétesés (vagy podzolosodás) folyamatához a redukció jelensége is társul (Stefanovits et al., 2010). Ennél fogva a talajerózió mellett a savanyú kémhatás is jelentős talajtermékenységet gátló tényező a megyében. Láptalajok esetén a láposodás, mint – speciális – talajképződési folyamat a meghatározó, mely során a talajképződés állandó vagy az év nagy részében fennálló vízborítás alatt játszódik le. Ennek következtében a láptalajok szervesanyag-tartalma kiugróan magas, mely akár 60–80% is lehet (Stefanovits et al., 2010).

IV.3. táblázat

Zala megye felszínfedettségi kategóriáinak területi megoszlása (CORINE LC 2012 alapján)

Felszínfedettség típus	Terület [km²]	Terület [%]
Szántóföld	1357,6	35,9
Szőlő	55,8	1,5
Gyümölcsös	199,0	5,3
Rét/legelő	428,4	11,3
Átmeneti erdős-cserjés	169,3	4,5
Erdő	1231,3	32,5
Tőzegláp	29,9	0,8
Mocsár	33,2	0,9
Nyílt víz	78,2	2,0
Mesterséges felszín	200,7	5,3
<i>Összesen</i>	<i>3783,4 km²</i>	<i>100%</i>

Zala megye földhasználata igen változatos, melyet főként domborzati viszonyainak köszönhet. A magasabban fekvő dombsági területek főként erdőkkel borítottak, melyek a megye egész területének mintegy 33%-át teszik ki (IV.3. táblázat). A lankásabb domboldalakon a szántók, a gyümölcsösök, illetve szőlőültetvények foglalnak helyet, míg a széles völgyekben természetközeli rétek, illetve legelők találhatók (IV.3. ábra). A dombsági táj nyugati területein a szántóföldi növénytermesztés kevésbé sikeres az alacsony termékenységű pszeudoglejes barna erdőtalajok révén. Ugyanakkor a keleti területek termékenyebb talajain már a kukorica is sikeresen megterem (Dövényi, 2010; Stefanovits et al., 2010).

V. ANYAG ÉS MÓDSZER

E fejezetben a felhasznált talajtani adatbázisok, segédinformációk, illetve – a célkitűzéseimben megfogalmazott – geostatisztikai megközelítésen alapuló digitális talajtérképezési és mintavétel optimalizációs módszertan bemutatására kerül sor. A fejezet első felében ismertetem a rendelkezésemre álló talajtani adatbázisokat, illetve segédinformációkat. A fejezet második felében részletesen bemutatom a célkitűzéseimben megjelölt módszereket. A fejezet végén részletesen írok a kialakított módszertan szinergiájáról. A fejezetben bemutatott képletek, egyenletek és függvények egységesítetten szerepelnek a dolgozatban. Az alkalmazott matematikai jelöléseket az 1. melléklet foglalja össze. A bemutatott geomatematikai és geostatisztikai számításokhoz, elemzésekhez, modellezésekhez, optimalizációkhoz szabad-hozzáférésű, nyílt forráskódú szoftvereket (*Free and Open Source Softwares, FOSSs*) használtam fel: *R software environment* (R Development Core Team, 2016), illetve *SAGA GIS* (Conrad et al., 2015). A következőkben bemutatásra kerülő eljárások, illetve algoritmusok szoftveres implementációit a 2. melléklet foglalja össze.

V. 1. Felhasznált talajtani adatbázisok és segédinformációk

V. 1. 1. Szálkai kisvízgyűjtő

A szálkai mintaterület talajtani adatbázisát a Szegedi Tudományegyetem, Természeti Földrajzi és Geoinformatikai Tanszék bocsátotta a rendelkezésemre. A mintaterület talajfelvételezésére, illetve a laboratóriumi vizsgálatokra 2009-ben került sor (Borcsik et al., 2011; Farsang, 2016). A talajfelvételezés fő célja a szálkai kisvízgyűjtőt sújtó vízerózió hatásainak nyomon követése, objektív értékelése és modellezése (Borcsik et al., 2011; Farsang, 2016). Ennek eredményeként az erdők – melyek kellő védelmet nyújtanak a vízerózió ellen (Thyll, 1997) – kissé alulreprezentáltak, míg a szántó területek – ahol is az erózió a leginkább ki tudja fejteni káros hatásait (Barta, 2004) – kissé felülreprezentáltak a mintavételi pontok számában. Az adatbázis 56 pontszerű észlelést tartalmaz, mely a feltalaj (0–10 cm) szervesanyag-tartalmára vonatkozik. Két mintavételi pont nem rendelkezik laboratóriumi vizsgálati eredményekkel, ezért kizártam őket a további feldolgozásokból.

A kisvízgyűjtőre digitális domborzatmodellt állítottam elő 10 méteres felbontással a rendelkezésemre álló 1:10.000-es méretarányú topográfiai térképek szintvonalainak és magasságpontjainak digitalizálásával. A digitális talajtérképezésben bevált gyakorlat, hogy a domborzatmodellből számos további morfometriai paramétert (például: lejtőszög, terep

egyenetlenség index, relatív lejtő pozíció) vezetünk le, melyek sikeresen alkalmazhatók a talajtulajdonságok becslésében (Moore et al., 1993; Odeh et al., 1994, 1995; Dobos et al., 1997, 2000; McBratney et al., 2003; Hengl, Reuter, 2008). A mintaterület digitális domborzatmodelljéből a következő paramétereket származtattam: lejtőszög, relatív lejtő pozíció, profil görbület, planáris görbület, LS tényező (Wischmeier, Smith, 1978), TWI (*topographic wetness index*), illetve a völgyek mélysége. További segédinformációként rendelkezésemre állt a mintaterületről 2005-ben készült légifelvétel. A talajok szervesanyag-tartalma és az egyes területhasználati típusok között kimutatható kapcsolat van: az erdők alatt a talajok szervesanyag-tartalma magasabb, míg a szántóterületeken a szervesanyag-tartalom alacsonyabb, mely az intenzív és monokultúras mezőgazdálkodással felborított szerves anyag mérlegnek, illetve a vízerózió pusztító hatásainak tulajdonítható (Szatmári, Barta, 2012, 2013; Centeri et al., 2014; Jakab, Takács, 2014; Jakab et al., 2014, 2016; Szalai et al., 2016). Ezért a mintaterület felszínfedettségi térképe jól alkalmazható a szervesanyag-tartalom térbeli becslésében. A 2005-ös légifelvétel alapján digitalizáltam a felszín fedettségi típusokat, melyekből egy felszínfedettségi térképet állítottam elő (IV.1. ábra), mely a következő kategóriákat tartalmazza: szántóföld, szőlő, rét/legelő, illetve erdő.

V. 1. 2. Előszállási szántóföldek

Az előszállási mintaterületről a Fejér Megyei Mezőgazdasági Szakigazgatási Hivatal Növény- és Talajvédelmi Igazgatósága bocsátott a rendelkezésemre talajtani adatokat. E talajtani információk részletes talajszelvény leírásokat, illetve laboratóriumi vizsgálati eredményeket foglaltak magukban, melyek az 1988–1989-es időszakban születtek termőföldértékelés céljából. A jegyzőkönyveket 2010-ben Dr. Barta Károllyal digitalizáltuk az Igazgatóság székhelyén (Velence) a diplomamunkám (Szatmári, 2011) keretei között. A mintaterületre 117 talajszelvény esett, melyek helyszíni, illetve laboratóriumi vizsgálati jegyzőkönyveit adatbázisba rendeztük. Az adatbázis tartalmazza a talajgenetikai szintek mélységét, a termőréteg vastagságát, illetve minden egyes genetikai szintre vonatkozó szervesanyag-tartalmat, pH-t, Arany-féle kötöttséget és szénsavas mésztartalmat. Az előszállási kutatási terület talajtani adatbázisát két részre osztottam: (1) az adatpontok közel 90%-át használtam fel a geostatistikai modellalkotáshoz, illetve (2) az adatpontok közel 10%-át – melyek kiválasztása véletlenszerűen történt – használtam fel az elkészült szervesanyag-tartalom térkép értékeléséhez. Az utóbbi, ún. kontroll adatok a geostatistikai modellalkotástól mindvégig függetlenek maradtak.

Az előszállási mintaterületre digitális domborzatmodellt állítottam elő 10 méteres felbontással a rendelkezésemre álló 1:10.000-es méretarányú topográfiai térképek szintvonalainak és magasságpontjainak digitalizálásával. A domborzatmodellből a következő paramétereket származtattam: lejtőszög, relatív lejtő pozíció, profil görbület, planáris görbület, lokális görbületek, LS tényező (Wischmeier, Smith, 1978), TWI, terep egyenetlenség index, völgyek mélysége, illetve potenciálisan beérkező direkt és diffúz napsugárzás. Az előszállási mintaterületről is rendelkezésemre álltak 2005-ben készült légifelvételek, melyek segítségével egy felszínfedettség térképet állítottam elő (IV.2. ábra), mely az alábbi típusokat tartalmazza: szántóföld, rét/legelő, átmeneti erdős-cserjés, erdő, mocsár, nyílt víz, mesterséges felszín.

V. 1. 3. Zala megye

Zala megye területére a Digitális Kreybig Talajinformációs Rendszer (DKTIR) feltalajra (0–30 cm) vonatkozó szervesanyag-tartalom adatait, illetve a DKTIR a talajok vízgazdálkodási tulajdonság talajtérképezési egységeit a Magyar Tudományos Akadémia Talajtani és Agrokémiai Intézete (MTA TAKI) bocsátotta a rendelkezésemre. A Kreybig-féle átnézetes talajismereti térképezés (Kreybig, 1934, 1937, 1950) 1934-től 1951-ig tartott. A térképeket és a felvételi adatokat mintegy 60 évvel később az MTA TAKI Környezetinformatikai Osztálya digitalizálta és rendezte koherens térbeli talajinformációs rendszerbe (Pásztor et al., 2010b, 2013b). Zala megyére összesen 1989 darab DKTIR szelvény esik. Az előszállási területhez hasonlóan a zalai mintaterület talajtani adatbázisát is két részre osztottam: (1) az adatpontok közel 90%-át használtam fel a geostatisztikai modellalkotáshoz, illetve (2) az adatpontok közel 10%-át használtam fel az elkészült szervesanyag-tartalom térkép értékeléséhez.

Segédinformációként rendelkezésemre állt az EU-DEM digitális domborzatmodell (Bashfeld, Keim, 2011) mintaterületre vonatkozó kivágata, melynek felbontása 25 méter. Az EU-DEM modellből a következő paramétereket származtattam: lejtőszög, relatív lejtő pozíció, profil görbület, planáris görbület, lokális görbületek, LS tényező (Wischmeier, Smith, 1978), TWI, terep egyenetlenség index, völgyek mélysége, illetve potenciálisan beérkező direkt és diffúz napsugárzás. A DKTIR adatok felvételezése több mint 60 évvel ezelőtt történt, ezért a jelenkori felszínfedettségi adatbázisokat (mint például: a CORINE Land Cover) nem használhattam segédinformációként lévén, hogy a megye területhasználata nagyban megváltozhatott az elmúlt 60 évben. Ezért a rendelkezésemre álló DKTIR a talajok vízgazdálkodási tulajdonság talajtérképezési egységeit használtam fel segédinformációként. A

digitális talajtérképezésben jól bevált gyakorlatnak számít az archív talajtérképek, talajtérképezési egységek segédinformációként való alkalmazása (Pásztor et al., 2016b; Minasny, McBratney, 2016). E talajtérképezési egységek implicit tartalmazzák a talajfelvételezők és –térképezők szakmai ismereteit és mentális talaj-táj modelljeit, melyek segíthetik a szervesanyag-tartalom térbeli modellezését.

V. 2. Térképezési módszertan

V. 2. 1. A becslési grid definiálása

A digitális talajtérképezés egyik sarkalatos pontja az eredmény térkép(ek) geometriai felbontásának a megválasztása (McBratney et al., 2000, 2003; Hengl, 2006). Habár a gyakorlat számára fontos, hogy már a térképezés(ek) előtt kiválasszuk a geometriai felbontást, ugyanakkor a szakirodalom csak igen kevés helyen tárgyalja az ezzel kapcsolatos megfontolásokat, kérdéseket, illetve döntési helyzeteket. Ezzel kapcsolatosan Hengl (2006) munkája nyújtja a legátfogóbb gyakorlati megfontolásokat, melyeket a dolgozatomban én is alkalmaztam.

A következő fejezetben bemutatásra kerülő regresszió krigelés (V.2.4. fejezet) adott pont(ok)ra vonatkozóan végez becslés(ek)e)t a térképezendő talajtulajdonság térbeli modellezésekor. Krigelés során a becslési pontokat hagyományosan egy becslési gridként definiálva adjuk meg (Deutsch, Journel, 1998; Hengl, 2006), ahol is a grid „háló” egyes „csomópontjai” (továbbiakban grid pontok) fogják jelenteni ezeket a pontokat. Egy 2D-s becslési grid definiálása a következő paraméterek alapján történik (Hengl, 2006): (1) az x és y koordináták minimum értékei (x_{min} , y_{min}), (2) a grid pontok közötti távolság az x és y tengely mentén (g), (3) a grid pontok száma az x és y tengely mentén (N_x , N_y), (4) az alkalmazandó koordináta rendszer, illetve (5) az ún. bennfoglaló doboz (*bounding box*). Az (1)–(3) paraméterek definiálása kötelező (Deutsch, Journel, 1998; Hengl, 2006), tulajdonképpen ezekből a paraméterekből állítható elő az adott grid topológiája. A (4) és (5) paraméterek definiálása opcionális (Hengl, 2006). Az (1) paraméter a „bal alsó” grid pont (vagy kezdőpont) koordinátái. A (2) paraméter a grid geometriai felbontása, míg az (5) paraméter a kutatási terület kiterjedését definiálja. Fontos kiemelni, hogy raszteres adatfájlok definiálása is hasonló paraméterek alapján történik, azonban ekkor a „bal alsó” koordináta az ott található raszter cella – mely nem egy pont – „bal alsó” sarkára vonatkozik (T. Hengl szóbeli közlése). Ebből az

következik, hogy a raszteres és grides adatállományok nem körültekintő felhasználása az adatértékek északkelet-délnyugat irányú „elcsúszását” eredményezheti.

A becslő gridek geometriai felbontását Hengl (2006) *inspection density* módszerével határoztam meg az egyes mintaterületekre:

$$g = 0,05 \cdot \sqrt{\frac{A}{n}}, \quad (\text{V.1.})$$

ahol g a becslő grid geometriai felbontása (méter dimenzióban), A a mintaterület kiterjedése/területe (négyzetméter dimenzióban) és n a mintavételi pontok száma. Ezek alapján előállítottam az egyes mintaterületekre vonatkozó becslési grid topológiáját, mely a becslési grid „modelljének” tekinthető. A definiált grid topológiákat és becslési grideket használtam fel a segédinformációk előfeldolgozása, a feltáró alapadat elemzés, illetve a térbeli becslés és szimuláció során.

V. 2. 2. Feltáró alapadat elemzés

A feltáró alapadat elemzés célja az adatok elsődleges értékelése, bizonyos alapvető statisztikai mennyiségek kiszámítása és az adatok grafikus reprezentációjának az elemzése (Geiger, 2007a). A feltáró alapadat elemzés során az egyes mintaterületekre rendelkezésemre álló szervesanyag-tartalom adatok szerkezetével ismerkedtem meg, mely az adatok alkalmazásának megfelelő módját, illetve bizonyos szintű korlátozását eredményezte. Az adatsorokon elvégzett vizsgálatok típusát, a felhasználásuk módját és az ezekből várt eredményeket az V.1. táblázat foglalja össze. A táblázat utolsó, „megjegyzések” c. oszlopa a vizsgálatok kivitelezésével kapcsolatos megjegyzéseimet tartalmazza.

Az adatsorok áttekintése céljából kiszámítottam az összefoglaló statisztikákat. A kiugró és extrém értékek szűrésére a box-plot eljárást választottam, mely igen gyakran alkalmazott eljárás a föld- és környezettudományokban (Geiger, 2007a). Az adatok eloszlásának normalitását grafikus és analitikus eszközökkel vizsgáltam. A normál eloszlásnak központi szerepe van a geostatistikai modellezésekben, mely annak tulajdonítható, hogy a legtöbb módszer (például: variográfia, krigelés, szekvenciális gaussi szimuláció) a többváltozós normál (vagy multi-gaussi) megközelítésen alapul (Geiger, 2012).

V.1. táblázat

A feltáró alapadat elemzés során alkalmazott eljárások és a várt eredmények tematikus összefoglalása

Vizsgálat tárgya	Alkalmazott eljárás	Várt eredmények/Felhasználás	Megjegyzések
- Adatok áttekintése,	- Összefoglaló statisztikák (átlag, medián, szórás, minimum, maximum, ferdeség),	- Az adathalmaz megismerése, - Következtetések a populációra vonatkozóan,	- Az Eredmények c. fejezetben a nyers és a szűrt adatsor egy táblázatban kerül bemutatásra,
- Kiugró és extrém értékek,	- Box-plot technika,	- Kiugró és extrém értékek azonosítása és kiszűrése,	—
- Normalitás vizsgálat,	- Grafikus eszk.: Hisztogram és Q-Q diagram, - Analitikus eszk.: Shapiro-Wilk, illetve Kolmogorov-Smirnov próba,	- Az adatok eloszlása szimmetrikus, normális?	- Ha a mintaszám alacsony (< 100): Shapiro-Wilk próba, - Ha a mintaszám magas (> 100): Kolmogorov-Smirnov próba,
- Az adatponti értékek térbeli eloszlása, változékonysága,	- Az adatponti értékek térképi ábrázolása,	- Az adatponti értékek térbeli eloszlásának megismerése, - Trend és térbeli folytonosság megfigyelése,	—
- Extrémítások földrajzi helyzete,	- Indikátor térképek az adatsorok decilisei alapján,	- Alacsony és magas értékek térbeli helyzetének megfigyelése, - Ezek térbeli kapcsolata,	—
- Adatpontok közötti interakciók,	- Üres-tér függvény módosított alakja,	- Az adatpontok közötti interakciók típusa, - Az üres-tér függvény szélsőségei,	- Baddeley és Turner (2005), - Az üres-tér függvény módosítása a becslési grid alapján,
- Adatpontok és a segédinformációk viszonya,	- back-to-back hisztogramok, - területi részesedés,	- Az adatpontok reprezentativitása a segédinformációkra nézve, - Extrapoláció veszélye az attribútum térben,	- Folytonos segédinformációk esetén back-to-back hisztogramok, - Kategorikus segédinformációk esetén területi részesedés,

A fenti feltáró alapadat elemzési eljárások érzéketlenek az adatpontok földrajzi helyzetére, ezért a következőkben olyan módszereket alkalmaztam, melyek képesek figyelembe venni ezt a sajátosságot is. Térképi alapon megjelenítettem az adatponti értékeket, melyekkel az adatok térbeli eloszlását és változékonyságát vizsgáltam. Segítségével megfigyelhetők az adatok átlagában megjelenő szisztematikus változások, illetve a folytonossági mintázatok (Isaaks, Srivastava, 1989).

Indikátor térképeket állítottam elő az adatponti értékek $\{z(\mathbf{u}_i), i = 1, \dots, n\}$ alapján. Az indikátor térképek z_k vágás értékeit az adatok decilisei jelentették:

$$i(\mathbf{u}_i; z_k) = \begin{cases} 1, & \text{ha } z(\mathbf{u}_i) \leq z_k, \\ 0, & \text{ha } z(\mathbf{u}_i) > z_k \end{cases}, \quad (\text{V.2.})$$

ahol $i(\mathbf{u}_i; z_k)$ az indikátor kódolt adatponti érték a z_k vágás érték alapján. A módszer alapján 9 térképet állítottam elő, melyek alkalmasak voltak arra, hogy megvizsgáljam az extremitások földrajzi helyzetét és kapcsolatát.

Az adatpontok közötti interakciók feltárására a pontmintázat analízisekben alkalmazott üres-tér függvényt (*empty space function*) alkalmaztam. Az $F(r)$ üres-tér függvény definíció szerint (Baddeley, Turner, 2005; Baddeley, 2008):

$$F(r) = P\{d(\mathbf{u}, X) \leq r\}, \quad (\text{V.3.})$$

ahol \mathbf{u} egy tetszőlegesen választott pont koordináta vektora a vizsgált területen, r a távolság és $d(\mathbf{u}, X)$ a legrövidebb távolság a tetszőlegesen választott pont és a hozzá legközelebb eső mintavételi pont között. Baddeley és Turner (2005) üres-tér függvény algoritmusát oly módon módosítottam, hogy a kiválasztott pontok a becslési gridek pontjai legyenek. Így a meghatározott $F(r)$ függvény a becslési pontokra specifikus, mellyel direkt információ nyerhető a krigelési szomszédságról és a legközelebbi krigelési szomszédról. A meghatározott $F(r)$ függvény empirikus alakja indikálni fogja az adatpontok közötti interakció típusát vagyis, hogy az adatpontok térbeli elhelyezkedése szabályos (egymást kizáró pontesemények), csoportosult (egymást nem kizáró pontesemények) vagy független. Utóbbi a Poisson folyamatnak (*Poisson process*) tekinthető. Az $F(r)$ függvény hazai alkalmazásait a föld- és környezettudományokban nem sikerült fellelnem. Ugyanakkor a nemzetközi szakirodalom

egyre gyakrabban alkalmazza a talajmintavételi stratégiák jellemzésére (például: Hengl, 2009; Lin et al., 2014), mely indokolja, hogy jelen kutatásaimban alkalmazzam a módszert.

Utolsó lépésként vizsgáltam az adatpontok és a rendelkezésemre álló segédinformációk kapcsolatát. Ezt az indokolja, hogy a geostatistikai modellezés során a szervesanyag-tartalom és a segédinformációk közötti számszerűsített kapcsolatot fogom felhasználni a térbeli becslésben. Ezért fontos kérdés, hogy az adatpontokra eső segédinformáció értékei mennyiben is tekinthetők reprezentatívnak az adott segédinformáció egészére vonatkozóan. Utóbbira úgy tekinthetünk, mint egy térben kimerítően mintázott adathalmazra. E vizsgálatokhoz az ún. back-to-back hisztogramokat, illetve az adatpontok összterületi részesedését alkalmaztam. A back-to-back hisztogramok folytonos segédinformációk esetén alkalmazandók. Esetében azt vizsgáltam, hogy az adatpontokra eső segédinformáció értékeinek eloszlása mennyiben is hasonlít a segédinformáció egészének eloszlására. Az adatpontok összterületi részesedése kategorikus segédinformációk esetén alkalmazandó. Esetében azt vizsgáltam, hogy az adatpontok részesedése az egyes kategóriákban mennyiben is arányos a kategóriák összterületi részesedésével. E módszerek segítségével megvilágíthatók a kutatási terület azon részei, melyeken – az attribútum térben vett – extrapolációra kell számítanunk.

V. 2. 3. Segédadatok előkészítése a geostatistikai modellezésekhez

A korábban bemutatott grid topológiáknak (V.2.1. fejezet) kitüntetett szerepe van a segédinformációk előkészítésében, mely a következő okokra vezethető vissza: (1) az alkalmazandó regresszió krigelés megkívánja, hogy minden egyes becslési pontban ismertek legyenek a segédinformációk értékei (V.2.4. fejezet), továbbá (2) az implementáció részéről elvárt, hogy a felhasználandó segédinformációk, illetve a becslő grid azonos grid topológiával rendelkezzenek. Ezért a rendelkezésemre álló segédinformációkat az egyes mintaterületekre előállított grid topológiákra „mintáztam” oly módon, hogy az eredeti adatértékek ne „sérüljenek”. E lépés segítségével technikailag előkészítettem az egyes mintaterületeken a további geostatistikai vizsgálatokat és modellezéseket. Ugyanakkor a segédinformációkon további előfeldolgozások is szükségesek.

Goovaerts (1997) szerint a rendelkezésünkre álló segédinformációkat a következőképpen csoportosíthatjuk:

- *folytonos segédinformációk*: A segédinformáció által felvehető értékek halmaza kontinuum számosságú,

- *kategorikus segédinformációk*: A segédinformáció által felvehető értékeknek a száma véges vagy megszámlálhatóan végtelen.

A folytonos segédinformációk egymással korrelálhatnak (Hengl, 2009); erre példa, hogy a legtöbb digitális talajtérképezési munka úgy állít elő folytonos segédinformációkat, hogy a rendelkezésre álló digitális domborzatmodellből származtat domborzati paramétereket (például: Odeh et al., 1994, 1995; McBratney et al., 2003; Hengl et al., 2003, 2004; Hengl, Reuter, 2008). Ezek szorosabban kapcsolód(hat)nak a vizsgált talajtulajdonság térbeli eloszlásához, így „pontosabb” térkép várható. Ugyanakkor ezek a domborzati paraméterek csakis az esetek kicsiny hányadában függetlenek egymástól (Hengl, Reuter, 2008). A digitális talajtérképezésben gyakran alkalmazott többszörös lineáris regresszió analízis – mely a következő fejezetben (V.2.4. fejezet) bemutatásra kerülő regresszió krigelésnek fontos komponense – azonban feltételezi a prediktor változók függetlenségét, melyet a korábbi megfigyelés cáfol. Ezért célom volt a multikollinearitás csökkentése oly módon, hogy az analízisbe bevont segédinformációkat függetlenné teszem. Ezt a segédinformációk főkomponens analízisével értem el. Az eljárás célja olyan új, mesterséges változók kialakítása, melyek egymástól függetlenek, de legalábbis páronként korrelálatlanok (Geiger, 2007a; Szabó, 2014). Továbbá igaz, hogy a főkomponensek kapcsolatrendszerének összvarianciája egyenlő a segédinformációk összvarianciájával, illetve, hogy az első főkomponens írja le az összvariancia legnagyobb hányadát, a második az összvariancia következő legnagyobb hányadát és végül az utolsó főkomponens írja le az összvariancia legkisebb hányadát (Geiger, 2007a). Ez azt is jelenti, hogy a főkomponens analízis segítségével sikeresen csökkenthető a dimenziószám, mellyel felgyorsíthatók a sokszor bonyolult és hosszadalmas mátrixszámítások.

A kategorikus segédinformációkat indikátor változókká transzformáltam. E transzformáció során olyan binárisan kódolt, ún. indikátor változókat állítottam elő, ahol az „1” jelzi az adott kategória jelenlétét, míg a „0” jelzi az adott kategória hiányát (Goovaerts, 1997).

V. 2. 4. Regresszió krigelés

A legáltalánosabb értelemben *a krigelés olyan lineáris interpolációs eljárás, mely a variogramot (vagy a kovarianciát) alkalmazza a térbeli változékonyság kifejezésére, minimalizálja a becsült érték hibájának varianciáját, melyet az előre jelzett értékek térbeli eloszlásából becsül* (Geiger, 2006a). A krigelés összes típusa visszavezethető a lineáris regresszió becslési egyenletére (Goovaerts, 1997):

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i(\mathbf{u}) [Z(\mathbf{u}_i) - m(\mathbf{u}_i)], \quad (\text{V.4.})$$

ahol $\lambda_i(\mathbf{u})$ a krigelési súly, melyet a $z(\mathbf{u}_i)$ adatponti értékhez rendeltünk (utóbbira úgy tekintünk, mint a $Z(\mathbf{u}_i)$ valószínűségi változó realizációja), $m(\mathbf{u})$ és $m(\mathbf{u}_i)$ a $Z(\mathbf{u})$ és $Z(\mathbf{u}_i)$ valószínűségi változók várható értékei és n az adatpontok száma. A krigelés célja, hogy minimalizálja a $\sigma_E^2(\mathbf{u})$ becslési varianciát azon megszorítás mellett, hogy a becslés egyértelmű legyen, vagyis minimalizáljuk:

$$\sigma_E^2(\mathbf{u}) = \text{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\}, \quad (\text{V.5.})$$

a következő megszorítás mellett:

$$E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0, \quad (\text{V.6.})$$

A krigelés típusa változik attól függően, hogy milyen modellt adaptálunk a $Z(\mathbf{u})$ valószínűségi függvényhez (Goovaerts, 1997). A digitális talajtérképezésben gyakori megközelítési mód, hogy a $Z(\mathbf{u})$ valószínűségi függvényt két komponensre bontjuk (Goovaerts, 1997; Hengl et al., 2004, 2007; Lark, 2012; Webster, 2015):

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}), \quad (\text{V.7.})$$

ahol $m(\mathbf{u})$ a trend komponens és $R(\mathbf{u})$ a reziduális (vagy sztochasztikus) komponens. A reziduális komponenst egy stacionárius valószínűségi függvényként modellezzük, melynek a tulajdonságai a következők (Goovaerts, 1997):

$$E\{R(\mathbf{u})\} = 0, \quad (\text{V.8.})$$

$$\text{Cov}\{R(\mathbf{u}), R(\mathbf{u} + \mathbf{h})\} = E\{R(\mathbf{u}) \cdot R(\mathbf{u} + \mathbf{h})\} = C_R(\mathbf{h}), \quad (\text{V.9.})$$

ahol $C_R(\mathbf{h})$ a reziduális komponens kovarianciája, mely csakis a pontokat elválasztó \mathbf{h} szeparációs vektortól függ. Mivel a reziduális komponens várható értéke nulla, ezért:

$$E\{Z(\mathbf{u})\} = m(\mathbf{u}). \quad (\text{V.10.})$$

A fentiek tükrében a regresszió krigelés a következőképp írható fel, visszavezetve azt a lineáris regresszió egyenletére (Goovaerts, 1997):

$$Z_{RK}^*(\mathbf{u}) - m_{SK}^*(\mathbf{u}) = \sum_{i=1}^{n(\mathbf{u})} \lambda_i^{SK}(\mathbf{u}) \cdot [Z(\mathbf{u}_i) - m_{SK}^*(\mathbf{u}_i)], \quad (\text{V.11.})$$

ahol $m_{SK}^*(\mathbf{u})$ és $m_{SK}^*(\mathbf{u}_i)$ a $Z(\mathbf{u})$ és $Z(\mathbf{u}_i)$ valószínűségi változók várható értékei, illetve $\lambda_i^{SK}(\mathbf{u})$ az egyszerű krigelés súlya, melyet a $r(\mathbf{u}_i)$ reziduum értékéhez rendeltünk (utóbbira úgy tekintünk, mint az $R(\mathbf{u}_i)$ realizációja). A fenti egyenletet átrendezve, illetve $m_{SK}^*(\mathbf{u})$ -t és $m_{SK}^*(\mathbf{u}_i)$ -t felbontva a következő formulát kapjuk:

$$Z_{RK}^*(\mathbf{u}) = \sum_{k=0}^p \beta_k^{GLS} \cdot q_k(\mathbf{u}) + \sum_{i=1}^{n(\mathbf{u})} \lambda_i^{SK}(\mathbf{u}) \cdot \left[Z(\mathbf{u}_i) - \sum_{k=0}^p \beta_k^{GLS} \cdot q_k(\mathbf{u}_i) \right], \quad (\text{V.12.})$$

ahol β_k^{GLS} a k -adik regressziós együttható, $q_k(\mathbf{u})$ k -adik független változó értéke és p a regresszióba bevont független változók száma. Kikötjük, hogy $k=0$ esetén $q_0(\mathbf{u}) \equiv 1$. A fentiek alapján elmondható, hogy *a regresszió krigelés szimultán alkalmazza a térképezendő talajtani változó és a rendelkezésünkre álló segédinformációk közötti többszörös lineáris regressziót, illetve a regresszió reziduumaira felépített egyszerű krigelési rendszert a célváltozó térbeli becslésére*. A könnyebb kezelhetőségért a fenti egyenlet (V.12. egyenlet) mátrixos formában is megadható:

$$Z_{RK}^*(\mathbf{u}) = \mathbf{q}^T(\mathbf{u}) \cdot \boldsymbol{\beta}_{GLS} + \boldsymbol{\lambda}_{SK}^T(\mathbf{u}) \cdot (\mathbf{z} - \mathbf{Q} \cdot \boldsymbol{\beta}_{GLS}), \quad (\text{V.13.})$$

ahol $\mathbf{q}(\mathbf{u})$ a független változók vektora, $\boldsymbol{\beta}_{GLS}$ a regressziós együtthatók vektora, $\boldsymbol{\lambda}_{SK}(\mathbf{u})$ az egyszerű krigelési súlyok vektora, \mathbf{z} az adatponti értékek vektora, \mathbf{Q} a független változók adatpontokra vonatkozó $(n \times p + 1)$ mátrixa. A regresszió krigelés minimalizálja a hiba varianciát azon megszorítás mellett, hogy a becslés egyértelmű legyen. A regresszió krigelés $\sigma_{RK}^2(\mathbf{u})$ becslési varianciája mátrixos formában a következő (Hengl et al., 2004, 2007):

$$\sigma_{RK}^2(\mathbf{u}) = C(0) - \mathbf{c}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u}) + [\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})]^T \cdot (\mathbf{Q}^T \cdot \mathbf{C}_R^{-1} \cdot \mathbf{Q})^{-1} \cdot [\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})] \quad (\text{V.14.})$$

ahol $C(0)$ a reziduális komponens varianciája és \mathbf{c} az adatpontok és az ismeretlen pont közötti kovariancia vektora.

Habár a talajtérképezési kutatásokban igen gyakran alkalmazott módszer a regresszió krigelés, ugyanakkor a nemzetközi szakirodalom kevés helyen tárgyalja a mögöttes koncepciókat, melyek figyelembevételre elengedhetetlen része a módszer helyes alkalmazásának. Definíció szerint regresszió krigelés során a trend leválasztása többszörös lineáris regresszióval történik, mely a többváltozós geomatematikai modellezésekben jól ismert

eljárás (Kovács, Szabó, 2015; Juhos et al., 2015, 2016). Az illesztett regressziós egyenlet alapján származtathatók a reziduumok, melyek a független változók értékeihez tartozó mért és becsült értékek különbségei (Geiger, 2007a). A reziduumok a következő tulajdonságokkal rendelkeznek (Webster, 1997): (1) a reziduumok ugyanazon valószínűségi változó realizációi, illetve (2) ezen valószínűségi változó eloszlása normál, melynek várható értéke nulla. A gyakorlatban sokszor megfigyelhető, hogy a reziduumok térben struktúrált kapcsolatot mutatnak (Webster, 2015). Ez sérti a valószínűségi változó egyik fontos sajátosságát, miszerint a valószínűségi változó realizációi függetlenek egymástól (II.1. fejezet). Ezért a GLS (*generalized least squares*) módszert alkalmaztam a többszörös lineáris regresszió paramétereinek becslésére, mely módszer képes számításba venni a reziduumok közötti kapcsolatot a reziduumok \mathbf{C}_R kovariancia mátrixán keresztül (Cressie, 1993):

$$\boldsymbol{\beta}_{GLS} = (\mathbf{Q}^T \cdot \mathbf{C}_R^{-1} \cdot \mathbf{Q})^{-1} \cdot \mathbf{Q}^T \cdot \mathbf{C}_R^{-1} \cdot \mathbf{z}, \quad (\text{V.15.})$$

A GLS módszerrel becsült többszörös lineáris regresszió paramétereinek a segítségével származtathatók az új reziduumok, melyek egy stacionárius valószínűségi függvényként modellezhetők. A térbeli becslésben felhasznált segédinformációkat „stepwise” módszerrel választottam ki. A gyakorlatban sokszor megengedettnek tekintik a hagyományos krigelés alkalmazását a reziduumok térbeli becslését illetően, mert bizonyos feltételek mellett hasonló eredményeket szolgáltat, mint az egyszerű krigelés (G.B.M. Heuvelink és T. Hengl szóbeli közlése). Ugyanakkor egyszerű krigelést alkalmazva szélsőséges esetekben is konzisztens eredményeket kapunk. Például olyan szélsőséges esetben, mikor is a reziduumok térbeli változékonyságát egy röghatás modellel tudnánk leírni, akkor az egyszerű krigelés minden egyes becslési pontra – leszámítva az adatpontokat, ahol megtartaná az adatponti értékeket – nulla értéket becsülne, vagyis visszakapnánk a két komponensű valószínűségi függvényünk (V.7. egyenlet) tulajdonságait (V.8. és V.9. egyenlet). A hagyományos krigeléstől ugyanez nem várható el.

A fent részletezett regresszió krigeléssel és korábban definiált becslési griddel modelleztem a szervesanyag-tartalom térbeli eloszlását az egyes mintaterületekre, melyekből szervesanyag-tartalom térképet szerkesztettem.

V. 2. 5. A regresszió krigeléssel előállított térképek hibája

A regresszió krigeléssel előállított szervesanyag-tartalom térképek az egyes mintaterületekre vonatkozóan nem „hibátlanok”; vagyis a $z_{RK}^*(\mathbf{u})$ becslült és a $z(\mathbf{u})$ valós értékek különbözhetnek egymástól, melynek számszerű értéke az $e(\mathbf{u})$ hiba:

$$e(\mathbf{u}) = z_{RK}^*(\mathbf{u}) - z(\mathbf{u}), \quad (\text{V.16.})$$

Egy térkép hibáját nem ismerhetjük kimerítően (Heuvelink, 2014). A fenti definícióból következik, hogy ehhez ismernünk kellene a térképezendő talajtulajdonság értékét minden egyes becslési pontra vonatkozóan, mely – reális keretek között – elképzelhetetlen. Ezért a gyakorlatban a térképezési eljárástól független, ún. kontroll pontok segítségével értékeljük a térkép hibáját. Az l darab kontroll pontba kiszámított hibákat különböző mértékekkel jellemezhetjük; ilyen például az átlagos hiba (*mean error*, ME):

$$ME = \frac{1}{l} \sum_{i=1}^l e(\mathbf{u}_i), \quad (\text{V.17.})$$

az átlagos abszolút hiba (*mean absolute error*, MAE):

$$MAE = \frac{1}{l} \sum_{i=1}^l |e(\mathbf{u}_i)|, \quad (\text{V.18.})$$

illetve az átlagos négyzetgyök hiba $\sigma(\mathbf{u}_i)$ krigelési szórással standardizált értéke ($RMNSE$) (Hengl, 2009):

$$RMNSE = \sqrt{\frac{1}{l} \sum_{i=1}^l \left[\frac{e(\mathbf{u}_i)}{\sigma(\mathbf{u}_i)} \right]^2}, \quad (\text{V.19.})$$

A vizsgálatok kezdetén elkülönített kontroll pontok révén kiszámítottam a fenti mértékeket az előszállási, illetve a Zala megyei szervesanyag-tartalom térképekre. A szátkai mintaterületre nem állt módomban kontroll pontokat kijelölni, melynek háttérében az alacsony mintaszám állt (V.1.1. fejezet). Ezért a szátkai szervesanyag-tartalom térkép értékelésére az LOOCV (*leave-one-out cross validation*) keresztvalidációs módszert alkalmaztam. Az LOOCV során az n darab adatpont közül ideiglenesen eltávolítunk egy adatpontot és a fennmaradó $n-1$ adatpont segítségével becslést végzünk – a korábban már felparaméterezett geostatistikai modellt felhasználva – az eltávolított adatpontba. Ezt a folyamatot az összes adatpontra elvégezzük (Isaaks, Srivastava, 1989). Ennek eredményeként minden egyes adatponti értékhez kapunk egy becslült értéket, melyből kiszámítható a hiba (V.16. egyenlet). Habár Goovaerts

(1997) vitatja az LOOCV ilyen célú felhasználását, azonban a nemzetközi szakirodalom ezt a módszert ajánlja azon esetekben, amikor az alacsony mintaszám miatt nem különíthető el kontroll adatbázis (Brus et al., 2011; Hengl et al., 2015; Samuel-Rosa et al., 2015).

