



Nyugat-magyarországi Egyetem

Simonyi Karoly Műszaki, Faanyagtudományi és Művészeti Kar

Informatikai és Gazdasági Intézet

A döntéstámogatás kapcsolata a Véletlen Összefüggéssel

Tézisfüzet

Bencsik Gergely

Konzulens:

Dr. Bacsárdi László

2016

1. Bevezetés

Az emberiség mindig is kereste a választ a honnan jövünk, mik vagyunk, hová megyünk kérdésekre. Paul Gauguin kérdéseire talán a tudomány fogja megadni a választ. Minden tudományterületen folynak elméleti és tapasztalati kutatások, hogy leírják a természetben zajló folyamatokat. A cél minden esetben, hogy jobban megismerjük a minket körbevevő univerzumot. A kutatók azonban átfogalmazzák Gauguin kérdéseit, úgy, mint például hogy összefügg-e két változó, két független adathalmaz korrelál-e egymással, egyik paraméternek van-e valamilyen hatása a másik paraméterre, független változók alapján milyen jóslást mutatnak az eredmények a függő változóra vonatkozóan. Ugyanakkor minden esetben a keresett tudás visszavezethető Gauguin kérdéseire.

A kísérletek során mért adatok kritikus szerepet töltenek be a tényeken alapuló döntések meghozatalában. Ezen adatsorok egységes tárolása nem minden esetben triviális, különösen, ha más céllal, ebből fakadóan más környezetben történt adatszerzésről van szó. Ebben az esetben az adatok struktúrájukban, integritási szintjükben mások lehetnek, amelyek nehezítik az egységes adatbázisba vagy adattárházba történő beillesztésüket. Az adatok kinyerése és értelmezése többlépcsős folyamat. A nyers adatokon értelmezett szűkítések és transzformációk ezután tipikusan már egy másik rendszerben kerülnek megvalósításra, végül a már transzformált adatokon értelmezzük a vizsgálati módszert, amely egyrészt adott elemzési területhez (matematika, statisztika, adatbányászat) kapcsolódik, másrészt általában ugyancsak különböző rendszerben implementálnak. Ugyanakkor egyértelmű igény van több tudományterületen mért adatok összevetésére, interdiszciplináris kapcsolatok kutatására.

Kutatásom első részében a fentiekből indultam ki és adtam egy általam kifejlesztett egységes megoldást. Létrehoztam egy univerzális adatbázis struktúrát, amely a különböző forrásokból érkező heterogén adatokat fogadni és összefűzni is képes. Az univerzális lekérdező felület segítségével az egyes módszerek különféle bemeneti struktúráját állíthatjuk össze. Kutatásom során tanulmányoztam az adatbázisok és adattárházak evolúciós vonalát, amelynek eredményeképpen egy új állomást definiáltam. Az univerzális döntéstámogató keretrendszer funkciói kibővítik a hagyományos adatbázis és adattárház műveleteket, amelyek: (1) új adatsor létrehozása [dinamikusan az adatbázis struktúrában], (2) adott adatsor összefűzése, (3) különböző adatsorok összefűzése adott szemantikai összerendelést követve. A logikai és megjelenítési szintű univerzalitás elérése nehézkes, ezért a szakirodalomban követett „add-on” technikákat alkalmaztam, ugyanakkor maximálisan törekedtem a könnyű bővíthetőségre. A rendszer képességeit három különböző tudományterületen végzett elemzéssel mutatom be.

Disszertációm második része az univerzális elemző keretrendszerben történő elemzések tapasztalatai alapján a módszerek és adatok karakterisztikájára vonatkozik. Az adatelemzési folyamat során a modell és a kapcsolatok megtalálása már önmagában eredményt jelent, ezután következik a – lehetőleg minél pontosabb – jóslás a jövőre nézve. A különböző adatelemzési módszerek önmagukban nem képesek értelmezni az eredményeket, vagyis az adott matematikai képlettel csak kiszámoljuk az eredményeket. A módszerek önmagukban nem ítélik meg, a megállapításokat, miszerint elfogadjuk vagy sem az adott összefüggést, mindig egy elemző személy teszi meg. Kutatásomban azt vizsgáltam, hogy mi történik akkor, ha az elemző által keresett összefüggés matematikailag bizonyítható, de az adott elemzési döntés mégsem helytálló, mivel a matematikai összefüggés a mért adatok olyan véletlenszerűségéből adódik, amelyet az elemző személy sem ismerhet. Az olyan összefüggéseket, amelyek a véletlenszerűség következményeként létrejönnek, Véletlen Összefüggéseknek neveztem el.

2. A probléma definiálása

2.1. Az általános probléma megfogalmazása

Manapság az adatok világában élünk, az adatgyűjtés javarészt automatizálódott. A state-of-the-art jellegű standard kutatási módszertanok definiálják a kísérletek, így az adatgyűjtés folyamatait is [1, 2]. Mivel az egyes tudományterületek eltérhetnek az általános kutatási folyamatoktól, ezért az egyes tudományterületeknek megfelelően specializálódott módszertanok is megjelentek [3, 4]. Általánosságban egy kutatási módszertan az adatelőkészítéssel (adatgyűjtés, tisztítás és/vagy adat transzformáció) kezdődik, ezután kiválasztjuk az elemző módszert, majd végül interpretáljuk az eredményeket. A megnövekedett adattömegek miatt egy új fogalom is megjelent Big Data elnevezéssel, amely még szélesebb körbe terjeszti ki az elemzési lehetőségeinket. Ugyanakkor kutatásom során azt tapasztaltam, hogy a tudományos irodalom mindezek ellenére nagyon sok egymásnak ellentmondó eredményt tartalmaz.