Az egyes mintaterületek szervesanyag-tartalom térképeire kiszámított *ME*, *MAE* és *RMNSE* értékeit a következő feltételek szerint értékeltem:

- *ME* és *MAE* várható értéke nulla (Isaaks, Srivastva, 1989),
- *RMNSE* várható értéke egy (Hengl, 2009),

Fontos kiemelni, hogy a fenti mértékekkel kapott eredmények – ahogy az a definícióikból is következik – a mintaterület egészére értelmezendők. Ezek a mértékek globális érvényűek az adott mintaterület felett, azaz nem lehetnek specifikusak az adott mintaterület egyes pontjaira vonatkozóan. A következőkben bevezetett sztochasztikus szimulációk (V.2.6. fejezet) ezzel szemben alkalmasak – többek között – a lokális bizonytalanság értékelésére is (Goovaerts, 2001).

V. 2. 6. A térbeli bizonytalanság modellezése szekvenciális sztochasztikus szimulációval

Regresszió krigelés során adott pontba a becslés a környező becslésektől függetlenül történik, mely becslés a „legjobb” abban az értelemben, hogy a lokális hiba variancia $\text{Var}\{Z_{RK}^*(\mathbf{u}) - Z(\mathbf{u})\}$ minimalizált. Ez a megállapítás általánosan igaz a krigelésre, melynek következményei (Goovaerts, 1997; Geiger, 2006a,b):

- A krigelés a célváltozó térbeli változékonyságának lokális részleteit kisimítja; vagyis az alacsony értékeket felülbecsli, míg a magas értékeket alulbecsli.
- Az előbbiből következik, hogy a krigelés a regionális tendenciákat hangsúlyozza; vagyis a krigelés alacsony frekvenciájú (*low-pass*) szűrő.
- A simító hatás függ az adatpontok lokális konfigurációjától; míg az adatpontokhoz közel a simítás minimális, addig az adatpontoktól távolabb egyre erősebbé válik.
- A krigeléssel előállított térképen a sűrűn mintázott területek változékonyságosabbak, mint a kevésbé sűrűn mintázott területeken; vagyis a térkép ún. artefaktumokat (*artifact*) mutat(hat).
- A krigeléssel előállított térkép varianciája általában jóval alacsonyabb, mint az adatponti értékek varianciája.

- A krigelt térképből számolt variogram jóval alacsonyabb röghatással rendelkezik, mint az adatponti értékekre számolt variogram.

A krigeléssel szemben a sztochasztikus szimulációk célja, hogy a regionalizált változónak olyan alternatív, de egyenlően valószínű, nagy felbontású térbeli modelljeit állítsák elő, melyek visszaadják a modell statisztikákat (például: hisztogram, variogram) (Deutsch, Journel, 1998). Ezeket a nagy felbontású térbeli modelleket realizációknak (vagy sztochasztikus képeknek) nevezzük (Geiger, 2006a,b). Tehát a sztochasztikus szimulációk bizonyos globális értelemben vett „legjobb becslést” adják az egyes pontokba (Goovaerts, 1997). Ez lehetőséget nyújt arra, hogy a realizációk révén modellezzük a térbeli bizonytalanságot (Goovaerts, 1997; Deutsch, Journel, 1998; Geiger, 2006b; Geiger, 2012). Ennek segítségével olyan döntési helyzetek is támogathatók (például: szennyezett területek lehatárolása), melyekre – a fent felsorolt következmények miatt – a krigelés alkalmatlan.

Vizsgálataimban a regresszió krigelésre felépített szekvenciális sztochasztikus szimulációt (Pebesma, Wesseling, 1998; Pebesma, 2004) alkalmaztam a térbeli bizonytalanság modellezésére. Ez a megoldás egyre szélesebb körben alkalmazott a nemzetközi szakirodalomban (például: Poggio et al., 2013; Poggio, Gimona, 2014). Ugyanakkor a hazai talajtanos szakirodalomban ennek alkalmazására nem találtam példát.

A szekvenciális sztochasztikus szimuláció paradigmája, Goovaerts (1997) munkája alapján a következőképpen foglалható össze. Legyen $\{Z(\mathbf{u}_j), j = 1, \dots, N\}$ olyan valószínűségi változók halmaza, mely N darab valószínűségi változót foglal magába, melyek \mathbf{u}_j pontokban helyezkednek el a vizsgált területen. Ezen N darab valószínűségi változó esetén célunk, hogy ezeknek olyan együttes realizációit hozzuk létre, melyek az n adatpontokra kondicionáltak. Az N -változós feltételes kumulatív eloszlás függvény (Goovaerts, 1997):

$$F(\mathbf{u}_1, \dots, \mathbf{u}_N; z_1, \dots, z_N | (n)) = P\{Z(\mathbf{u}_1) \leq z_1, \dots, Z(\mathbf{u}_N) \leq z_N | (n)\}, \quad (\text{V.20.})$$

Ugyanakkor ez utóbbi N -változós eloszlás függvény felírható N darab egyváltozós eloszlás függvény szorzataként is (Goovaerts, 1997):

$$F(\mathbf{u}_1, \dots, \mathbf{u}_N; z_1, \dots, z_N | (n)) = F(\mathbf{u}_N; z_N | (n + N - 1)) \cdot \dots \cdot F(\mathbf{u}_2; z_2 | (n + 1)) \cdot F(\mathbf{u}_1; z_1 | (n)) \quad (\text{V.21.})$$

ahol is az egyváltozós feltételes eloszlás függvények kondicionálva vannak az n adatpontokra és minden korábban szimulált értékre. Ezen alapulva egy realizáció N darab szukcesszív lépésből állítható elő (Goovaerts, 1997):

- Modellezzük a feltételes kumulatív eloszlás függvényt az \mathbf{u}_1 pontban, mely az n adatpontra van kondicionálva,
- Válasszunk ki véletlenszerűen egy értéket ebből az eloszlás függvényből és rendeljük ezt az értéket az \mathbf{u}_1 ponthoz (ez utóbbi lesz a szimulált értékünk), mely bekerül a kondicionáló adathalmazba is,
- Modellezzük a feltételes kumulatív eloszlás függvényt az \mathbf{u}_i pontban, mely kondicionálva van az n adatpontra és minden $(i - 1)$ korábban szimulált értékre,
- Az utóbbi két lépést addig ismételjük, míg minden N helyre nincs szimulált érték.

Az eredményül kapott „térkép” a $Z(\mathbf{u})$ valószínűségi függvény egy realizációja. További realizációk úgy állíthatók elő, hogy a fentebbi lépéseket ismételjük további véletlenszerűen kiválasztott bejárési útvonalakon keresztül (Goovaerts, 1997). Szekvenciális gaussi szimuláció esetén az egyes pontokban a feltételes kumulatív eloszlás függvény átlaga és varianciája az egyszerű krigelés átlagával és varianciájával határozandó meg (Deutsch, Journel, 1998). Ugyanakkor lehetőségünk van arra is, hogy a feltételes kumulatív eloszlás függvénynek az átlagát egyéb, például: a regresszió krigelés vagy a krigelés külső drifftel eljárások átlagával határozzuk meg (Goovaerts, 1997). Fontos kiemelni, hogy ebben az esetben is a feltételes eloszlás függvények varianciája az egyszerű krigelés varianciájával határozandó meg (Goovaerts, 1997). Ezért Pebesma és Wesseling, (1998), illetve Pebesma (2004) algoritmusát úgy módosítottam, hogy az egyszerű krigelés becslési varianciáját alkalmazza. Az ily módon előállított realizációk visszaadják a reziduumok variogramját és hisztogramját, továbbá visszaadják az adatponti értékeket.

Vizsgálataim során – Goovaerts (2001), Geiger és Mucsi (2005), Geiger (2006b), illetve Rossi és Deutsch (2014) munkái alapján – az egyes mintaterületekre 100-100 realizációt állítottam elő a fent részletezett algoritmussal. A realizációk segítségével vizuálisan és kvantitatívan is értékelhető a térbeli bizonytalanság (Goovaerts, 1997); vagyis a realizációk közötti különbség az együttes térbeli bizonytalanságot jeleníti meg, mely egyben lehetőséget is ad annak mérésére. E feladat megoldására a „hagyományos” krigelési eljárások – konstrukciójukból kifolyólag – alkalmatlanok.

V. 3. Modell alapú talajmintavétel optimalizációs módszertan

V. 3. 1. Spatial simulated annealing

Az SSA algoritmust van Groenigen és Stein (1998) dolgozta ki, mely definíciójuk szerint olyan iteratív, kombinatorikus, modell alapú mintavétel optimalizációs algoritmus, mely a lehetséges, alternatív mintavételi elrendezések (ún. kombinációk) szekvenciáját oly módon állítja elő, hogy minden új mintavételi kombináció az azt megelőző kombináció véletlenszerű megváltoztatásából jön létre. Az optimalizáció során minden új mintavételi kombináció esetén kiszámítjuk a minőségi mérték értékét, melyet összehasonlítunk az azt megelőző kombinációra kiszámított értékével. A Metropolis kritérium (Metropolis et al., 1953) határozza meg, hogy az új mintavételi kombinációt elfogadjuk a következő kombináció alapjaként vagy visszautasítjuk azt, és az azt megelőző kombinációt tartjuk meg a következő kombináció alapjának:

$$\begin{cases} P(D_i \rightarrow D_{i+1}) = 1, & \text{ha } \phi(D_{i+1}) \leq \phi(D_i) \\ P(D_i \rightarrow D_{i+1}) = \exp\left(\frac{\phi(D_i) - \phi(D_{i+1})}{T}\right), & \text{ha } \phi(D_{i+1}) > \phi(D_i) \end{cases}, \quad (\text{V.22.})$$

ahol D_{i+1} , illetve D_i jelöli az új és az azt megelőző mintavételi kombinációt, T az ún. pozitív kontroll paraméter (vagy „rendszer hőmérséklet”), melynek értéke csökken az optimalizáció előrehaladtával, és $\phi(\cdot)$ a célfüggvény. Ha D_{i+1} mintavételi kombinációt elfogadjuk, akkor ez nyújtja a következő D_{i+2} mintavételi kombináció alapját. Mindazonáltal, ha visszautasítjuk D_{i+1} mintavételi kombinációt, akkor a következő D_{i+2} mintavételi kombináció alapját D_i kombináció fogja jelenteni. Adott mintavételi kombináció véletlenszerű megváltoztatásán azt értjük, hogy az adott kombinációnak egy véletlenszerűen kiválasztott – „leendő” mintavételi – pontját egy véletlenszerűen megválasztott ún. transzformációs vektor mentén „elmozdítjuk” (van Groenigen et al., 1999), mely vektor abszolút értéke csökken az optimalizáció előrehaladtával. Az előbb vázolt folyamat mindaddig folytatódik ciklikusan, míg a legutolsó mintavételi kombinációval el nem érjük a $\phi(\cdot)$ célfüggvény globális minimumát. Az ehhez tartozó mintavételi elrendezést (vagy konfigurációt) tekintjük az optimális mintavételi konfigurációnak (van Groenigen et al., 1999).

Ahogy az a Metropolis kritériumból (V.22. egyenlet) következik, az optimalizációs folyamat elején megengedett, hogy magasabb $\phi(\cdot)$ értékű mintavételi kombinációkat is elfogadjunk. Ennek hátterében az a megfontolás húzódik, miszerint az optimalizáció során elkerülendő, hogy valamely lokális minimumba „csapdázódjunk”; vagyis az optimalizációs

folyamat elején elfogadott „rosszabb” kombinációk elő fogják segíteni a $\phi(\cdot)$ globális minimumának „megtalálását”. A folyamat végén a „rosszabb” kombinációk elfogadási valószínűsége gyakorlatilag elhanyagolható.

Elméletben az optimalizáció „végtelen hosszú”, hiszen $\phi(\cdot)$ globális minimumát megtalálva újabb és újabb kombinációkat hoznánk létre, melyeket mindig visszautasítanánk. Ezért szükségünk van egy olyan beállításra, mellyel megállítható a folyamat. Ennek két megoldása lehet (Heuvelink et al., 2007): (1) előre definiált iteráció szám elérése után az optimalizáció „elhal”, és az utolsó kombinációt tekintjük az optimális mintavételi konfigurációnak, illetve (2) az algoritmusba épített ún. „stop kritérium” segítségével adott számú új kombináció sorozatos visszautasítása után „megállítjuk” az optimalizációt, és a legutolsó – még vissza nem utasított – kombinációt tekintjük az optimális konfigurációnak. Dolgozatomban e kettő kombinációját választottam; vagyis 50.000 iterációs számot adtam meg, illetve az általam definiált stop kritérium 1.000 volt, melyek jól bevált értékek (D.J. Brus szóbeli közlése). Az 50.000-es iterációs szám elég nagy ahhoz, hogy kellően „sok ideje” legyen az algoritmusnak megtalálni az optimális mintavételi konfigurációt. Továbbá az 1.000-es „stop kritérium” szintén kellően magas ahhoz, hogy elkerüljük a lokális minimumokban való „csapdázódást”, illetve, hogy megállítsuk a folyamatot az optimális mintavételi konfigurációnál (D.J. Brus szóbeli közlése). A futtatások során a kezdeti „rosszabb” kombinációnak az elfogadási valószínűsége 0,8 volt, mely a nemzetközi szakirodalom szerint jól bevált érték (Brus, Heuvelink, 2007). A rendszer „hűlésének”, valamint a transzformációs vektor abszolút értékének a csökkenéséhez exponenciális lefutást választottam, mely szintén megfelel a nemzetközi gyakorlatnak (van Groenigen et al., 1999; van Groenigen, 2000; Brus, Heuvelink, 2007).

Az SSA algoritmus felépítéséből következik, hogy számos mintavételi megszorítást is számításba tudunk venni, mely megszorítások a „hagyományos” módszerek használata esetén problémásnak tekinthetők (van Groenigen, 1999); ilyenek például: (1) a rendelkezésünkre álló segédinformációk figyelembe vétele, (2) korábban gyűjtött mintavételi pontok figyelembe vétele, (3) a vételezhető mintavételi pontok száma, (4) a mintavételezés számára hozzá nem férhető vagy irreleváns területek, (5) a mintavétel költségkerete, illetve (6) az előre definiált megengedhető bizonytalanság.

V. 3. 2. A regresszió krigelés becslési varianciája, mint az optimalizáció minőségi mértéke

Az SSA algoritmus gyakorlati alkalmazásának egyik legfontosabb momentuma a mintavételezés – és vele együtt a talajtérképezés – céljainak megfelelő minőségi mérték megválasztása. Az ebből képzett $\phi(\cdot)$ célfüggvény globális minimumát adó konfiguráció jelenti azt a mintavételi elrendezést, mely az adott valószínűségi modell és az előírányzott mintavételi megszorítások mellett optimális lesz. Az alapvetően regresszió krigelésre felépített térképezési módszertannak (V.2. fejezet) két fontos követelménye van: (1) a mintavételi pontoknak a lehető leghívebben kell reprezentálniuk a térképezés során felhasználandó segédinformációkat, illetve (2) a regresszió reziduumaikra felépített egyszerű krigelés révén a mintavételi pontoknak kvázi egységes földrajzi lefedettséget kell biztosítaniuk. E kettős követelményt tovább nehezíti azon – a gyakorlati tervezésekben teljesen megszokott – elvárás, hogy a minőségi mértéknek már azelőtt becsülhetőnek kell lennie még mielőtt a konkrét mintavételezés megtörténne. A fentieket összegezve a térképezési módszertanom olyan minőségi mértéket kíván meg, mely (1) szimultán képes figyelembe venni a mintavételi pontok földrajzi és attribútum térbeli „lefedettségét”, illetve (2) független az adatponti értékektől. Utóbbi a módszertan idő- és költséghatékonyágát fogja erősíteni.

E kívánalmaknak a regresszió krigelés $\sigma_{RK}^2(\mathbf{u})$ becslési varianciája felel meg, mely definíció szerint (Hengl et al., 2004, 2007):

$$\sigma_{RK}^2(\mathbf{u}) = C(0) - \mathbf{c}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u}) + [\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})]^T \cdot (\mathbf{Q}^T \cdot \mathbf{C}_R^{-1} \cdot \mathbf{Q})^{-1} \cdot [\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})] \quad (\text{V.23.})$$

ahol $C(0)$ a reziduális komponens varianciája és \mathbf{c} az adatpontok és az ismeretlen pont közötti kovariancia vektora, $\mathbf{q}(\mathbf{u})$ a független változók vektora, $\boldsymbol{\lambda}_{SK}(\mathbf{u})$ az egyszerű krigelési súlyok vektora, \mathbf{Q} a független változók adatpontokra vonatkozó $(n \times p + 1)$ mátrixa és \mathbf{C}_R a reziduumok kovariancia mátrixa. A $\sigma_{RK}^2(\mathbf{u})$ becslési variancia első komponense $\{C(0) - \mathbf{c}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})\}$ az egyszerű krigelés becslési varianciájával egyenlő, míg a második komponense $\{[\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})]^T \cdot (\mathbf{Q}^T \cdot \mathbf{C}_R^{-1} \cdot \mathbf{Q})^{-1} \cdot [\mathbf{q}(\mathbf{u}) - \mathbf{Q}^T \cdot \boldsymbol{\lambda}_{SK}(\mathbf{u})]\}$ a trend hiba varianciájával egyenlő. A $\sigma_{RK}^2(\mathbf{u})$ becslési variancia a következőktől függ:

- Az egyszerű krigelés becslési varianciája részéről (Geiger, 2006a):
 - Az adatpontoknak az ismeretlen területhez viszonyított geometriájától,

- Az adatpontok egymáshoz viszonyított helyzetétől,
- A félvariogram típusától,
- A krigelési súlyoktól,
- A trend hiba varianciája részéről (Hengl, 2009):
 - Az MSE (*mean square residual error*) értékétől,
 - Az adatpontok attribútum térbeli eloszlásától (*spread*),
 - A becslési pontnak az attribútum tér centrumától vett távolságától.

Heuvelink és munkatársai (2007) szerint a $\sigma_{RK}^2(\mathbf{u})$ becslési variancia korábban említett „kettőssége” azt eredményezi, hogy minőségi mértékként alkalmazva szimultán fogja minősíteni a mintavételi pontok földrajzi és attribútum térbeli lefedettségét. Továbbá, ahogy az definíciójából is következik (V.23. egyenlet) független az adatponti értékektől. Számításának feltétele a többszörös lineáris regresszió struktúrájának, illetve a reziduumok variogram modelljének az apriori ismerete. Ezek a feltételek a hazai archív talajtani adatbázisok és szakértői ismeretek segítségével kielégíthetők.

A minőségi mérték megnevezése után fontos feladat volt, hogy a választott mértékből olyan célfüggvényt képezzek, mely a mintavételezés – és vele együtt a térképezés – céljainak megfelel. Jól bevált gyakorlat a krigelési variancia összterületi átlagának az alkalmazása (például: Bárdossy, Bogárdi, 1983; Szidarovszky, 1983a,b; Bogárdi et al., 1985; van Groenigen et al., 1999); ezért a $\sigma_{RK}^2(\mathbf{u})$ becslési variancia összterületi átlagát definiáltam célfüggvényként:

$$\phi(D_i) = \frac{1}{N_{grid}} \sum_{j=1}^{N_{grid}} \sigma_{RK}^2(\mathbf{u}_j | D_i), \quad (V.24.)$$

ahol N_{grid} a definiált becslési grid pontjainak a száma. E célfüggvény értelmében az SSA algoritmus a regresszió krigelés $\sigma_{RK}^2(\mathbf{u})$ becslési varianciájának összterületi átlagát fogja minimalizálni és e függvény globális minimumát adó mintavételi konfigurációt fogom optimálisnak tekinteni.

A félreértést elkerülve, fontosnak tartom kiemelni a krigelési variancia alapvető sajátosságait. A nemzetközi talajtérképezési szakirodalomban elfogadottnak tekintik a regresszió krigelés – és általában véve az összes krigelési eljárás – becslési varianciájának az alkalmazását a lokális becslési pontosság számszerűsítésére (például: Lark, Lapworth, 2012; Heuvelink, 2014; Kempen et al., 2014). Habár a krigelési eljárásokat a legjobb lineáris egyértelmű

becslőként (*best linear unbiased estimator, BLUE*) említi a szakirodalom (például: Isaaks, Srivastava, 1989; Webster, Oliver, 2007; Hengl, 2009), mely „legjobb” abban az értelemben, hogy minimalizálja a becsült érték hibájának varianciáját. Azonban a hiba variancia független az adatponti értékektől. E tény mellett a becslési variancia csakis a következő, szigorú feltételek teljesülése esetén alkalmazható a lokális bizonytalanság jellemzésére (Isaaks, Srivastava, 1989): (1) az $e(\mathbf{u}) = z^*(\mathbf{u}) - z(\mathbf{u})$ hiba egy normál eloszlású valószínűségi változó realizációjaként modellezhető, illetve (2) a $\sigma_E^2(\mathbf{u})$ hiba variancia független az adatponti értékektől. Ezek a feltételek a gyakorlatban többnyire nem adóttak (Journel, 1986; Isaaks, Srivastava, 1989; Journel, Rossi, 1989; Goovaerts, 1997, 1999a); ezért *általában véve a geostatistika nem tekinti a krigelési varianciát a lokális becslési pontosság mértékének* (Journel, 1986; Journel, Rossi, 1989; Geiger, 2006a). *Ugyanakkor definíciójából kifolyólag alkalmas az egyes alternatív geometriai adat elrendezések rangsorolására* (Journel, 1986; Journel, Rossi, 1989; Geiger, 2006a), mely – esetemben – az optimalizáció alapját jelenti.

Fontos megjegyezni, hogy kategorikus segédinformációk esetén az SSA futtatása problémába ütközik. A $\sigma_{RK}^2(\mathbf{u})$ becslési variancia – mint választott minőségi mérték – számítása feltételezi, hogy a modellbe vont kategorikus segédinformációk minden kategóriája legalább egy mintavételi ponttal rendelkezik. Ellenkező esetben a $\sigma_{RK}^2(\mathbf{u})$ becslési variancia nem határozható meg. E feltétel teljesülése ugyanakkor nem garantálható az optimalizáció teljes folyamata alatt lévén, hogy az SSA által létrehozott alternatív mintavételi elrendezések szekvenciája véletlenszám generátor segítségével kerül előállításra. Ezért a kategorikus segédinformációkat nem indikátor változóként, hanem ún. faktor változóként definiáltam az optimalizáció során.

V. 3. 3. Mintavétel optimalizációs scénáriók beállítása a módszertan tesztelésére

A mintavétel optimalizációs módszertan tesztelésére az előszállási mintaterületet választottam. A mintavételi scénáriókat úgy állítottam be, hogy azok a lehető leghívebben reprezentálják az SSA algoritmus alkalmazhatóságát, illetve azokat a legfontosabb talajmintavételezési kérdéseket, melyekkel a szakemberek találkozhatnak. Ennek érdekében négy scénáriókat állítottam be:

- 1. *scenárió*: 100 darab új mintavételi pont optimális elhelyezése „érintetlen területen”,

- 2. *szcenárió*: Új mintavételezés optimalizációja adott (előzetes) minőségi mérték elérése érdekében „érintetlen területen”,
- 3. *szcenárió*: 65 darab új mintavételi pont optimális elhelyezése a korábbi felmérésekből származó mintavételi pontok figyelembe vétele mellett,
- 4. *szcenárió*: Kiegészítő mintavételezés optimalizációja adott (előzetes) minőségi mérték elérése érdekében oly módon, hogy figyelembe vesszük a korábbi felmérésekből származó mintavételi pontokat,

Az 1. *szcenárió* esetében a megadott mintavételi pontok száma (100 db) megegyezik a geostatistikai modellalkotáshoz használt adatpontok számával (V.1.2. fejezet). A 3. és 4. *szcenárió* esetén a meglévő mintavételezésből származó mintavételi pontok száma 35 volt. Ezek a mintavételi pontok két különböző elrendezést követtek: (1) csoportosult, illetve (2) szabályos mintavételezés. A 2. és 4. *szcenárió* esetében a legfőbb cél az volt, hogy meghatározzam azt az ún. kalibrációs görbét, mely a mintaméret függvényében ábrázolja a célfüggvény értékét. A szerkesztett kalibrációs görbe segítségével meghatározható, hogy adott (előzetes) minőségi mérték eléréséhez mekkora mintaméret szükséges. Az egyes *szcenáriók* során további mintavételi megszorításként vettem figyelembe a mintaterület azon részeit, melyek a mintavételezés, illetve térképezés szempontjából „hozzáférhetetlenek” vagy irrelevánsnak tekinthetők. Ilyenek voltak: (1) nyílt vizek, (2) mocsarak, illetve (3) mesterséges felszínek (IV.2. ábra). Az optimalizációk során megkötésem volt, hogy ezekre a területekre nem eshet mintavételi pont.

V. 3. 4. Többváltozós mintavétel optimalizációs módszertan és tesztelése

Az SSA algoritmus egyik legnagyobb hátránya, hogy csakis egy adott talajtulajdonságra képes optimalizálni a mintavételezést (Vašát et al., 2010; Szatmári et al., 2015b); azonban a gyakorlatban számos talajtulajdonság egyidejű térképezését célozzuk meg. Ezért kardinális módszertani kérdés, hogy miként is lehetne az alapvetően egyváltozós SSA algoritmust kiterjeszteni olyan többváltozós algoritmussá, mely képes lenne a talajmintavételi konfigurációt kettő vagy több talajtulajdonságra szimultán optimalizálni. A térképezési módszertant (V.2. fejezet) tekintve a legfőbb problémát az jelenti, hogy az egynél több talajtulajdonságra történő mintavétel optimalizáció során: (1) az egyes talajtulajdonságok más és más segédinformáció kombinációt alkalmaznak a térbeli becslés során, továbbá (2) a reziduumok variogramjainak a

modelljei szintén más és más karakterisztikával bírnak. A probléma megoldását e korlátozó tényezők „közös nevezőre” hozása jelenti.

Adott talajtulajdonság térbeli becsléséhez felhasználandó segédinformációkra – vagyis az adott talajjellemző és a rendelkezésünkre álló segédinformációk közötti regresszió struktúrájára – úgy is tekinthetünk, mint a talaj-környezet kapcsolati rendszer által meghatározott sokváltozós attribútum tér azon részalmazára, mely effektíve befolyásolja az adott talajtulajdonság térbeli eloszlását. Például: egy domboldalon található talajtakaró egészének térbeli változékonyságát számos környezeti paraméter határozza meg, azonban e talajtakaró szervesanyag-tartalmának térbeli változékonyságát – jóval – kevesebb környezeti paraméter határozza meg. Ez abból következik, hogy egy adott alrendszer (talajjellemző) nem lehet komplexebb, mint az őt magába foglaló rendszer (talajtakaró). A regresszió krigelés első lépése, hogy a rendelkezésünkre álló segédinformációkból kiválasztjuk azon – az előzőnél kisebb számú – segédinformációkat, melyek effektíve meghatározzák a kérdéses talajjellemző térbeli változékonyságát (V.2.4. fejezet). Az egyváltozós SSA mintavétel optimalizáció során arra törekszünk, hogy ezt az ún. részalmazt optimálisan lefedjük (V.3.2. fejezet). Kettő vagy több talajtulajdonság esetén a fent említett sokváltozós attribútum térnek kettő vagy több részalmazára lesz, melyek effektíve befolyásolják e talajtulajdonságok térbeli eloszlását. Ezeknek a részalmazoknak lesz közös részük és lesz olyan részük is, mely csakis egy talajtulajdonság térbeli változékonyságához kapcsolódik. Ha optimalizálni akarjuk a mintavételt ezekre a talajjellemzőkre, akkor a mintavételi pontokat úgy kell elhelyeznünk, hogy azok optimálisan fedjék le e részalmazok unióját. A gyakorlatban ez úgy kivitelezhető, hogy – a legkisebb közös többszörös elvéhez hasonlóan – a regressziós struktúrákból kiválasztjuk azokat a segédinformációkat, melyek uniója teljesen meghatározza a vizsgált talajjellemzők térbeli változékonyságát. Az ily módon felparaméterezett SSA algoritmus arra fog „törekedni”, hogy optimálisan „fedje le” ezeket a segédinformációkat.

A következő módszertani kérdés, hogy hozhatók „közös nevezőre” a reziduumok variogram modelljei. Füst és Geiger (2010) munkáját követve a válasz a domináns paraméter megtalálásában rejlik. Kutatásaik szerint az optimális mintavételi távolságot mindenesetben a „mértékadóra” kell beállítanunk, melynek térbeli változékonysága a legnagyobb, azaz a legkisebb hatástávolsággal rendelkezik. E megfontolás szerint a földrajzi tér optimális lefedettségét a legnagyobb térbeli változékonysággal rendelkező talajtulajdonságra kell méretezni. Ez azt jelenti, hogy az SSA algoritmust azon talajjellemző hatástávolságával kell felparaméterezni, amely a legrövidebb. Fennmaradó kérdés, hogy az eltérő talajtulajdonságok

akár nagyságrendileg is eltérő variogram paramétereit (például: tető és röghatás) miként is kell(ene) számításba vennünk. Van Groenigen (2000) szerint nulla röghatás mellett a tetőnek nincs ráhatása az optimalizált mintavételi konfigurációra lévén, hogy kétszeres nagyságú tető kétszeres nagyságú becslési varianciát fog eredményezni, de ettől még az optimális mintavételi elrendezés változatlan marad. Ezért kutatásai szerint a relatív röghatás – vagyis a röghatás és a tető hányadosa –, amit szem előtt kell tartanunk. Vizsgálatai szerint a 0,75-öt meghaladó relatív röghatás van érdemi hatással az optimális konfigurációra. Továbbá a variogram modell típusának (például: szférikus, exponenciális stb.) sincs érdemi ráhatása az optimalizált mintavételi konfigurációra (van Groenigen, 2000). E megfontolások szerint az SSA felparaméterezése során a domináns talajtulajdonság variogram modellje közvetben alkalmazható.

A fenti beállítások szerint olyan „pszeudo-talajjellemzőt” definiálunk az optimalizáció során, mely a „legkedvezőtlenebb” helyzetet tükrözi a mintavételezés szempontjából; hiszen (1) a legtöbb segédinformációval írható le és (2) a legnagyobb térbeli változékonysággal rendelkezik. E „pszeudo-talajjellemzőre” kiszámítható a korábban definiált $\phi(\cdot)$ célfüggvény (V.24. egyenlet), mely segítségével optimalizálható a mintavételezés a vizsgált talajtulajdonságokra. Ugyanakkor kihangsúlyozandó, hogy a $\phi(\cdot)$ célfüggvény értékei képzetes minőségi értékek, melyek csakis kizárólag az egyes mintavételi kombinációk rangsorolására alkalmazhatók.

A többváltozós mintavétel optimalizációs módszertan tesztelésére szintén az előszállási mintaterületet választottam, ahol is a szervesanyag-tartalom adatokon túl rendelkezésemre álltak egyéb talajjellemzők mért értékei is (V.1.2. fejezet). Ebből a célból a termőréteg vastagságot vettem második talajjellemzőnek. A többváltozós módszertan tesztelésére két scenáriót állítottam be:

- 5. *scenário*: 100 darab új mintavételi pont optimális elhelyezése „érintetlen területen”,
- 6. *scenário*: 65 új mintavételi pont optimális elhelyezése a korábbi felmérésekből származó mintavételi pontok figyelembe vétele mellett,

Az 5. *scenário* esetében a mintavételi pontok száma (100 db) megegyezik a geostatistikai modellalkotáshoz használt adatpontok számával (V.1.2. fejezet). A 6. *scenário* esetében a meglévő mintavételezésből származó mintavételi pontok száma 35 volt. Ezek a mintavételi pontok – a 3. és 4. *scenário*hoz hasonlóan – két különböző elrendezést követtek: (1)

csoportosult, illetve (2) szabályos mintavételezés. E scenáriók során is figyelembe vettem a mintavételezés számára „hozzá nem férhető” területeket (nyílt vizek, mocsarak, közutak, tanyák stb.). Az optimalizációk során megkötésem volt, hogy ezekre a területekre mintavételi pontok nem eshetnek.

V. 3. 5. Az optimalizált mintavételi elrendezések értékelése

Az egyes scenáriókra optimalizált mintavételi elrendezések értékelésére az $F(r)$ üres-tér függvény módosított alakját, a krigelési szomszédok térképeit, illetve a Kolmogorov-Smirnov próbát alkalmaztam. Az $F(r)$ függvény segítségével megvilágíthatók: (1) az optimalizált mintavételi helyek közötti interakció típusa, (2) a legközelebbi krigelési szomszéd távolsága, minden egyes becslési pontra vonatkozóan, illetve (3) a relatíve kevés krigelési szomszéddal rendelkező területek elhelyezkedése. Dolgozatomban krigelési szomszédnak tekintem azon mintavételi pontokat, melyek távolsága az adott becslési ponttól kisebb, mint a variogram hatástávolsága (figyelembe véve az anizotrópiát). A krigelési szomszédok további vizsgálatára a becslési grid pontjaihoz tartozó összes krigelési szomszédot leválogattam. Ezekből olyan térképet szerkesztettem, mely a becslési gridre vonatkozóan ábrázolja a krigelési szomszédok számát. Minimum követelmény, hogy minden egyes becslési pontnak a „szomszédságában” legalább egy olyan mintavételi pont legyen, melynek távolsága az adott becslési ponttól kisebb, mint a variogram hatástávolsága (Füst, Geiger, 2010). A kiszámított $F(r)$ függvény, illetve a krigelési szomszédok térképével objektíven értékelhető a földrajzi tér lefedettsége. Az attribútum tér lefedettségét Kolmogorov-Smirnov próbával vizsgáltam. A próba során azt a hipotézist teszteltem, hogy az optimalizált mintavételi pontokra eső folytonos segédinformációk értékeinek eloszlása, illetve a segédinformációk rasztereinek eloszlása ugyanazon folytonos eloszlásból származik.

V. 4. A módszertanok szinergiája

A regresszió krigelés, illetve az SSA alkalmazásaira számos példát találhatunk (Hengl et al., 2004, 2007; Brus, Heuvelink, 2007; Heuvelink et al., 2007; Webster, Lark, 2013), azonban a szakirodalom sehol sem tesz említést e módszerek szinergiájáról, mely a gyakorlatban kiemelkedő fontosságú lehet. Ez utóbbi abból következik, hogy a III.3. fejezetben definiált mintavételi stratégiának két fontos eleme van (Särndal, 1978; Hansen et al., 1983; de Gruijter

et al., 2006): (1) a mintavételi helyek kiválasztásának módszere, illetve (2) a keresett paraméter becslésére szolgáló eljárás. E két komponensnek összhangban kell lennie, hogy a lehető legjobban és legcélszerűbben használjuk fel a talajfelvételezésekre és –térképezésekre szánt pénzügyi erőforrásokat illetve, hogy a lehető leghívebb statisztikai következtetéseket vonhassuk le a vételezett talajminták alapján.

Magyarországon a közeljövőben új országos léptékű talajfelvételezésre nemigen látszik lehetőség (Szabó et al., 2005). Ezért napjainkban a talajtani információkkal szembeni igények kielégítése a rendelkezésünkre álló archív talajtani adatbázisok mind alaposabb kihasználásával történik (Pásztor et al., 2013b, 2014b, 2015). Ugyanakkor ez nem zárja ki annak lehetőségét, hogy a jövőben ne lenne szükség, ha nem is országos, de kisebb területeken részletes talajfelvételezésre, illetve –térképezésre (például: Szalóki, Somodi, 2014; Kovács et al., 2016; Szabó et al., 2016), ahol is a fent említett szinergia kedvező hatással lesz. Továbbá a jelenleg rendelkezésünkre álló talajtani adatbázisok kiegészítésére irányuló mintavételezések esetén is sikeresen alkalmazható e megközelítés.

Dolgozatomban a térképezési (V.2. fejezet), illetve a mintavétel optimalizációs módszertan (V.3. fejezet) alapját a regresszió krigelés jelenti. Azonban ez nem jelenti azt, hogy a kialakított digitális talajtérképezési és mintavétel optimalizációs módszertan ne lenne alkalmazható akkor, ha más modellt adaptálnánk a $Z(u)$ valószínűségi függvényhez. Tegyük fel, hogy a szervesanyag-tartalom térképezését illetően nem áll rendelkezésünkre segédinformáció és/vagy nem kell „gyanakodnunk” trend jelenlétére. Ebben az esetben jó megoldás lenne a humusztartalom térbeli modellezése hagyományos krigeléssel. E módszer mellett a kialakított módszertan regresszió krigeléshez kapcsolódó komponensei kicserélendők a hagyományos krigelés megfelelőire, vagyis (1) a térbeli becslő eljárás a hagyományos krigelés lenne, (2) a minőségi mértéket a hagyományos krigelés becslési varianciája jelentené, továbbá (3) a célfüggvényt e becslési variancia összterületi átlaga képezné. Ugyanez igaz a másik szélsőséges esetre is, mikor a modell (V.7. egyenlet) reziduális komponense elhanyagolható vagy röghatás modellel „írható le”. Ebben az esetben a többszörös lineáris regresszió nyújtana jó megoldást a térképezésre. Ekkor a kialakított módszertanban: (1) a térbeli becslő eljárás a többszörös lineáris regresszió lenne, (2) a minőségi mértéket a trend hiba varianciája jelentené, továbbá (3) a célfüggvényt e hiba variancia összterületi átlaga képezné. Ennek bemutatására az előszállási mintaterületet vettem, ahol is e két szélsőséges esetben mutatom be a kialakított digitális talajtérképezési és mintavétel optimalizációs módszertan flexibilitását. Erre két scenáriót állítottam be:

- 7. *szcenárió*: 50 darab új mintavételi pont optimális elhelyezése „érintetlen területen” (hagyományos krigelés),
- 8. *szcenárió*: 50 darab új mintavételi pont optimális elhelyezése „érintetlen területen” (többszörös lineáris regresszió).

E *szcenáriók* során is figyelembe vettem a mintavételezés számára „hozzá nem férhető” területeket (nyílt vizek, mocsarak, közutak, tanyák stb.). Az optimalizációk során ez esetben is megkötésem volt, hogy ezekre a területekre mintavételi pontok nem eshetnek.

VI. EREDMÉNYEK BEMUTATÁSA ÉS ÉRTÉKELÉSE

A kutatásaimmal kapcsolatos eredmények ismertetése az előző fejezetben (V. fejezet) részletezett digitális talajtérképezési és mintavétel optimalizációs módszertan egyes lépései alapján történik. A disszertáció terjedelmi korlátai miatt nem törekedhettem, hogy mindhárom mintaterületen részleteiben bemutassam és értékeljem a módszertant. A fejezetben bemutatott eredményeket a VI.1. táblázat foglalja össze.

VI.1. táblázat
A bemutatásra kerülő eredmények táblázatos összefoglalása

Mintaterületek	EDA	RK + értékelés	SSRK + értékelés	SSA + értékelés	MV-SSA + értékelés	Módszertanok szinergiája
Szátkai kiszívgyűjtő	★	★				
Előszállási szántóföldek	★	★	★	★	★	★
Zala megye	★	★				

Rövidítések: EDA: feltáró alapadat elemzés, RK: térképezés regresszió krigeléssel, SSRK: térbeli bizonytalanság modellezése szekvenciális szimulációval, SSA: mintavétel optimalizáció, MV-SSA: többváltozós mintavétel optimalizáció.

A fejezetben – és a hozzá kapcsolódó mellékletekben – bemutatott eredményeket (ábrák, diagramok, térképek stb.) szabad-hozzáférésű, nyílt forráskódú szoftverek segítségével készítettem el. Kizárólag a szervesanyag-tartalom térképek kartografáláshoz használtam egyéb szoftvert (*ArcGIS*). Habár hazánkban a *tizedesvessző* használata az elfogadott, az általam használt nyílt forráskódú szoftverekben az erre foganatosított jelölés a *tizedespont*, mely a nemzetközi igényekhez és követelményekhez igazodik. Ezért az eredmények elkészítése során nem törekedhettem ezek korrigálására. A táblázatokra ez a megszorítás nem vonatkozik.

VI. 1. A definiált becslési gridek

Az *inspection density* (Hengl, 2006) módszerrel kiszámoltam a becselő gridek geometriai felbontását. A számított *g* értékeket kerekítettem, így a ténylegesen definiált geometriai felbontásokat a VI.2. táblázat tartalmazza. Az *x* és *y* tengely menti grid pontok számát úgy határoztam meg, hogy az előállított becslési gridek maradéktalanul fedjék le a mintaterületeket.

Ezek alapján előállítottam az egyes mintaterületekre vonatkozó becslési gridek topológiáját. E mellett beállítottam az egyes topológiáknak a hazánkban használatos egységes országos vetületi (EOV) rendszert. Az egyes mintaterületekre előállított becslő gridek főbb jellemzőit a VI.2. táblázat foglalja össze. A definiált grid topológiákat és becslési grideket használtam fel a segédinformációk előfeldolgozása, a feltáró alapadat elemzés, illetve a térbeli becslés és sztochasztikus szimuláció során.

VI.2. táblázat

Az egyes mintaterületekre definiált becslési gridek főbb jellemzői

(Megjegyzés: a táblázatban piros kerettel jelöltem a grid topológia definiálásához szükséges paramétereket)

Becslési gridek főbb jellemzői	Száлка	Előszállás	Zala megye
<i>koordináta rendszer</i>	EOV	EOV	EOV
x_{min}	618.991,7	627.989,4	445.033,0
y_{min}	105.137,8	160.390,3	107.427,0
g	10 × 10 méter	20 × 20 méter	75 × 75 méter
N_x	105 db	323 db	1067 db
N_y	184 db	251 db	1117 db
N_{grid}^*	13.221 db	44.359 db	671.679 db
$N_{\text{összes}}$	19.320 db	81.073 db	1.191.839 db

* A mintaterületre eső grid pontok száma $\{N_{grid} \in A\}$.

VI. 2. A szervesanyag-tartalom adatok elsődleges értékelése

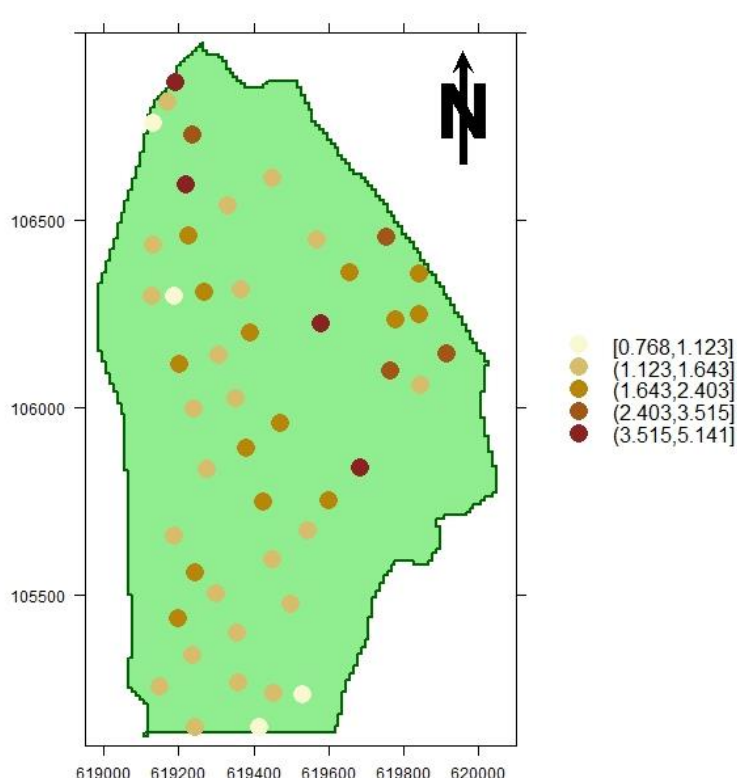
VI. 2. 1. Szálkai kisvízgyűjtő

A mintaterületről származó szervesanyag-tartalom adatokra kiszámolt alapvető statisztikai mennyiségeket a VI.3. táblázat foglalja össze. Az adatbázis három olyan megfigyelést tartalmaz, ahol a szervesanyag-tartalom magasabb, mint 6%. A mintaterületen folytatott intenzív mezőgazdálkodás, a területet sújtó vízerózió, illetve a mintaterületre jellemző talajtípusok (IV.1. fejezet) azonban cáfolják, hogy ilyen kiugróan magas lehessen a feltalaj (0–10 cm) humusztartalma. Ezért ezeket az adatokat kizártam a további vizsgálatokból. Továbbá két, erodált szántóföldre eső megfigyelés szervesanyag-tartalma is magasabb volt, mint 5%. Ez szintén tévesnek vélhető az erózió okozta károk, illetve az intenzív mezőgazdasági gyakorlat révén felgyorsított szerves anyag mineralizáció miatt. Ezért ezeket az adatokat is kizártam a további feldolgozásokból.

VI.3. táblázat
A szálkai szervesanyag-tartalom adatok összefoglaló statisztikája

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Nyers adatsor (n = 54)	2,25	1,64	0,76	7,55	1,53	1,89
Szűrt adatsor (n = 49)	1,85	1,58	0,76	5,14	0,87	2,06

A szerves anyag adatsorra elkészített box-plot-ot, hisztogramot, illetve Q-Q diagramot a 3. melléklet tartalmazza. A mellékletben található hisztogram és Q-Q diagram a szűrt adatsorra készült el. A szervesanyag-tartalom adatok pozitív ferde eloszlást mutatnak az elkészített hisztogram és Q-Q diagram, illetve a kiszámított ferdeségi paraméter alapján. Ez meglehetősen gyakori a talajtanban (Webster, Oliver, 2007). A Shapiro-Wilk próba során a megfigyelt szignifikancia szint $1,8 \cdot 10^{-7}$ -nek adódott, mely szerint 5%-os szignifikancia szint mellett visszautasítottam a próba nullhipotézisét vagyis, hogy az adatok normál eloszlású populációból származnak.



VI.1. ábra
A szálkai szervesanyag-tartalom adatok (szűrt adatsor) térképi megjelenítése
(Megjegyzés: a jelmagyarázat logaritmikus skálázású)

Térképi alapon megjelenítettem az adatponti értékeket (VI.1. ábra), mellyel az adatok térbeli eloszlását és változékonyságát vizsgáltam. A kisvízgyűjtő északi, délkeleti, illetve nyugati területei alulreprezentáltak a mintavételi pontok számában. A szervesanyag-tartalom adatok igen szélsőséges értékek között változnak (0,76–5,14%), mely a térbeli változékonyságukban is drasztikusan megmutatkozik. A szervesanyag-tartalom adatok (VI.1. ábra), illetve a felszínfedettségi típusok (IV.1. ábra) szoros kapcsolatot mutatnak: míg a szántóföldek, az erodált szántóföldek és a szőlőültetvények szervesanyag-tartalma igen alacsony (középső, déli és délkeleti területek), addig a rétek, a legelők, illetve az erdők alatt a humusztartalom jelentősen magasabb (északi és északkeleti területek). Ezért a szervesanyag-tartalom térbeli változékonyságában jól visszatükröződik a felszínfedettségi típusok nagyfokú térbeli változatossága. Ez trend jelenlétére utal, ahol is a trend az adatok átlagában megjelenő szisztematikus változás. A trend jelenléte cáfolja a geostatistikában ismeretes belső hipotézist, mely a „leggyengébb” stacionaritási típus (Füst, Geiger, 2010). Ez azt jelenti, hogy el kell távolítani a trendhatást az adatok átlagából és a reziduumokon végezhetem el a geostatistikai vizsgálatokat (Füst, Geiger, 2010; Szatmári, Barta, 2013). Ez indokolja a regresszió krigelés alkalmazását a térbeli becslés során.

Az adatsor decilisei alapján előállított indikátor térképeket a 3. melléklet tartalmazza. Az első két térkép (D1, D2) alapján jól látható, hogy a szervesanyag-tartalom minimumai a kisvízgyűjtő északi és déli – szántó hasznosítású – területeire esnek. Ezeknek a minimumoknak az elhelyezkedése igen szórványos. Ennek az oka, hogy ezek az adatpontok erősen erodált talajfoltokra esnek, melyek megjelenése igen mozaikos a talajerózió dinamikájából, a talajművelési gyakorlatból, a termesztett növények rossz talajvédő hatásából, a táblák méretéből és elhelyezkedéséből, illetve a domborzat mikro-heterogenitásából kifolyólag. Az utolsó két térkép (D8, D9) szerint a maximumok a kutatási terület északi és északkeleti – erdőkkel borított – részein helyezkednek el. Ahogy azt a D1 és D9 térképek is mutatják, a vízgyűjtő északi részén egy minimum és egy maximum meglehetősen közel esik egymáshoz. Ez azzal magyarázható, hogy míg az északabbra található adatpont erdő, addig a délebbre elhelyezkedő adatpont szántó területhasználatra esik. Ugyanakkor az is megfigyelhető (D7), hogy a szántóföldek alatt a humusztartalom nem emelkedik magasabbra, mint 2%, melynek hátterében az erózió okozta szerves anyag veszteség, illetve az intenzív mezőgazdasági gyakorlattal felgyorsított szerves anyag mineralizáció áll.

A módosított $F(r)$ üres-tér függvény algoritmussal kapott eredményt a 3. melléklet tartalmazza, mellyel a mintavételi pontok közötti kapcsolat jellegét tártam fel. A tapasztalati

$\hat{F}(r)$ üres-tér függvény közel áll a Poisson folyamatot jelentő elméleti függvényhez, melyhez 95%-os konfidencia sávot szerkesztettem. Ennek alapját – a kutatási terület karakterisztikaival azonos – Poisson (pont-)folyamatra felépített szimuláció 100 realizációja jelentette. A tapasztalati $\hat{F}(r)$ függvény teljes lefutásában a realizációk alapján meghatározott konfidencia sávon belül marad, melyből azt a következtetést vontam le, hogy a mintavételi helyek függetlenek egymástól. E szerint a mintavételi pontok térbeli elhelyezkedése se nem szabályos, se nem csoportosult (Baddeley, Turner, 2005). A függvény meghatározásának alapját jelentő, ún. távolság térkép a 3. mellékletben található; összefoglaló statisztikáját a VI.4. táblázat mutatja be. A távolság térkép a becslési gridre (VI.1. fejezet) vonatkozóan azt a távolságot (méter dimenzióban) mutatja be, mely az adott grid pont és a hozzá legközelebb eső adatpont között van. A térkép alapján a kutatási terület északi, délkeleti, illetve nyugati területei alulreprezentáltak. A távolságok itt meghaladják a 200 métert, esetenként a 300 métert is (északi és délkeleti mintaterület határok). A legkisebb távolság gyakorlatilag nulla méter, míg a legnagyobb távolság közel 350 méter. A becslési pontok fele nincs távolabbra a hozzájuk legközelebb eső adatponttól, mint 72,6 méter (VI.4. táblázat).

VI.4. táblázat

A szálkai adatpontok alapján számított távolság térkép összefoglaló statisztikája*

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Távolság térkép ($N_{\text{grid}} = 13.221$)	85,7	72,6	0,02	349,8	54,7	1,3

* A távolság térkép a 3. mellékletben található

Utolsó lépésként vizsgáltam a mintavételi pontok, illetve a felhasználni kívánt segédinformációk kapcsolatát. A mintavételi pontok és a folytonos segédinformációk alapján elkészített back-to-back hisztogramokat a 3. melléklet tartalmazza. A hisztogramok alapján elmondható, hogy a mintavételi pontok jól reprezentálják a profil és planáris görbületet, melyek jelentős szereppel bírnak a vízerózió kialakulásában és dinamikájában. A mintavételi pontok alacsony száma miatt a többi hisztogram értékelése nehéz („fűrész” hisztogramok). Ugyanakkor jól látható, hogy az LS faktor magasabb értékei – melyek a meredek lejtőket, illetve a hosszan elnyúló lejtőoldalakat jelentik – alulreprezentáltak; ugyanez figyelhető meg a lejtőszög esetén is. Ennek hátterében az áll, hogy ezek a területek többnyire erdőkkel fedettek, melyek a mintavételezés során alul voltak reprezentálva (V.1.1. fejezet). A korábbi megfigyeléseimmel összhangban vannak a kategória típusú segédinformációkra kiszámított

összterületi részesedések. A VI.5. táblázatban látható, hogy a mintavételi pontok számában a szántók jelentősen felül, míg az erdők jelentősen alulreprezentáltak. A szőlőültetvények esetén is elmondható, hogy habár összterületük közel 10%-a az egész kisvízgyűjtőnek, addig „csupán” három mintavételi pont esik e fészíntfedettségi típusra (VI.5. táblázat). A rét/legelő területek – a szántókhoz hasonlóan – kissé felülreprezentáltak az adatpontok számában.

VI.5. táblázat

A szátkai kategrikus segédinformációkra kiszámított összterületi részesedése a mintavételi pontoknak

Felszíntfedettségi típus	Mintavételi pontok száma		Terület	
	[db]	[%]	[ha]	[%]
Erdő	3	6,1	32,7	24,9
Rét/legelő	11	22,5	21,2	16,1
Szőlő	3	6,1	12,6	9,5
Szántóföld	32	65,3	65,2	49,5
<i>Összesen</i>	<i>49 db</i>	<i>100 %</i>	<i>131,7 ha</i>	<i>100 %</i>

VI. 2. 2. Előszállási szántóföldek

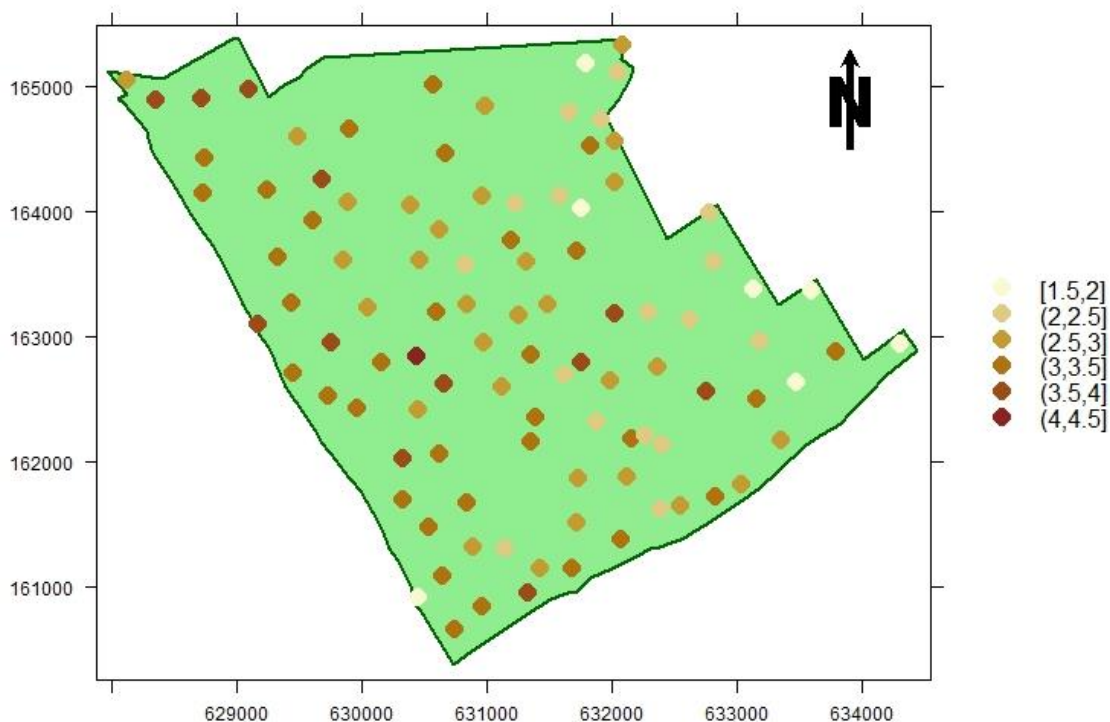
Az előszállási adatbázis szervesanyag-tartalmára vonatkozó összefoglaló statisztikát a VI.6. táblázat mutatja be. Az adatbázis összes adatpontjára rendelkezésre állt mért humusztartalom érték. Az adatbázis két olyan adatpontot tartalmaz – melyek helyileg szántó területekre esnek –, ahol a szervesanyag-tartalom magasabb, mint 5%. Ezek az értékek bizonyos szempontból elfogadhatók, bizonyos szempontból visszautasíthatók. Elfogadhatók, mivel az előszállási területre jellemző mészlepedékes csernozjom talajokra a humuszosodás a jellemző talajképződési folyamat, mely révén szerves anyaggal gazdagon ellátottak (Stefanovits et al., 2010). Másrészt visszautasíthatók mivel a területre jellemző intenzív monokultúrás mezőgazdasági gyakorlat felgyorsítja a szántóföldekben raktározott szerves anyag mineralizációját, amit tovább súlyosbít a vízerózió okozta szerves anyag veszteség is (Szatmári, Barta, 2013). Az adatokra kiszámított box-plot (4. melléklet) alapján kizártam ezt a két adatpontot a további feldolgozásokból. Az adatsorból véletlenszerűen kiválasztottam 15 adatpontot, melyeket független kontroll pontként használtam fel a későbbi térképezési eredmények értékelésére. A fennmaradó 100 adatpontot használtam fel a geostatistikai modellalkotáshoz.

A szerves anyag adatokra elkészített box-plot-ot, hisztogramot és Q-Q diagramot a 4. melléklet tartalmazza. A mellékletben található hisztogram és Q-Q diagram a geostatistikai modellalkotásra szánt 100 adatra készült el. Az elkészített hisztogram (4. melléklet), illetve a

kiszámított ferdeségi paraméter (VI.6. táblázat) a szervesanyag-tartalom adatok közel szimmetrikus eloszlásáról tanúskodnak. A Shapiro-Wilk próba során a megfigyelt szignifikancia szint 0,117-nek adódott, mely szerint 5%-os szignifikancia szint mellett nem utasíthatom vissza a próba nullhipotézisét vagyis, hogy az adatok normál eloszlású populációból származnak.

VI.6. táblázat
Az előszállási szervesanyag-tartalom adatok összefoglaló statisztikája

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Nyers adatsor (n = 117)	2,94	2,95	1,51	5,12	0,62	0,36
Szűrt adatsor (n = 115)	2,90	2,91	1,51	4,44	0,56	-0,25



VI.2. ábra
Az előszállási szervesanyag-tartalom adatok térképi megjelenítése

Térképi alapon megjelenítettem a 100 adatponti értéket (VI.2. ábra), mellyel az értékek térbeli eloszlását és változékonyságát, illetve a trendhatás jelenlétét vizsgáltam. Az adatpontok közel egységesen fedik le a kutatási területet. A humusztartalom tág határok között váltakozik (1,51–4,44%), melynek háttérében a területen folytatott intenzív monokultúrás mezőgazdálkodás és a területet érintő talajerózió áll (Szatmári, Barta, 2013). Délnyugat-

északkeleti irányban az adatponti értékek markáns csökkenése figyelhető meg. Továbbá relatíve alacsony humusztartalom (~1,7%) adatok jelentkeznek a terület északkeleti határa mentén. E területek talajtani viszonyait a domborzat undulációja határozza meg. Ezért az itt jelentkező talajanyag áthalmozódás humuszban szegény földes kopár talajokat hozott létre, melyek mozaikosan jelennek meg a kutatási terület e részein. Ezek a megfigyelések trend jelenlétére utalnak, mely indokolja a regresszió krigelés alkalmazását.

Az előszállási adatok decilisei alapján előállított indikátor térképeket a 4. melléklet tartalmazza. A szervesanyag-tartalom minimumai (D1) az északkeleti határ mentén jelentkeznek, melyek háttérében a talajerózió okozta szerves anyag áthalmozódás áll. A D1 térképen egy minimum jelenik meg a mintaterület dél sarkában, mely a löszplatóból a meredek völgybe hajló perem „túlszántásából” adódik (Szatmári, Barta, 2012). A humuszban gazdagabb területek (D6–D9) főként a kutatási terület délnyugati határa mentén található, melyek a löszplató „tetején” helyezkednek el, ahol védettek a vízerózió pusztító hatásaival szemben. A mintaterület közepén található magasabb szervesanyag-tartalmú adatpontok erdő, illetve rét/legelő felszínfedettségi kategóriák alá esnek. Ezek kellő védelmet tudnak nyújtani a talajerózióval szemben (Thyll, 1997; Barta, 2004). Továbbá egyensúlyba hozzák a talajban a humusz anyagok szintézisének és mineralizációjának mérlegét, melyet az intenzív és monokultúrás növénytermesztés „felborít” a mineralizáció javára (Szatmári, Barta, 2013). A decilisek térképein jól megfigyelhető, hogy északnyugat-délkeleti irányban az adatok nagy(obb) térbeli folytonosságot mutatnak, mint a rá merőleges irányban, mely a VI.2. ábrán is megfigyelhető. Ez utóbbi fontos információ lesz a feltáró variográfia során, mikor is a lehetséges anizotrópiák feltárására és modellezésére kerül sor (VI.3.2. fejezet).

A módosított $F(r)$ üres-tér függvény algoritmussal kapott eredményt a 4. melléklet tartalmazza. A tapasztalati $\hat{F}(r)$ üres-tér függvény meredekebb lefutású, mint a Poisson folyamatot jelentő elméleti függvény. Az elméleti függvényhez szerkesztett 95%-os konfidencia sávon kívül esik a tapasztalati $\hat{F}(r)$ függvény az $r \in [180, 425]$ intervallumon, melyből azt a következtetést vontam le, hogy a mintavételi pontok elhelyezkedései nem függetlenek egymástól. A kapcsolat típusára a két függvény meredeksége alapján következtethetünk (Baddeley, Turner, 2005): a tapasztalati függvény meredekebb lefutású, mint az elméleti függvény, ezért a mintavételi pontok közötti kapcsolat jellege kizáró. Ez azt jelenti, hogy a mintavételi pontok elhelyezkedése a szabályos felé tart (Baddeley, Turner, 2005). A tapasztalati függvény meghatározásának alapját jelentő távolság térkép a 4. mellékletben

található; összefoglaló statisztikáját a VI.7. táblázat mutatja be. A számítás alapját a becslési grid jelentette (VI.1. fejezet). A távolság térkép szerint a kutatási terület északi részei kissé alulreprezentáltak a mintavételi pontok számában. Az északi határ mentén a számított távolságok meghaladják az 500 métert, esetenként a 600 métert. Szintén megfigyelhető egy kisebb „hiátus” a mintaterület északkeleti határának közelében. A legkisebb távolság valamivel több, mint 2 méter, míg a legnagyobb távolság közel 630 méter (VI.7. táblázat). A becslési pontok fele nincs távolabbra a hozzájuk legközelebb eső adatponttól, mint 174,2 méter.

VI.7. táblázat

Az előszállási adatpontok alapján számított távolság térkép összefoglaló statisztikája*

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Távolság térkép ($N_{\text{grid}} = 44.359$)	185,4	174,2	2,4	630,3	94,6	0,8

* A távolság térkép a 4. mellékletben található

VI.8. táblázat

Az előszállási kategorikus segédinformációkra kiszámított összterületi részesedése a mintavételi pontoknak

Felszínfedettségi típus	Mintavételi pontok száma		Terület	
	[db]	[%]	[ha]	[%]
Szántóföld	87	87,0	1401,2	84,3
Rét/legelő	5	5,0	79,9	4,8
Átmeneti erdős-cserjés	3	3,0	46,0	2,8
Erdő	5	5,0	124,8	7,5
Mocsár*	0	0	3,0	0,2
Nyílt víz*	0	0	4,1	0,3
Mesterséges felszín*	0	0	2,0	0,1
Összesen	100 db	100%	1661,0 ha	100%

* A talajfelvételezések szempontjából irreleváns felszínfedettségi típusok

Utolsó lépésként vizsgáltam a mintavételi pontok és a felhasználni kívánt segédinformációk kapcsolatát. A mintavételi pontok és a folytonos segédinformációk alapján elkészített back-to-back hisztogramokat a 4. melléklet tartalmazza. A hisztogramokat elemezve elmondható, hogy a mintavételi pontok jól reprezentálják a felhasználni kívánt folytonos segédinformációkat, hiszen a mintavételi pontok alapján számított hisztogramok és a segédinformációk raszterei alapján számított hisztogramok nagyban hasonlítanak egymásra. A tengerszint feletti magasság, a lejtőszög, az LS faktor, illetve a TWI esetén megfigyelhető, hogy

a mintavételi pontok kissé felül reprezentáltak e morfometriai paraméterek alacsony értékeiben. A kategorikus segédinformációkra kiszámított összterületi részesedéseket a VI.8. táblázat foglalja össze. A táblázat szerint a mintavételi pontok megoszlása az egyes felszínfedettségi kategóriákban jóval kiegyenlítettebb, mint a szátkai mintaterület esetén (VI.5. táblázat). A talajfelvételezések szempontjából irreleváns területekre (például: nyílt vizek, mesterséges felszínek) nem estek mintavételi pontok (VI.8. táblázat).

VI. 2. 3. Zala megye

A Zala megyei szervesanyag-tartalom észlelések összefoglaló statisztikáját a VI.9. táblázat tartalmazza. Összesen 1989 adatpontot foglal magába a DKTIR adatbázis Zala megyére vonatkozóan. Az adatponti értékek áttekintése – nagy számuk révén – nehéz feladat. Ezért a DKTIR talajok vízgazdálkodási tulajdonság talajtérképezési egységeit (5. melléklet) használtam fel az adatponti értékek áttekintésére és szűrésére.

VI.9. táblázat
A Zala megyei szervesanyag-tartalom adatok összefoglaló statisztikája

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Nyers adatsor (n = 1989)	2,72	2,23	0,43	29,34	3,03	5,79
Szűrt adatsor (n = 1911)	2,59	2,23	0,43	29,34	2,79	6,67

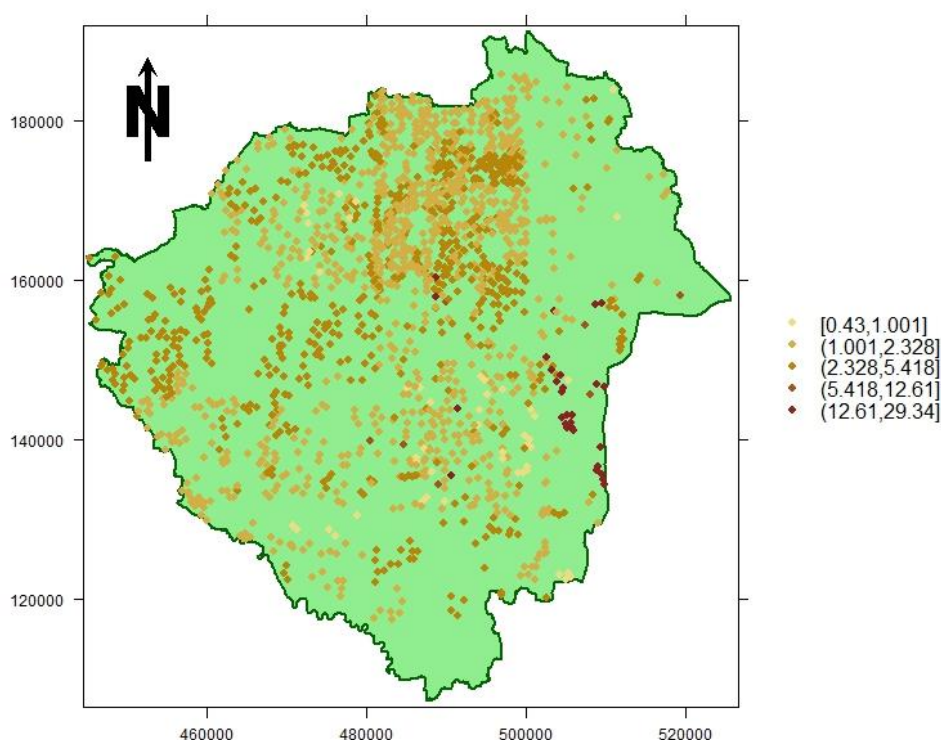
A talajtérképezési egységek implicit tartalmazzák a talajtérképezők szakmai ismereteit, illetve mentális talaj-táj modelljeit, ezért jól alkalmazhatók az adatok előzetes áttekintésére. Elsőként a DKTIR „erdő” poligonjait használtam fel az adatok szűrésére lévén, hogy a Kreybig-féle talajismereti térképezések során az erdők nem kerültek felmérésre (Kreybig, 1934, 1937, 1950). Összesen 31 adatpont esett erdő területhasználat alá, melyeket kizártam a további adatfeldolgozásból. A Kreybig-féle talajtérképezések során lehatárolásra kerültek a „tőzeges talajok”, illetve az „időszakosan vízállásos, vízjárta területek” (Kreybig, 1934, 1937, 1950), mely talajtérképezési egységek fontos információkkal rendelkeznek a talajok szerves anyag forgalmára vonatkozóan. A láptalajok szervesanyag-tartalma kiugróan magas, meghaladhatja a 10%-ot is (Stefanovits et al., 2010). Ugyanakkor voltak olyan adatpontok, melyeknek a szervesanyag-tartalma alacsonyabb volt, mint 3%. Ezeket az adatpontok (összesen 10 db) kizártam a további adatfeldolgozásokból lévén, hogy ilyen alacsony szervesanyag-tartalom

nem jellemző a láptalajokra. Az „időszakosan vízállásos, vízjárta területek” talajai szintén humuszban igen gazdagok a megyében. Ezekre a területekre koncentrálnak a megye lápos és réti talajai, melyek szervesanyag-tartalma – összefoglalva a pangóvízes, illetve az agyagbemosódásos barna erdőtalajokra jellemző értékekkel – relatíve magas. Azonban találhatók olyan adatpontok, mely szervesanyag-tartalma igen alacsony (kisebb, mint 1,5%). Ezeket az adatpontokat (5 db) kizártam a további vizsgálatokból. Az egyéb talajtérképezési egységekbe tartozó adatpontok a megyére jellemző pangóvízes, illetve agyagbemosódásos barna erdőtalajokra esnek. Ezek a talajok humusszal közepesen, illetve szegényen ellátottak. Ennek háttérben a talajba kerülő savas karakterű szerves anyagok állnak, melynek eredményeként az itt képződő humuszanyagok könnyen mobilizálhatók és alacsony polimerizáltsági fokkal jellemezhetők, melyek a kilúgozás révén vertikálisan elmozdulnak a szelvény mélyebb szintjei felé (Stefanovits et al., 2010). Ebből az is következik, hogy a szerves anyag minősége is eltér a réti és lápos talajok szerves anyagától, hiszen utóbbiak esetében az időszakosan vízzel telített talajokban a szerves anyag mineralizációja akadályozott. Ez alapján 32 adatpontot zártam ki a további vizsgálatokból. A szűrt adatsorból véletlenszerűen kiválasztottam 200 adatpontot, melyeket független kontroll pontokként használtam fel a térképezési eredmények értékelésére. A fennmaradó 1711 adatpont jelentette a geostatistikai modellalkotás alapját.

A szerves anyag adatokra elkészített box-plot-ot, hisztogramot és Q-Q diagramot az 5. melléklet tartalmazza. A mellékletben található hisztogram és Q-Q diagram a geostatistikai modellalkotásra szánt 1711 adatpontra készült el. Az elkészített hisztogram (5. melléklet), illetve a kiszámított ferdeségi paraméter (VI.9. táblázat) alapján az adatok eloszlása pozitív ferde. A Kolmogorov-Smirnov próba során 5%-os szignifikancia szint mellett visszautasítottam a próba nullhipotézisét vagyis, hogy az adatok normál eloszlású populációból származnak.

Térképi alapon megjelenítettem az adatponti értékeket (VI.3. ábra), mellyel az adatok térbeli eloszlását és változékonyságát vizsgáltam. A mintavételi pontok térbeli eloszlása „csoportosult”. Megfigyelhető, hogy a megye északi területeire több mintavételi pont esik, mint a déli és keleti területekre, ahol a Kreybig-féle térképszelvények szerint kiterjedt erdőségek találhatók (5. melléklet). Ez jól megfigyelhető a Göcsejben, melynek 40%-a erdővel borított. A szervesanyag-tartalom értékek rendkívül tág határok között változnak (0,43–29,34%). Az adatponti értékek alapján jól kirajzolódnak a meridioniális völgyek, ahol a megye réti és lápos talajai találhatók (VI.3. ábra). Ezek a talajok szerves anyaggal gazdagon ellátottak. Az adatponti értékek szintén jól kirajzolják a Felső-Zala-völgy „futását”, ahol a talajképződést ez esetben is

vízhatás kíséri, melynek következtében szintén szerves anyagban gazdag talajokat találunk (VI.3. ábra). A legmagasabb szervesanyag-tartalmú adatpontok a Kis-Balaton medence területére esnek, ahol is a különböző vastagságú tőzegrétegeken kialakult síkláptalajokat találunk. Ezek szervesanyag-tartalma meghaladja a 20%-ot is (VI.3. ábra). A fentiekkel kontrasztban, jól megfigyelhetők a megye közepes és alacsony humusztartalmú területei. Itt többnyire pangóvízes és agyagbemosódásos barna erdőtalajokat találunk. Ezek a talajok a megyét hosszában „felszelő” meridionális völgyek közötti domboldalakon helyezkednek el (VI.3. ábra), ahol talajdegradációs folyamatként jelenik meg a vízerózió és a talajok savanyodása. Előbbi degradációs folyamat tovább differenciálja a szervesanyag-tartalom térbeli eloszlását, hiszen a vízerózió okozta károk Zala megye területének közel 52%-át érintik (Thyll, 1997). A dombtetők többnyire erdővel fedettek, ezért ezekre a területekre nem estek mintavételi pontok. A talajképződési folyamatok e nagymértékű változékonysága igen jellegzetes „kontrasztos” mintázatot ad a megye talajainak, mely a szervesanyag-tartalom térbeli változékonyságában is jelentkezik.



VI.3. ábra

*A Zala megyei szervesanyag-tartalom adatok térképi megjelenítése
(Megjegyzés: a jelmagyarázat logaritmikus skálázású)*

A zalai adatsor decilisei alapján előállított indikátor térképeket az 5. melléklet tartalmazza. A humusztartalom minimumai (D1) főként a megyét „felszelő” meridionális

völgyek meredek lejtésű peremein helyezkednek, ahol a vízerózió akadályozza a humuszanyagok talajban való felhalmozódását. A Göcsej északi területein is megjelennek ezek a minimumok, ahol az erózióval erősen sújtott hosszú és meredek lejtésű domboldalak találhatók. A szervesanyag-tartalom maximumai (D9) a Kis-Balaton medencében, az Felső- és Alsó-Zala-völgyben, illetve a Principális-völgyben találhatók, ahol szerves anyagban gazdag réti és lápos talajokat találunk. További humusztartalom maximumok jelennek meg a Kerka-vidék és az Egerszeg-Letenyei-dombság alacsony lejtésű völgytalpi területein, ahol a talajképződést vízhatás kíséri, mely akadályozza a talajba kerülő szerves anyag mineralizációját.

VI.10. táblázat

A Zala megyei adatpontok alapján számított távolság térkép összefoglaló statisztikája*

Összefoglaló statisztika	átlag	medián	min.	max.	szórás	ferdeség
Távolság térkép ($N_{\text{grid}} = 671.679$)	1.196,9	861,3	0,9	11.231,9	1.165,5	2,89

* A távolság térkép az 5. mellékletben található

A módosított $F(r)$ üres-tér függvény algoritmussal kapott eredményt az 5. melléklet tartalmazza. A tapasztalati $\hat{F}(r)$ üres-tér függvény lefutása kisebb meredekségű, mint a Poisson folyamatot jelentő elméleti függvény. Az elméleti függvényhez szerkesztett 95%-os konfidencia sávon kívül esik a tapasztalati $\hat{F}(r)$ függvény, mely megerősíti a korábbi megállapításomat, miszerint a mintavételi pontok csoportosult mintázatot mutatnak. A tapasztalati függvény meghatározásához használt távolság térkép az 5. mellékletben található; összefoglaló statisztikáját a VI.10. táblázat tartalmazza. A számítás alapját a becslési grid jelentette (VI.1. fejezet). A távolság térképen jól megfigyelhető, hogy a mintavételi pontok számában Zala megye északi területei felül, míg a keleti és déli területei alulreprezentáltak. A déli és keleti területeken a számított távolság értékek meghaladják az 5 km-t. A megye déli részén ezek a távolságok esetenként meghaladják a 10 km-t is (5. melléklet). A legkisebb távolság közel 1 méter, míg a legnagyobb távolság meghaladja a 11.200 métert (VI.10. táblázat). A becslési pontok fele nincs távolabbra a hozzájuk legközelebb eső adatponttól, mint 861,3 méter.

VI.11. táblázat

A DKTIR talajok vízgazdálkodási tulajdonság térképezési egységeire kiszámított összterületi részesedése a mintavételi pontoknak

Talajok vízgazdálkodási tulajdonságai	Mintavételi pontok száma		Terület	
	[db]	[%]	[km ²]	[%]
Jó víztartó és vízvezetőképességű talajok	833	48,7	1302	50,4**
Közepes vízvezetőképességű, vizet erősebben tartó talajok	32	1,9	74	2,9**
Gyenge vízvezetőképességű, vizet erősen tartó talajok	3	0,2	33	1,3**
Nagy vízvezetőképességű, még jó víztartó talajok	637	37,2	549	21,2**
Igen nagy vízvezetőképességű, gyengén víztartó talajok	107	6,2	142	5,5**
Köves, kavicsos	0	0	2	0,1**
Tőzeges talajok	18	1,1	55	2,2**
Időszakosan vízállásos, vízjárta területek	81	4,7	422	16,4**
Erdők*	0	0	1030	-
Nyílt vizek, nádasok*	0	0	57	-
Település*	0	0	110	-
Összesen	1711 db	100%	3781 km²	100%

* A talajfelvételezések szempontjából irreleváns térképezési egységek;

** A talajfelvételezések és térképezések szempontjából releváns térképezési egységek összterületéhez viszonyított százalékok;

Rövidítés: DKTIR: Digitális Kreybig Talajinformációs Rendszer

Utolsó lépésként vizsgáltam a mintavételi pontok és a felhasználni kívánt segédinformációk kapcsolatát. A mintavételi pontok és a folytonos segédinformációk alapján elkészített back-to-back hisztogramokat az 5. melléklet tartalmazza. A hisztogramokat elemezve elmondható, hogy a mintavételi pontok jól reprezentálják a felhasználni kívánt –

digitális domborzatmodellből levezetett – segédinformációkat lévén, hogy a mintavételi pontok alapján és a segédinformációk raszterei alapján elkészített hisztogramok nagyban hasonlítanak egymásra. A TWI és a völgymélységek esetén ugyanakkor megfigyelhető, hogy a mintavételi pontok kissé felülreprezentáltak e morfometriai paraméterek alacsony értékeiben. A DKTIR talajok vízgazdálkodási tulajdonságai térképezési egységekre kiszámított összterületi részesedéseket a VI.11. táblázat foglalja össze. A jó víztartó és vízvezetőképességű talajok, a közepes, a gyenge, illetve az igen nagy vízvezetőképességű talajok, továbbá a tözeges talajok kissé felülreprezentáltak a mintavételi pontok számában. Az időszakosan vízállásos, vízjárta területekre eső mintavételi pontok száma jelentősen magasabb, mint ahogy azt a térképezési egység összterületi részesedése megkívná. A nagy vízvezetőképességű talajok alul vannak reprezentálva a mintavételi pontok számában.