Biológia területén a tintahalak méretének kutatása ellentmondó eredményeket állít. Jackson és Moltaschaniwskyj azt állítják, hogy a tintahalak mérete nagyobb, mint korábban [5]. Egy másik kutatás azonban azt állítja, hogy ezen tintahalak mérete csökken [6]. Zavaleta et al. azt állítják, hogy a füves területek nedvesebbek, mint korábban [7]. Liu et al. szerint a füves területeknek kevesebb a nedvességtartalma, mint korábban [8]. Church és White a tengerszintre vonatkozóan szignifikáns növekedést mutattak ki [9]. Ezzel szemben Houston és Dean a tengerszint csökkenését állítják [10]. Egy kutatócsoport szerint az indiai rizsföldek növekednek [11], míg egy másik szerint csökkennek [12].

Az orvostudományban a só egészségre gyakorolt hatásának kutatása mindig ellentmondó eredményeket generál. Vannak olyan tanulmányok, amelyek ajánlják a só fogyasztást, és nem hozzák összefüggésbe olyan betegségekkel, mint például a magas vérnyomás [13]. Mások szerint a sófogyasztás nem csak magas vérnyomást, hanem vesebetegséget is okoz [14]. Egy Kelet-afrikai ország maláriával erősen fertőzött terület. Két ellentmondó eredményt publikáltak a maláriás betegek számát illetően. Egy beszámoló arról ír, hogy a maláriás betegek száma nő [15], míg Nkurunziza és Pilz ennek ellenkezőjét állítják [16]. További maláriához kapcsolódó kutatásokat is végeztek globális szinten. Martens et al. 160 millió maláriás beteget jósolnak világszerte 2080-ra [17], míg más kutatások a maláriás betegek számának visszaszorításáról számolnak be [18].

Az erdőzetben Fowler és Ekström azt állítják, hogy az Egyesült Királyság területein több eső esett az elmúlt években, mint korábban [19]. Burke et al. nem csak szárazságról írnak az Egyesült Királyságra vonatkozóan, hanem további szárazságot jósolnak a következő évekre [20]. Held et al. azt állítják, hogy Sahel tartományban kevesebb a csapadék [21]. Sahel egy átmeneti terület a Szahara és a szavanna között. Ugyanakkor, ugyanerre a területre vonatkozóan egy másik kutatócsoport több csapadékot állapított meg [22]. Giannini eredménye szerint Sahel tartomány kevesebb és több csapadékot is kaphat [23]. Crimmins et al. azt állították, hogy a növények az alacsonyabb lejtők felé mozdulnak el [24], miközben Grace et al. ellentétes eredményre jutottak: a növények felfelé mozognak [25]. Dueck et al. a növényi eredetű metán gázok kibocsátásával foglalkoztak. Eredményük szerint a növények metán kibocsátása nem szignifikáns [26]. Keppler et al. szerint a növények metán kibocsátása szignifikáns és a növényeket a Föld globális metán mennyiségének jelentős okozóiként nevezték meg [27]. Ellentétes eredmények tapasztalhatóak a levélindex tekintetében is. Siliang et al. levél index növekedést mutattak ki [28], miközben egy másik kutatás a levél index csökkenéséről számol be [29]. Jaramillo et al. szerint a Latin-amerikai esőerdők szén-dioxid szintje növekszik [30]. Ezzel szemben Salazar et al. azt jósolják, hogy ezen őserdők szén-dioxid szintje

csökkenni fog [31]. Egy kutatócsoport több esőzészről számol be Afrikát illetően [32], miközben egy másik kutatás szerint csökken az esőzés mértéke Afrikában [33]. Flannigen et al. szerint a boreális erdőkben az erdőtüzek száma csökken [34], ugyanakkor Kasischke et al. kutatási eredményeiben az erdőtüzek számának növekedéséről olvashatunk ezen területeket illetően [35]. A madarak vándorlásáról három egymásnak ellenmondó eredményt olvashatunk. Az egyik kutatás szerint a madarak vándorlási ideje rövidül [36]. A második szerint növekszik [37]. A harmadik kutatás azt állítja, hogy a madarak vándorlási idejének meghatározásának módszere idejét múlt [38]. Két ellentétes eredmény született az Amazonas esőerdeire vonatkozóan is [39, 40].

Földtudományok területén Schindell et al. azt állították, hogy az északi féltekén a telek egyre melegebbek [41]. Egy másik vélemény szerint ugyanezen féltekén a telek egyre hidegebbek [42]. Knippertz et al. a szelek sebességével foglalkoztak, és a sebesség növekedését állapították meg [43]. Egy másik kutatócsoport szerint a szelek sebessége 10-15%-kal csökkent [44]. Egy harmadik kutatás szerint a szelek sebessége először nő, aztán csökken [45]. Több kutatás foglalkozik az Alpok törmelékfolyásainak elemzésével. Az egyik kutatás szerint a törmelékfolyások mérete növekedhet [46], ugyanakkor egy másik szerint csökkenhet [47]. Egy harmadik kutatás szerint csökkenthet, majd növekedhet [48].