VI. 3. A szervesanyag-tartalom térképek és értékelésük

VI. 3. 1. Szálkai kisvízgyűjtő

A feltáró alapadat elemzésen átesett adatsort használtam fel a kisvízgyűjtő szervesanyag-tartalmának geostatisztikai alapú térképezéséhez. A szálkai adatok pozitív ferde eloszlást mutattak (3. melléklet és VI.3. táblázat), ezért logaritmus transzformációt végeztem a szűrt adatsoron, mely segítségével normál eloszlásúvá tettem az adatsort. A térbeli becslésben felhasználandó segédinformációkon főkomponens analízist végeztem a multikollinearitás csökkentésére. Az eredményül kapott főkomponenseket használtam fel a többszörös lineáris regresszió analízisben, ahol a függő változó a szervesanyag-tartalom volt. A regressziós modellbe bevont független változókat „stepwise” módszerrel választottam ki, ahol az alkalmazott szignifikancia szint 5% volt. A regresszió analízis eredményét a VI.12. táblázat foglalja össze. A „stepwise” módszer összesen 6 független változót vont be a regressziós modellbe: három főkomponens, illetve három felszínfedettségi típust. Az erodált szántóföldekre becsült regressziós együttható értéke negatív előjelű. Ebből arra következtettem, hogy az erózióval sújtott szántókon a humusztartalom alacsonyabb, mint más felszín fedettségi kategóriák alatt. Ez megegyezik a feltáró alapadat elemzés során tett megfigyeléseimmel (VI.2.1. fejezet). Az erdő, illetve a rét/legelő típusokra becsült regressziós együtthatók értékei pozitív előjelűek, melyből arra következtettem, hogy e kategóriák alatt magasabb a szervesanyag-tartalom. Ez szintén összhangban van a korábbi megfigyeléseimmel (VI.2.1. fejezet). A modellbe vont első főkomponens (SPC-1) kialakításában a tengerszint feletti

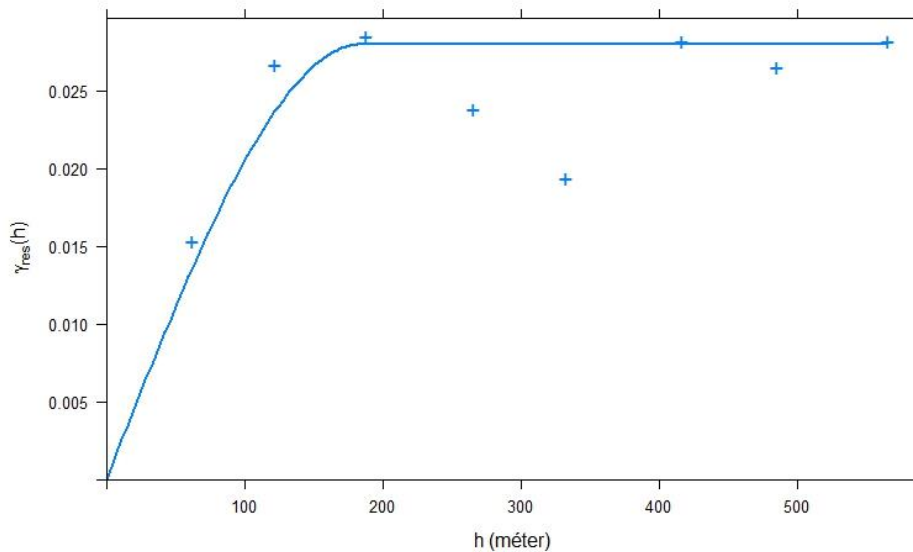
magasság töltött be döntő szerepet, a hetedik főkomponens (SPC-7) kialakításában a völgyek mélysége töltött be fontos szerepet, míg a nyolcadik főkomponensre (SPC-8) a lejtőszög és az LS faktor volt hatással. Ezek a domborzati jellemzők a vízerózió kiváltó tényezői közé sorolhatók, melyek ily módon döntő szerepet játszanak a szerves anyag áthalmozódásában (Thyll, 1997; Centeri et al., 2014; Jakab et al., 2014, 2016; Szalai et al., 2016). A regressziós modell determinációs együtthatója (R^2): 79%; e szerint a humusztartalom térbeli változékonyságának közel 80%-át tudtam leírni az adatpontokra illesztett többszörös lineáris regressziós modellel. Ez azt is jelenti, hogy a fennmaradó, közel 20%-nyi térbeli változékonyságot kell egyszerű krigeléssel modellezni.

VI.12. táblázat

A szálkai mintaterületre végzett többszörös lineáris regresszió analízis eredménye

Független változók	Reg. együtthatók	p
SPC-1	-0,047	0,007
SPC-7	2,798	$3,10 \cdot 10^{-4}$
SPC-8	0,510	0,009
Erodált szántó	-0,224	0,001
Erdő	0,800	$4,01 \cdot 10^{-7}$
Rét/legelő	0,576	$1,58 \cdot 10^{-8}$
Ordinátatengely-metszet	0,317	$3,43 \cdot 10^{-8}$
Determinációs együttható (R^2):	0,79	
Modellre számított p:	$4,88 \cdot 10^{-11}$	

Az adatpontokra illesztett többszörös lineáris regresszióból származtattam a regresszió reziduumaikat, melyek egy stacionárius valószínűség függvényként modellezhetők. Gyenge stacionaritás feltételezése mellett a variogram és a kovariancia ekvivalens eszközöknek tekinthetők a térbeli változékonyság kifejezésére (Füst, Geiger, 2010). A reziduumokra kiszámított irányfüggetlen tapasztalati variogramot a VI.4. ábra mutatja be. Az alacsony mintaszám miatt nem vállalkoztam a lehetséges anizotrópiák modellezésére lévén, hogy a kiszámított irányfüggő variogramok nagy bizonytalansággal terheltek. Ez „ellehetetleníti” az anizotrópia korrekt modellezését (R.M. Lark szóbeli közlése). Ezért izotróp variogram modellt illesztettem a kiszámított variogramra (VI.4. ábra). Az izotróp modell típusa szférikus, a tető értéke 0,028, a hatástávolsága 185 méter és a röghatás 0.

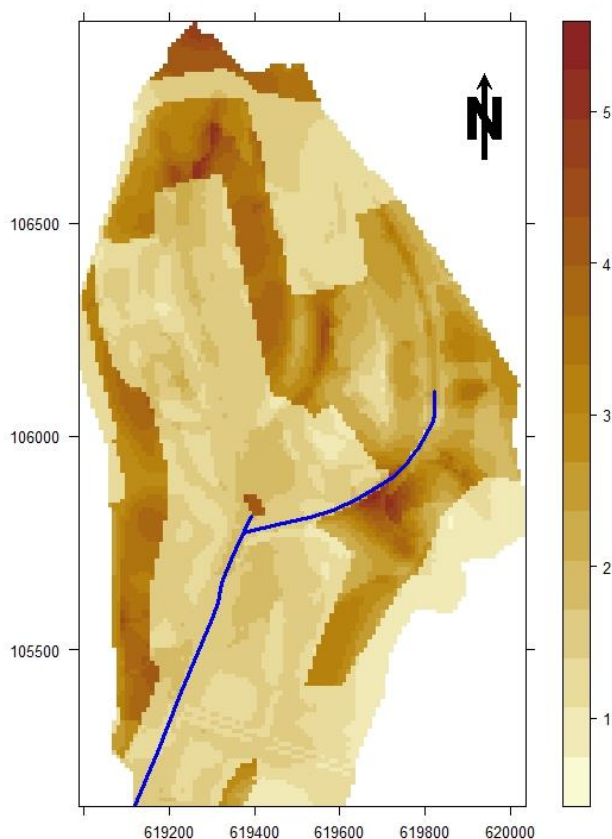


VI.4. ábra

A szátkai reziduumok irányfüggetlen tapasztalati variogramja („+”) és az illesztett izotróp variogram modell (folytonos vonal)

A többszörös lineáris regresszió modell és a regresszió reziduumaira illesztett variogram modell segítségével elvégezhető a térbeli becslés. A regresszió krigeléssel adott pontszerű térbeli becslést és a kapcsolódó becslési varianciát a VI.5. ábra és a 6. melléklet mutatja be. A kartografált szervesanyag-tartalom térképet a 6. melléklet tartalmazza. A megszerkesztett szervesanyag-tartalom térkép jól reprezentálja a humusztartalom térbeli változékonyságát. A térkép jól igazodik a felszínfedettségi típusokhoz, miszerint a feltalaj humusztartalma jelentősen alacsonyabb a szántóföldeken, az erősen erodált szántóföldeken és a szőlőültetvényeken. Ennek háttérében a vízerózió káros hatása áll, melynek eredményeként a humuszban gazdagabb talajanyag elszállítódik a területről és a helyén a humuszban jóval szegényebb altalaj marad. A térkép szintén jól visszaadja az erózió eredményeként megjelenő, mozaikos elhelyezkedésű kopár „foltokat”, ahol – a talajtakaró szinte teljes lehordódásával – a felszínre került a szerves anyagban szegény talajképző közet. Kiemelendő, hogy a térkép jól reprezentálja a Lajvér-patak nyomvonalát, mely egyúttal a helyi erózióbázis. A felszínen meginduló vízlepel elszállítja a talaj „könnyen” oldódó, gyengén polimerizálódott, savas karakterű humuszsavait, melyek a lerakott üledékben „dúsulnak” (Borcsik et al., 2011). A kutatási terület erdővel, illetve réttel fedett területein a feltalaj humusztartalma jelentősen magasabb, mint a korábbi felszínfedettségi típusok esetén. Ez azzal magyarázható, hogy e területhasználati típusok kellő védelmet tudnak nyújtani a vízerózió káros hatásaival szemben (Thyll, 1997; Barta, 2004), továbbá az erdővel és réttel fedett területeken sokkal

kiegyenlítettebb a szerves anyagok szintézise és mineralizációja, mely egyensúly az intenzív mezőgazdálkodású szántóterületeken felborul a mineralizáció javára (Szatmári, Barta, 2013).



VI.5. ábra

*A szálkai mintaterület szervesanyag-tartalmának térbeli becslése regresszió krigeléssel
(Megjegyzés: kék szín jelöli a Lajvér-patakot)*

A szervesanyag-tartalom hibájának értékelésére az LOOCV keresztvalidációs módszert alkalmaztam. Az LOOCV révén kiszámítható volt a hiba, mely a becslés és valós érték különbsége. A hibák alapján meghatároztam az átlagos hibát (ME), az átlagos abszolút hibát (MAE), illetve az átlagos négyzetgyök hiba standardizált értékét (RMNSE). Az eredményeket a VI.13. táblázat foglalja össze. Az ME és RMNSE jól közelítik azok várható értékét, vagyis a térkép átlagos hibája közel nulla, míg az RMNSE értéke közel egy. Az ME értéke negatív előjelű, ami azt jelenti, hogy „általában véve” alul becsülte a térkép a tényleges szervesanyag-tartalmat. Azonban ennek mértéke a gyakorlatban jóformán elhanyagolható. A gyakorlati alkalmazásokat (például: 90/2008. FVM rendelet) tekintve a szervesanyag-tartalom értékeket legfeljebb egy tizedesjegyre használják. A ME értékével kapcsolatban megjegyzendő, hogy értéke akkor is nulla – vagy nullához igen közeli –, ha például: sok negatív hiba mellett kevés pozitív nagy hiba van (Isaaks, Srivastava, 1989). Ezért kiszámítottam az átlagos abszolút hibát

(MAE), mely szigorúbb mérték, mint az átlagos hiba (ME). Az MAE értéke kizárólag nem negatív szám lehet. A gyakorlati alkalmazásokat és a területre jellemző szélsőséges értékek közötti nagyfokú térbeli változékonyságot tekintve a MAE értéke is alacsonynak mondható.

VI.13. táblázat
A szálkai szervesanyag-tartalom térkép értékelése

ME	MAE	RMNSE
-0,024	0,193	1,033

Rövidítések: ME: átlagos hiba, MAE: átlagos abszolút hiba, RMNSE: az átlagos négyzetgyök hiba standardizált értéke.

VI. 3. 2. Előszállási szántóföldek

A feltáró alapadat elemzésen átesett adatsort használtam fel az előszállási mintaterület szervesanyag-tartalmának térképezéséhez. Az előszállási adatsor közel normál eloszlást mutatott. A térbeli becslésben felhasználandó segédinformációkon főkomponens analízist végeztem a multikollinearitás csökkentésére. Az eredményül kapott főkomponenseket használtam fel a többszörös lineáris regresszió analízisben, ahol a függő változó a szervesanyag-tartalom volt. A regressziós modell független változóit „stepwise” módszerrel választottam ki, ahol az alkalmazott szignifikancia szint 5% volt. A regresszió analízis eredményét a VI.14. táblázat foglalja össze. A „stepwise” módszer nyolc független változót vont be a regressziós modellbe: öt főkomponenst és három területhasználati típust. A felszín fedettségi kategóriákra becsült regressziós együtthatók értékei negatív előjelűek. A szántók, illetve az erodált szántók esetén ebből arra következtettem, hogy e kategóriák alatt a humusztartalom alacsonyabb. Ez jól magyarázható az intenzív és monokultúras mezőgazdasági gyakorlattal, illetve a talajerózió káros hatásaival. A rét/legelő kategória esetén a negatív előjel azzal magyarázható, hogy a rétek a mintaterületen húzódó löszvölgyek talpaiban helyezkednek el, melyet a vízerózió kényszerített ki (Szatmári, Barta, 2012). A modellbe vont első főkomponens (SPC-1) kialakításában a mintaterület domborzatának konvexitását kifejező paraméterek (profil görbület, planáris görbület stb.) játszottak döntő szerepet. A második főkomponensre (SPC-2) a lejtőszög volt hatással. A harmadikra (SPC-3) a beérkező napsugárzás gyakorolt hatást. A tizedik főkomponensre (SPC-10) a topográfiai nedvesség index hatott leginkább, míg a tizenharmadik főkomponenst (SPC-13) az LS faktor befolyásolta. E domborzati paraméterek mindegyike szorosan kapcsolódik a talajerózió kiváltó és befolyásoló tényezőihez. A domborzat konvexitása alapvetően meghatározza, hogy a talajerózió a lejtő mely szakaszán tud kialakulni (Thyll, 1997); ez természetesen szoros kapcsolatban van a

szerves anyag térbeli átrendeződésével is (Jakab et al., 2014, 2016; Szalai et al., 2016). A déli kitettségű lejtőkre érkező napsugárzás a feltalaj kiszáradását és csapadékesemény esetén „morzsarobbanást” vált ki, mely a felületi (vagy réteg) erózió egyik lehetséges formája (Thyll, 1997). A lejtőszög az egyik legfontosabb komponense az eróziót kiváltó tényezőknek; meredek lejtőn összefüggő vízlepel tud kialakulni, mely elszállítja a termékeny és humuszból gazdag talajréteget a lejtőoldalról (Barta, 2004). A topográfiai nedvesség index a „rétiesedéshez” kapcsolódik, mely a mintaterületet felszabdáló löszvölgyek talpaiban jelentkezik. Az LS faktor szorosan kapcsolódik az erózióval veszélyeztetett lejtőoldalakhoz; fontos komponense az erózió veszélyeztetettség térképezésének (például: Thyll, 1997; Centeri, 2001; Centeri et al., 2003, 2008; Szatmári, Barta, 2012). A regressziós modell determinációs együtthatója (R^2): 62%. E szerint a szervesanyag-tartalom térbeli változékonyságának valamivel több, mint 60%-át tudtam leírni a segédinformációk segítségével. Ez azt is jelenti, hogy a további, közel 40%-nyi változékonyságot kell egyszerű krigeléssel modellezni.

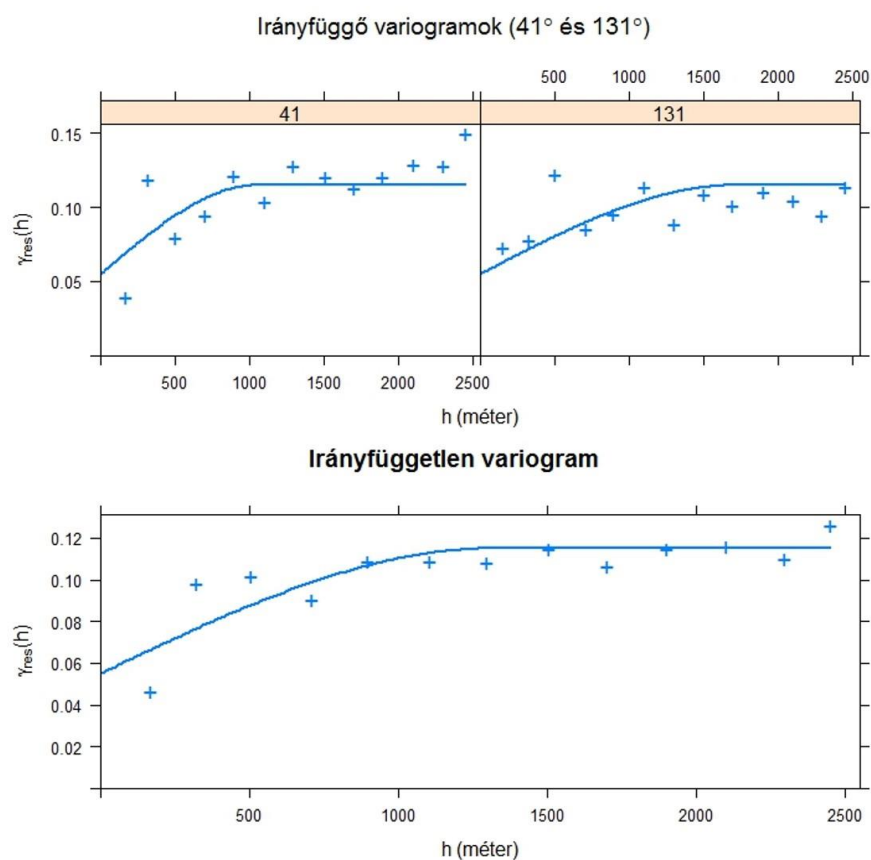
VI.14. táblázat

Az előszállási mintaterületre végzett többszörös lineáris regresszió analízis eredménye

Független változók	Reg. együtthatók	p
SPC-1	0,067	$3,13 \cdot 10^{-4}$
SPC-2	0,139	$1,01 \cdot 10^{-5}$
SPC-3	0,050	0,04
SPC-10	0,309	$6,07 \cdot 10^{-5}$
SPC-13	-0,899	$8,62 \cdot 10^{-9}$
Erodált szántó	-0,619	$7,25 \cdot 10^{-8}$
Szántó	-1,003	$8,43 \cdot 10^{-7}$
Rét/legelő	-0,668	0,002
Ordinátatengely-metszet	3,904	$< 2 \cdot 10^{-16}$
Determinációs együttható (R^2):	0,62	
Modellre számított p:	$1,61 \cdot 10^{-13}$	

A szervesanyag-tartalom adatokra illesztett többszörös lineáris regresszióból származtattam a regresszió reziduumaikat, melyek egy stacionárius valószínűségi függvényként modellezhetők. A reziduumok irányfüggő és irányfüggetlen tapasztalati variogramjait a VI.6. ábra szemlélteti. A feltáró alapadat elemzés során (VI.2.2. fejezet) megfigyelt térbeli folytonosság a reziduumok irányfüggő variogramjain is visszatükröződik (VI.6. ábra). Habár az anizotrópia korrekt modellezéséhez a rendelkezésemre álló 100 adatpont továbbra is alacsony számúnak mondható, mindazonáltal a feltáró variográfia nem nélkülözheti az adott

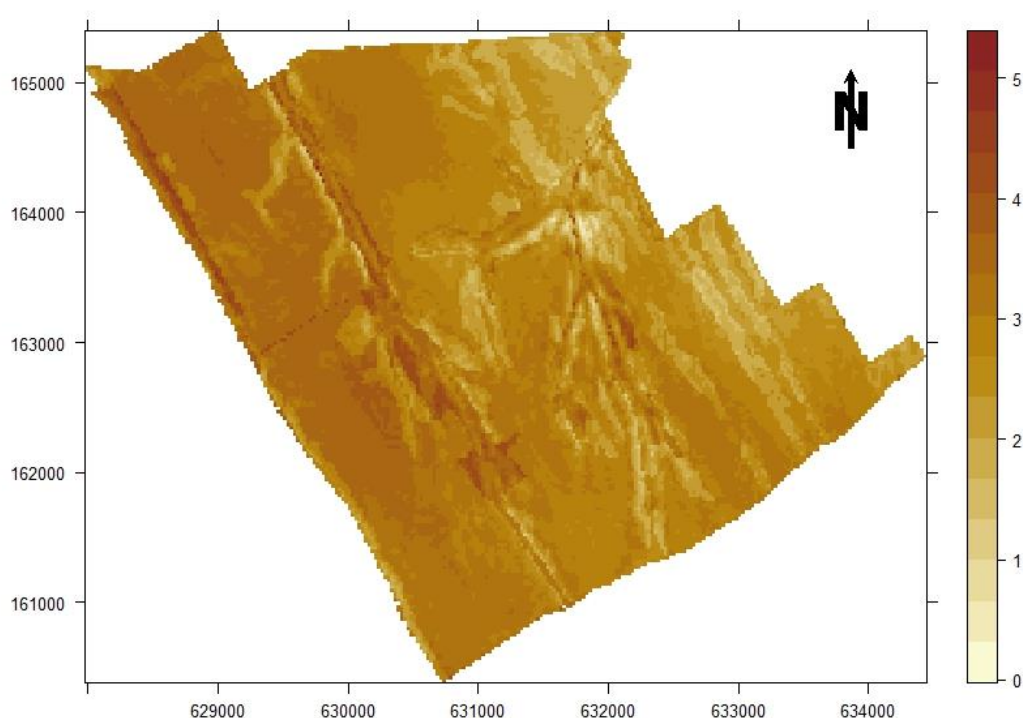
változó térbeli sajátásaival kapcsolatos szakértői ismereteket (Gringarten, Deutsch, 2001; Bárdossy, 2006; Geiger, 2006a; Webster, Oliver, 2007; Szatmári, Pásztor, 2016). Ezért anizotróp variogram modellt illesztettem a reziduumok variogramjaira, hogy a térbeli változékonyságukat a lehető leghívebben vegyem számításba. A fő térbeli folytonosságot 131°-nál (északnyugat-délkelet) állapítottam meg, mely egyezik a feltáró alapadat elemzés során tett megfigyeléseimmel. Ez az irányultság egybeesik a löszplató magasán fekvő, alacsony lejtésű területeivel, melyek északnyugat-délkeleti irányban terülnek el. Az illesztett variogram modell típusa szférikus, a tető értéke 0,115, a röghatás 0,055, a hatástávolság 1750 méter a fő folytonossági irányban, az anizotrópia hányados 0,6.



VI.6. ábra

Az előszállási reziduumok irányfüggő és irányfüggetlen variogramjai („+”) és az illesztett anizotróp variogram modell (folytonos vonal)

A többszörös lineáris regresszió és a reziduumokra illesztett variogram modell segítségével elvégezhető volt a térbeli becslés. A regresszió krigeléssel adott pontszerű becslést és a kapcsolódó becslési varianciát a VI.7. ábra és a 7. melléklet mutatja be. Az előszállási mintaterület kartografált szervesanyag-tartalom térképét a 7. melléklet tartalmazza.



VI.7. ábra

Az előszállási mintaterület szervesanyag-tartalmának térbeli becslése regresszió krigeléssel

A kartografált szerves anyag térkép (7. melléklet) jól reprezentálja a feltalaj szervesanyag-tartalmának térbeli eloszlását, mely igazodik a domborzathoz és a területhasználathoz. Délnyugat-északkeleti irányában a szerves anyag markáns csökkenése figyelhető meg, melyet az adatponti értékek is jeleztek (VI.2.2. fejezet). Az északkeleti határ mentén megjelennek azok a mozaikos elrendeződésű, erősen erodált talajfoltok, ahol a feltalaj humusztartalma igen alacsonyra csökkent az erózió okozta talajanyag lehordódás következtében. Továbbá megfigyelhető, hogy ezek a foltok – melyek a területen előforduló földes kopárokkal azonosíthatók – egymással párhuzamosan futnak, mely a felszín undulációjának eredménye. A löszvölgyek peremeit borító erdők jól prezentálják talajvédő hatásukat, mely a szervesanyag-tartalom relatíve magas értékeiben mutatkozik meg. Ugyanakkor szintén jól kivehető a löszvölgyek közvetlen közelében jelentkező markáns szerves anyag csökkenés, mely a löszplatóból a meredek völgyekbe hajló peremek „túlszántásából” adódik. A löszvölgyek talpaiban jól megfigyelhetők a réti csernozjom talajok. A mintaterületet felszelő középső löszvölgy baloldalán felismerhetők a vonalas erózió formái, melyek a helytelen mezőgazdasági gyakorlat (lejtő irányú talajművelés és rossz talajvédő hatású monokultúrák termesztése) következtében alakultak ki. A térkép alapján az prognosztizálható, hogy ha nem változtatnak a jelenlegi – talajvédelmi megfontolásokat

nélkülöző – mezőgazdasági gyakorlaton, akkor az itt található vonalas formák később kiterjedt vízmosásokká fejlődnek. Ezek akadályozni fogják a mezőgazdasági művelést és később a terület mezőgazdasági „felhagyását” fogják kikényszeríteni, ahogy azt korábban a területet felszelő – mára már főként akáccal beerdősödött – löszvölgyek is tették (Szatmári, Barta, 2012).

Az előszállási szerves anyag térkép értékelését 15 kontroll pont alapján végeztem el, melyek mindvégig függetlenek maradtak a geostatistikai modellalkotástól. A hibák alapján kiszámítottam az átlagos hibát (ME), az átlagos abszolút hibát (MAE), illetve az átlagos négyzetgyök hiba standardizált értékét (RMNSE). Az eredményeket a VI.15. táblázat foglalja össze. Az ME értéke igen közel áll annak várható értékéhez, így a számított ME a térkép pontosságáról tanúskodik. Szintén ugyanez mondható el az MAE értékéről is, mely 0,3%-os átlagos abszolút hibát jelez, mely a gyakorlati alkalmazásokat tekintve alacsonynak mondható. Az RMNSE értéke kissé magasabb, mint a várható értéke.

VI.15. táblázat
Az előszállási szervesanyag-tartalom térkép értékelése

ME	MAE	RMNSE
0,039	0,319	1,332

Rövidítések: ME: átlagos hiba, MAE: átlagos abszolút hiba, RMNSE: az átlagos négyzetgyök hiba standardizált értéke.

VI. 3. 3. Zala megye

A feltáró alapadat elemzésen átesett adatsort használtam fel Zala megye feltalajainak szervesanyag-tartalom térképezéséhez. Az adatok pozitív ferde eloszlást mutattak, melynek hátterében a megye területén található igen magas szervesanyag-tartalmú lápos talajok állnak. A box-plot kiugró értékeként azonosította ezeket az adatokat (5. melléklet), ugyanakkor a megye szervesanyag-tartalom „profiljának” kiemelt jelentőségű tagjai. Ezért az adatok normál érték (*normal score*) transzformációját végeztem el, mely transzformáció jól bevált módszer a geostatistikában (például: Goovearts, 1997; Geiger, 2006b, 2012; Wágenhoffer, 2013). A transzformáció eredményeként az adatsor standard normál eloszlást mutatott.

A térbeli becslésben felhasználandó folytonos segédinformációkon főkomponens analízist végeztem és az eredményül kapott főkomponenseket használtam fel a többszörös lineáris regresszió analízisben, ahol a függő változót a transzformált szervesanyag-tartalom adatsor jelentette. A független változókat „stepwise” módszerrel választottam ki. A regresszió

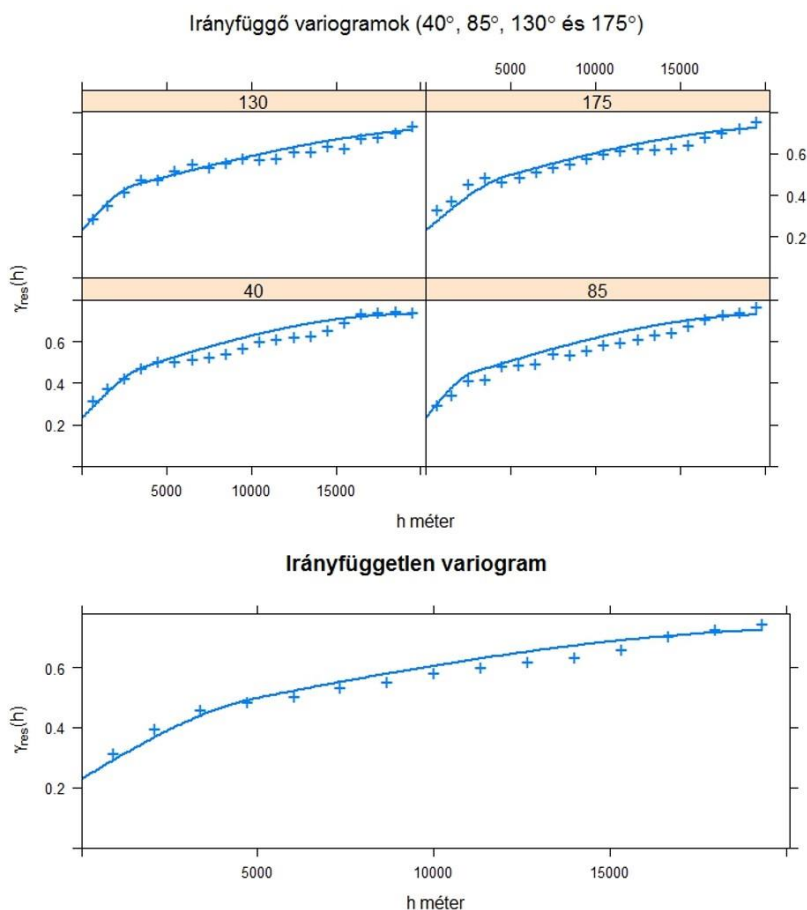
analízis eredményét a VI.16. táblázat mutatja be. A „stepwise” módszer összesen 12 független változót vont be a regressziós modellbe: 8 főkomponenst és 4 talajtérképezési egységet. A talajtérképezési egységek közül a „tőzeges talajok”, az „időszakosan vízállásos, vízjárta területek”, illetve a „jó víztartó és vízvezetőképességű talajok” esetén a becsült regressziós együttthatók értékei pozitív előjelűek. Ebből arra következtettem, hogy e talajtérképezési egységek alatt magasabb a talajok szervesanyag-tartalma. A „tőzeges talajok”, illetve az „időszakosan vízállásos, vízjárta területek” térképezési egységek talajai szerves anyagban igen gazdagok, melynek hátterében: (1) a „tőzeges talajok” esetén a láposodás, mint talajképződési folyamat, illetve (2) a réti jellegű talajok esetén talajképződési folyamatokat kísérő vízhatás áll. Míg az „igen nagy vízvezetőképességű, gyengén víztartó talajok” fizikai félesége homok, addig a „jó víztartó és vízvezetőképességű talajok” fizikai félesége vályog. A talajok fizikai félesége döntő befolyással van a talaj vízháztartására, a beszivárgásra, a kilúgozás intenzitására, a nedvességtartalomra és ezeken keresztül a humuszképződés folyamatára (Szatmári et al., 2013). A modellbe vont 1. főkomponens (SPC-1) kialakításában a lejtőszög játszott döntő szerepet. A 6. és 7. főkomponensre (SPC-6 és SPC-7) a profil és planáris görbület volt nagy hatással. A 9. főkomponens (SPC-9) kialakításában az LS faktor játszott fontos szerepet. A 10., 11. és 12. főkomponensre (SPC-10, SPC-11 és SPC-12) a topográfiai nedvesség index, a völgyek mélysége, illetve a tengerszint feletti magasság volt nagy hatással; míg a 14. főkomponensre (SPC-14) a relatív lejtő pozíció gyakorolt érdemi hatást. E morfológiai paraméterek szorosan kapcsolódnak a megye talajait sújtó vízerózió kiváltó és befolyásoló tényezőihez. A topográfiai nedvesség index és a völgyek mélysége jól reprezentálják a megye azon területeit, ahol a talajvízszint közelsége, illetve a felszínen összegyülekező vizek révén a talajképződést vízhatás kíséri. A lejtőszög nagysága fordítottan arányos a humusztartalommal lévén, hogy a meredek lejtésű területek erősen erodáltak (Thyll, 1997). A profil és planáris görbületek, mint morfológiai paraméterek előre jelzik, hogy a domboldalak mely szakaszain tud kialakulni felszíni vízmozgás, mely révén elszállítódik a szerves anyagban gazdagabb feltalaj. A regressziós modell determinációs együttthatója (R^2) közel 30% (VI.16. táblázat). Ez az érték jóval alacsonyabb, mint Szálka, illetve Előszállás esetén. Ennek hátterében a segédinformációk és a szervesanyag-tartalom adatok közötti nem lineáris kapcsolat, illetve a talajképződési folyamatok nagyfokú térbeli változékonysága és komplexitása áll. Az illesztett többszörös lineáris regressziós modell ezért „leegyszerűsítette” a segédinformációk és a szervesanyag-tartalom adatok közötti kapcsolatot.

VI.16. táblázat
Zala megyére végzett többszörös lineáris regresszió analízis eredménye

Független változók	Reg. együtthatók	p
SPC-1	-0,053	$8,84 \cdot 10^{-10}$
SPC-6	0,069	0,001
SPC-7	0,080	0,002
SPC-9	-0,074	0,017
SPC-10	0,166	$1,59 \cdot 10^{-6}$
SPC-11	-0,081	0,045
SPC-12	-0,272	$7,76 \cdot 10^{-9}$
SPC-14	0,272	$3,00 \cdot 10^{-4}$
Jó víztartó és vízvezetőképességű talajok	0,201	$1,68 \cdot 10^{-5}$
Igen nagy vízvezetőképességű, gyengén víztartó talajok	-0,693	$4,05 \cdot 10^{-14}$
Tőzeges talajok	2,368	$< 2 \cdot 10^{-16}$
Időszakosan vízállásos, vízjárta területek	1,178	$< 2 \cdot 10^{-16}$
Ordinátatengely-metszet	-0,067	0,005
Determinációs együttható (R^2):	0,28	
Modellre számított p:	$2,20 \cdot 10^{-16}$	

Az illesztett regressziós modell segítségével származtattam a regressziós reziduumokat, melyeket egy stacionárius valószínűségi függvényként modelleztem. A reziduumok irányfüggő és irányfüggetlen variogramjait a VI.8. ábra szemlélteti. Mind az irányfüggő, mind az irányfüggetlen variogramokon jól megfigyelhető, hogy a reziduumok térbeli változékonyságára összetett variogram struktúra illeszthető, mely bevált gyakorlat a geostatistikában (például: Goovaerts, 1997; Webster, Oliver, 2007; Geiger et al. 2016). Az első (belső) struktúra típusa szférikus, a rész-tető értéke 0,15, a fő folytonossági irányban (85°) a hatástávolság 4.900 méter, az anizotrópia hányados 0,56. A második (külső) struktúra típusa szintén szférikus, a rész-tető értéke 0,35, a fő folytonossági irányban (130°) a hatástávolság 24.000 méter, az anizotrópia hányados 0,8. Az illesztett modell röghatása 0,23. A belső struktúra észak-déli irányú fő

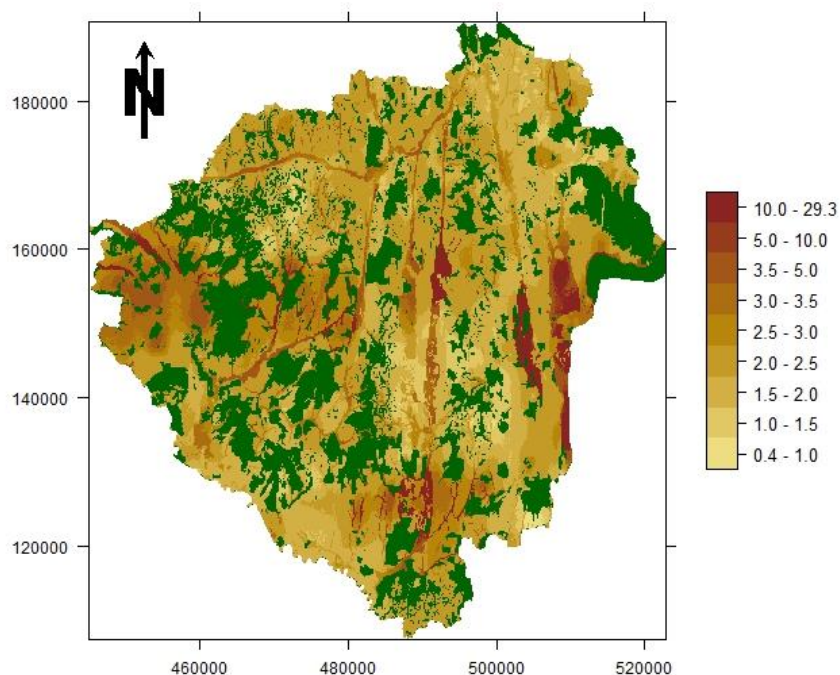
folytonosságot mutat, melynek hátterében a megyét „felszelő” meridionális völgyek – mint az Alsó-Zala-völgy, illetve a Principális-völgy – és a völgyek közötti domboldalak észak-déli irányú elhelyezkedése áll. A külső struktúra fő folytonossága északnyugat-délkeleti irányultságot mutat, mely párhuzamos a megye éves csapadék eloszlásával. Az éves csapadék mennyiség direkt módon hatással van a talajok vízforgalmára, illetve a talajképződési folyamatokra és ez által a talajban található szerves anyag mennyiségére és minőségére.



VI.8. ábra

A Zala megyei reziduumok irányfüggő és irányfüggetlen variogramjai („+”) és az illesztett variogram modell (folytonos vonal)

A többszörös lineáris regresszió és az illesztett variogram modell segítségével elvégeztem a térbeli becslést. A regresszió krigeléssel kapott pontszerű térbeli becslést és a kapcsolódó becslési varianciát a VI.9. ábra és a 8. melléklet mutatja be. Zala megye kartografált szervesanyag-tartalom térképét a 8. melléklet tartalmazza. Az előállított térképekből „kimaszkoltam” a Kreybig-féle térképszelvények szerinti erdővel borított területeket, illetve a Balatont.



VI.9. ábra

*Zala megye szervesanyag-tartalmának térbeli becslése regresszió krigeléssel
(Megjegyzés: zöld szín jelöli a Digitális Kreybig Talajinformációs Rendszer szerint erdővel borított területeket, illetve a Balatont)*

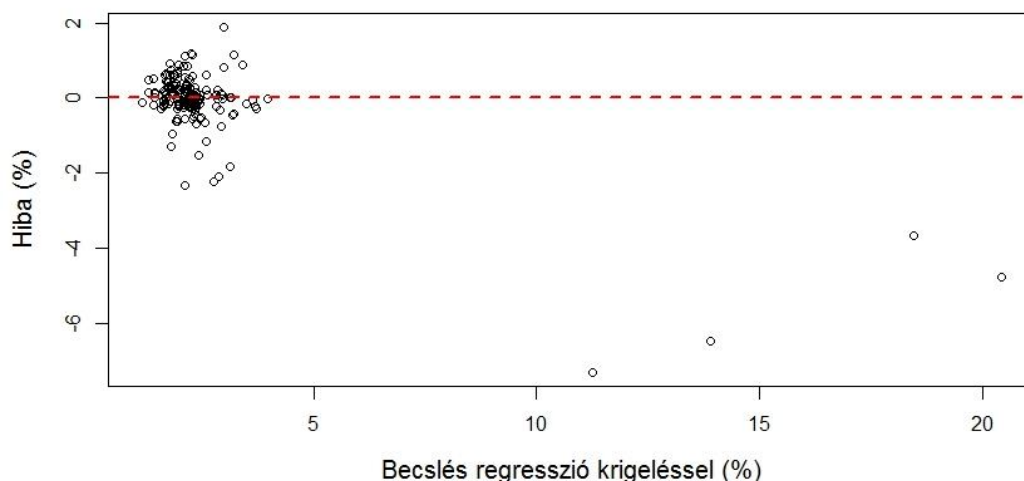
A kartografált térkép (8. melléklet) jól reprezentálja a megyére jellemző „kontrasztos” mintázatot, melyet az adatponti értékek is jeleztek. Jól megfigyelhetők a réti és lápos talajok kiugróan magas szerves anyaga a meridionális völgyekben, illetve a Felső-Zala-völgyben, ahol a láposodás – mint talajképződési folyamat –, továbbá a réti talajokra jellemző túlnedvesedés okozta levegőtlenesség következtében a talajba kerülő szerves anyag mineralizációja akadályozott. A legmagasabb szervesanyag-tartalom értékek a Kis-Balaton medence területére esnek, ahol a – különböző vastagságú tőzegrétegen képződött – síkláptalajokat találjuk. A térkép szerint e területeken a szervesanyag-tartalom meghaladja a 20%-ot is, mely megfelel az adatponti értékeknek (VI.2.3. fejezet), illetve a szakirodalmi adatoknak (Stefanovits et al., 2010). Ezek mellett a Kerka-vidék alacsony lejtésű völgytalpi területein is megfigyelhető a vízhatás okozta magasabb humusztartalom, melyet az adatponti értékek szintén előre jeleztek (VI.2.3. fejezet). Az előzőekkel „kontrasztban” jól megfigyelhetők a völgyhálózatokkal mozaikosan „feldarabolt” domboldalak, ahol a humusszal közepesen, illetve szegényen ellátott barna erdőtalajokat találjuk. A legalacsonyabb szervesanyag-tartalmú területek a meredek és hosszú domboldalakon találhatók, ahol a vízerózió következtében áthalmozódó talajmennyiséggel együtt a szerves anyag is távozik a domboldalakról.

VI.17. táblázat

A Zala megyei szervesanyag-tartalom térkép értékelése

ME	MAE	RMNSE
-0,095	0,437	0,998

Rövidítések: ME: átlagos hiba, MAE: átlagos abszolút hiba, RMNSE: az átlagos négyzetgyök hiba standardizált értéke.



VI.10. ábra

A Zala megyei kontroll pontokba adott regresszió krigeléses becslés és a kiszámított hibák szórásdiagramja

Zala megye szervesanyag-tartalom térképének az értékelését 200 független kontroll pont segítségével végeztem el, melyek mindvégig függetlenek maradtak a geostatistikai modellalkotástól. A hibák alapján kiszámítottam az átlagos hibát (ME), az átlagos abszolút hibát (MAE), illetve az átlagos négyzetgyök hiba standardizált értékét (RMNSE). Az eredményeket a VI.17. táblázat foglalja össze. A kiszámított ME és RMNSE értéke relatíve közel áll azok várható értékéhez. Az MAE értéke magasabb, mint a korábbi két térkép esetében (VI.13. és VI.15. táblázat). Megvizsgáltam a térkép feltételes torzítatlanságát (*conditional unbiasedness*) lévén, hogy ez esetben kellően nagyszámú kontroll pont állt a rendelkezésemre. Ezért előállítottam a becslési értékek és a kiszámított hibák szórásdiagramját (VI.10. ábra). Ideális esetben a hibák az $y = 0$ egyenes mentén „szóródnak” (Isaaks, Srivastava, 1989). Az ábra alapján a magas szervesanyag-tartalmú becslések abszolút értelemben vett hibája magasabb, mint az alacsony értékű becslések esetén. A hibák előjeléből az is következik, hogy a magas szervesanyag-tartalmú becslések jelentősen alulbecslik a tényleges szervesanyag-tartalmat. Ezért újra kiszámítottam az ME és az MAE értékét úgy, hogy kizártam a VI.10. ábrán jól elkülönülő 4 kontroll pontot. A kiszámított hiba mértékek: 0,017 és 0,331. Ezek jelentősen alacsonyabbak a VI.17. táblázatban közölt értékeknél, vagyis a kontroll pontok alapján az

alacsonyabb humusztartalmú becslések hibája jelentősen alacsonyabb, mint a magasabb szervesanyag-tartalmú becsléseké. Ennek hátterében a talajokban raktározott szerves anyag mennyiségét befolyásoló talajképződési folyamatok komplexitása és kölcsönös egymásra hatása áll, melyet a trend leválasztásra alkalmazott többszörös lineáris regresszió „leegyszerűsített”.

VI. 4. A térbeli bizonytalanság modellezése és értékelése az előszállási szántóföldeken

Az előszállási kutatási területre korábban felépített regresszió krigelési modell (VI.3.2. fejezet) jelentette az alapját a szekvenciális sztochasztikus szimulációnak, melynek segítségével 100 alternatív, egyenlően valószínű realizációt generáltam (9. melléklet). A realizációk segítségével vizuálisan és kvantitatívan is értékelhető a térbeli bizonytalanság (Goovaerts, 1997); vagyis a realizációk közötti különbség az együttes térbeli bizonytalanságot jeleníti meg, mely egyben lehetőséget is ad annak mérésére (Geiger, 2006b). A szekvenciális szimulációval kapcsolatban kiemelendő, hogy az algoritmus a megismerés folyamatát modellezi azzal, hogy minden egyes grid pont szimulációja a korábbi eredmények ismeretében történik (V.2.6. fejezet). Ezért az egyes realizációk a valóság „ilyen is lehet” megjelenítését adják (Geiger, 2006b). Elvárjuk, hogy a sztochasztikus szimuláció reprodukálja a modell statisztikákat (Goovaerts, 1997). Ennek vizsgálatára kiszámítottam az egyes realizációk reziduális variogramjait, melyeket a 9. melléklet tartalmazza. A származtatott tapasztalati variogramok jól visszaadják a szimuláció során alkalmazott variogram modellt (VI.6. ábra). Ugyanakkor megfigyelhető, hogy bizonyos realizációk „kevésbé” jól adják vissza az eredeti modellt (például: 12., 41. és 88. realizáció). Ennek hátterében az ergodikus fluktuációk (*ergodic fluctuations*) állnak (Goovaerts, 1997; Deutsch, Journel, 1998). A szekvenciális szimuláció(k)nak egyik fontos tulajdonsága, hogy a variogramot csak várható értékében reprodukálja, vagyis a „sok” realizációnak az átlaga fogja reprodukálni az alkalmazott variogram modellt (Goovaerts, 1997). Ez okozza, hogy nagyobb fluktuációkat is találhatunk az egyes realizációra kiszámított variogramokban.

A realizációk közötti különbségek a szervesanyag-tartalom térbeli leképezhetőségének a bizonytalanságát fejezik ki. A 9. mellékletben bemutatott 100 realizáció alapján elmondható, hogy a mintaterület alacsony lejtésű – kvázi sík – területein a szimulált értékek „alig” különböznek egymástól. Ugyanakkor megfigyelhető, hogy a területet felszelő két löszvölgy mentén az egyes realizációk közötti különbség sokkal szembetűnőbb, mely a löszvölgyek

meredek oldalaival magyarázható. Ezek a megállapítások a rendelkezésre álló adatok térbeli kiterjesztésének bizonytalanságával függenek össze (Geiger, 2006b).

Az előállított 100 sztochasztikus kép 100 szimulált értéket jelent minden egyes grid pontra vonatkozóan, mely kellően nagy számú ahhoz, hogy megadjuk az egyes grid pontok feletti gyakorisági eloszlást. Ezeket a gyakorisági eloszlásokat felhasználva megválaszolhatók olyan kérdések, melyekre a „hagyományos” krigelési algoritmusok – konstrukciójuk miatt – nem alkalmasak (Geiger, 2006a,b). A gyakorisági eloszlásokat felhasználva előállítottam az egyes grid pontok várható érték típusú becslését, melynek térképét a 9. melléklet tartalmazza. Geiger (2006b) szerint, a realizációk sorozatának várható értéke az a legjellemzőbb térbeli eloszlás lesz, mely a kisléptékű heterogenitást a leginkább megjeleníti. Ennek értelmében összehasonlítva a regresszió krigeléssel kapott eredmény térképet (7. melléklet), illetve a várható érték típusú becslésből szerkesztett térképet (9. melléklet) elmondható, hogy utóbbi részletgazdagabban mutatja be a kisléptékű heterogenitást. Azonban fontos megemlíteni, hogy ennek gyakorlati alkalmazásáról/alkalmazhatóságáról megoszlanak a vélemények. A szénhidrogén tárolók háromfázisú dinamikus modellezése kapcsán az utóbbi megközelítés már bizonyított (Geiger, 2006b), ugyanakkor a digitális talajtérképezésben többnyire elhatárolódnak a megközelítéstől. A digitális talajtérképezésben a területi lehatárolásokat – és az ezzel kapcsolatos döntéseket – szerencsésebbnek látják a „krigelt” térképen megtenni (T. Hengl és G.B.M. Heuvelink szóbeli közlése). Szintén előállítottam a várható érték típusú becslés 95%-os konfidencia intervallumának szélességét, melynek térképét a 9. melléklet tartalmazza. Ahogy a geomatematikában, úgy a geostatisztikában is igaz, hogy annál szűkebb a várható érték körüli konfidencia intervallum, minél kisebb bizonytalanságú („stabilabb”) a várható érték becslése (Geiger, 2006b). E megfontolások alapján a 95%-os konfidencia intervallum szélességének térképi ábrázolásával direkt információt nyerhetünk a várható érték típusú becslés bizonytalanságáról. A térkép összhangban van a 100 realizáció összehasonlítása során tett megállapításaimmal. A várható érték becslése stabilabb a mintaterület alacsony lejtésű – közel sík – területein; ugyanakkor a becslés nagyobb bizonytalanságú a mintaterületen húzódó löszvölgyek mentén. Ezek a területek meredekebb lejtésűek és több esetben is szántóföldi területhasználat alatt állnak, ahol az erózió erőteljesebben tudja kifejteni káros hatásait. Nem utolsósorban ezekre a területekre koncentrálnak a mintaterület réti csernozjom talajai is. Összességében elmondható, hogy ezeken a területeken a szervesanyag-tartalom térbeli leképezhetőségének a bizonytalansága nagyobb, melynek „talajtanos háttérében” a humuszszódás – mint talajképző folyamat – komplexitása és a vele szoros összefüggésben álló

bonyolult, komplex, egymásra kölcsönösen pozitív vagy negatív irányú hatást gyakorló fizikai, kémiai, illetve biológiai folyamatok eredői állnak.

Az egyes grid pontokra megadott gyakorisági eloszlások olyan feladatok megoldását és térképezését is támogatják, mint például: mi annak a valószínűsége, hogy a humusztartalom 2 és 4% közé esik. E feladatok megoldására a „hagyományos” krigelési eljárások alkalmatlanok (V.2.6. fejezet). Ugyanakkor a megadott gyakorisági eloszlások alapján ezek a célspecifikus „lekérdezések” gond nélkül térképezhetők. A 9. melléklet négy valószínűség térképet tartalmaz, melyek azt képezik le – a térben –, hogy mi annak a valószínűsége, hogy:

- A szervesanyag-tartalom kisebb vagy egyenlő, mint 2%,
- A szervesanyag-tartalom magasabb, mint 2%, de kisebb vagy egyenlő, mint 3%,
- A humusztartalom magasabb, mint 3%, de kisebb vagy egyenlő, mint 4%,
- A humusztartalom magasabb, mint 4%.

A térképeken jól megfigyelhető a feltalaj szerves anyagának térbeli változékonysága. A 2%-nál alacsonyabb humusztartalmú talajok többnyire a vízerózióval leginkább sújtott területeken találhatók. Jól nyomon követhetők a réteg erózió okozta szervesanyag-tartalom csökkenések, melyek többnyire földes kopár talajok formájában jelentkeznek a területen. A kutatási terület közepén húzódó löszvölgy mentén jól kivehetők a vonalas erózió formái. Szintén megfigyelhető azon drasztikus szervesanyag-tartalom csökkenés, melynek háttérében a löszvölgyek meredek völgyekbe hajló peremeinek „túlszántása” áll. A második és harmadik térképen jól nyomon követhető a szervesanyag-tartalom délnyugat-északkeleti irányú markáns csökkenése, melyet az adatponti értékek is előre jeleztek. A második térképen jól kirajzolódnak a középső löszvölgyhöz „becsatlakozó” vonalas erózió nyomai. A harmadik és negyedik térkép alapján elmondható, hogy a legmagasabb humusztartalmú talajok a mintaterület délnyugati határa mentén húzódó löszplaton található. Az utolsó térképen jól megfigyelhetők az erdővel fedett területek, ahol a szervesanyag-tartalom meghaladja a 4%-ot.

VI. 5. A mintavétel optimalizációs scenáriók és értékelésük

Ebben a fejezetben a beállított mintavétel optimalizációs scenáriók bemutatására és értékelésére kerül sor. Az 1., 2., 3., és 4. *scenárió* célja, hogy az előszállási mintaterületen optimalizáljam a mintavételi helyeket a feltalaj szervesanyag-tartalmának térképezése céljából eltérő mintavételi megszorítások mellett. Az 5. és 6. *scenárió* célja a többváltozósra kiterjesztett optimalizációs módszertan tesztelése eltérő mintavételi megszorítások mellett.

Ebben az esetben a mintavételezés és a térképezés célja a szervesanyag-tartalom, illetve a termőréteg vastagság együttes térképezése. Ezért elvárt, hogy az 5. és 6. *szcenárió* során eredményül kapott mintavételi elrendezések optimálisak legyenek *mindkét* talajtulajdonság térképezésére az adott mintavételi megszorítások mellett. A beállított mintavétel optimalizációs *szcenáriókat*, illetve a választott minőségi mértékeket a VI.18. táblázat foglalja össze. A táblázat szintén tartalmazza a 7. és 8. *szcenárió*t, melyek a következő fejezetben (VI.6. fejezet) kerülnek bemutatásra. A választott minőségi mérték meghatározásának az alapját a korábbi fejezetben (VI.3.2. fejezet) bemutatott regresszió krigelési modell jelentette; vagyis felhasználtam: (1) a szervesanyag-tartalomra és a segédinformációkra illesztett regresszió struktúráját, illetve (2) a reziduumok variogram modelljét.

VI.18. táblázat
A mintavétel optimalizációs scenáriók táblázatos összefoglalása

	Mintavételezés célja	Térbeli becslő eljárás	Minőségi mérték	Mintavételi megszorítás*				
				1	2	3	4	5
1. scenárió	SOM	RK	RKV	★			★	★
2. scenárió	SOM	RK	RKV		★		★	★
3. scenárió	SOM	RK	RKV	★		★	★	★
4. scenárió	SOM	RK	RKV		★	★	★	★
5. scenárió	SOM + RD	RK	K-RKV	★			★	★
6. scenárió	SOM + RD	RK	K-RKV	★		★	★	★
7. scenárió	SOM	OK	OKV	★			★	
8. scenárió	SOM	MLR	REV	★			★	★

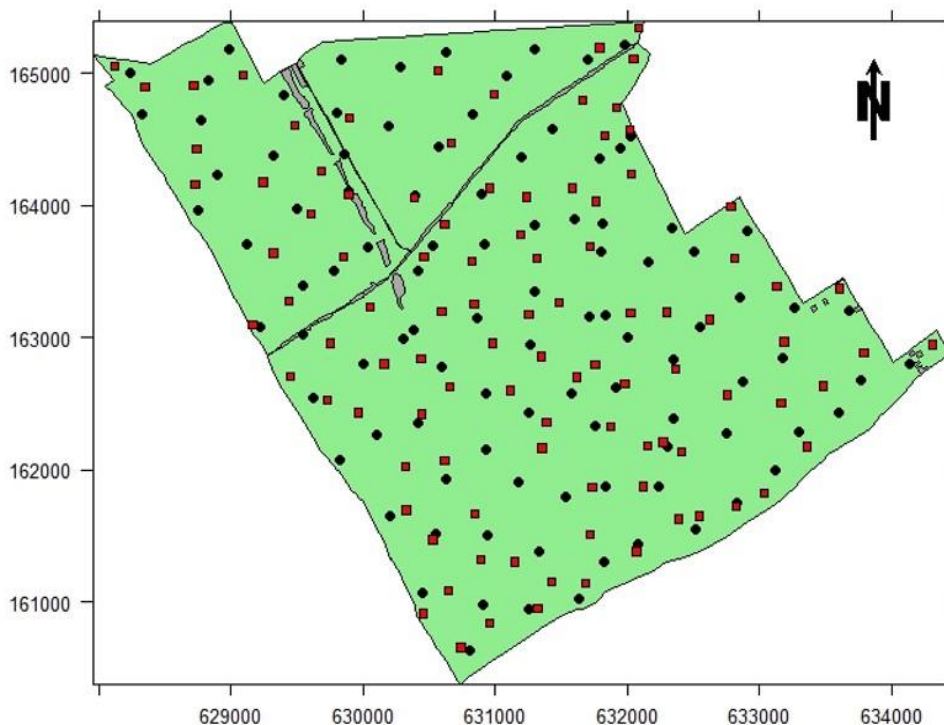
* Mintavételi megszorítások: 1: vételezhető minták száma, 2: előre definiált (előzetes) minőségi mérték, 3: korábbi mintavételezés(ek) figyelembe vétele, 4: mintavételezés és térképezés szempontjából irreleváns területek (nyílt vizek, mocsarak, épületek, utak stb.) kihagyása és 5: segédinformációk figyelembe vétele.

Rövidítések: SOM: szervesanyag-tartalom, RD: termőréteg vastagság, RK: regresszió krigelés, OK: hagyományos krigelés, MLR: többváltozós lineáris regresszió, RKV: regresszió krigelés becslési varianciája, K-RKV: többváltozós optimalizációhoz definiált képzetes minőségi mérték, OK: hagyományos krigelés becslési varianciája és REV: regresszió hiba varianciája.

VI. 5. 1. Egyváltozós scenáriók: 1. és 2.

Az 1. és 2. *szcenárió* esetén megkötés volt, hogy a mintavételezés, illetve térképezés szempontjából irreleváns (vagy „hozzáférhetetlen”) területekre nem eshetnek mintavételi

pontok. Ilyen területek voltak: (1) nyílt vizek, (2) mocsarak, illetve (3) mesterséges felszínek (IV.2. ábra). További mintavételi megszorítás volt a rendelkezésemre álló segédinformációk figyelembe vétele.

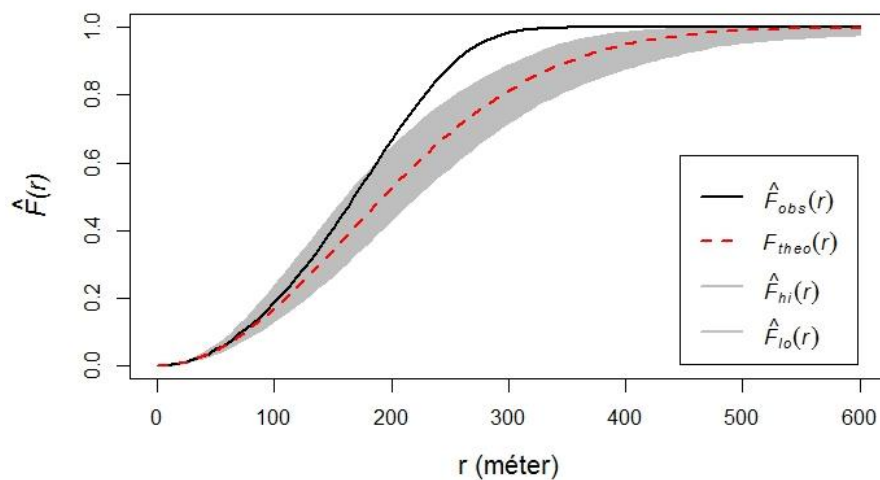


VI.11. ábra

*Az 1. scenárióra optimalizált mintavételi konfiguráció (fekete pontok), illetve az előszállási adatpontok (piros négyzetek)
(Megjegyzés: az ábra szürke színnel jelöli a mintavételezés és térképezés szempontjából irreleváns területeket)*

Az 1. scenárió során a vételezhető minták száma adott volt. Az SSA algoritmus által optimalizált mintavételi konfigurációt a VI.11. ábra mutatja be. Az ábrán szintén jelöltem a geostatistikai modellezésekhez felhasznált eredeti adatpontokat (továbbiakban: eredeti mintavételezés). A két mintavételi elrendezést összehasonlítva elmondható, hogy az SSA algoritmussal optimalizált mintavételi konfiguráció jobb térbeli lefedettséget adott, mint az eredeti mintavételezés (VI.11. ábra). A feltáró alapadat elemzés (VI.2.2. fejezet) során megállapítottam, hogy a kutatási terület északi részét az eredeti mintavételezés kissé alulreprezentálta. Ezzel szemben az 1. scenárióra optimalizált konfiguráció egyenletesebb térbeli lefedettséget biztosított. Az optimalizált konfiguráció e mellett egységesebben fedte le a határmenti régiókat is, ami a mintaterület északi és délnyugati határain a legszembetűnőbb. Ennek megkülönböztetett jelentősége van a térbeli becslések szempontjából, hiszen e területeken kell a leginkább extrapolációra számítanunk (Journel, Rossi, 1989; Kovács, Szanyi,

2005; Webster, Oliver, 2007). A mintaterületet „felszelő” löszvölgyek mentén megfigyelhető, hogy az optimalizált mintavételi pontok közelebb helyezkednek el egymáshoz, mint a kvázi síkvidéki területeken (VI.11. ábra). Ennek háttérében a mintaterület karakterisztikája, illetve a választott minőségi mérték kettős tulajdonsága áll. A mintaterület jobbára homogénnek mondható a domborzati, illetve a felszínfedettség kategóriák tekintetében (Szatmári, Barta, 2012, 2013). A célfüggvény szimultán keresi a földrajzi és attribútum tér optimális lefedettségét. Ezért az SSA algoritmus a löszvölgyek területén – melyek a legváltozatosabbak – a mintavételi pontok „besűrítésével” éri el, hogy optimálisan fedje le a trendmentesítéshez szükséges attribútum teret.

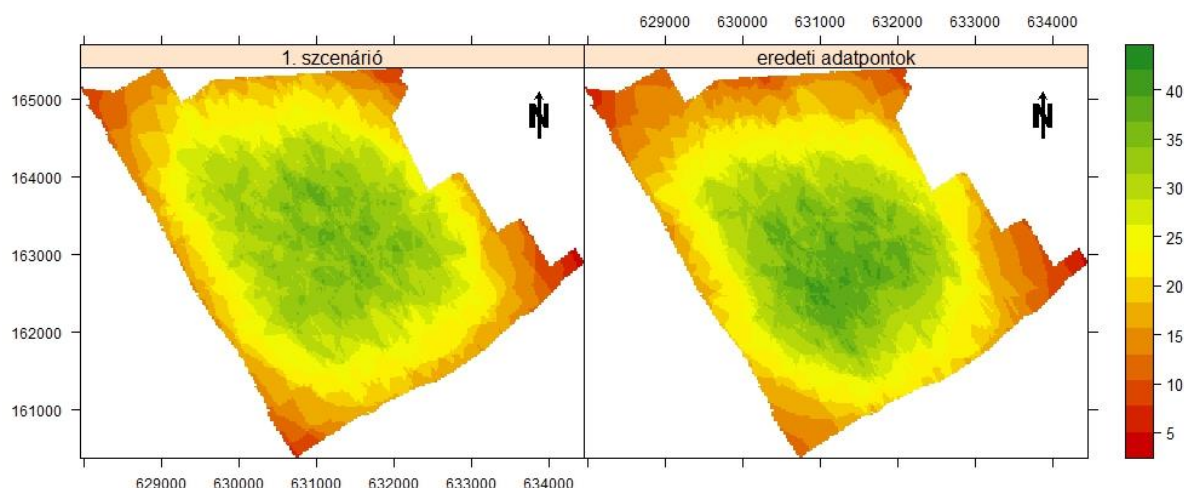


VI.12. ábra

Az 1. scenárióra optimalizált mintavételi konfiguráció tapasztalati üres-tér függvénye (fekete vonal) és a Poisson folyamatra (piros szaggatott vonal) szimulált 95%-os konfidencia sáv (szürke színnel)

Az $F(r)$ üres-tér függvény algoritmussal kapott eredményt a VI.12. ábra mutatja be. A tapasztalati függvény meredekebb lefutású, mint az elméleti függvény, melyből arra következtettem, hogy az optimalizált mintavételi pontok nem függetlenek egymástól, közöttük kizáró kapcsolat áll fent. E szerint az optimalizált mintavételi pontok térbeli elhelyezkedését leginkább az alkalmazott variogram modell határozta meg (Heuvelink et al., 2007). Az $\hat{F}(r)$ függvény alapján elmondható, hogy 350 méteren belül a becslési grid (VI.1. fejezet) minden pontjának legalább egy krigelési szomszédja van. Felhasználva a variogram hatásterületi ellipszisést előállítottam az eredeti mintavételezés, illetve az optimalizált mintavételezés alapján a becslési pontok krigelési szomszédság térképeit (VI.13. ábra). A térképek az egyes becslési pontokra adják meg a krigelési szomszédok számát. A két térképeket összehasonlítva elmondható, hogy az 1. scenárióra optimalizált konfiguráció egyenletesebben fedi le a mintaterületet, mint az eredeti mintavételezés (VI.13. ábra). Ideális esetben minden becslési

pontnak a szomszédságában legalább egy olyan mintavételi pont van, melynek távolsága nem nagyobb, mint az alkalmazott variogram modell hatástávolsága (Füst, Geiger, 2010). Az optimalizált konfigurációra elkészített térkép szerint minden egyes becslési pontnak legalább 5 krigelési szomszédja van (VI.13. ábra). Ebből azt a következtetést vontam le, hogy az *1. scenárióra* optimalizált mintavételi konfiguráció megfelelően fedi le a kutatási terület által meghatározott földrajzi teret.



VI.13. ábra

Az 1. scenárióra optimalizált mintavételi konfiguráció (bal oldalon) és az eredeti adatpontok (jobb oldalon) krigelési szomszédság térképe az alkalmazott anizotróp variogram modell alapján

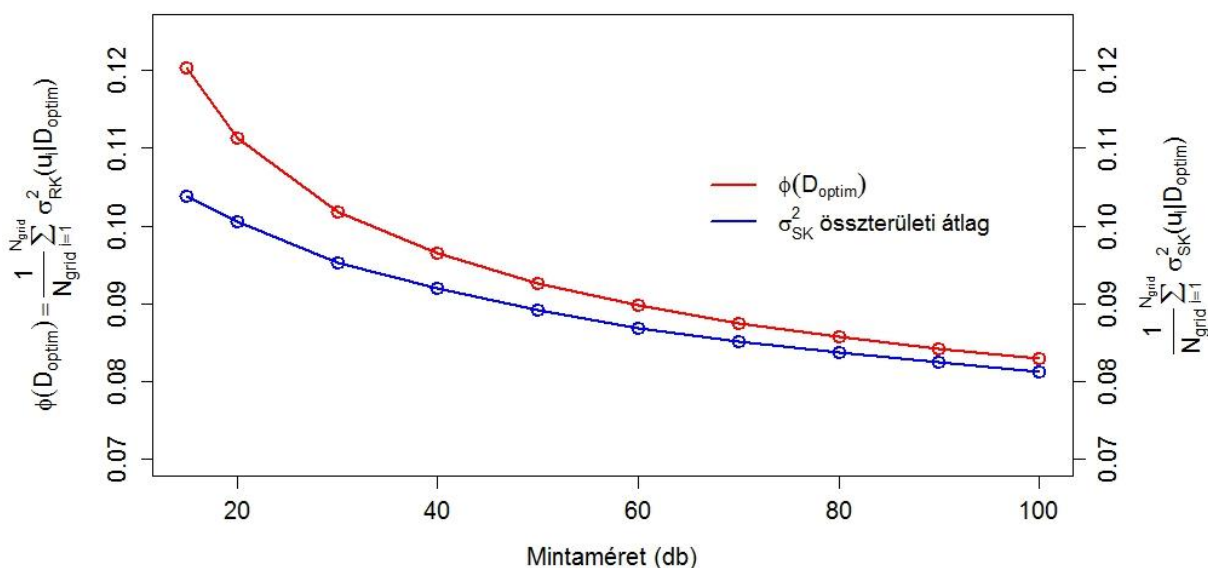
VI.19. táblázat

Az 1. scenárióra végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei (p)

Mintaméret (db)	Segédinformációk				
	SPC-1	SPC-2	SPC-3	SPC-10	SPC-13
p	0,375	0,559	0,879	0,415	0,131

A szervesanyag-tartalom térbeli becslése szempontjából releváns segédinformációk lefedettségét Kolmogorov-Smirnov próbával vizsgáltam. A próba nullhipotézise szerint a mintavételi pontokra eső folytonos segédinformációk értékeinek eloszlása, illetve a segédinformációk rasztereinek eloszlása ugyanazon folytonos eloszlásból származik. A választott szignifikancia szint 5% volt. A próba eredményeit a VI.19. táblázat foglalja össze. A statisztikai próba eredményei szerint egyik segédinformáció esetén sem lehet visszautasítani a próba nullhipotézisét, melyből azt a következtetést vontam le, hogy az *1. scenárióra* optimalizált konfiguráció teljesen lefedi a releváns segédinformációkat. A fenti eredményeket összegezve elmondható, hogy az *1. scenárióra* optimalizált mintavételi konfiguráció

megfelelően fedi le a szervesanyag-tartalom térképezése szempontjából releváns segédinformációkat és a mintaterület által meghatározott földrajzi teret.



VI.14. ábra

A 2. szcenárióra szerkesztett kalibrációs görbe (piros vonal), illetve az egyszerű krigelés becslési varianciájának összterületi átlaga (kék vonal)

A 2. szcenárió esetén a cél az volt, hogy előállítsam azt az ún. kalibrációs görbét, melynek segítségével meghatározható, hogy adott (előzetes) minőségi mérték eléréséhez mekkora mintaméret szükséges. A kalibrációs görbe a mintaméret függvényében ábrázolja az optimalizált mintavételekre kiszámolt célfüggvény értékét, vagyis a $\sigma_{\text{RK}}^2(\mathbf{u})$ becslési variancia összterületi átlagát. A kalibrációs görbe analitikus meghatározására nem törekedhettem lévén, hogy a választott minőségi mérték számos – és egymással kölcsönösen kapcsolatban álló – módon függ a vételezhető minták számától, a segédinformációk eloszlásától és felbontásától, a mintaterület alakjától és kiterjedésétől, az alkalmazott variogram modelltől és a regressziós modell struktúrájától. Ezért a kalibrációs görbe numerikus megközelítését választottam. A görbe megszerkesztéséhez egy optimalizációs „ütemtervet” készítettem. Az „ütemterv” során szisztematikusan növeltem az optimalizálandó minták méretét és az SSA algoritmussal optimalizáltam ezek térbeli elhelyezkedését. Az optimalizációk során a mintavételi megszorítások (például: mintavételezés számára irreleváns területek), illetve az alapbeállítások (például: stop kritérium, iterációs szám, exponenciális rendszer hűlés) nem változtak, melyekkel az azonos körülményeket kívántam biztosítani. Az optimalizált konfigurációkra kiszámítottam a célfüggvény értékeit és a mintaméret függvényében grafikusán ábrázoltam (VI.14. ábra). A kalibrációs görbe megszerkesztéséhez használt (optimalizált) mintavételi

konfigurációkat, továbbá ezek krigelési szomszédság térképeit a 10. melléklet tartalmazza. A megszerkesztett kalibrációs görbe segítségével meghatározható, hogy egy előre definiált (előzetes) minőségi mérték eléréséhez mekkora mintaméret szükséges; például: ha a várt érték 0,09, akkor a görbe alapján az ehhez szükséges mintavételi pontok száma 60. A kalibrációs görbe „vice versa” alkalmazható. Ugyanakkor fontos szem előtt tartani a minőségi mérték gyakorlati alkalmazásával kapcsolatban tett korábbi megállapításokat (V.3.2. fejezet).

A kalibrációs görbe jellegzetesen exponenciális lefutású (Yfantis et al., 1987). A görbe alapján megadható olyan küszöbérték, melyen túl csak a mintaméret növelésével tovább már effektíve nem csökkenthető a minőségi mérték értéke. A görbe esetén ez az érték 50–60 között található (VI.14. ábra). Minden optimalizált konfigurációra kiszámítottam az egyszerű krigelés $\sigma_{SK}^2(\mathbf{u})$ becslési varianciáját. A trend hiba varianciája a $\sigma_{RK}^2(\mathbf{u})$ és a $\sigma_{SK}^2(\mathbf{u})$ becslési varianciák különbsége (V.23. egyenlet). A kalibrációs görbe mellett ábrázoltam a $\sigma_{SK}^2(\mathbf{u})$ becslési varianciájának összterületi átlagát (VI.14. ábra), mellyel azt vizsgáltam, hogy a választott minőségi mérték miként befolyásolja az optimalizált mintavételi konfigurációkat. A VI.14. ábrán jól látható, hogy a trend hiba varianciája jelentősen lecsökken 70 mintavételi pont után, majd 90 mintavételi pont után csaknem elhanyagolhatóvá válik. Ez az optimalizált konfigurációkon is jól megfigyelhető (10. melléklet). Az első öt optimalizált konfigurációt (15, 20, 30, 40 és 50 mintaméret) egyaránt megszabja a variogram anizotrópiája, illetve a területet „felszelő” löszvölgyek jelentette domborzati változatosság. E szerint az optimalizáció arra „irányul”, hogy a trend hiba varianciájának átlagát és az egyszerű krigelés becslési varianciájának átlagát szimultán csökkentse. Az ezeket követő optimalizált konfigurációk (60, 70, 80, 90 és 100 mintaméret) esetén az anizotróp variogram dominanciája figyelhető meg. Ugyanakkor szintén jól kivehetők azok a mintavételi pontok, melyek kardinális fontosságúak a trend hiba varianciájának hathatós csökkentése céljából. Ezek a löszvölgyek legmeredekebb, leghosszabb és legösszetettebb lejtőire esnek.

A kalibrációs görbe meghatározásához használt, optimalizált mintavételi konfigurációk alapján meghatározható az a minimális mintaméret, mely teljes földrajzi és attribútum térbeli lefedettséget nyújt. Utóbbiak „teljes lefedettsége” alatt azt értem, hogy: (1) minden tervezett becslési pontnak van legalább egy krigelési szomszédja (Füst, Geiger, 2010) és (2) a mintavételi pontok jól reprezentálják a segédinformációk rasztereit. Az attribútum tér lefedettségét Kolmogorov-Smirnov próbával vizsgáltam, melynek eredményeit a VI.20. táblázat foglalja össze. A statisztikai próba során az alkalmazott szignifikancia szint ezúttal is 5% volt. A

táblázat szerint 30 mintavételi pont felett nem utasítható vissza az egyes segédinformációkra a statisztikai próba nullhipotézise. E szerint 30 mintavételi pont az a minimális mintaméret, mellyel már optimálisan lefedhetők a szerves anyag térképezése szempontjából releváns segédinformációk. A krigelési szomszédság térképek (10. melléklet) alapján meghatározható az a minimális mintaszám, mely mellett teljesen lefedhető a földrajzi tér. A szomszédság térképek szerint szintén 30 mintavételi pont elegendő a mintaterület által meghatározott földrajzi tér teljes lefedésére. A fentiek szerint 30 mintavételi az a minimális mintaméret, mely mellett a releváns segédinformációk által definiált attribútum tér, illetve a mintaterület által meghatározott földrajzi tér teljesen lefedhető.

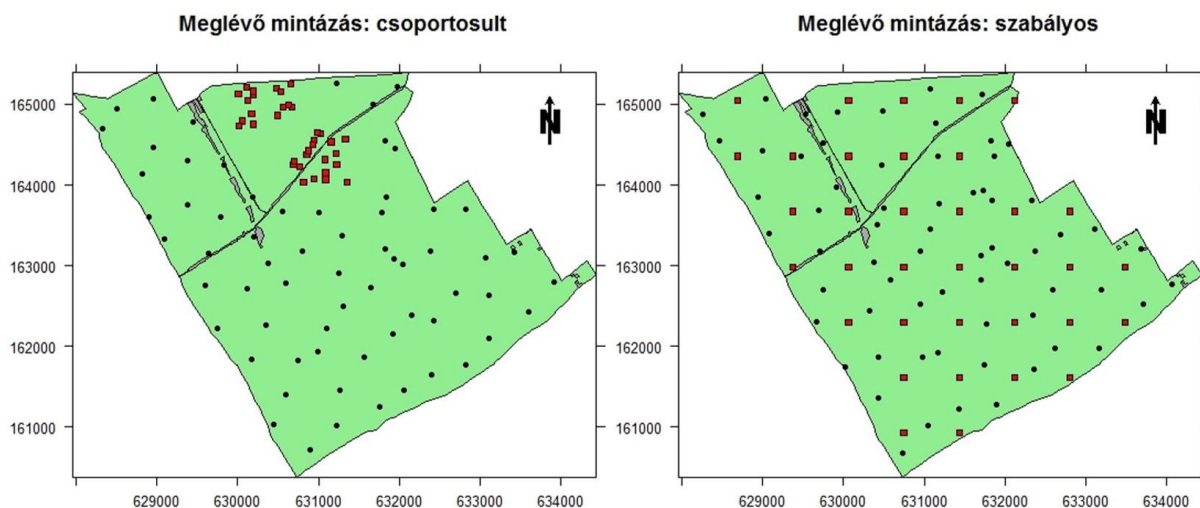
VI.20. táblázat

*A 2. scenárióra végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei
(Megjegyzés: dőlt betűtípussal jelöltem azon értékeket, melyek mellett a statisztikai próba nullhipotézisét visszautasítottam)*

Mintaméret (db)	Segédinformációk				
	SPC-1	SPC-2	SPC-3	SPC-10	SPC-13
15	0,076	0,004	0,291	0,003	0,007
20	0,028	0,030	0,125	0,036	0,022
30	0,181	0,172	0,569	0,210	0,131
40	0,076	0,172	0,125	0,210	0,131
50	0,375	0,172	0,879	0,093	0,131
60	0,375	0,082	0,569	0,210	0,270
70	0,375	0,172	0,569	0,210	0,131
80	0,375	0,172	0,569	0,210	0,270
90	0,660	0,329	0,569	0,093	0,270
100	0,375	0,559	0,879	0,415	0,131

VI. 5. 2. Egyváltozós scenáriók: 3. és 4.

Az előző fejezethez (VI.5.1. fejezet) hasonlóan a 3. és 4. scenárió esetén is megkötés volt, hogy a mintavételezés, illetve térképezés szempontjából irreleváns területekre nem eshetnek mintavételi pontok. További mintavételi megszorítás volt a korábbi mintavételezésekből származó mintavételi pontok figyelembe vétele, mely mintavételezések egy csoportosult, illetve egy szabályos mintavételezést követtek. Mindkét esetben a meglévő mintavételi pontok száma 35 volt.