Nosek et al. 98 + 2 pszichológiai kutatás eredményeit akarták reprodukálni [49]. A kísérletek 39%-ban kaptak az eredeti eredményekhez hasonló eredményeket. Minden egyéb esetben más, ellentmondó eredményeket tapasztaltak az eredeti eredmények és a reprodukált eredmények között. Az eredeti szerzők is szerepet vállaltak a reprodukáltság kísérletében, így biztosítva a régi és az új elemzések azonos kutatási módszertanát. A 270 szerzős cikk főbb konklúziói a következők voltak:

- *A legtöbb tudományos folyóirat bírálati folyamata kevésbé hatékony.* Kevésbé azt ellenőrzik, hogy az eredmények jók vagy rosszak, és sokszor nem szeretnék meghazudtolni az eredményeket. Ez a megközelítés a mi megközelítésünkhöz is kapcsolódik: nem vonom kétségbe tényleges összefüggések meglétét.
- *Több csalás gyanús eredményeket fedeztek fel.* A szerzők munkája egy nagyobb volumenű projekt része. A projekt további részeinél több csalás gyanús eredményt találtak.
- *A többi tudományterületen is probléma az eredmények reprodukálhatóságának problémája, nem csak a pszichológiában.* Kutatásunkban több ellentmondó eredményt is találtunk, csakúgy, mint a szerzők is több olyan eredményt mutattak be, amelyek megkérdőjelezhetőek.
- *A szerzők a kutatók közötti kooperációra buzdítanak.* Tudományos adattárházak építését javasolják, ahol az eredményekhez használt adatok, az adatokat elemző módszerek mindenki számára elérhetőek. Mivel az általam fejlesztett Univerzális Elemző Keretrendszer (Universal Decision Support System, UDSS) koncepció támogatja különböző tudományterületek adatainak integrálását, ezért az UDSS alkalmas lehet tudományos adattárházak építésére.

A fenti kutatások azonos tématerületeket érintenek, de különböző, sokszor egymásnak ellentmondó eredményeket mutatnak. Ez a tény is azt mutatja, hogy a döntéstámogatási folyamatok egyáltalán nem triviálisak. Kutatásomban az inkonzisztens eredmények keletkezésének lehetséges körülményeire fókuszáltam. Ez nem azt jelenti, hogy egy adott problémát nem lehet több nézőpontból megvizsgálni. Én inkább azt fogalmaztam meg, hogy léteznek olyan esetek, amikor az eredmények szimplán véletlenszerűségből adódnak. Más szavakkal, bizonyos paraméterek mentén, az adatok alapján (pl.: mért adatok tartománya, várható érték, szórás) és az elemző módszerek alapján (módszerek száma, szélsőséges adatelemek szűrése) olyan környezetet teremtünk, ahol a döntés eleve meghatározott (például,

adatsorok, amelyek hatással vannak egymásra, avagy sem, összefüggnek vagy nem függenek össze). A céloim ezen ellentmondó eredmények vizsgálata volt, arra a kérdésre kerestem a választ, hogy hol és hogyan keletkezhetnek ilyen ellentétes eredmények. Az elért eredményeim fényében egy új fogalmat, a Véletlen Összefüggéseket (Random Correlations, RC) vezettem be.

RC mindegyik tudományterületen jelentkezik. Ahhoz, hogy a különböző adatforrásokon alapuló tudományos eredmények ilyen fajta véletlenszerű viselkedést analizálni tudjam, egy univerzális elemző keretrendszert (UDSS) kell építeni. Általánosságban egy döntéstámogató rendszer (Decision Support System, DSS) építése a következő állomásokat foglalja magában: (1) a probléma definiálása, (2) az adatgyűjtés módszereinek megtervezése, amely kihat az adatbázis struktúrára, (3) a DSS funkcióinak megtervezése, (4) DSS implementáció, és (5) tesztelési és validációs folyamatok. A DSS építését nagyban megkönnyíti a folyamatosan fejlődő technológiai háttér, amellyel egyre gyorsabban lehet egyre jobb rendszereket építeni. A DSS építési ideje tehát rövidül, a komponens szemlélet támogatja ezen rendszerek magasabb absztrakciós szinten történő definiálását, ugyanakkor mindezek ellenére sem lehet minden módosítást és fejlesztést könnyen kivitelezni. Az adatok természetüknél fogva eltérnek, a döntési célok is mások lehetnek, valamint a kutatási környezet heterogén tulajdonsága is problémákat eredményez. Amennyiben egy adattárházat használunk kutatásunk során, annak struktúráját meg kell változtatni, amennyiben egy új adatsort kezdünk mérni. Minden új módosítás egy új projektet igényel az adattárház szempontjából. Például ha egy vállalatban belül egy új termelő eszköz kerül használatba, akkor ez új folyamatok bevezetését is eredményezi. Új adatsorok mérése kerül bevezetésre, amelyek az aktuális döntéstámogató rendszer teljes, vagy részleges újratervezését igénylik. A különböző forrásokból érkező adatok egységes kezelése nehéz feladat. Az egyes tudományterületek különböznek a mért adatok struktúráját illetően, de különböznek az elemzési módszerek használatában is. Különböznek továbbá az adattárolási módokban, az adatok lekérdezésekben, adat transzformációs eljárásokban, egyszóval az egész kutatási folyamat kivitelezésében. Ugyanakkor ahhoz, hogy RC elemzéseket tudjak végezni, ezen különbségeket kezelnem és egységesítenem kell. Ezért építettem egy Elemző Keretrendszert univerzális célokkal.