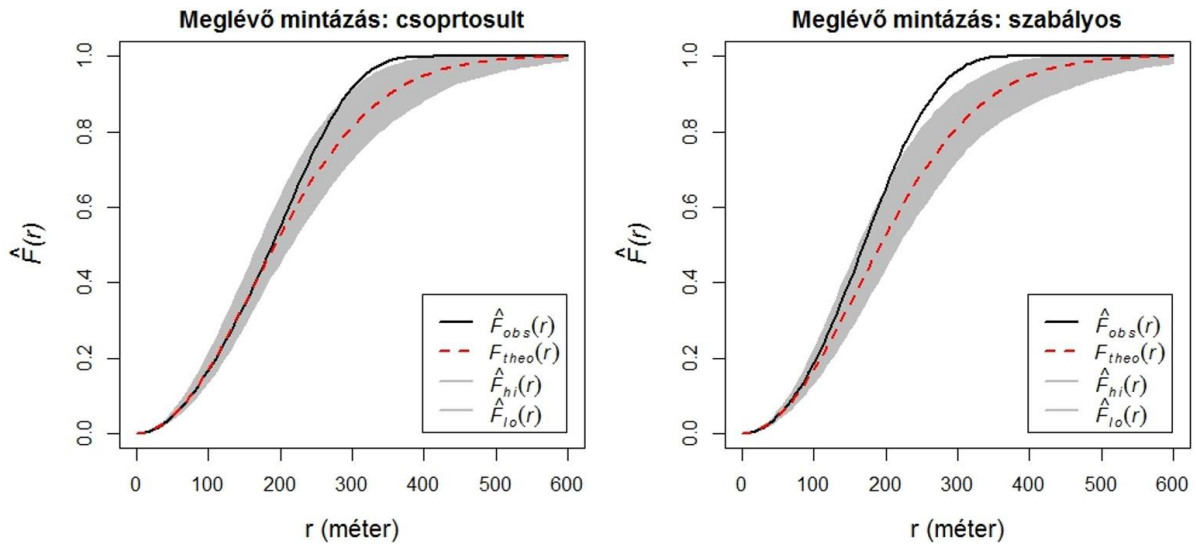
VI.15. ábra

A 3. szcenárióra optimalizált mintavételi konfigurációk (fekete pontok), illetve a meglévő mintavételi pontok (piros négyzetek)

(Megjegyzés: az ábra szürke színnel jelöli a mintavételezés szempontjából irreleváns területeket)

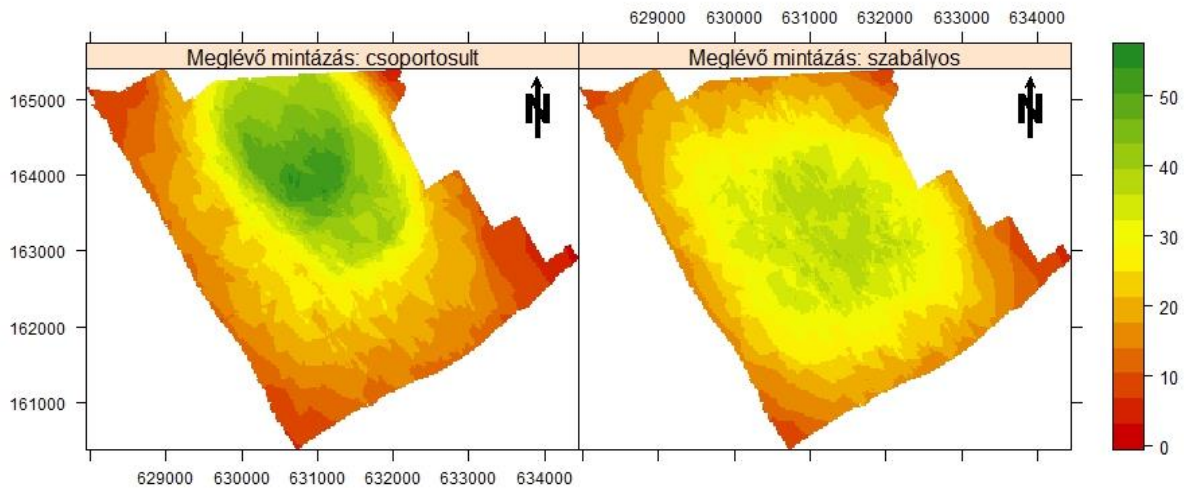
A 3. szcenárió során az optimalizálandó mintavételi pontok száma 65 volt. Az optimalizált mintavételi konfigurációkat a VI.15. ábra mutatja be. A két konfiguráció jelentősen különbözik egymástól, melynek hátterében a meglévő mintavételezések eltérő volta áll. A csoportosult mintavételezés – a maga 35 pontjával – elégtelenül fedi le (1) a kutatási terület által meghatározott földrajzi teret, illetve (2) a szervesanyag-tartalom térképezése szempontjából releváns segédinformációkat (VI.15. ábra). Ezért a csoportosult mintavételezésre kiszámított célfüggvény értéke igen magas volt. Ennek eredményeként a 65 mintavételi pont optimalizációja során az SSA algoritmusnak szimultán kellett csökkentenie a trend hiba varianciáját, illetve az egyszerű krigelés becslési varianciáját. A csoportosult mintavételezéssel ellentétben, a meglévő szabályos mintázás közel teljes földrajzi lefedettséget nyújtott (VI.15. ábra). Továbbá ezek a mintavételi pontok nem csak szántóföldi területhasználatra, illetve síkvidéki területekre esnek. Ezért az SSA algoritmus elsődleges célja az volt, hogy az optimalizálandó 65 mintavételi pont révén lefedje azokat a földrajzi, illetve attribútum térbeli „hiányosságokat”, melyeket a szabályos mintázás „kihagyott”. Ezek a „hiányosságok” jól megfigyelhetők például: a határmenti területeken, illetve a területet „felszelő löszvölgyek” mentén, ahol a domborzat a legváltozatosabb (VI.15. ábra).

Meghatároztam az $F(r)$ üres-tér függvényt az egyes optimalizált konfigurációkra (VI.16. ábra) oly módon, hogy figyelembe vettem a korábbi mintavételezésekből származó mintavételi pontokat is. A tapasztalati üres-tér függvényeket összehasonlítva elmondható, hogy szabályos mintázás esetén a függvény lefutása meredekebb, mint a csoportosult esetén.



VI.16. ábra

A 3. scenárióra optimalizált mintavételi konfigurációk tapasztalati üres-tér függvénye (fekete vonal) és a Poisson folyamatra (piros szaggatott vonal) szimulált 95%-os konfidencia sáv (szürke szín)



VI.17. ábra

A 3. scenárióra optimalizált mintavételi konfigurációk krigelési szomszédság térképei az alkalmazott anizotróp variogram modell alapján

E szerint előbbinél a mintavételi pontok elhelyezkedése „szabályosabb” (Baddeley, Turner, 2005). A krigelési szomszédok térképeit a VI.17. ábra mutatja be. A csoportosult mintavételezés kiegészítésére optimalizált konfiguráció esetében megfigyelhető, hogy a krigelési szomszédok száma egyenlőtlenül alakul a mintaterületen, melynek háttérében a meglévő mintavételezés áll. A szabályos mintavételezés kiegészítésére optimalizált konfiguráció ugyanakkor egységesen fedi le a kutatási területet. Esetében minden egyes becslési ponthoz legalább 5 krigelési szomszéd tartozik (VI.17. ábra).

A szervesanyag-tartalom térbeli becslése szempontjából releváns segédinformációk optimális lefedettségét ezúttal is Kolmogorov-Smirnov próbával vizsgáltam. A statisztikai próba eredményeit a VI.21. táblázat foglalja össze. A megfigyelt szignifikancia szintek alapján egyik segédinformáció esetén sem lehet visszautasítani a próba nullhipotézisét, melyből azt a következtetést vontam le, hogy a 3. *szcenárióra* optimalizált mintavételi konfigurációk jól reprezentálják a releváns segédinformációkat. A fenti eredményeket összegezve elmondható, hogy a 3. *szcenárióra* optimalizált mintavételi konfigurációk megfelelően fedik le a szervesanyag-tartalom térképezése szempontjából releváns segédinformációkat, illetve a mintaterület által meghatározott földrajzi teret.

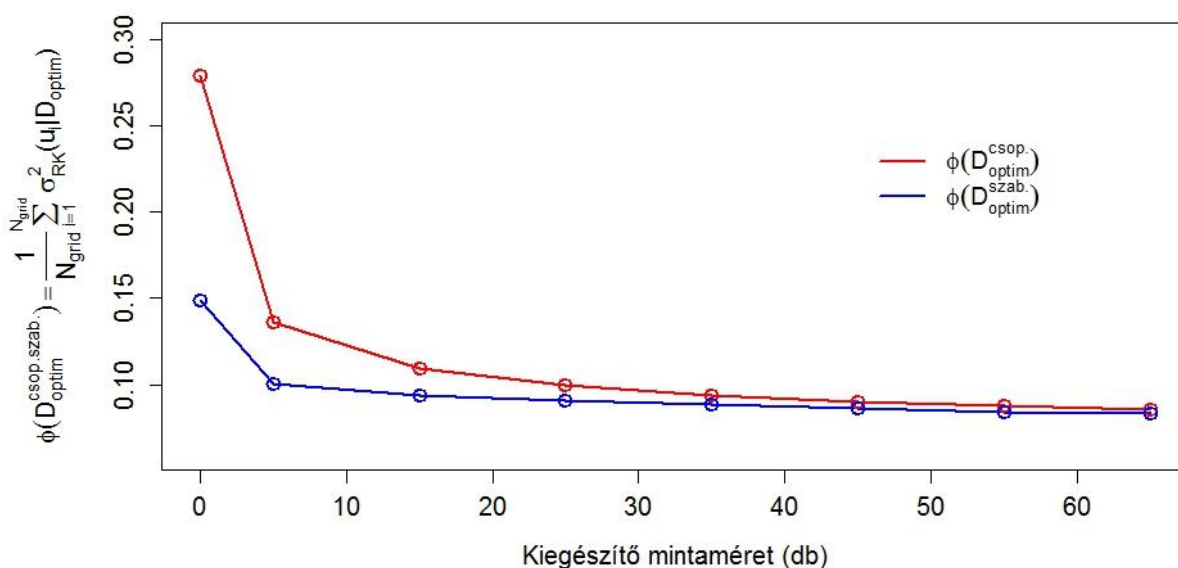
VI.21. táblázat

A 3. scenárióra végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei

Meglévő mintázás	Segédinformációk				
	SPC-1	SPC-2	SPC-3	SPC-10	SPC-13
szabályos	0,660	0,559	0,879	0,210	0,270
csoportosult	0,660	0,329	0,879	0,211	0,057

A 4. *szcenárió* során a meglévő csoportosult, illetve szabályos mintázásokra vonatkozó kalibrációs görbéket szerkesztettem meg. A görbék segítségével – a korábbi mintavételezések figyelembe vétele mellett – meghatározható, hogy adott (előzetes) minőségi mérték eléréséhez mekkora kiegészítő mintaméret szükséges. Kiegészítő mintaméreten a korábban felvételezett (már meglévő) mintavételt kiegészítő, új mintavételi pontok számát értem. A kalibrációs görbéket ez esetben is numerikusan közelítettem. A 4. *szcenárió* esetén is egy optimalizációs „ütemtervet” dolgoztam ki, mely során szisztematikusan növeltem a kiegészítő minták méretét és az SSA algoritmussal optimalizáltam ezek térbeli elhelyezkedését a korábbi mintavételezések figyelembe vétele mellett. Az optimalizációk során a mintavételi megszorítások, illetve az alapbeállítások változatlanok maradtak. Az optimalizált konfigurációkra kiszámítottam a célfüggvény értékeit és a mintaméret függvényében grafikusán ábrázoltam (VI.18. ábra). A kalibrációs görbék megszerkesztéséhez használt (optimalizált) mintavételi konfigurációkat, illetve ezek krigelési szomszédság térképeit a 11. melléklet tartalmazza. A megszerkesztett kalibrációs görbék segítségével meghatározható, hogy egy előre definiált (előzetes) minőségi mérték eléréséhez mekkora kiegészítő mintaméret szükséges a meglévő minták számításba vétele mellett; például: ha a várt érték 0,09, akkor szabályos mintavételezés esetén 25, míg csoportosult mintavételezés esetén 45 kiegészítő mintavételi pont felvétele szükséges (VI.18. ábra). A kalibrációs görbék ez esetben is „vice

versa” alkalmazhatók. A minőségi mérték gyakorlati alkalmazásával kapcsolatban fontos ez esetben is szem előtt tartani a korábbi megállapításokat (V.3.2. fejezet).



VI.18. ábra

A 4. szcenárióra szerkesztett kalibrációs görbék: kék színnel a szabályos, illetve piros színnel a csoportosult mintavételezésre szerkesztett kalibrációs görbe

A két kalibrációs görbét összehasonlítva látható, hogy meglévő csoportosult mintavétel esetén a görbe meredekebb lefutású, mint a szabályos mintavételezés esetén. A görbék kezdő értékei – mikor csakis a korábbi felmérésekből származó mintavételi pontok alapján határoztam meg a célfüggvény értékét – a szabályos mintavételezésre számolt érték csaknem a fele a csoportosultul számoltak (VI.18. ábra). Ennek hátterében a korábban ismertetett okok állnak, miszerint a csoportosult mintavételezés elégtelenül fedte le mind a földrajzi, mind az attribútum teret, míg a szabályos mintavételezés közel teljesen lefedte a földrajzi teret. Ez az optimalizált konfigurációkon is jól nyomon követhető (11. melléklet). A csoportosult mintavételezés esetén az SSA algoritmusnak szimultán kellett csökkentenie a trend hiba varianciáját, illetve az egyszerű krigelés becslési varianciáját, míg a szabályos mintavételezés esetén a kiegészítő mintákkal azon „hiányosságok” lefedésére került sor, melyeket a meglévő mintázás „kihagyott” (11. melléklet). Ezek természetesen hatással voltak a görbék kezdeti „lefutására” is, vagyis csoportosult mintavételezés mellett több kiegészítő mintavételi pont felvétele szükséges, mint szabályos mintavételezés esetén, hogy adott (előzetes) minőségi mértéket elérjünk (VI.18. ábra). Ez a különbség 35 kiegészítő mintavételi pont után elhanyagolható (VI.18. ábra).

VI.22. táblázat

A 4. scenárióra végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei
(Megjegyzés: dőlt betűtípussal jelöltem azon értékeket, melyek mellett a statisztikai próba nullhipotézisét visszautasítottam)

Kiegészítő mintaméret (db)	Meglévő mintázás	Segédinformációk				
		SPC-1	SPC-2	SPC-3	SPC-10	SPC-13
5	szabályos	0,028	0,034	0,569	0,121	0,007
	csoportosult	0,028	0,004	0,290	0,003	0,001
15	szabályos	0,375	0,172	0,291	0,210	0,022
	csoportosult	0,076	0,004	0,569	0,036	0,007
25	szabályos	0,181	0,329	0,879	0,210	0,057
	csoportosult	0,076	0,081	0,569	0,210	0,022
35	szabályos	0,181	0,329	0,879	0,415	0,057
	csoportosult	0,375	0,172	0,879	0,210	0,131
45	szabályos	0,181	0,172	0,879	0,210	0,057
	csoportosult	0,375	0,172	0,879	0,415	0,131
55	szabályos	0,181	0,559	0,879	0,093	0,131
	csoportosult	0,375	0,329	0,879	0,210	0,131
65	szabályos	0,660	0,559	0,879	0,210	0,270
	csoportosult	0,660	0,329	0,879	0,211	0,057

A kalibrációs görbék meghatározásához használt, optimalizált mintavételi konfigurációk alapján meghatározható mindkét korábbi mintavételezésre az a minimális kiegészítő mintaméret, mely mellett a földrajzi és attribútum terület teljes lefedettségét biztosíthatjuk a meglévő és az újonnan elhelyezett (kiegészítő) mintavételi pontokkal. Szabályos, illetve csoportosult mintavételezés esetén már 15-15 kiegészítő mintavételi pont optimalizációja teljes földrajzi lefedettséget biztosít a krigelési szomszédság térképek alapján (11. melléklet). Az attribútum terület lefedettségét Kolmogorov-Smirnov próbával vizsgáltam, melynek eredményeit a VI.22. táblázat foglalja össze. A megfigyelt szignifikancia szintek szerint meglévő szabályos mintavételezés esetén 25 az a minimális kiegészítő mintaméret, mely mellett a releváns segédinformációk által meghatározott attribútum terület teljesen lefedhető. Ugyanez az érték csoportosult mintavételezés esetén magasabb; a statisztikai próba eredményei szerint 35 az a minimális kiegészítő mintaméret, mely mellett a szerves anyag szempontjából releváns segédinformációk teljesen lefedhetők (VI.22. táblázat). A különbség hátterében a korábban már részletezett okok állnak. Mivel mindkét meglévő mintázás esetén más-más a földrajzi, illetve attribútum terület lefedéséhez szükséges minimális kiegészítő mintaméret, ezért kompromisszumot

kell kötni és azt a legkisebb mintaszámot választani, mely mellett a földrajzi és attribútum tér is teljes lefedésre kerül. A fentiek alapján szabályos mintavételezés esetén 25, míg csoportosult mintavételezés esetén 35 kiegészítő mintavételi pont az a minimális kiegészítő mintaméret, mely mellett az attribútum és földrajzi tér teljesen lefedhető.

VI. 5. 3. Többváltozós scenáriók: 5. és 6.

Az 5. és 6. *scenáriók* esetén a cél az volt, hogy teszteljem az SSA többváltozósra kiterjesztett algoritmusát (V.3.4. fejezet). Ezesetben a mintavételezés és a térképezés célja a feltalaj szervesanyag-tartalmának, illetve a termőréteg vastagság együttes térképezése. E scenáriók esetén is megkötés volt, hogy a mintavételezés, illetve térképezés szempontjából irreleváns területekre nem eshetnek mintavételi pontok.

A többváltozós mintavételi scenáriók futtatása előtt származtatnom kellett a termőréteg vastagság regressziós struktúráját, illetve a reziduumok variogramját. Az előszállási talajadatbázis minden egyes mintavételi pontra rendelkezik mért termőréteg vastagság értékkel (V.1.2. fejezet). Többszörös lineáris regresszió analízist végeztem, ahol a független változók a segédinformációk főkomponensei voltak, míg a független változót a termőréteg vastagság töltötte be. Az illesztett többszörös lineáris regresszió struktúráját a VI.23. táblázat foglalja össze. Az illesztett regresszióból származtattam a reziduumokat, melyekből kiszámítottam az irányfüggő, illetve irányfüggetlen tapasztalati variogramokat. A szervesanyag-tartalomhoz hasonlóan a termőréteg vastagság reziduumai is északnyugat-délkeleti irányban nagyobb térbeli folytonosságot mutattak, melynek hátterében – ez esetben is – a domborzat, illetve a területhasznosítás áll. Ezért anizotróp modellt illesztettem a reziduumok variogramjaira, mely modell típusa szférikus, a tető értéke 167,8, a röghatás 67, a hatástávolság 1120 méter a fő folytonossági irányban, az anizotrópia hányados 0,7.

VI.23. táblázat

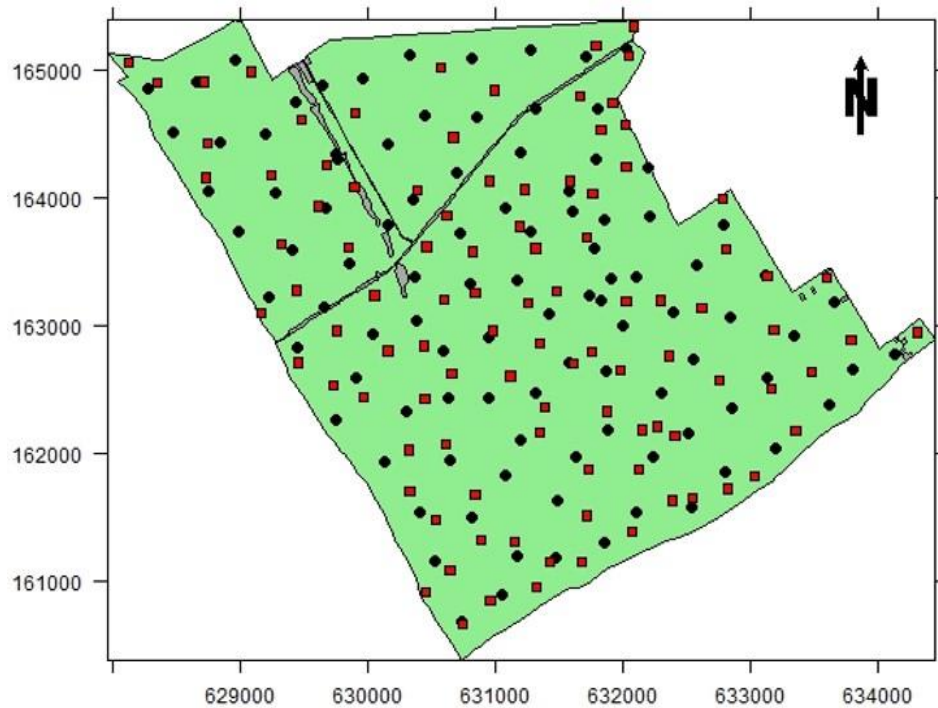
A szervesanyag-tartalom és a termőréteg vastagság regressziós struktúrái, illetve a többváltozós optimalizációban felhasznált „összetett” struktúra

Talajtulajdonság	Független változók száma	Független változók
Szervesanyag-tartalom	8 db	SPC-1, SPC-2, SPC-3, SPC-10, SPC-13, Erodált szántó, Szántó, Rét/legelő,
Termőréteg vastagság	8 db	SPC-1, SPC-3, SPC-5, SPC-13, SPC-16, Erodált szántó, Szántó, Rét/legelő,
„Összetett” struktúra	10 db	SPC-1, SPC-2, SPC-3, SPC-5, SPC-10, SPC-13, SPC-16, Erodált szántó, Szántó, Rét/legelő,

A kiterjesztett SSA algoritmus felparaméterezéséhez szükség van (V.3.4. fejezet): (1) a szervesanyag-tartalom és a termőréteg vastagság térképezése szempontjából releváns segédinformációkra, illetve a (2) domináns talajtulajdonság variogram modelljére. A VI.23. táblázat tartalmazza a két talajjellemző regressziós struktúráját, melyek nagyfokú egyezést mutatnak. Azonban megfigyelhető, hogy míg a szervesanyag-tartalom szempontjából a 2. főkomponenes (SPC-2) releváns segédinformáció, addig a termőréteg vastagság térképezése szempontjából nem tekinthető relevánsnak (VI.23. táblázat). Mindazonáltal a termőréteg vastagság struktúrájában is található olyan segédinformációk, melyek a humusztartalom térképezése szempontjából mondhatók irrelevánsnak, mint például: az 5. és 16. főkomponens (SPC-5 és SPC-16). Ezért fontos, hogy a regressziós struktúrákból kiválasszuk azokat a segédinformációkat, melyek uniója teljesen meghatározza a vizsgált talajtulajdonságok térbeli változékonyságát; ezt az „összetett” struktúrát használtam fel az SSA algoritmus felparaméterezésére (VI.23. táblázat). A domináns talajtulajdonság azonosításához összehasonlítottam a humusztartalomra, illetve a termőréteg vastagság reziduumaiba illesztett variogram modelleket. A két variogram modell ugyanazon irányban mutatnak fő folytonosságot. A hatástávolságokat összehasonlítva elmondható, hogy a termőréteg vastagság reziduuma nagyobb térbeli változékonysággal rendelkeznek. Ezek alapján a termőréteg vastagság a domináns talajtulajdonság, melyre a mintavételezést optimalizálni kell. Ezért az SSA felparaméterezésére a termőréteg vastagság reziduumaiba illesztett variogram modellt használtam. Ezek alapján kiszámítható a választott minőségi mértékből képzett célfüggvény, mely segítségével rangsorolhatók a lehetséges, alternatív mintavételi elrendezések és optimalizálható a mintavételezés a választott két talajtulajdonságra. Az SSA további beállításai – mint az iterációs szám, a rendszer „hűlés”, illetve a „stop kritérium” – megegyeztek az 1., 2., 3. és 4. *szcenáriók* során alkalmazott beállításokkal.

Az 5. *szcenárió* során további mintavételi megszorítás volt a vételezhető minták száma. A kiterjesztett SSA algoritmus által optimalizált mintavételi konfigurációt a VI.19. ábra mutatja be. Az ábrán szintén jelöltem a geostatistikai modellezésekhez felhasznált eredeti mintavételezést. A két mintavételi konfigurációt összehasonlítva elmondható, hogy az optimalizált konfiguráció egységesebben fedte le a mintaterületet, mint az eredeti mintavételezés (VI.19. ábra). Ez jól megfigyelhető – az eredeti mintavételezés során alulreprezentált – északi területeken, illetve a határmenti régiókon. Utóbbinak megkülönböztetett szerepe van a térbeli becslés során, hiszen e területeken kell a leginkább extrapolációra számítanunk (Journel, Rossi, 1989; Kovács, Szanyi, 2005; Webster, Oliver,

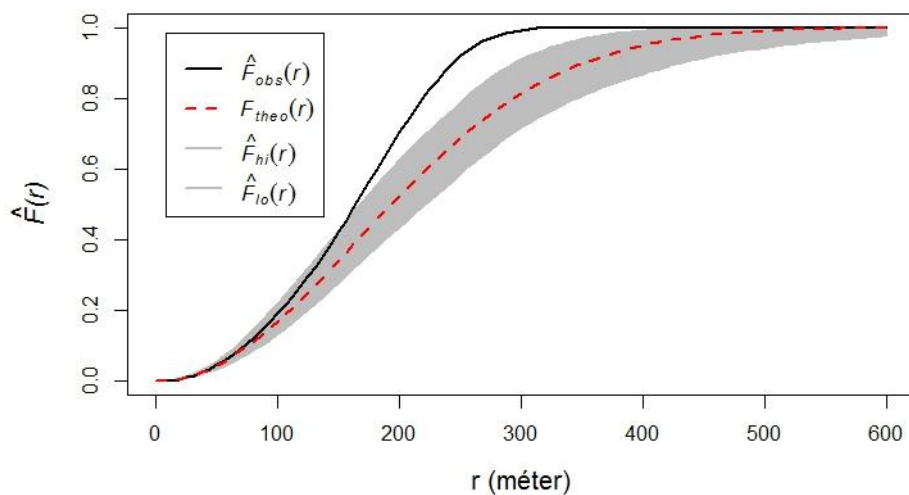
2007). Ugyanakkor jól megfigyelhető, hogy a löszvölgyek mentén az optimalizált mintavételi pontok közelebb helyezkednek el egymáshoz, mint a kvázi síkvidéki területeken (VI.19. ábra). Ennek háttérében, ez esetben is a mintaterület karakterisztikája és a választott minőségi mérték kettős tulajdonsága áll.



VI.19. ábra

Az 5. szcenárióra optimalizált mintavételi konfiguráció (fekete pontok), illetve az előszállási adatpontok (piros négyzetek)

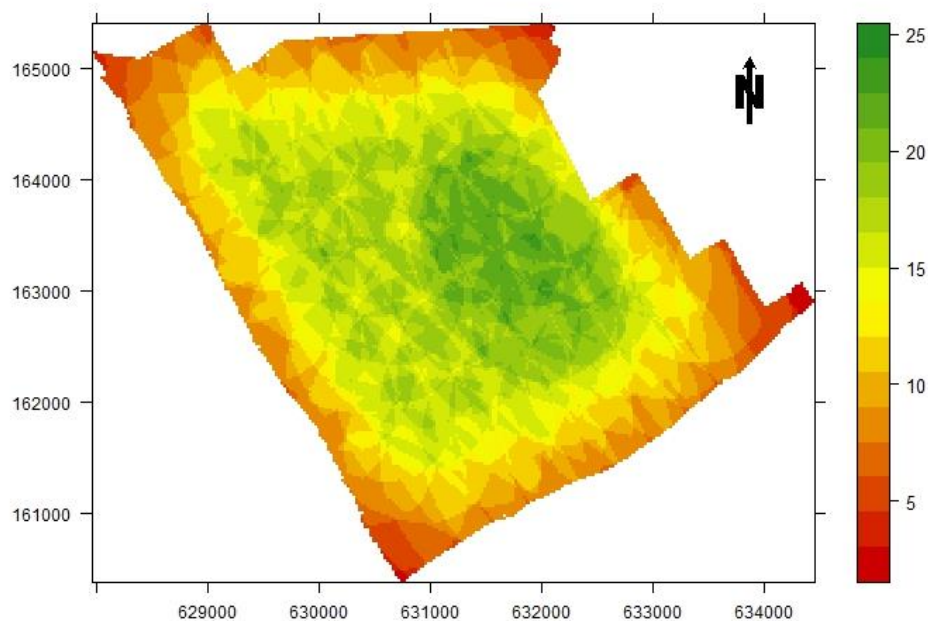
(Megjegyzés: az ábra szürke színnel jelöli a mintavételezés szempontjából „hozzáférhetetlen” területeket)



VI.20. ábra

Az 5. szcenárióra optimalizált mintavételi konfiguráció tapasztalati üres-tér függvénye (fekete vonal) és a Poisson folyamatra (piros szaggatott vonal) szimulált 95%-os konfidencia sáv (szürke szín)

A módosított $F(r)$ üres-tér függvény algoritmussal kapott eredményt a VI.20. ábra mutatja be. A tapasztalati függvény meredekebb lefutású, mint az elméleti függvény. E szerint az optimalizált mintavételi pontok között kizáró kapcsolat van, vagyis a mintavételi pontok térbeli elhelyezkedése a szabályos felé tart (Baddeley, Turner, 2005). Ebből arra következtethetünk, hogy az optimalizált mintavételi pontok térbeli elhelyezkedését leginkább az alkalmazott variogram modell határozta meg (Heuvelink et al., 2007). Az 5. *szcenárióra* optimalizált mintavételi konfiguráció és a termőréteg vastagság anizotróp variogram modellje alapján elkészítettem a becslési pontok krigelési szomszédság térképét (VI.21. ábra). A krigelési szomszédság térkép alapján elmondható, hogy az optimalizált mintavételi konfiguráció egységesen fedi le a mintaterület által meghatározott földrajzi teret. A térkép alapján minden egyes tervezett becslési pont legalább 3 krigelési szomszéddal rendelkezik, melyből azt a következtetést vontam le, hogy az optimalizált konfiguráció megfelelően fedi le a földrajzi teret.



VI.21. ábra

*Az 5. *szcenárióra* optimalizált mintavételi konfiguráció krigelési szomszédság térképe a domináns talajtulajdonság variogram modellje alapján*

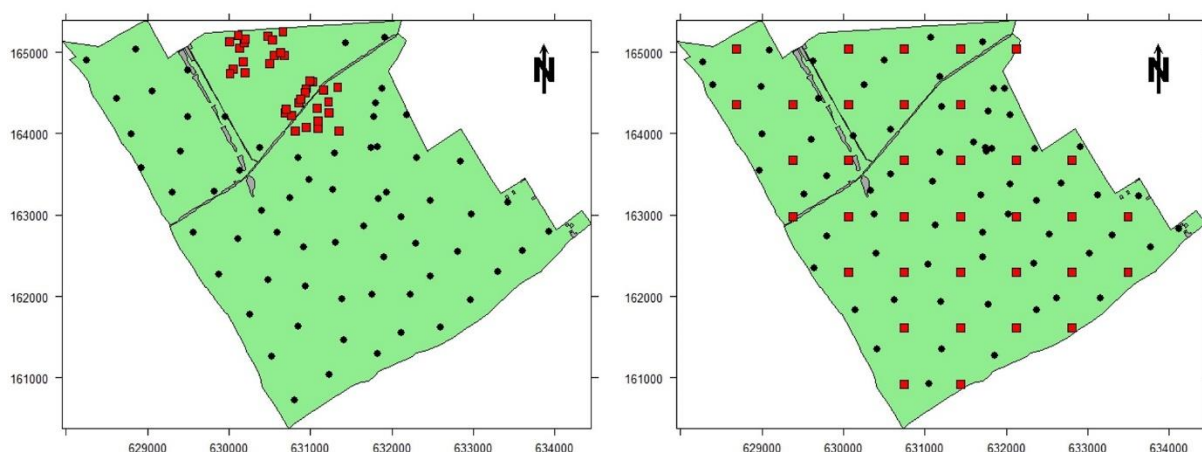
A szervesanyag-tartalom és a termőréteg vastagság térbeli becslése szempontjából releváns segédinformációk optimális lefedettségét Kolmogorov-Smirnov próbával vizsgáltam, melynek eredményeit a VI.24. táblázat foglalja össze. A próba eredményei szerint 5%-os szignifikancia szint mellett egyik segédinformáció esetén sem lehet visszautasítani a próba nullhipotézisét. Ebből azt a következtetést vontam le, hogy az 5. *szcenárióra* optimalizált

konfiguráció teljesen lefedi a releváns segédinformációkat. A fentieket összegezve elmondható, hogy a többváltozós kiterjesztett SSA algoritmussal optimalizált mintavételi konfiguráció teljesen lefedi a szervesanyag-tartalom és a termőréteg vastagság térképezése szempontjából releváns segédinformációkat és a mintaterület által meghatározott földrajzi teret.

VI.24. táblázat

Az 5. scenárióra végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei (p)

Mintaméret (db)	Segédinformációk						
	SPC-1	SPC-2	SPC-3	SPC-5	SPC-10	SPC-13	SPC-16
p	0,375	0,329	0,879	0,125	0,210	0,057	0,081



VI.22. ábra

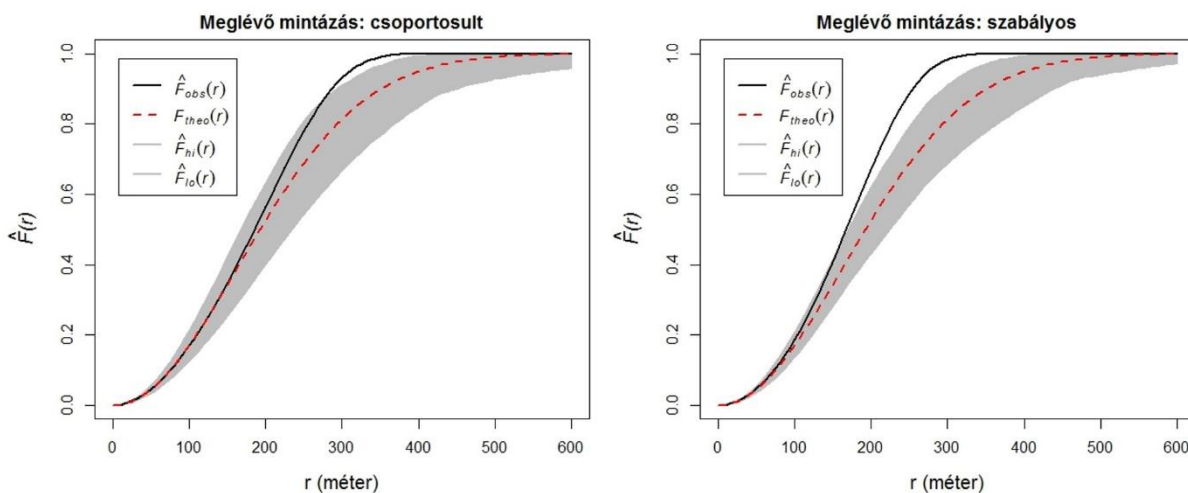
A 6. scenárióra optimalizált mintavételi konfigurációk (fekete pontok), illetve a meglévő mintavételi pontok (piros négyzetek)

(Megjegyzés: az ábra szürke színnel jelöli a mintavételezés szempontjából „hozzáférhetetlen” területeket)

A 6. scenárió során az optimalizálandó mintaméret 65-65 volt. A scenárióban figyelembe vettem a korábbi mintavételezésekből származó 35-35 adatpontot, melyek csoportosult, illetve szabályos elrendezést követtek. A 6. scenárióra optimalizált mintavételi konfigurációkat a VI.22. ábra mutatja be. A két optimalizált konfiguráció jelentősen különbözik egymástól, melynek háttérében ez esetben is a meglévő mintavételezések állnak. A csoportosult mintavételezés – a maga 35 pontjával – elégtelenül fedi le (1) a kutatási terület által meghatározott földrajzi teret, illetve (2) a szervesanyag-tartalom és a termőréteg vastagság térképezése szempontjából fontos segédinformációk által meghatározott attribútum teret. Ennek eredményeként a 65 mintavételi pont optimalizációja során a kiterjesztett SSA algoritmusnak szimultán kellett csökkentenie a minőségi mérték mindkét komponensét. A meglévő csoportosult mintavételezéssel ellentétben, a szabályos mintázás önmagában is közel

teljes földrajzi lefedettséget nyújtott (VI.22. ábra). Ezért a kiterjesztett SSA algoritmus az optimalizálandó 65 mintavételi pont révén azon földrajzi, illetve attribútum térbeli „hiányosságokat” fedte le, melyeket a szabályos mintavételezés „kihagyott”. Ezek jól megfigyelhetők a határmenti területeken, illetve a területet „felszelő löszvölgyek” mentén, ahol a domborzat a legváltozatosabb (VI.22. ábra).

Az egyes optimalizált mintavételi konfigurációkra kiszámított tapasztalati $\hat{F}(r)$ üres-tér függvényt a VI.23. ábra mutatja be. A számítások során figyelembe vettem a korábbi mintavételezésekből származó mintavételi pontokat. A tapasztalati $\hat{F}(r)$ függvényeket összehasonlítva elmondható, hogy szabályos mintázás esetén a függvény lefutása meredekebb, vagyis az optimalizált mintavételi pontok elhelyezkedése „szabályosabb” (Baddeley, Turner, 2005). A krigelési szomszédok térképeit a VI.24. ábra mutatja be. A csoportosult mintavételezés kiegészítésére optimalizált konfiguráció esetében megfigyelhető, hogy a krigelési szomszédok száma egyenlőtlenül oszlik el a mintaterületen; ugyanakkor minden becslési pont rendelkezik legalább egy krigelési szomszéddal. A meglévő szabályos mintavételezés kiegészítésére optimalizált konfiguráció egységesebben fedi le a kutatási terület. Esetében minden egyes becslési ponthoz legalább 3 krigelési szomszéd tartozik (VI.24. ábra).

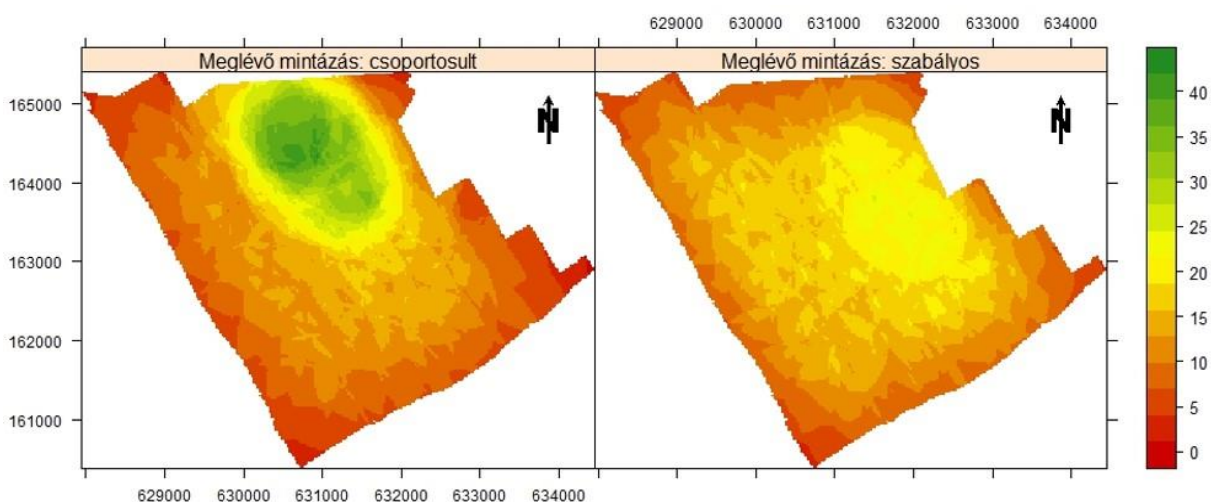


VI.23. ábra

A 6. scenárióra optimalizált mintavételi konfigurációk tapasztalati üres-tér függvénye (fekete vonal) és a Poisson folyamatra (piros szaggatott vonal) szimulált 95%-os konfidencia sáv (szürke színnel)

A szervesanyag-tartalom és a termőréteg vastagság térbeli becslése szempontjából releváns segédinformációk optimális lefedettségét Kolmogorov-Smirnov próbával vizsgáltam. A statisztikai próba eredményeit a VI.25. táblázat tartalmazza. A megfigyelt szignifikancia szintek alapján egyik segédinformáció esetén sem lehet visszautasítani a próba nullhipotézisét

5%-os szignifikancia szint mellett. Ebből azt a következtetést vontam le, hogy a 6. *szenárióra* optimalizált mintavételi konfigurációk teljesen lefedik a szervesanyag-tartalom és a termőréteg vastagság térképezése szempontjából releváns segédinformációkat. A fentiek alapján elmondható, hogy a kiterjesztett algoritmussal optimalizált mintavételi konfigurációk teljesen lefedik a szerves anyag és a termőréteg vastagság szempontjából releváns segédinformációkat, illetve a mintaterület által meghatározott földrajzi teret.