Az Univerzális Elemző Keretrendszer és a Véletlen Összefüggések képezik dolgozatomban két fő részét.

2.2. Specifikus kutatási célok meghatározása

Az adatok kritikus szerepet játszanak a Véletlen Összefüggések szempontjából. A felhasznált adathalmazok tanulmányozása során megállapítottam, hogy az adatok mennyiségileg statisztikai sokaságnak tekinthetőek, ugyanakkor nem olyan nagyméretűek, hogy a Big Data környezet kritériumainak megfeleljenek. Ezért az adatok környezetét „Big Data inspirált” („Big Data inspired”) fogalomnak definiáltam. Ezért a standard relációs adatbázis-kezelő rendszerek képesek ezen adatokat kezelni. A teljesítmény kevésbé kritikus, ezért a standard SQL alapú adatbázis struktúrát választottam, mint adattárolási forma. Miután létrehoztam az Univerzális Elemző Keretrendszert, a Véletlen Összefüggések elemzését kezdtem meg.

1. Probléma. *Univerzális Elemző Keretrendszer koncepció és architektúra.* Középpontba helyezve az univerzalitást, új tervezési módszereket kellett alkalmaznom. A fő probléma, hogy sok hasonló rendszer speciális problémákra ad választ, és kevésbé lehet általánosítani nagymértékű koncepcionális, logikai és fizikai változtatások nélkül. Az adatbázis struktúrája univerzális kell, hogy legyen, ami azt jelenti, hogy minden különböző struktúrájú adatot egy egységes struktúrába kell eltárolni. Az adatok eredeti struktúrája és az új univerzális struktúra között mindig léteznie kell egy egyértelmű megfeleltetésnek. Az

adatokat nem csak egy adott módszerrel elemezzük, hanem több különböző elemzési módszert használhatunk. Az elemző módszerek halmaza bővíthetőnek kell, hogy legyen, ugyanakkor, ha egy új módszert fejlesztünk, az UDSS többi része érintetlen kell, hogy maradjon. Mivel az elemzés komplex lehet, ezért több elemzési módszert lehet egymásután végrehajtani az elemzési keretrendszerben. Hasonlóan, az eredmények megjelenítésének különbözőségeit is kezelni kell.

2. Probléma. Véletlen Összefüggések. Az UDSS környezetben végzett elemzések eredményei alapján arra a következtetésre jutottam, hogy egymásnak ellentmondó eredményeket lehet generálni. Ugyanazon adathalmazok esetén képes voltam úgy manipulálni az elemzési folyamatot, hogy az általam előre meghatározott, adott eredményt kapjam. Egy másik elemzési folyamatot használva ugyanazon adathalmazon, egy ellentétes eredményt kaptam. Mindkét eredmény matematikailag pontos, a standard pontos elemzési módszertanokat követtem. Ugyanakkor, a különböző eredmények ellentétes döntésekhez vezettek. Azokat a körülményeket határoztam meg, ahol ilyen ellentmondó eredménypárok kialakulhatnak. Mivel egy elemzés során több módszert hajthatunk végre egymás után, és sok módszert különböző értékekkel paraméterezhetünk, az elemzési lehetőségek száma nagyon nagy. Arra a kérdésre kerestem a választ, hogy az ilyen nagy mennyiségű elemzési lehetőség mellett, valamint a „Big Data inspirált” környezet milyen hatást gyakorol az eredmények megbízhatóságára. Köszönhetően a növekvő adatmennyiségnek és a nagyszámú elemzési lehetőségeknek, lehetséges, hogy az eredmények csak véletlenszerűen keletkeznek. Amennyiben ez így van, úgy az UDSS azon nézete, miszerint segítsük egyre komplexebb elemzések kivitelezését illetve a Véletlen Összefüggések szemben állnak egymással.

3. Univerzális Elemző Keretrendszer

3.1. Architektúra

Az UDSS általános architektúrája a háromrétegű architektúrán alapszik. A rendszer architektúráját az 1. ábra foglalja össze.

Definíció. *Univerzális Adatbázis (Universal Database Structure, UDB).* Az összes különböző forrásból származó adat tárolására alkalmas struktúra, vagy legalábbis minden adat esetén létezik egy egyértelmű összerendelő szabály a régi adat struktúra és az UDB struktúra között.

Definíció. *Adatintegrációs Modul (Data Integration Module, DIM).* Szabály alapú interfész, amely biztosítja transzformációs szabályok definiálását az eredeti adatstruktúra és az UDB között.

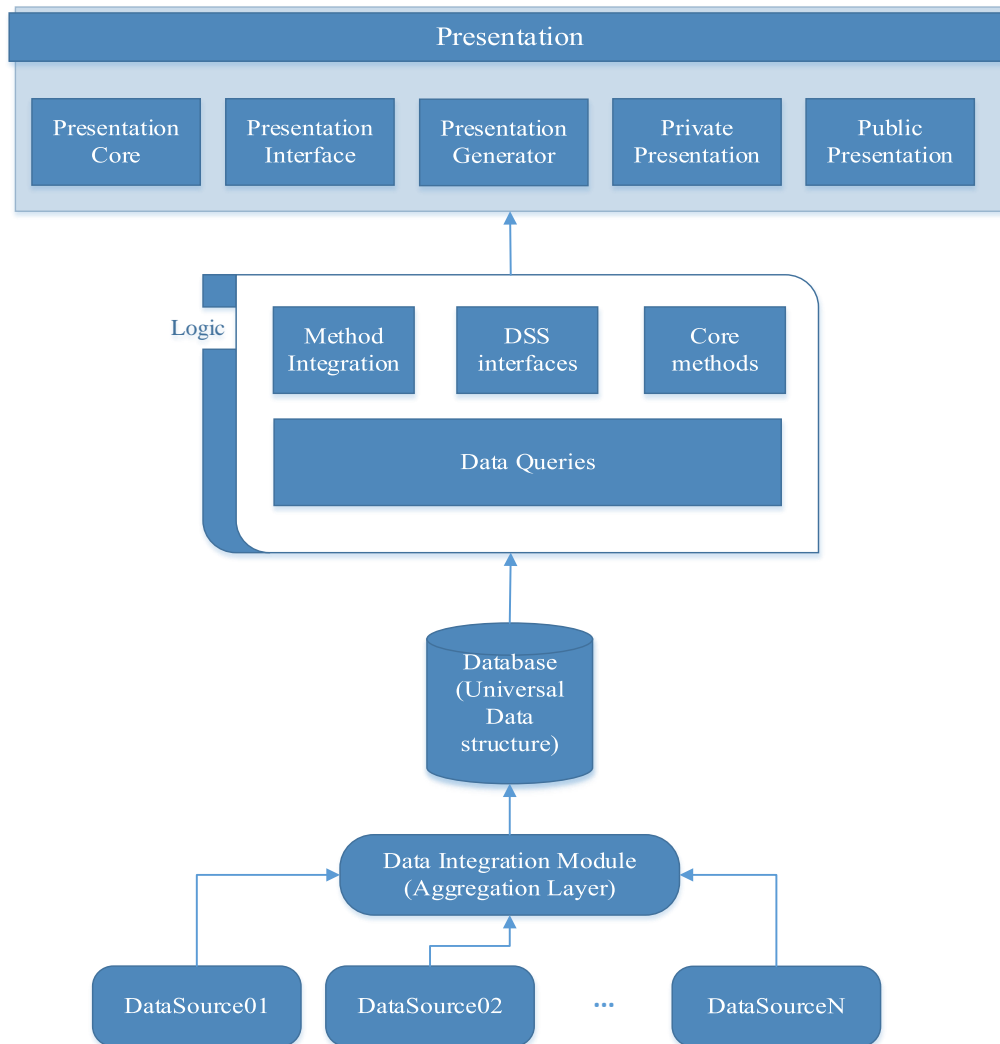
Definíció. *Adatlekérdezés Modul (Data Queries, DQ).* A kívánt adatok adatbázisból (UDB) történő kinyerését teszi lehetővé.

Definíció. *Adatmanipulációs Modul (Data Manipulation Module, DMM).* az elemző módszerek halmaza.

Definíció. *Belső Módszerek (Core Methods, CM).* Azon algoritmusok halmaza, amelyek elérhetőek a rendszeren belül.

Definíció. *Módszerek Integrálása (Method Integration, MI).* A rendszeren kívüli elemző módszerek betöltését támogatja.

Definíció. *Döntéstámogató Rendszer Interfészek (Decision Support System Interfaces).* Biztosítják azon elemző módszerek hívását, amelyeket egy másik rendszerben implementáltak.



1. ábra: Univerzális Elemző Keretrendszer koncepciójának architektúrája

Definíció. *Belső Nézetek* (Presentation Core, PM). A rendszeren belüli nézetek halmaza.

Definíció. *Nézet Generátor* (Presentation Generator, PG). Az elemzés alatt használt módszerek kimeneti struktúrájának felhasználó által definiált nézeteinek megvalósítását támogatja.

Definíció. *Nézet Interfészek* (Presentation Interface, PI). A külső rendszerekben megvalósított nézetek rendszerbeli futtatásáért felelősek.

Definíció. *Felhasználói Interfész Modul* (User Interface Module, UIM). Biztosítja a kommunikációt a rendszer és a felhasználó között.

3.2. Validáció és eredmények

3.2.1. Első esettanulmány: UDSS működése meglévő implementációkkal

Az elkészült UDSS implementációt a ForAndesT rendszer képességeivel validáltam. A ForAndesT egy erdészeti döntéstámogató rendszer. A tudományterületnek megvannak a maga speciális karakterisztikái.