VI.24. ábra

A 6. *szenárióra* optimalizált mintavételi konfigurációk krigelési szomszédság térképei a domináns talajtulajdonság variogram modellje alapján

VI.25. táblázat

Az 6. *szenárióra* végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei

Meglévő mintázás	Segédinformációk						
	SPC-1	SPC-2	SPC-3	SPC-5	SPC-10	SPC-13	SPC-16
szabályos	0,660	0,329	0,879	0,145	0,415	0,131	0,081
csoportosult	0,660	0,559	0,879	0,057	0,093	0,057	0,131

VI. 6. A módszertanok szinergiája

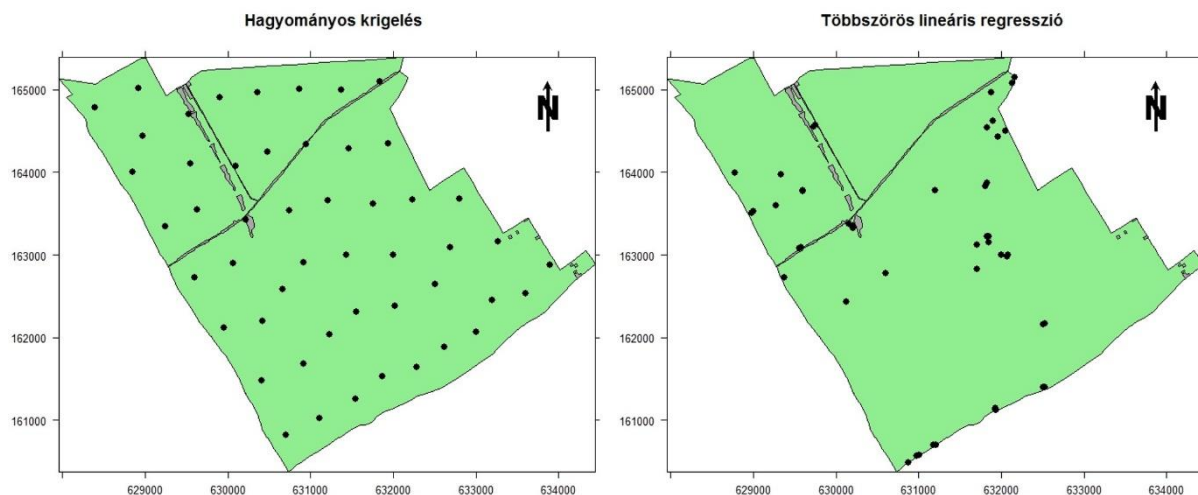
A korábban bemutatott *szenáriók* rámutattak arra, hogy az SSA és a minőségi mértékként választott $\sigma_{RK}^2(\mathbf{u})$ becslési variancia együttese jól alkalmazható a gyakorlatban mintavételezések optimalizációjához egy-, illetve többváltozós esetekben. A mintavételezési módszertant tovább erősíti, hogy az SSA-val optimalizált mintavételi pontok megfelelnek a regresszió krigelés követelményeinek (Hengl, 2009): (1) teljesen lefedik a segédinformációk által definiált attribútum teret, illetve (2) a mintaterület által meghatározott földrajzi teret.

Emellett számos további mintavételi megszorítást figyelembe tud venni, mint például: a vételezhető minták számát, a korábbi felmérésekből származó mintavételi pontokat, illetve a mintavételezés és térképezés számára irreleváns vagy hozzáférhetetlen területeket. Nem utolsósorban a választott minőségi mérték független az adatpontoktól, ezért idő- és költséghatékonyan alkalmazható a gyakorlatban. Ezért a regresszió krigelés, illetve az SSA összekapcsolása olyan módszertant alkot, mely a talajfelvételezések és –térképezések terén költséghatékonyan és effektíven alkalmazható.

Fontos kiemelni, hogy e módszertan akkor is jól alkalmazható a gyakorlatban, ha más modellt adaptálunk a $Z(\mathbf{u})$ valószínűségi függvényhez. Ezt mutatja be a 7. és 8. *szcenárió*, mely során két különböző modellt adaptáltam a $Z(\mathbf{u})$ valószínűségi függvényhez. A 7. *szcenárió* esetén feltételeztem, hogy a szervesanyag-tartalom térképezéséhez nem áll rendelkezésünkre segédinformáció, illetve nem kell „trendhatára” gyanakodnunk, míg a 8. *szcenárió* esetén feltételeztem, hogy a szervesanyag-tartalom térképezését illetően lineáris kapcsolat áll fent a humusztartalom és a segédinformációk között, továbbá a reziduális komponens nagysága elhanyagolható. Ezekre a feltételezésekre a választott becslési eljárás a hagyományos krigelés, illetve a többszörös lineáris regresszió lenne (Hengl, 2009). Az optimalizáció minőségi mértékét ehhez mérten választottam meg, vagyis a hagyományos krigelés becslési varianciáját, illetve a regresszió hiba varianciáját, melyek szintén függetlenek az adatponti értékektől. A definiált célfüggvények ez esetben is e becslési, illetve hiba varianciák összterületi átlaga volt.

A 7. és 8. *szcenárióra* optimalizált mintavételi konfigurációkat a VI.25. ábra mutatja be, mely során szintén mintavételi megszorításként vettem figyelembe a mintavételezés szempontjából irreleváns területeket. A két konfiguráció alapjaiban tér el egymástól. Hagyományos krigelés esetén a cél, hogy minél egységesebben fedjük le a mintaterület által meghatározott földrajzi teret (Webster, Oliver, 2007). Többszörös lineáris regresszió esetén a cél, hogy a releváns segédinformációk által definiált attribútum teret fedjük le (Hengl, 2009). A 7. *szcenárióra* optimalizált mintavételi konfiguráción megfigyelhető a variogram modell anizotrópiájának a hatása. A mintavételi pontok közötti távolság a fő folytonossági irányban nagyobb, mint a rá merőleges irányban. A mintaterület délnyugati, illetve északkeleti határa mentén jól megfigyelhető, hogy a mintavételi pontok „kihúzódnak” a határ menti területekre, ahol a leginkább kell extrapolációra számítanunk. A 8. *szcenárióra* optimalizált mintavételi konfiguráció csoportosult mintázatot mutat. Esetében nem szükséges, hogy a minták egységesen fedjék le a kutatási terület által meghatározott földrajzi teret, csakis a releváns

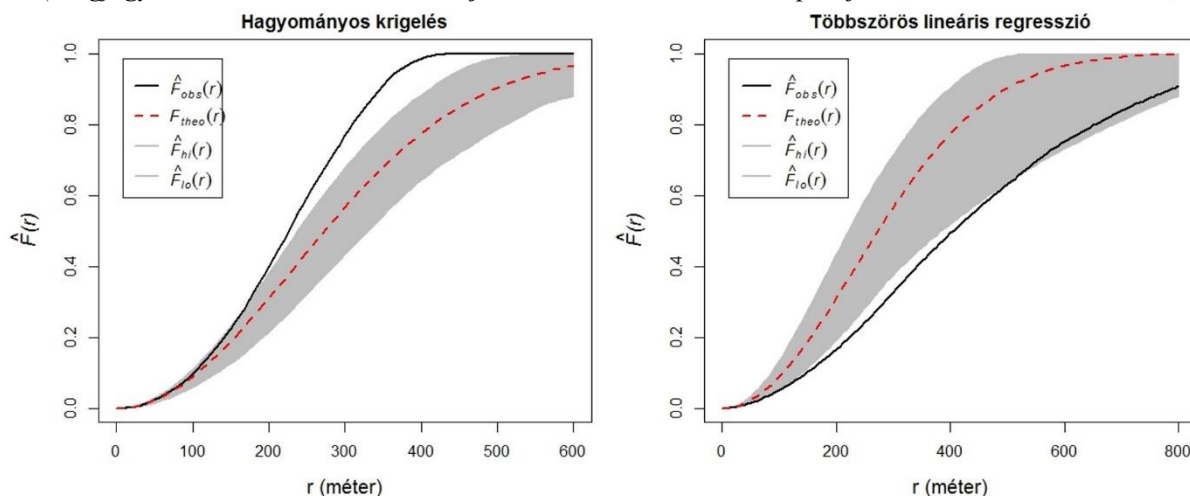
segédinformációk szerinti attribútum teret. Ezért a minták „egységes” földrajzi elhelyezkedését nem szabályozza. A térképen jól megfigyelhető, hogy az optimalizált konfiguráció lefedi a löszvölgyek nyújtotta domborzati változatosságokat, melyek kritikusak a regresszió hiba varianciájának hathatós csökkentése céljából.



VI.25. ábra

A 7. scenárióra (baloldalon), illetve a 8. scenárióra (jobboldalon) optimalizált mintavételi konfigurációk (fekete pontok).

(Megjegyzés: az ábra szürke színnel jelöli a mintavételezés szempontjából irreleváns területeket)



VI.26. ábra

A 7. és 8. scenárióra optimalizált mintavételi konfigurációk tapasztalati üres-tér függvénye (fekete vonal) és a Poisson folyamatra (piros szaggatott vonal) szimulált 95%-os konfidencia sáv (szürke színnel)

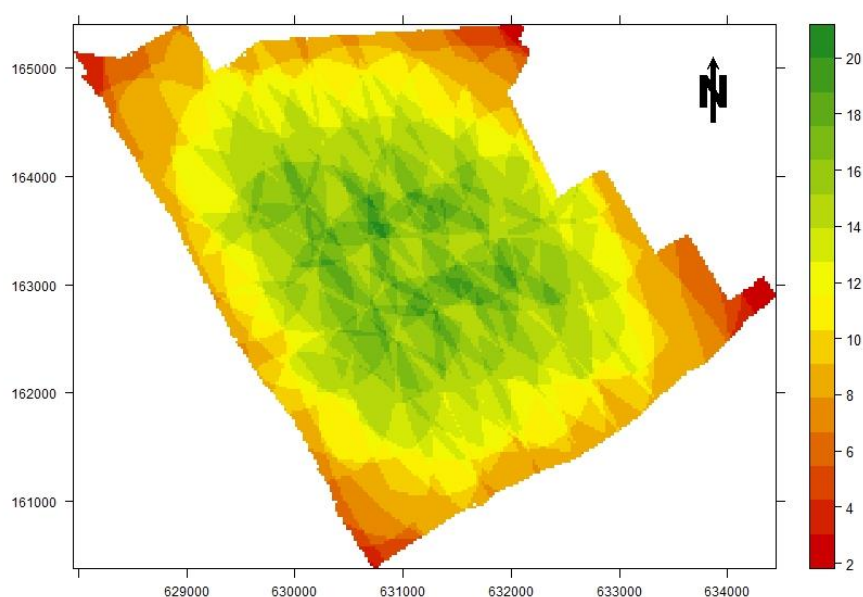
A tapasztalati $\hat{F}(r)$ üres-tér függvényt a VI.26. ábra mutatja be. A 8. scenárióra optimalizált konfigurációnál meredekebb lefutású az elméleti függvény, melyből következik, hogy a mintavételi pontok csoportosult mintázatot mutatnak (Heuvelink et al., 2007). A 7. scenárió esetén a tapasztalati függvény meredekebb lefutású, mint az elméleti függvény, miszerint a 7. scenárióra optimalizált konfiguráció szabályos mintázatot mutat. A 7.

szcenárióra optimalizált mintavételi konfiguráció, illetve az alkalmazott anizotróp variogram modell alapján előállítottam a tervezett becslési pontok krigelési szomszédság térképét (VI.27. ábra). A térkép alapján elmondható, hogy a mintavételezés egységesen fedte le a kutatási területet, továbbá minden becslési pont legalább 2 krigelési szomszáddal rendelkezik. Ezek alapján a 7. *szcenárióra* optimalizált mintavételezés megfelelően fedi le a mintaterület által meghatározott földrajzi teret. A 8. *szcenárió* esetén az attribútum tér lefedettségét Kolmogorov-Smirnov próbával vizsgáltam, melynek eredményeit a VI.26 táblázat tartalmazza. A megfigyelt szignifikancia szintek alapján egyik segédinformáció esetén sem utasítható vissza a próba nullhipotézise. E szerint a 8. *szcenárióra* optimalizált mintavételi pontok teljesen lefedik a szervesanyag-tartalom térképezése szempontjából releváns segédinformációkat.

VI.26. táblázat

A 8. *szcenárióra* végzett Kolmogorov-Smirnov próba megfigyelt szignifikancia szintjei (*p*)

Mintaméret (db)	Segédinformációk				
	SPC-1	SPC-2	SPC-3	SPC-10	SPC-13
<i>p</i>	0,181	0,181	0,172	0,291	0,126



VI.27. ábra

A 7. *szcenárióra* optimalizált mintavételi konfiguráció krigelési szomszédság térképe az alkalmazott variogram modell alapján

VII. ÖSSZEFOGLALÁS

Doktori kutatásaimban három mintaterületen alkalmaztam és értékeltem a kidolgozott, koherens, geostatistikai megközelítésen alapuló digitális talajtérképezési és mintavétel optimalizációs módszertant. A választott mintaterületek mind kiterjedésükben, mind az ott érvényre jutó talajképződési folyamatok tükrében lényeges különbségeket mutattak. A vizsgálatok célváltozójaként a talajok szervesanyag-tartalmát választottam, mely talajtulajdonság kiemelkedő jelentőséggel bír nemcsak a talajtan és az agrárium, hanem a környezetvédelem (például: éghajlatváltozás, szén biogeokémia körforgalma) és földtudományok számára is. A dolgozat legfőbb eredményeit, a bevezetésben kitűzött célokkal és az elért eredményekkel összhangban, a következő pontokban foglalom össze:

- 1) Kialakítottam egy geostatistikai megközelítésen alapuló, koherens digitális talajtérképezési és mintavétel optimalizációs módszertant, mely a térképezendő talajtulajdonságo(ka)t sztochasztikus folyamatként modellezi. E megközelítés szerint a talajtulajdonságok a talajnak olyan jellemzői, melyek a mérési hely végtelen kicsiny sugarú környezetében valószínűségi változónak tekinthetők, ugyanakkor megadható olyan térben értelmezett függvény, mely ezeket a valószínűségi változókat összekapcsolja. Az ilyen tulajdonságokat a geostatistika regionalizált változóknak nevezi. Az eredmények alapján elmondható, hogy e megközelítésen alapulva a kidolgozott módszertan keretet nyújt a talaj-mintavételezések optimalizációjára, a digitális talajtérképezésre, a térbeli bizonytalanság modellezésére és a térképek értékelésére. Az eredmények rávilágítottak arra, hogy a módszertan lépései konzekvensek és kompatibilisek egymással. Rámutattam továbbá arra, hogy a kidolgozott módszertan komponensei szinergizálnak. Az eredmények alapján elmondható, hogy a kidolgozott módszertan különböző léptékekben alkalmazható a talajtulajdonságok térképezésére. A módszertan képes számításba venni a rendelkezésünkre álló folytonos és kategória típusú segédinformációkat a térbeli becslés, a térbeli bizonytalanság modellezése, illetve a mintavétel optimalizáció során. A módszertan szerves részét képezi az eredmény térkép hibájának meghatározása, illetve a térbeli bizonytalanság modellezése, mellyel kielégíthetők a talajtani információkkal szembeni aktuális igények. Nem utolsó sorban rámutattam arra, hogy a módszertan egyes komponensei az adaptált valószínűségi függvény alapján módosíthatók. Ezért a módszertan az adott célokhoz, szakértői ismeretekhez, illetve a

rendelkezésre álló információkhoz igazítható, mely a módszertan flexibilitásáról tanúskodik.

- 2) A kialakított digitális talajtérképezési és mintavétel optimalizációs módszertant szabad-hozzáférésű, nyílt forráskódú szoftver (*Free and Open Source Software, FOSS*) környezetben dolgoztam ki. A segédinformációk előkészítéshez SAGA GIS (Conrad et al., 2015), a geomatematikai és geostatistikai elemzésekhez, illetve modellezésekhez R (R Development Core Team, 2016) szoftver környezetet használtam. Emellett továbbfejlesztettem, illetve kidolgoztam több eljárást (például: módosított üres-tér függvény, többváltozós mintavétel optimalizációs algoritmus, krigelési szomszédok térképezése), mely a kialakított módszertanhoz igazodik. A szabad-hozzáférésű, nyílt forráskódú szoftver környezetben legfőbb előnye, hogy a módszertan bárki számára szabadon elérhető, alkalmazható, illetve módosítható. Ez utóbbi lehetőséget biztosít a további fejlesztésekre.
- 3) A regresszió krigeléssel előállított szervesanyag-tartalom térképek és a meghatározott hibamértékek (ME, MAE, RMNSE) alapján elmondható, hogy a regresszió krigelés különböző léptékekben alkalmazható a talajtulajdonságok térbeli modellezésére. A módszer a rendelkezésünkre álló (folytonos és kategória típusú) segédinformációk alapján apriori távolítja el az adatok átlagában megjelenő szisztematikus változást. Az alkalmazott többváltozós lineáris regresszió együtthatóit GLS (*generalized least squares*) módszerrel határoztam meg, mely képes számításba venni a reziduumok kovariancia mátrixát. A regresszió analízis során a multikollinearitás csökkentésére főkomponens analízist végeztem. A regresszió reziduumait egy stacionárius valószínűségi függvényként modelleztem. Emellett rámutattam arra, hogy a regresszió reziduumai egyszerű krigeléssel modellezhetők, mely eljárás megtartja az adaptált valószínűségi függvény tulajdonságait. A Zala megyei szervesanyag-tartalom térkép esetében rámutattam arra, hogy a nem lineáris kapcsolat a segédinformációk és a térképezendő talajtulajdonság között feltételes torzítást okoz.
- 4) Regresszió krigelésen alapuló szekvenciális sztochasztikus szimulációt alkalmaztam az előszállási mintaterületen a térbeli bizonytalanság modellezésére. Az algoritmust az egyszerű krigelés becslési varianciájával módosítottam, melynek eredményeként a realizációk visszaadták a reziduumokra illesztett variogram modellt. A 100 darab alternatív, egyenlően valószínű realizáció eredményei alapján kimutattam, hogy a sztochasztikus szimulációval olyan – a talajtani számára kiemelt jelentőségű – kérdések is megválaszolhatók, melyekre a krigelési eljárások konstrukciójuk miatt

alkalmatlanok. Az előállított 100 realizáció segítségével vizuálisan és kvantitatívan is értékelhető volt a térbeli bizonytalanság. A realizációk alapján megadtam az egyes grid pontok feletti gyakorisági eloszlásokat, melyekből származtattam: (1) a szervesanyag-tartalom várható érték típusú becslését, (2) a hozzátartozó 95%-os konfidencia intervallumot, illetve (3) adott szervesanyag-tartalom értékekhez tartozó valószínűségi térképeket. Az eredmények alapján elmondható, hogy a szekvenciális sztochasztikus szimuláció jól alkalmazható a térbeli bizonytalanság modellezésére. Továbbá rámutattam arra, hogy az egyes grid pontok feletti gyakorisági eloszlások segítségével célspecifikus talajtérképek származtathatók, mellyel kielégíthetők a talajtani információkkal szembeni aktuális igények, illetve hatékonyan támogathatók a szakmai döntések.

- 5) Kialakítottam egy olyan modell alapú talajmintavétel optimalizációs módszertant, melynek az alapját az SSA (*spatial simulated annealing*) algoritmus jelenti. A módszertan minőségi mértékének a regresszió krigelés becslési varianciáját választottam, mely szimultán keresi a földrajzi és attribútum tér optimális lefedettségét. Ennek következtében a lehető legjobban támogatja a regresszió krigelésre felépített térképezési módszertant. A választott minőségi mérték független az adatponti értékektől, ezért a módszertan idő- és költséghatékony. Ugyanakkor rámutattam arra is, hogy a regresszió krigelés becslési varianciája általában véve nem alkalmazható a lokális becslési pontosság mértékeként. A módszertan tesztelésére hat scenáriót állítottam be. A scenáriók eredményei rámutattak arra, hogy a kidolgozott talajmintavétel optimalizációs módszertan képes számításba venni számos mintavételi megszorítást (például: vételezhető minták száma, korábbi felvételezésekből származó mintavételi pontok, mintavételezés számára irreleváns vagy hozzáférhetetlen területek, segédinformációk, előre meghatározott minőségi mérték értéke, a mintaterület szabálytalan határai) a mintavételi pontok optimális elhelyezése során. A scenáriók eredményei alapján szintén elmondható, hogy a módszertan jól alkalmazható a talajmintavételi elrendezések optimalizációjára. A scenáriók segítségével rámutattam arra is, hogy a talajmintavétel optimalizációs módszertan hozzáigazítható egyéb térbeli becslő eljárásokhoz (például: hagyományos krigelés, többszörös lineáris regresszió), mely a módszertan flexibilitásától tanúskodik.
- 6) Kidolgoztam az SSA többváltozós módszerét, mely alkalmas, hogy szimultán optimalizálja a mintavételi helyek elrendezését a térképezendő talajtulajdonságokra vonatkozóan, melyre az egyváltozós SSA algoritmus nem alkalmas. A kiterjesztett

algoritmus a legkedvezőtlenebb helyzetre optimalizálja a talaj-mintavételezést. A kiterjesztett módszer tesztelésére két scenáriót állítottam be, ahol a hipotetikus cél a szervesanyag-tartalom és a termőréteg vastagság térképezése volt. Az optimalizáció minőségi mértékét a regresszió krigelés becslési varianciája jelenti, mely a térképezendő talajtulajdonságok és a rendelkezésünkre álló segédinformációk közötti regressziós struktúrákból, illetve a domináns talajtulajdonság variogram modelljéből határozható meg. A scenáriók eredményei alapján elmondható, hogy a kapott mintavételi konfigurációk optimálisak a választott talajtulajdonságok térképezésére az adott mintavételi megszorítások mellett. Továbbá a kiterjesztett módszertan megtartotta az egyváltozós SSA, illetve a regresszió krigelés becslési varianciájának előnyeit. Így a mintavételi helyek optimalizálása során képes számításba venni a vételezhető minták számát, a korábbi mintavételezésekből származó mintavételi pontokat, az elérhető segédinformációkat, a mintaterület szabálytalan határait, illetve a mintavételezés számára irreleváns vagy hozzáférhetetlen területeket. Emellett rámutattam arra, hogy az alkalmazott minőségi mérték értékei csakis kizárólag a mintavételi konfigurációk rangsorolására alkalmazhatók.

VIII. SUMMARY

During the course of my doctoral research, I have elaborated a coherent, digital soil mapping and sampling optimization methodology based on geostatistical approaches, which methodology was applied and evaluated on three Hungarian pilot areas. The selected study sites show significant differences in their extensions (scales), as well as the affecting soil forming processes. I have selected the soil organic matter content to test the developed, coherent methodology, which soil property has a great importance not only for soil science and agriculture, but for earth and environmental sciences (e.g. global climate change, carbon biogeochemical cycle) as well. Hereby I summarize the result of my research in accordance with the goals previously set as follows:

- 1) I have elaborated a coherent, digital soil mapping and sampling optimization methodology based on geostatistical approaches, which methodology proceeds on the assumption that the soil properties of interest are the outcomes of random processes. According to that assumption, locally the point values of the soil property are considered as random variables, which variables are not independent but are related by a correlation expressing the spatial structure of the phenomenon. These (soil) properties are referred to as regionalized variables by geostatistics. According to the results, the elaborated methodology provides a framework for soil sampling optimization, digital soil mapping, spatial uncertainty modelling and the assessment of the resulted map(s). The results revealed that the components of the elaborated methodology are consequent and compatible with each other. Furthermore, the results pointed out that the components of the developed methodology synergize with each other that is they use the most effective way the available resources for soil sampling and digital soil mapping. The results of the digital soil mapping processes showed that the methodology can be applied to map soil properties in different scales. The elaborated methodology can take into account the available continuous and categorical secondary (or auxiliary) information in the course of spatial prediction, spatial uncertainty modelling, as well as soil sampling optimization. An essential part of the elaborated methodology is to determine the error of the map using different measures, as well as to model the spatial uncertainty. These can be used to satisfy the recent demands on soil information. Moreover, I pointed out that the components of the elaborated methodology are exchangeable in accordance with the adopted random function model. Hence, the methodology can gear to the actual

goals, expert knowledge and available information, which makes the methodology more flexible.

- 2) I have developed the digital soil mapping and sampling optimization methodology with free and open source software (FOSS) environment. SAGA GIS (Conrad et al., 2015) was applied to prepare secondary information. Furthermore, R (R Development Core Team, 2016) was applied to geomathematical and geostatistical analyses and modelling processes. In addition, I improved and elaborated several methods (e.g. modified empty space function, multivariate sampling design optimization algorithm, kriging neighbours mapping) using R software environment, which methods act on the elaborated methodology. The greatest advantage of using FOSS is that the methodology and its components are freely available, applicable and variable. Last one has the advantages to improve further on the elaborated methodology.
- 3) According to the created soil organic matter content map and the calculated error measures (i.e. ME, MAE, RMNSE), I can state that regression kriging (RK) can be applied to model the spatial variabilities of soil properties in different scales. RK removes a priori the systematic change from the mean values of the data using the available continuous and categorical secondary information. I applied the GLS (generalized least squares) method to estimate the coefficients of the multiple linear regression, which method able to take the covariance matrix of the regression residuals into account. Furthermore, I applied principal component analysis in order to decrease the multicollinearity. I modelled the residuals as a stationarity random function. In addition, I pointed out that the residuals have to be modelled with simple kriging because it is able to retain the properties of the adopted random function model. In case of Zala County, I pointed out that non-linearity between the target soil property and the available secondary information causes conditional bias.
- 4) I have applied the sequential stochastic simulation algorithm based on regression kriging in Előszállás to model the spatial uncertainty. I have modified the simulation algorithm with the simple kriging variance in order to reproduce the applied variogram model. The generated alternative and equally probable realizations reproduced the variogram, which confirmed the necessity of the previous modification. Based on the generated 100 realizations, I established that one can use stochastic simulation in order to answer such kind of highly relevant soil related tasks, which cannot be answered by kriging methods because of their designs. The generated 100 realizations provided a visual and quantitative measure of spatial uncertainty. I applied the 100 realizations to

calculate the cumulative distribution around an infinitesimally small neighbourhood of each grid point. Using these cumulative distributions I derived: (1) the E-type estimation of the soil organic matter content, (2) the corresponding 95% confidence interval and (3) probability maps for given soil organic matter values. Based on these results I can state that the sequential stochastic simulation algorithm based on regression kriging is able to model the spatial uncertainty. Furthermore, I have pointed out that the calculated cumulative distributions around the grid points can be applied to create goal oriented digital soil maps. These can be used to satisfy the recent demands on soil information and to support various decision-making.

- 5) I have developed a model based soil sampling optimization method, which is based on the spatial simulated annealing (SSA) algorithm. I selected the regression kriging variance as the quality measure of the optimization method, which measure endeavours SSA to optimize the sampling configuration in both geographic and feature space. Hence, regression kriging variance as quality measure of the optimization procedure valets the digital soil mapping method based on regression kriging. The selected quality measure is independent from the observed values. Therefore, it can be calculated before the actual sampling takes place, which can be considered as a beneficial property in point of costs and time. In addition, I pointed out that the regression kriging variance – in general – is not measure of the local prediction accuracy. Six scenarios were set to test the soil sampling optimization method. According to the results of the scenarios, I can state that the sampling optimization method is able to consider numerous sampling constraints, such as the number of new observations, taking previously collected samples into account, irrelevant or inaccessible areas for sampling, taking secondary information into account, budget and/or accuracy constraints. Based on the results of the scenarios, I can also state that the developed, model based soil sampling optimization method is able to optimize the soil sampling configuration. Furthermore, I pointed out that the developed method can be modulated to other spatial prediction methods (e.g. ordinary kriging, multiple linear regression), which makes the methodology more flexible.
- 6) I have elaborated a multivariate generalization of the univariate SSA. The extended algorithm is able to simultaneously optimize the sampling configuration for the target soil properties. The algorithm optimizes the soil sampling design for the most unfavorable situation. Two scenarios were set to test the multivariate SSA algorithm, where the hypothetical goal was to map soil organic matter content together with rooting

depth. I selected the regression kriging variance as the quality measure of the optimization, which was derived from the structures of the regression models and the variogram of the dominant soil property. According to the results of the scenarios, I can state that the resulted sampling configurations are optimal for the digital mapping of target soil properties considering the predefined sampling constraints. Furthermore, the extended algorithm retained the advantageous properties of the univariate SSA and regression kriging variance. Hence, the extended method is able to consider numerous sampling constraints, such as the number of new observations, taking previously collected sampling points into account, irrelevant or inaccessible areas for sampling, budget constraints and taking secondary information into account. Moreover, I pointed out that the values of the applied quality measure can be used only to rank the alternative sampling configurations.

IX. KÖSZÖNETNYILVÁNÍTÁS

Mindenekelőtt szeretnék köszönetet mondani témavezetőmnek, *Dr. Barta Károlynak*, aki nemcsak a doktori éveim alatt, hanem már hallgató koromtól kezdve figyelemmel kísérte tanulmányaimat és mindvégig támogatott. Köszönöm, hogy a szakdolgozatomon és diplomamunkámon túl a doktori kutatásaimat is a témavezetése alatt végezhettem. Ezúton is szeretném megköszönni lelkiismeretes munkáját.

Szeretnék köszönetet mondani *Dr. Pásztor Lászlónak*, hogy elvállalta a doktori kutatásaim konzulensi feladatait. Köszönöm, hogy 2015. szeptemberétől a kutatócsoportjában dolgozhatok, mellyel olyan új perspektíváit ismerhettem meg a digitális talajtérképezésnek, melyek nagyban hozzájárultak a dolgozat mindalaposabb kidolgozásához. Ezúton is szeretném megköszönni türelmét és lelkiismeretes munkáját.

Ezúton szeretném megköszönni a Szegedi Tudományegyetem, Természeti Földrajzi és Geoinformatikai Tanszék vezetőjének, *Prof. Dr. Mezősi Gábornak*, illetve a Magyar Tudományos Akadémia, Talajtani és Agrokémiai Intézet igazgatójának, *Dr. Szabó Józsefnek*, hogy helyet, infrastruktúrát és anyagi forrást biztosítva lehetővé tették, hogy doktori kutatásaimat elvégezhessem.

Köszönöm volt doktorandusztársaimnak azt a felhőtlen légkört, melyben volt szerencsém dolgozni három évig Szegeden. Köszönöm jelenlegi munkatársaimnak, a Talajtérképezési és Környezetinformatikai Osztály dolgozóinak, hogy doktori kutatásaimban mindvégig támogattak, segítettek és jótanácsokkal láttak el.

A disszertációmban felhasznált adatbázisok terén köszönettel tartozom *Antal Kristófnak* (FMKH NTI), *Dr. Bakacsi Zsófiának* (MTA TAKI), *Borcsik Zoltánnak* (CSMKH NTO), *Dr. Farsang Andreának* (SZTE TFGT), *Dr. Illés Gábornak* (NAIK ERTI), illetve *Szabóné Kele Gabriellának* (FMKH NTI). Ezúton is szeretném megköszönni hozzájárulásukat a doktori kutatásaimhoz.

Hálás köszönettel tartozom *szüleimnek*, akik töretlenül bíztak bennem és támogattak céljaim elérésében. Nem utolsósorban hálás köszönettel tartozom páromnak, *Boginak*, aki mindvégig mellettem állt és támogatott.

X. IRODALOMJEGYZÉK

- Baddeley, A., 2008. Analysing spatial point patterns in R (Workshop Notes). CSIRO. 171 p.
- Baddeley, A., Turner, R., 2005. spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software* 12(6), 1–42.
- Bakacsi, Zs., Pásztor, L., Szabó, J., Kuti, L., Laborcz, A., 2012. 3D textúra adatbázis létrehozása indikátor-krigeléssel, talajtani és agrogeológiai adatbázisok egységesítésével. *Agrárinformatika* 3(1), 46–51.
- Bárdossy, A., Bogárdi, I., 1983. Network design for the spatial estimation of environmental variables. *Applied Mathematics and Computation* 12(4), 339–365.
- Bárdossy, A., Molnár, Z., 2003. A Kisalföld talajvízjárásának geostatistikai értékelése. *Hidrológiai Közlöny* 83(4), 214–220.
- Bárdossy, A., Molnár, Z., 2004. Felszín alatti víz észlelőhálózat optimalizációjának módszere. *Hidrológiai Közlöny* 84(1), 56–63.
- Bárdossy, Gy., 2006. Geologic and geostatistical evaluation of spatial variability. *Acta Geologica Hungarica* 49, 89–101.
- Barta, K., 2004. Talajeróziós modellépítés a EUROSEM modell nyomán (Ph.D. értekezés). Szegedi Tudományegyetem, Szeged. 82 p.
- Bashfeld, A., Keim, A., 2011. Continent-wide DEM creation for the European Union. 34th International Symposium on Remote Sensing of Environment, 10–15 April 2011, Sydney.
- Baume, O.P., Gebhardt, A., Gebhardt, C., Heuvelink, G.B.M., Pilz, J., 2011. Network optimization algorithms and scenarios in the context of automatic mapping. *Computers & Geosciences* 37, 289–294.
- Baumgardner, M.F., 2011. Soil databases. In: Huang, P.M., Li, Y., Sumner, M.E. (Eds.) *Handbook of Soil Sciences: Resource, Management and Environmental Impacts*. CRC Press, Boca Raton. pp. 21–35.
- Baxter, S.J., Oliver, M.A., 2005. The spatial prediction of soil mineral N and potentially available N using elevation. *Geoderma* 128, 325–339.
- Bechler, A., Romary, T., Jeannée, N., Desnoyers, Y., 2013. Geostatistical sampling optimization of contaminated facilities. *Stochastic Environmental Research and Risk Assessment* 27(8), 1967–1974.
- Benyhe, B., 2013. Agrogén hatásra kialakuló felszínformák és folyamatok vizsgálata eltérő geomorfológiai adottságú területeken (Ph.D. értekezés). Szegedi Tudományegyetem, Szeged. 123 p.
- Bogárdi, I., Bárdossy, A., Duckstein, L., 1985. Multicriterion network design using geostatistics. *Water Resources Research* 21(2), 199–208.
- Borcsik, Z., Farsang, A., Barta, K., Kitka, G., 2011. Humuszanyagok mennyiségi és minőségi eróziójának mérése a Tolna megyei Szálka település melletti vízgyűjtőn. *Talajvédelem (különszám)*, 127–137.
- Bourennane, H., Dére, Ch., Lamy, I., Cornu, S., Baize, D., van Oort, F., King, D., 2006. Enhancing spatial estimates of metal pollutants in raw wastewater irrigated fields using a topsoil organic carbon map predicted from aerial photograph. *Science of the Total Environment* 361, 229–248.
- Bourennane, H., King, D., 2003. Using multiple external drifts to estimate a soil variable. *Geoderma* 114, 1–18.
- Brus, D.J., 2015. Balanced sampling: A versatile sampling approach for statistical soil surveys. *Geoderma* 253–254, 111–121.
- Brus, D.J., de Gruijter, J.J., 1993. Design-based versus model-based estimates of spatial means: Theory and application in environmental soil science. *Environmetrics* 4(2), 123–152.
- Brus, D.J., de Gruijter, J.J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* 80, 1–44.
- Brus, D.J., Heuvelink, G.B.M., 2007. Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* 138, 86–95.
- Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *European Journal of Soil Science* 62, 394–407.
- Burgess, T.M., Webster, R., 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. *Journal of Soil Science* 31, 315–331.
- Burgess, T.M., Webster, R., 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. *Journal of Soil Science* 31, 333–341.
- Burgess, T.M., Webster, R., McBratney, A.B., 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. *Journal of Soil Science* 32, 643–659.
- Carter, M.R., Gregorich, E.G. (Eds.), 2007. *Soil Sampling and Methods of Analysis* (2nd Ed.). Taylor & Francis Group, London. 1264 p.
- Centeri, Cs., 2001. Az általános talajveszteség becslési egyenlet (USLE) K tényezőjének vizsgálata (Ph.D. értekezés). Szent István Egyetem, Gödöllő. 125 p.
- Centeri, Cs., Barcsi, A., Grónás, V., Joó, K., Néráth, M., Vona, M., Pataki, R., Zsembéry, Z., Kristóf, D., Skutai, J., Belényesi, M., Penksza, K., 2008. Talajeróziós térképezés térinformatikai eszközökkel hazai

- mintaterületeken. In: Simon, L. (Szerk.) Talajtani Vándorgyűlés 2008 (talajvédelem különszám). Talajvédelmi Alapítvány, Nyíregyháza. pp. 569–576.
- Centeri, Cs., Pataki, R., Bíró, Zs., Császár, A., 2003. Az eróziós térképek kategóriáinak értékelése. *Agrokémia és Talajtan* 52(3–4), 443–454.
- Centeri, Cs., Szabó, B., Jakab, G., Kovács, J., Madarász, B., Szabó, J., Tóth, A., Gelencsér, G., Szalai, Z., Vona, M., 2014. State of Soil Carbon in Hungarian Sites: Loss, Pool and Management. In: Margit, A. (Ed.) *Soil Carbon: Types, Management Practices and Environmental Benefits*. Nova Science Publishers, Inc., New York. pp. 91–117.
- Chu, H.-J., Lin, Y.-P., Jang, C.-S., Chang, T.-K., 2010. Delineating the hazard zone of multiple soil pollutants by multivariate indicator kriging and conditioned Latin hypercube sampling. *Geoderma* 158, 242–251.
- Cochran, W.G., 1977. *Sampling Techniques* (3rd Ed.). John Wiley & Sons, Inc., Chichester. 448 p.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development* 8, 1991–2007.
- Cressie, N., 1990. The origins of kriging. *Mathematical Geology* 22(3), 239–252.
- Cressie, N., Gotway, C.A., Grondona, M.O., 1990. Spatial prediction from networks. *Chemometrics and Intelligent Laboratory Systems* 7, 251–271.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data* (Revised Ed.). John Wiley & Sons, Inc., New York. 901 p.
- Csillag, F., Kertész, M., 1989. Spatial variability: Error in natural resource maps? *Agrokémia és Talajtan* 38(3–4), 715–726.
- Csillag, F., Kertész, M., Kummert, Á., 1996. Sampling and mapping of heterogeneous surfaces: Multi-resolution tiling adjusted to spatial variability. *Geographical Information Systems* 10, 851–875.
- Csillag, F., Kummert, Á., Kertész, M., 1992. Resolution, accuracy and attributes: Approaches for environmental geographical information systems. *Computers, Environment and Urban Systems* 16, 289–297.
- de Gruijter, J.J., Brus, D.J., Bierkens, M.F.P., Knotters, M., 2006. *Sampling for Natural Resource Monitoring*. Springer, Berlin. 334 p.
- de Gruijter, J.J., ter Braak, C.J.F., 1990. Model-free estimation from spatial samples: A reappraisal of classical sampling theory. *Mathematical Geology* 22(4), 407–415.
- de Gruijter, J.J., ter Braak, C.J.F., 1992. Design-based versus model-based sampling strategies: Comment on R.J. Barnes' „bounding the required sample size for geologic site characterization”. *Mathematical Geology* 24(7), 859–864.
- Delmelle, E.M., Goovaerts, P., 2009. Second-phase sampling designs for non-stationary spatial variables. *Geoderma* 153, 205–216.
- Deutsch, C.V., Journel, A.G., 1998. *GSLIB: Geostatistical Software Library and User's Guide* (2nd Ed.). Oxford University Press, New York. 369 p.
- Dobai, A., Holndorfer, P., Dobos, E., 2012. Egy digitális talajtérképezési módszertan bemutatása meglévő adatbázisok felhasználásával egy választott mintaterületen. In: Lóki, J. (Szerk.) *Az elmélet és a gyakorlat találkozása a térinformatikában III*. Debreceni Egyetemi Kiadó, Debrecen. pp. 135–142.
- Dobos, E., Carré, F., Hengl, T., Reuter, H.I., Tóth, G., 2006. Digital Soil Mapping as a support to production of functional maps. Office for Official Publications of the European Communities, Luxembourg. 68 p.
- Dobos, E., Michéli, E., Baumgardner, M.F., 1997. Digitális magassági modell és a felszíntagoltsági mutató használata kis méretarányú talajtérképezésben. *Agrokémia és Talajtan* 46(1–4), 311–326.
- Dobos, E., Michéli, E., Baumgardner, M.F., Biehl, L., Helt, T., 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma* 97, 367–391.
- Dobos, E., Michéli, E., Montanarella, L., 2007. The population of a 500-m resolution soil organic matter spatial information system for Hungary. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.) *Digital Soil Mapping: An Introductory Perspective*. Elsevier, Amsterdam. pp. 487–495.
- Dokucsajev, V.V., 1883. *The Russian Chernozem* (Kaner, N., Trans.). Israel Program for Scientific Translations, Jerusalem.
- Douaik, A., Van Meirvenne, M., Tóth, T., 2004. Spatio-temporal kriging of soil salinity rescaled from bulk soil electrical conductivity. In: Sanchez-Vila, X., Carrera, J., Gómez-Hernández, J.J. (Eds.) *geoENV IV – Geostatistics for Environmental Applications*. Kluwer Academic Publishers, Dordrecht. pp. 413–424.
- Douaik, A., Van Meirvenne, M., Tóth, T., 2005. Soil salinity mapping using spatio-temporal kriging and Bayesian maximum entropy with interval soft data. *Geoderma* 128, 234–248.
- Dövényi, Z. (Szerk.), 2010. *Magyarország Kistájainak Katasztere*. MTA Földrajztudományi Kutatóintézet, Budapest. 876 p.
- Farkas, Cs., Rajkai, K., Kertész, M., Bakacsi, Zs., Meirvenne, M., 2008. Spatial variability of soil hydro-physical properties: A case study in Herceghalom, Hungary. In: Krasilnikov, P., Carré, F., Montanarella, L. (Eds.) *Soil Geography and Geostatistics: Concepts and Applications*. Office for Official Publications of the European Communities, Luxembourg. pp. 107–128.