Kérdéseket fogalmazhatunk meg az adott döntéstámogató rendszernek, majd ezekre a rendszer eredményként egy választ generál. Az erdészetben a következő kérdéseket lehet feltenni:

- „What” kérdés. Az adott környezeti tényezők mellett (erdőművelési típus), milyen lesz a földegységek teljesítménye. A földművelési típus ebben az esetben az adott fafaj.
- „What if” kérdés. Milyen lesz a kiválasztott földegységek teljesítménye, amennyiben az erdőművelési típust a régiről egy újra változtatjuk. Például egy adott fafaj helyett egy másikat telepítünk.
- „Where” kérdés. A felhasználó által definiált szabályok mellett melyik földegységek adják a legjobb teljesítményt.

A dolgozatomban a „Where” kérdés megválaszolására adtam példát UDSS környezetben. Az elemzés mögött lévő módszer az ún. Iterative Ideal Point Threshold (IIP) technika, amelyet Annelies vezetett kutatócsoport fejlesztett ki [50]. Ez az algoritmus iteratív módon találja meg a megfelelő földegységeket.

Land Unit	Runoff Production, Minimize, weight: 0,3	Sediment Production, Maximize, weight: 0,7	Solution
No optimal Land unit	0	0	optimal
New Values:	8.62529731665596	96.3044321345206	
240	1.50007404427479	99.8585145735604	suboptimal 1
417	1.3256194379127	96.772798545704	suboptimal 1
New Values:	16.9055717749045	92.7503496954807	
130	5.92883065770639	96.1932785936415	suboptimal 2
219	4.53678869138401	96.0827691328372	suboptimal 2
240	1.50007404427479	99.8585145735604	suboptimal 2
254	5.70413321724629	95.8837977405627	suboptimal 2
326	10.0924342532874	94.8952888334829	suboptimal 2
398	5.16915240613264	95.4801406109371	suboptimal 2
408	11.0120377648205	96.5015343195076	suboptimal 2
417	1.3256194379127	96.772798545704	suboptimal 2

2. ábra: IIP eredmény

A módszer magját alkotó gondolatot a K1 képlet foglalja össze.

$$goal_value = optimal_value \pm iteration_nr * \left(\frac{max_weight}{weight_ES} \right) * \left(\frac{range}{\#iteration} \right), \quad (K1)$$

ahol

optimal_value jelöli a minimum vagy a maximum teljesítmény értéket a felhasználó által definiált feltételeknek megfelelően.

Iteration_nr az iterációk száma.

max_weight a maximális súly a kiválasztott attribútumok súlyai közül.

range az adott attribútum numerikus tartományának minimum és maximum értékének a különbsége.

#iteration az aktuális iteráció száma.

Az iterációk számát a felhasználó adja meg. Az IIP algoritmust végrehajtva részoptimális eredményeket kapunk. Az eredmények azon földegységeket tartalmazzák, amelyek a legjobban közelítik a felhasználó által definiált feltételek értékeit. Ritka eset, amikor az első iteráció eredményeképpen egyből megtaláljuk a keresett földegységeket. Egy IIP futás eredményét mutatja be a 2. ábra.

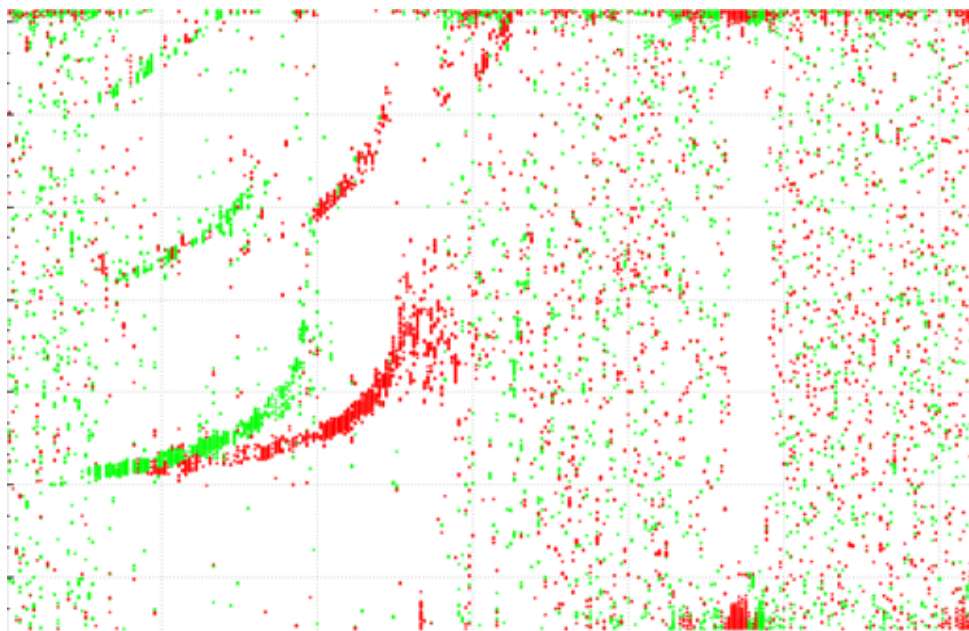
A 2. ábrán látható, hogy nincs olyan földegység, amely a felhasználó által definiált feltételeket teljesítené. Ugyanakkor részoptimális földegységeket találtunk.

Az UDSS megoldás előnyei:

- Az összes adat feltölthető az Univerzális Adatbázisba.
- Ha egy új adatsort mérünk, akkor azt a meglévő erdészeti mérésekhez csatolhatjuk.
- Amennyiben egy új módszert szeretnénk használni (azaz egy másik típusú kérdést akarunk megválaszolni), akkor csak a módszert kell implementálni, vagy egy másik rendszerből meghívni.
- Az adatok és az elemzési módszerek könnyen kombinálhatóak.
- Többszempontú döntéshozatal támogatott a rendszeren belül.

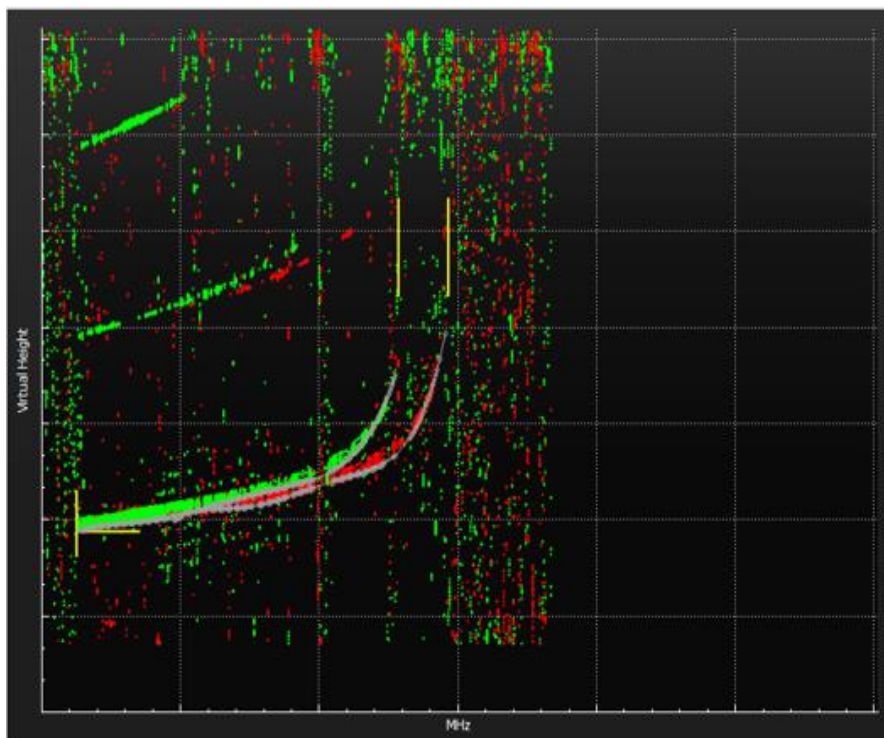
3.2.2. Második esettanulmány: Ionogramok feldolgozása

UDSS nem csak erdészeti kérdésekre képes válaszolni, hanem használható földtudományi kérdések megválaszolására is. A rendszert ionogramok releváns területeinek meghatározására használtuk. Az ionoszféra a Földet körülvevő légkör egy eleme, ahol az elemek ionizált állapotban vannak. Az ionizáció többnyire a napsugarak miatt megy végbe. A folyamat során a neutronok pozitív vagy negatív töltésű elemekké alakulnak, attól függően, hogy elektront vesznek fel vagy vesztenek el. Az ionoszféra tovább rétegződik. A legalsó réteg a *D-réteg*, és ahogy egyre magasabbra megyünk, úgy következnek sorjában az *E-*, *F1-* és *F2-rétegek*. Az ionoszféra aktuális állapotát az ún. ionoszonda méri. Ennek kimenete tulajdonképpen egy bináris kép, amely a számunkra fontos releváns területek mellett sok zajt is tartalmaz. A releváns terület két részből áll: (1) *ordinary* komponens és (2) *extraordinary* komponens. Egy ionogram példa látható a 3. ábrán, ahol a zöld az *ordinary*, piros pedig az *extraordinary* komponens.



3. ábra: Ionogram példa

Az ionogram feldolgozásának fő célja a két komponens zajtól történő elkülönítése. Ezért a feldolgozásnak két fázisa: (1) adattisztítás és (2) analizálni és meghatározni a két komponens. A másik problémát az ionogramok sokfélesége jelenti: általános modell kevésbé definiálható. Ezért nem triviális az ionogramok teljesen automatikus feldolgozása. Léteznek részleges megoldások, de egyik sem általánosítható az összes ionogramra.



4. ábra: Feldolgozott ionogram

Általánosságban az Ionogramok esetében a komponensekre legjobban illeszkedő görbéket keresik. A legkisebb módszerek illeszkedést használtam az elemzés során. Ez a döntéstámogatási része az elemzésnek. A legjobb illeszkedések alapján fel lehet írni az ionosféra állapotát az adott időben. A komponensek formája nagyon változó lehet, ezért lehetséges, hogy elsőre nem találjuk meg a legjobb illeszkedést. Például egy adott ionogram esetében nem a negyedfokú, hanem egy ötöd- vagy nagyobb fokú illesztés a legjobb. Ugyanakkor a sok zaj miatt, az elemzés torz eredményt is adhat. Ekkor iterált elemzési folyamatokkal, valamint az UDSS rendszer szűrő módszereivel, mint DMM módszerek, az illesztés megfelelő lesz. A 4. ábrán látható fehér vonalak a két főkomponens legjobb göbeillesztését mutatják. Az ionogram elemzés rámutat, hogy az UDSS fél-strukturált elemzési folyamatot is képes támogatni.