- Farsang, A., 2016. A víz- és szélerózió szerepe a talaj humusz- és elemtartalmának horizontális átrendeződésében (MTA doktori értekezés). Szegedi Tudományegyetem, Szeged. 182 p.
- Fehér, Zs.Z., 2015a. A spatiotemporal stochastic framework of groundwater fluctuation analysis on the south-eastern part of the Great Hungarian Plain. *Journal of Environmental Geography* 8(3–4), 41–52.
- Fehér, Zs.Z., 2015b. Talajvízkészletek változásának geostatistikai alapú elemzése – a rendelkezésre álló információk természete és feldolgozása. *Hidrológiai Közlöny* 95(2), 15–31.
- Flatman, G.T., Yfantis, A.A., 1984. Geostatistical strategy for soil sampling: The survey and the census. *Environmental Monitoring and Assessment* 4, 335–349.
- Füst, A., 1982. A minták hatástávolsága és kutatási hálóméret kapcsolata az üledékes kőzetek lelőhelyein. *Bányászati és Kohászati Lapok. Bányászat* 115(3), 187–190.
- Füst, A., 1984. Adalékok az instacionárius paramétermezők kutatási optimumának meghatározásához. A Nehézipari Műszaki Egyetem Közleményei I. sorozat, *Bányászat* 32(1–4), 93–112.
- Füst, A., 1997. *Geostatisztika* (kézirat). Eötvös Kiadó, Budapest. 427 p.
- Füst, A., 2011. A természeti folyamatok monitoring hálózatainak tervezése és működtetése. *Bányászati és Kohászati Lapok. Bányászat* 144(2), 19–25.
- Füst, A., 2012. Calibration of monitoring systems. In: Geiger, J., Pál-Molnár, E., Malvić, T. (Eds.) *New Horizons in Central European Geomathematics, Geostatistics and Geoinformatics. GeoLitera*, Szeged. pp. 11–15.
- Füst, A., Geiger, J., 2010. Monitoringtervezés és -értékelés geostatistikai módszerekkel I. Szakértői véleményen alapuló, „igazoló” mintázás geostatistikai támogatása. *Földtani Közlöny* 140(3), 303–312.
- Gandin, L.S., 1965. *Objective Analysis of Meteorological Fields* (Translated from the Russian). Israel Program for Scientific Translations, Jerusalem. 242 p.
- Géczy, G., 1968. Magyarország Mezőgazdasági Területe. Akadémiai Kiadó, Budapest. 307 p.
- Geiger, J., 2006a. *Geostatisztika* (kézirat). Szegedi Tudományegyetem, Földtani és Őslénytani Tanszék, Szeged. 77 p.
- Geiger, J., 2006b. Szekvenciális gaussi szimuláció az övzátonytestek kisléptékű heterogenitásának modellezésében. *Földtani Közlöny* 136(4), 527–546.
- Geiger, J., 2007a. *Geomatematika*. JATEPress, Szeged. 116 p.
- Geiger, J., 2007b. *Statistikai alapú földtani mintavétel*. JATEPress, Szeged. 102 p.
- Geiger, J., 2012. Some thoughts on the pre- and post-processing in sequential gaussian simulation and their effects on reservoir characterization. In: Geiger, J., Pál-Molnár, E., Malvić, T. (Eds.) *New Horizons in Central European Geomathematics, Geostatistics and Geoinformatics. GeoLitera*, Szeged. pp. 17–34.
- Geiger, J., 2015. Some applications of Markov-type sequential Gaussian co-simulations. In: Horváth, J., Cvetković, M., Hatvani, I.G. (Eds.) 7th Croatian - Hungarian and 18th Hungarian Geomathematical Congress: "The Geomathematical Models: The Mirrors of Geological Reality or Science Fictions?". Hungarian Geological Society, Mórahalom. pp. 58–65.
- Geiger, J., Jakab, N., Csökmei, B., Horváth, Zs., Gellért, B., 2016. Statistical and geostatistical study of Rn and hydrocarbon components of a soil gas monitoring system: an application to surface hydrocarbon exploration. *Geologia Croatica* 69(2), 255–268.
- Geiger, J., Mucsi, L., 2005. A szekvenciális sztochasztikus szimuláció előnyei a talajvízszint kisléptékű heterogenitásának térképezésében. *Hidrológiai Közlöny* 85(2), 37–47.
- Glinka, K.D., 1908. *Soil Science: With 32 Figures and a Soil Map of the Globe*. A.F. Devrien Publishing House, St. Petersburg. 568 p.
- Goovaerts, P., 1998c. Ordinary cokriging revisited. *Mathematical Geology* 30(1), 21–42.
- Goovaerts, P., 1992. *Multivariate Geostatistical Tools for Studying Scale Dependent Correlation Structures and Describing Space-time Variations* (Ph.D. Thesis). Université Catholique de Louvain, Louvain-la-Neuve, 233p.
- Goovaerts, P., 1993. Spatial orthogonality of the principal components computed from coregionalized variables. *Mathematical Geology* 25(3), 281–302.
- Goovaerts, P., 1994. Comparative performance of indicator algorithms for modeling conditional probability distribution functions. *Mathematical Geology* 26(3), 389–411.
- Goovaerts, P., 1996. Stochastic simulation of categorical variables using a classification algorithm and simulated annealing. *Mathematical Geology* 28(7), 909–921.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York. 483 p.
- Goovaerts, P., 1998a. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of Soils* 27(4), 315–334.
- Goovaerts, P., 1998b. Accounting for estimation optimality criteria in simulated annealing. *Mathematical Geology* 30(5), 511–534.
- Goovaerts, P., 1999a. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89, 1–45.
- Goovaerts, P., 1999b. Using elevation to aid the geostatistical mapping of rainfall erosivity. *Catena* 34, 227–242.

- Goovaerts, P., 2000. Estimation or simulation of soil properties? An optimization problem with conflicting criteria. *Geoderma* 97, 165–186.
- Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. *Geoderma* 103, 3–26.
- Goovaerts, P., 2009. AUTO-IK: A 2D indicator kriging program for the automated non-parametric modeling of local uncertainty in earth sciences. *Computers & Geosciences* 35, 1255–1270.
- Goovaerts, P., Webster, R., Dubois, J.-P., 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environmental and Ecological Statistics* 4, 31–48.
- Gringarten, E., Deutsch, C.V., 2001. Teacher's aide variogram interpretation and modeling. *Mathematical Geology* 33(4), 507–534.
- Hägerstrand, T., 1965. A Monte Carlo approach to diffusion. *Européenes de Sociologie* 6, 43–67.
- Hagyó, A., Rajkai, K., 2004. A talajnedvesség-tartalom alakulása egy bükkös erdőben és benne kialakított lékekben. *Agrokémia és Talajtan* 53(1–2), 17–34.
- Hansen, M.H., Madow, W.G., Tepping, B.J., 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association* 78(384), 776–793.
- Hartemink, A.E., McSweeney, K. (Eds.), 2014. *Soil Carbon*. Springer, Switzerland. 506 p.
- Hatvani, I.G., Kovács, J., Korponai, J., 2012. Mintavételezési gyakoriság optimalizálása variogram függvénnel a Kis-Balaton Vízvédelmi Rendszer példáján. *Természetvédelmi Közlemények* 18, 202–210.
- Hengl, T., 2003. *Pedometric mapping: Bridging the gaps between conventional and pedometric approaches* (Ph.D. Thesis). Wageningen University, Wageningen. 214 p.
- Hengl, T., 2006. Finding the right pixel size. *Computers & Geosciences* 32, 1283–1298.
- Hengl, T., 2009. *A Practical Guide to Geostatistical Mapping* (2nd Ed.). University of Amsterdam, Amsterdam. 294 p.
- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLoS ONE* 10(6), 1–26.
- Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. *Computers & Geosciences* 33, 1301–1315.
- Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120, 75–93.
- Hengl, T., Reuter, H.I. (Eds.), 2008. *Geomorphometry: Concepts, Software, Applications*. Elsevier, Amsterdam. 796 p.
- Hengl, T., Rossiter, D.G., Stein, A., 2003. Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research* 41, 1403–1422.
- Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., McBratney, A.B. (Eds.) *GlobalSoilMap: Basis of the Global Spatial Soil Information System*. Taylor & Francis Group, London. pp. 335–340.
- Heuvelink, G.B.M., Brus, D.J., de Gruijter, J.J., 2007. Optimization of sample configurations for digital mapping of soil properties with universal kriging. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.) *Digital Soil Mapping: An Introductory Perspective*. Elsevier, Amsterdam. pp. 137–151.
- Heuvelink, G.B.M., Griffith, D.A., Hengl, T., Melles, S.J., 2013. Sampling design optimization for space-time kriging. In: Mateu, J., Müller, W.G. (Eds.) *Spatio-Temporal Design: Advances in Efficient Data Acquisition*. John Wiley & Sons, Ltd., Chichester. pp. 207–230.
- Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: past, present, and future. *Geoderma* 100, 269–301.
- Illés, G., Kovács, G., Heil, B., 2011. Nagyfelbontású digitális talajtérképezés a Vaskereszt erdőrezervátumban. *Erdészettudományi Közlemények* 1, 29–43.
- Isaaks, E.H., Srivastava, R.M., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, New York. 561 p.
- Jakab, G., Szabó, J., Madarász, B., Mészáros, E., Szalai, Z., 2014. Az erózió által mozgatott szén szerepe a szénkőrforgalomban. In: Jakab, G., Szalai, Z. (Szerk.) *Talajpusztulás Térben és Időben*. MTA CSFK FTI, Budapest. pp. 9–17.
- Jakab, G., Szabó, J., Szalai, Z., Mészáros, E., Madarász, B., Centeri, Cs., Szabó, B., Németh, T., Sipos, P., 2016. Changes in organic carbon concentration and organic matter compound of erosion-delivered soil aggregates. *Environmental Earth Sciences* 75(2), 1–11.
- Jakab, G., Takács, L., 2014. A területhasználat változásának felszínfejlődési vonatkozásai egy gödöllői mintaterület példáján. *Tájökológiai Lapok* 12(1), 49–61.
- Jenny, H., 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. McGraw-Hill, New York. 281 p.
- Journel, A.G., 1986. Geostatistics: Models and tools for the earth sciences. *Mathematical Geology* 18(1), 119–140.
- Journel, A.G., Huijbregts, Ch.J., 1978. *Mining Geostatistics*. Academic Press, New York. 600 p.

- Journel, A.G., Rossi, M.E., 1989. When do we need a trend model in kriging? *Mathematical Geology* 21(7), 715–739.
- Juhos, K., Szabó, Sz., Ladány, M., 2015. Influence of soil properties on crop yield: a multivariate statistical approach. *International Agrophysics* 29, 433–440.
- Juhos, K., Szabó, Sz., Ladány, M., 2016. Explore the influence of soil quality on crop yield using statistically-derived pedological indicators. *Ecological Indicators* 63, 366–373.
- Kantey, B.A., Williams, A.A.B., 1962. The use of engineering soil maps for road projects. *Transactions of the South African Institute of Civil Engineers* 4, 149–159.
- Kempen, B., 2011. Updating soil information with digital soil mapping (Ph.D. Thesis). Wageningen University, Wageningen. 218 p.
- Kempen, B., Heuvelink, G.B.M., Brus, D.J., Walvoort, D.J.J., 2014. Towards GlobalSoilMap.net products for The Netherlands. In: Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., McBratney, A.B. (Eds.) *GlobalSoilMap: Basis of the Global Spatial Soil Information System*. Taylor & Francis Group, London. pp. 85–90.
- Kerry, R., Oliver, M.A., 2003. Variograms of ancillary data to aid sampling for soil surveys. *Precision Agriculture* 4, 261–278.
- Kerry, R., Oliver, M.A., 2007. The analysis of ranked observations of soil structure using indicator geostatistics. *Geoderma* 140, 397–416.
- Kohán, B., 2014. GIS-alapú vizsgálat a Duna-Tisza közti homokhátság szárazodásának témakörében (Ph.D. értekezés). Eötvös Loránd Tudományegyetem, Budapest. 138 p.
- Kohán, B., Szalai, J., 2014. Spatial analysis of groundwater level monitoring network in the Danube-Tisza Interfluvium using semivariograms. *Hungarian Geographical Bulletin* 63, 379–400.
- Kovács, B., Szanyi, J., 2005. Hidrodinamikai és transzport modellezés II: Processing Modflow és Surfer for Windows környezetben. Miskolci Egyetem Műszaki Földtudományi Kar, Miskolc. 213 p.
- Kovács, J., Korponai, J., Kovácsné, Sz.I., Hatvani, I.G., 2012. Introducing sampling frequency estimation using variograms in water research with the example of nutrient loads in the Kis-Balaton Water Protection System (W Hungary). *Ecological Engineering* 42, 237–243.
- Kovács, K., Dobos, E., Gál Szabó, L., Vadnai, P., 2016. Talajképző laza üledékek térképezése domborzatmodell alapján. In: Balázs, B. (Szerk.) *Az elmélet és a gyakorlat találkozása a térinformatikában VII*. Debrecen Egyetemi Kiadó, Debrecen. pp. 245 – 250.
- Kovács, Z., Szabó, Sz., 2015. Talajtulajdonságok becslése a LUCAS talajadatbázis spektrumkönyvtárával. In: Boda, J. (Szerk.) *Az elmélet és a gyakorlat találkozása a térinformatikában VI*. Debrecen Egyetemi Kiadó, Debrecen. pp. 253–258.
- Kreybig, L., 1934. A talajtérképezés a tervgazdálkodás szolgálatában. *Mezőgazdasági Közlöny* 7, 217–228.
- Kreybig, L., 1937. A M. Kir. Földtani Intézet talajfelvételi, vizsgálati és térképezési módszere. *M. Kir. Földtani Intézet Évkönyve* 31, 147–244.
- Kreybig, L., 1950. A tájtermesztési talajtérképezés célja és irányelvei. *Agrokémia* 2, 17–28.
- Krige, D.G., 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa* 52, 119–139.
- Krige, D.G., 1960. On the departure of ore-value distributions from the lognormal model in South African gold mines. *Journal of The Southern African Institute of Mining and Metallurgy* 61, 231–244.
- Kummert, Á., Csillag, F., Szabó, J., Várallyay, Gy., Zilahy, P., 1989. A geographical information system for soil analysis and mapping: HunSIS (concepts and functionality). *Agrokémia és Talajtan* 38(3–4), 822–835.
- Lark, R.M., 2000a. Designing sampling grids from imprecise information on soil variability, an approach based on the fuzzy kriging variance. *Geoderma* 98, 35–59.
- Lark, R.M., 2000b. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science* 51, 717–728.
- Lark, R.M., 2002. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. *Geoderma* 105, 49–80.
- Lark, R.M., 2012. Towards soil geostatistics. *Spatial Statistics* 1, 92–99.
- Lark, R.M., Ferguson, R.B., 2004. Mapping risk of soil nutrient deficiency or excess by disjunctive and indicator kriging. *Geoderma* 118, 39–53.
- Lark, R.M., Lapworth, D.J., 2012. Quality measures for soil surveys by lognormal kriging. *Geoderma* 173–174, 231–240.
- Lin, Y.-P., Lin, W.-C., Li, M.-Y., Chen, Y.-Y., Chiang, L.-C., Wang, Y.-C., 2014. Identification of spatial distributions and uncertainties of multiple heavy metal concentrations by using spatial conditioned Latin hypercube sampling. *Geoderma* 230–231, 9–21.
- Marbut, C.F., Bennett, H.H., Lampham, J.E., Lapham, M.H., 1913. *Soils of the United States*. USDA Bureau of Soils Bulletin No.96., Washington D.C. 791 p.

- Matérn, B., 1960. Spatial Variation (Report). Reports of the Forest Research Institute of Sweden (Meddelanden från Statens Skogsforskningsinstitut) 49, 144 p.
- Matheron, G., 1963. Principles of geostatistics. *Economic Geology* 58, 1246–1266.
- Matheron, G., 1965. *Les Variables Régionalisées et Leur Estimation*. Masson, Paris. 306 p.
- Matheron, G., 1971. *The Theory of Regionalized Variables and Its Applications*. École nationale supérieure des mines, Paris. 211 p.
- McBratney, A.B., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203–213.
- McBratney, A.B., Mendonca-Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- McBratney, A.B., Odeh, I.O.A., Bishop, T.F.A., Dunbar, M.S., Shatar, T.M., 2000. An overview of pedometric techniques for use in soil survey. *Geoderma* 97, 293–327.
- McBratney, A.B., Webster, R., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables–II.: Program and Examples. *Computers & Geosciences* 7(4), 335–365.
- McBratney, A.B., Webster, R., 1983. Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy. *Journal of Soil Science* 34, 137–162.
- McBratney, A.B., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables–I.: Theory and method. *Computers & Geosciences* 7(4), 331–334.
- Melles, S.J., Heuvelink, G.B.M., Twenhöfel, C.J.W., Stöhlker, U., 2008. Sampling optimization trade-offs for long-term monitoring of gamma dose rates. In: Gervasi, O., Murgante, B., Laganá, A., Tanar, D., Mun, Y., Gavrilova, M.L. (Eds.) *Computational Science and Its Applications – ICCSA 2008*. Springer, Berlin. pp. 444–458.
- Melles, S.J., Heuvelink, G.B.M., Twenhöfel, C.J.W., van Dijk, A., Hiemstra, P.H., Baume, O., Stöhlker, U., 2011. Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computers & Geosciences* 37, 280–288.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics* 21(6), 1087–1092.
- Meul, M., Van Meirvenne, M., 2003. Kriging soil texture under different types of nonstationarity. *Geoderma* 112, 217–233.
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32, 1378–1388.
- Minasny, B., McBratney, A.B., 2007a. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. *Geoderma* 140, 324–336.
- Minasny, B., McBratney, A.B., 2007b. Latin hypercube sampling as a tool for digital soil mapping. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.) *Digital Soil Mapping: An Introductory Perspective*. Elsevier, Amsterdam. pp. 153–165.
- Minasny, B., McBratney, A.B., 2010. Conditioned Latin hypercube sampling for calibrating soil sensor data to soil properties. In: Viscarra Rossel, R.A., McBratney, A.B., Minasny, B. (Eds.) *Proximal Soil Sensing*. Springer, Dordrecht. pp. 111–119.
- Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, 301–311.
- Molnár, Z., 2009. Felszín alatti vizek modellezési tartományának és bearányosítási időszakának kijelölése geostatistikai alapokon (Ph.D. értekezés). Budapesti Műszaki és Gazdaságtudományi Egyetem, Budapest. 108 p.
- Moore, I.D., Gessler, P.E., Nielsen, G.A., Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal* 57(2), 443–452.
- Morse, R.K., Thornburn, T.H., 1961. Reliability of soil units. *Proceedings of the 5th International Conference on Soil Mechanics and Foundation Engineering* 1, 259–262.
- Mucsi, L., Geiger, J., Malvić, T., 2013. The advantages of using sequential stochastic simulations when mapping small-scale heterogeneities of the groundwater level. *Journal of Environmental Geography* 6(3–4), 39–47.
- Myers, D.E., 1989. To be or not to be... Stationarity? That is the question. *Mathematical Geology* 21(3), 347–362.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma* 63, 197–214.
- Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* 67, 215–226.
- Oertli, J.J., Rajkai, K., 1988. Spatial variability of soil properties and plant cover-age on alkali soils of the Hungarian Puszta. In: Adam, M., Szabolcs, I., Miljkovic, N., Plamenac, N. (Eds.) *Proc. of the International Symposium on Solonch Soil: Problems, Properties, Utilization*. Yugoslav Society of Soil Science, Osijek. pp. 156–161.
- Oliver, M.A., Webster, R., 1986. Combining nested and linear sampling for determining the scale of spatial variation of regionalized variables. *Geographical Analysis* 18, 227–242.

- Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113, 56–69.
- Oliver, M.A., Webster, R., 2015. *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer, Cham. 100 p.
- Pahlavan Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., Bogaert, P., 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma* 232–234, 97–106.
- Pannatier, Y., 1996. *Variowin: Software for Spatial Data Analysis in 2D*. Springer, New York. 91 p.
- Papritz, A., Schwierz, C., 2016. Package „georob”. CRAN. 64 p.
- Pásztor, L., Bakacsi, Zs., Laborczi, A., Szabó, J., 2013a. Kategória típusú talajtérképek térbeli felbontásának javítása kiegészítő talajtani adatok és adatbányászati módszerek segítségével. *Agrokémia és Talajtan* 62, 205–218.
- Pásztor, L., Dobos, E., Szatmári, G., Laborczi, A., Takács, K., Bakacsi, Zs., Szabó, J., 2014b. Application of legacy soil data in digital soil mapping for the elaboration of novel, countrywide maps of soil conditions. *Agrokémia és Talajtan* 63(1), 79–88.
- Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Bakacsi, Zs., Szabó, J., 2016b. Variations for the Implementation of SCORPAN's „S”. In: Zhang, G.-L., Brus, D.J., Liu, F., Song, X.-D., Lagacherie, P. (Eds.) *Digital Soil Mapping Across Paradigms, Scales and Boundaries*. Springer, Singapore. pp. 331–342.
- Pásztor, L., Laborczi, A., Takács, K., Szatmári, G., Dobos, E., Illés, G., Bakacsi, Zs., Szabó, J., 2015. Compilation of novel and renewed, goal oriented digital soil maps using geostatistical and data mining tools. *Hungarian Geographical Bulletin* 64(1), 49–64.
- Pásztor, L., Szabó, J., Bakacsi, Zs., 2002. GIS Processing of Large-scale Soil Maps in Hungary. *Agrokémia és Talajtan* 51(1–2), 273–282.
- Pásztor, L., Szabó, J., Bakacsi, Zs., 2010a. Application of the Digital Kreybig Soil Information System for the delineation of naturally handicapped areas in Hungary. *Agrokémia és Talajtan* 59, 47–56.
- Pásztor, L., Szabó, J., Bakacsi, Zs., 2010b. Digital processing and upgrading of legacy data collected during the 1:25.000 scale Kreybig soil survey. *Acta Geodaetica et Geophysica Hungarica* 45, 127–136.
- Pásztor, L., Szabó, J., Bakacsi, Zs., Laborczi, A., 2013b. Elaboration and applications of spatial soil information systems and digital soil mapping at Research Institute for Soil Science and Agricultural Chemistry of the Hungarian Academy of Sciences. *Geocarto International* 28(1), 13–27.
- Pásztor, L., Szabó, J., Bakacsi, Zs., Laborczi, A., Dobos, E., Illés, G., Szatmári, G., 2014a. Elaboration of novel, countrywide maps for the satisfaction of recent demands on spatial, soil related information in Hungary. In: Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A., McBratney, A.B. (Eds.) *GlobalSoilMap: Basis of the Global Spatial Soil Information System*. Taylor & Francis Group, London. pp. 207–212.
- Pásztor, L., Szabó, K.Zs., Szatmári, G., Laborczi, A., Horváth, Á., 2016a. Mapping geogenic radon potential by regression kriging. *Science of the Total Environment* 12, 883–891.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30, 683–691.
- Pebesma, E.J., Wesseling, C.G., 1998. gstat: A program for geostatistical modelling, prediction and simulation. *Computers & Geosciences* 24(1), 17–31.
- Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation – An example from Scotland. *Geoderma* 232–234, 284–299.
- Poggio, L., Gimona, A., Brewer, M., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. *Geoderma* 209–210, 1–14.
- Poggio, L., Gimona, A., Brown, I., Castellazzi, M., 2010. Soil available water capacity interpolation and spatial uncertainty modelling at multiple geographical extents. *Geoderma* 160, 175–188.
- Prescott, J.A., 1933. The soil zones of Australia. *Soil Resource* 3, 133–145.
- R Development Core Team, 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rajkai, K., 2001. Modellézés és modellhasználat a talajtani kutatásban (Szemle). *Agrokémia és Talajtan* 50(3–4), 469–508.
- Rajkai, K., Marchand, D., Oertli, J.J., 1988. Study of the Spatial Variability of Soil properties on Alkali Soils. In: Adam, M., Szabolcs, I., Miljkovic, N., Plamenac, N. (Eds.) *Proc. of the International Symposium on Solonchak Soil: Problems, Properties, Utilization*. Yugoslav Society of Soil Science, Osijek. pp. 150–155.
- Rajkai, K., Rydén, B.E., 1992. Measuring areal soil moisture distribution with the TDR method. *Geoderma* 52, 73–85.
- Rakonczai, J., Fehér, Zs.Z., 2015. A klímaváltozás szerepe az Alföld talajvízkészleteinek időbeli változásaiban. *Hidrológiai Közlemények* 95, 1–15.
- Romary, T., de Fouquet, Ch., Malherbe, L., 2011. Sampling design for air quality measurement surveys: An optimization approach. *Atmospheric Environment* 45, 3613–3620.
- Rossi, M.E., Deutsch, C.V., 2014. *Mineral Resource Estimation*. Springer, Dordrecht. 332 p.

- Samuel-Rosa, A., Heuvelink, G.B.M., Vasques, G.M., Anjos, L.H.C., 2015. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma* 243–244, 214–227.
- Särndal, C.-E., 1978. Design-based and model-based inference in survey sampling. *Scandinavian Journal of Statistics* 5(1), 27–52.
- Simbahan, G.C., Dobermann, A., Goovaerts, P., Ping, J., Haddix, M.L., 2006. Fine-resolution mapping of soil organic carbon based on multivariate secondary data. *Geoderma* 132, 471–489.
- Stefanovits, P., Filep, Gy., Fülek, Gy., 2010. Talajtan (második kiadás). Mezőgazda Kiadó, Budapest. 470 p.
- Stefanovits, P., Szűcs, L., 1961. Magyarország genetikai talajtérképe. OMMI, Budapest. 103 p.
- Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., de Remy de Courcelles, V., Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M., 2013. The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems and Environment* 164, 80–99.
- Szabó, J., Pásztor, L., Bakacsi, Zs., 2005. Egy országos, átnézetes, térbeli talajinformációs rendszer kiépítésének igénye, lehetősége és lépései. *Agrokémia és Talajtan* 54, 41–58.
- Szabó, J., Pásztor, L., Koós, S., László, P., Bakacsi, Zs., Laborcz, A., Takács, K., Szatmári, G., Pirkó, B., Nagy, A., Bialkó, T., Dobos, E., Szabóné Kele, G., Havasné Tátrai, É., Tihanyi, K., Mandula, R., Lukácsy, Gy., 2016. Termőhely felvételezés és térképezés Tokaj Hegyalján. In: Kátai, J., Sándor, Zs., Szász, G. (Szerk.) Talajtani Vándorgyűlés: „Okszerű talajhasználat - Talajvédelem”; Az előadások és a posztterek összefoglalója. Magyar Talajtani Társaság, Debrecen. pp. 33.
- Szabó, Sz., 2014. Tájmetriai módszerek kritikai alkalmazása a tájanalízisben (MTA doktori értekezés). Debreceni Egyetem, Debrecen. 115 p.
- Szabolcs, I. (Szerk.), 1966. A genetikai üzemi talajtérképezés módszerekönyve. OMMI, Budapest. 428 p.
- Szalai, Z., Szabó, J., Kovács, J., Mészáros, E., Albert, G., Centeri, Cs., Szabó, B., Madarász, B., Zacháry, D., Jakab, G., 2016. Redistribution of soil organic carbon triggered by erosion at field scale under subhumid climate, Hungary. *Pedosphere* 26(5), 652–665.
- Szalóki, A., Somodi, Zs., 2014. Talajok nehézfém-tartalmának heterogenitás-vizsgálata geostatistikai módszerek segítségével. In: Balázs, B. (Szerk.) Az elmélet és a gyakorlat találkozása a térinformatikában V. Debrecen Egyetemi Kiadó, Debrecen. pp. 359–368.
- Szalóki, A., Szabó, Sz., 2016. Talajtulajdonságok térbeli eloszlásának vizsgálata determinisztikus és sztochasztikus módszerek segítségével. In: Balázs, B. (Szerk.) Az elmélet és a gyakorlat találkozása a térinformatikában VII. Debrecen Egyetemi Kiadó, Debrecen. pp. 457–462.
- Szatmári, G., 2011. Az USLE talajerózió modell alkalmazása mezőföldi mintaterületen (Diplomamunka). Szegedi Tudományegyetem, Szeged. 65 p.
- Szatmári, G., 2014. Optimization of sampling configuration by spatial simulated annealing for mapping soil variables. In: Cvetković, M., Novak Zelenika, K., Geiger, J. (Eds.) 6th Croatian-Hungarian and 17th Hungarian Geomathematical Congress: „Geomathematics – from Theory to Practice”. Croatian Geological Society, Zagreb. pp 105–111.
- Szatmári, G., Barta, K., 2012. Az erózió, az erózió-veszélyeztetettség és a területhasznosítás kapcsolata mezőföldi területen. *Agrokémia és Talajtan* 61(1), 41–56.
- Szatmári, G., Barta, K., 2013. Csernozjom talajok szervesanyag-tartalmának digitális térképezése erózióval veszélyeztetett mezőföldi területen. *Agrokémia és Talajtan* 62(1), 47–60.
- Szatmári, G., Barta, K., Farsang, A., Pásztor, L., 2015a. Testing a sequential stochastic simulation method based on regression kriging in a catchment area in Southern Hungary. *Geologia Croatica* 68(3), 273–283.
- Szatmári, G., Barta, K., Pásztor, L., 2015b. An application of a spatial simulated annealing sampling optimization algorithm to support digital soil mapping. *Hungarian Geographical Bulletin* 64(1), 35–48.
- Szatmári, G., Barta, K., Pásztor, L., 2016. Multivariate sampling design optimization for digital soil mapping. In: Zhang, G.-L., Brus, D.J., Liu, F., Song, X.-D., Lagacherie, P. (Eds.) *Digital Soil Mapping Across Paradigms, Scales and Boundaries*. Springer, Singapore. pp. 77–87.
- Szatmári, G., Laborcz, A., Illés, G., Pásztor, L., 2013. A talajok szervesanyag-készletének nagyléptékű térképezése regresszió krigeléssel Zala megye példáján. *Agrokémia és Talajtan* 62(2), 219–234.
- Szatmári, G., Pásztor, L., 2016. Geostatistika a talajtérképezésben (Szemle). *Agrokémia és Talajtan* 65(1), 95–114.
- Szidarovszky, F., 1983a. Multiobjective observation network design for regionalized variables. *International Journal of Mining Engineering* 1, 331–342.
- Szidarovszky, F., 1983b. Optimal observation network in geostatistics and underground hydrology. *Applied Mathematical Modelling* 7(1), 25–32.
- Thyll, Sz. (Szerk.), 1997. Talajvédelem és vízrendezés dombvidéken. Mezőgazda Kiadó, Budapest. 350 p.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit Region. *Economic Geography* 46, 234–240.

- Tóth, G., Hengl, T., Hermann, T., Makó, A., Kocsis, M., Tóth, B., Berényi Üveges, J., 2015. Magyarország mezőgazdasági területeinek talajtulajdonosság-térképei (JRC Technical Reports). Office for Official Publications of the European Communities, Luxembourg. 63 p.
- Tóth, G., Hermann, T., Szatmári, G., Pásztor, L., 2016. Maps of heavy metals in the soils of the European Union and proposed priority areas for detailed assessment. *Science of the Total Environment* 565, 1054–1062.
- Tóth, T., Kertész, M., 1996. Application of soil-vegetation correlation to optimal resolution mapping of solonchic rangeland. *Arid Soil Research and Rehabilitation* 10, 1–12.
- Tóth, T., Várallyay, Gy., 2001. Egy mintaterület talajának variabilitása a sófelhalmozódás tényezői szerint. *Agrokémia és Talajtan* 50, 19–34.
- van Groenigen, J.W., 1999. Constrained Optimisation of Spatial Sampling: A Geostatistical Approach (Ph.D. Thesis). Wageningen Agricultural University and ITC, Enschede. 149 p.
- van Groenigen, J.W., 2000. The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma* 97, 223–236.
- van Groenigen, J.W., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. *Geoderma* 87, 239–259.
- van Groenigen, J.W., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* 27, 1078–1086.
- Várallyay, Gy., 1989a. Soil mapping in Hungary. *Agrokémia és Talajtan* 38(3–4), 695–714.
- Várallyay, Gy., 1989b. Mapping of hydrophysical properties and moisture regime of soils. *Agrokémia és Talajtan* 38(3–4), 800–817.
- Várallyay, Gy., 2012. Talajtérképezés, talajtani adatbázisok. *Agrokémia és Talajtan* 61, 249–268.
- Várallyay, Gy., Szűcs, L., 1978. Magyarország új, 1:100 000 méretarányú talajtérképe és felhasználási lehetőségei. *Agrokémia és Talajtan* 27(3–4), 267–288.
- Várallyay, Gy., Szűcs, L., Murányi, A., Rajkai, K., Zilahy, P., 1979. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100 000 méretarányú térképe I. *Agrokémia és Talajtan* 28(3–4), 363–384.
- Várallyay, Gy., Szűcs, L., Murányi, A., Rajkai, K., Zilahy, P., 1980a. Magyarország termőhelyi adottságait meghatározó talajtani tényezők 1:100 000 méretarányú térképe II. *Agrokémia és Talajtan* 29(1–2), 35–76.
- Várallyay, Gy., Szűcs, L., Rajkai, K., Zilahy, P., Murányi, A., 1980b. Magyarországi talajok vízgazdálkodási tulajdonságainak kategóriarendszere és 1:100 000 méretarányú térképe. *Agrokémia és Talajtan* 29(1–2), 77–112.
- Vašát, R., Heuvelink, G.B.M., Borůvka, L., 2010. Sampling design optimization for multivariate soil mapping. *Geoderma* 155, 147–153.
- Wágenhoffer, A., 2013. 3D facies analysis with combination of classic and stochastic sedimentological approach. In: Geiger, J., Pál-Molnár, E., Malvić, T. (Eds.) *Theories and Applications in Geomathematics*. GeoLitera, Szeged. pp. 151–160.
- Waltner, I., 2013. Hazai térképi és más talajadatok nemzetközi digitális térképekbe és adatbázisokba illesztésének megalapozása (Ph.D. értekezés). Szent István Egyetem, Gödöllő. 159 p.
- Wang, J.-F., Stein, A., Gao, B.-B., Ge, Y., 2012. A review of spatial sampling. *Spatial Statistics* 2, 1–14.
- Warrick, A.W., Myers, D.E., 1987. Optimization of sampling locations for variogram calculations. *Water Resources Research* 23(3), 496–500.
- Webster, R., 1989. Recent achievements in geostatistical analysis of soil. *Agrokémia és Talajtan* 38(3–4), 519–536.
- Webster, R., 1994. The development of pedometrics. *Geoderma* 62, 1–15.
- Webster, R., 1997. Regression and functional relations. *European Journal of Soil Science* 48(3), 557–566.
- Webster, R., 2000. Is soil variation random? *Geoderma* 97, 149–163.
- Webster, R., 2015. Technological developments for spatial prediction of soil properties, and Danie Krige's influence on it. *Journal of The Southern African Institute of Mining and Metallurgy* 115, 165–172.
- Webster, R., Beckett, P.H.T., 1970. Terrain classification and evaluation using air photography: a review of recent work at Oxford. *Photogrammetria* 26, 51–75.
- Webster, R., Burgess, T.M., 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. *Journal of Soil Science* 31, 505–524.
- Webster, R., Cuanalo de la C., H.E., 1975. Soil transect correlograms of north Oxfordshire and their interpretation. *Journal of Soil Science* 26, 176–194.
- Webster, R., Lark, R.M., 2013. *Field Sampling for Environmental Science and Management*. Routledge Taylor & Francis Group, London. 192 p.
- Webster, R., Oliver, M.A., 1989. Optimal interpolation and isarithmic mapping of soil properties. VI. Disjunctive kriging and mapping the conditional probability. *Journal of Soil Science* 40, 497–512.
- Webster, R., Oliver, M.A., 2007. *Geostatistics for Environmental Scientists* (2nd Ed.). John Wiley & Sons Ltd., Chichester. 318 p.
- Whitney, M., 1909. *Soils of the United States*. USDA Bureau of Soils Bulletin No.55., Washington D.C. 243 p.

- Wischmeier, W.H., Smith, D.D., 1978. Predicting rainfall erosion losses: A guide to conservation planning. U.S. Department of Agriculture, Washington. 58 p.
- Yfantis, E.A., Flatman, G.T., Behar, J.V., 1987. Efficiency of kriging estimation for square, triangular, and hexagonal grids. *Mathematical Geology* 19(3), 183–205.

XI. MELLÉKLETEK