UDSS rendszert használva az előnyök a következők:

- UDB képes ionogram adatok tárolására is.
- Az eredeti képek megőrződnek, a módosítások elmenthetőek vagy visszaállíthatóak.
- Az összes ionogram típus analizálható.
- Több különböző algoritmus alkalmazható minden elemzési fázisban. (Kutatásomban a Connected-Component Labeling algoritmust használtam.)
- Automatikus és kézi kiértékelés is támogatott. (Fél-strukturált probléma).

3.2.3. Harmadik esettanulmány: beszállítói teljesítmény elemzése

A harmadik esettanulmányban az UDSS optimalizációs képességeit mutatom be. Egy faanyagokkal foglalkozó vállalatnál a deszkák színének meghatározása kritikus, mert a termelési folyamatok a pontos színmeghatározáson alapulnak. A faanyagokat a vállalathoz a beszállítói szállítják. A szállítás előtt a faanyagok feldolgozása is kritikus tényező, mert például a fa szárítási ideje befolyásolhatja a színt. Ezért a deszkák színének meghatározása a vállalat termelési folyamatának első lépése.

Ebben az esetben az a beszállító, aki jobban biztosítja a deszkák előre meghatározott színét, az feljebb kerül a szállítói ranglistán. Az alkalmazott módszer az ANOVA statisztikai teszt. Ez a teszt meghatározza, hogy az egyes beszállítók átlagos deszkaszíne eltér-e avagy sem. A tesztnek két feltételét kell megvizsgálni: (1) az adatelemeknek követnie kell a normális eloszlást és (2) a szórásnégyzeteknek meg kell egyeznie. Ezen feltételeket ellenőrizni kell az ANOVA módszer használata előtt. A beszállítók átlagait és szórásait az 1. táblázat mutatja.

1. táblázat: Beszállítók átlagai és szórásai

Beszállító	Átlag	Szórásnégyzet	Elemzés
A	162.77	8.93	30593
B	166.53	8.96	56731
C	164.11	10.97	11776
D	157.82	12.18	11418
E	174.52	10.55	35758
F	162.9	11.38	7484
G	168.51	12.5	194004
H	160.2	11.54	60779
I	164.83	10.85	77569
J	162.95	11.74	427304
M	162	11.53	15870
N	166.99	12.83	41754
O	161.39	10.13	33223
P	165.41	11.08	9454

Az 1. táblázat adatai alapján feltételezhetjük a normális eloszlást, valamint a szórásnégyzet egyenlőséget. Ugyanakkor léteznek tesztek ezen karakterisztikák bizonyítására. A klasszikus Khí-négyzet próba azt mutatta ki, hogy az adatelemek nem követik a normális eloszlást. A kritikus érték 2692,289 volt, a Khí kritikus érték az adott szabadságfok mellett $\alpha = 0.05$ szignifikancia szinten pedig 55,76. A próbastatisztika értéke jóval nagyobb, mint a kritikus érték, bár az 1. táblázat adatai nem ezt sugallják. Ellenőriztem a normalitás tulajdonságot a D'Agostino Pearson teszttel is, amely az adott eloszlás formája alapján elemzi a normalitás tulajdonságot. Ezen teszt esetében is óriási értékeket kaptunk. A szórásnégyzetek egyenlőségét Bartlett teszttel ellenőriztem. A próbastatisztika értéke 15773,06 volt, ami jóval nagyobb a kritikus értéknél, amely 59,334 ebben az esetben.

Vannak olyan megközelítések is, hogy a feltételek elhagyhatóak, amennyiben úgy ítéljük meg, hogy az adatelemek robusztusak. Ezért végrehajtottam magam az ANOVA módszert is. Az eredményeket a 2. táblázat foglalja össze.

2. táblázat: ANOVA eredmények

<i>Paraméterek</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p érték</i>	<i>F_{kritikus}</i>
Csoportok között	10389968.79	13	799228.4	5998.189	0	1.720166
Csoportokon Belül	135071078.7	1013705	133.245			
Teljes eltérés	145461047.5	1013718				

A 2. táblázat F értéke sokkal magasabb, mint a kritikus érték. Feltételezhetjük, hogy a beszállítók által szállított faanyagok színének átlagos értéke eltér (a legkisebb érték $157,82$ a legmagasabb $174,52$ az 1. táblázatban). ANOVA támogatja is ezt a megközelítést, ugyanakkor az F érték több ezres nagyságrend. Az adatokat nézve ekkora eltérés azonban nem indokolt. Tovább elemezve a beszállítókat, a *Duncan* tesztet alkalmaztam. Míg az ANOVA azt mondja meg, hogy van-e olyan beszállító (akár egy is), aki eltér az átlagos teljesítménytől (fák színe), addig a Duncan teszt azon csoportokat azonosítja, amelyek tagjai egymástól nem térnek el átlagban, de a csoportok eltérnek egymástól. Végrehajtva a tesztet, az összes csoportba csak egyetlenegy beszállító került. Ugyanakkor az F , J és M beszállítónak majdnem azonos átlaga van, indokolt lenne, az egy csoportba történő tartozás.

Ebben az esettanulmányban az UDSS használatának előnyei a következők:

- Beszállítói adatok tárolására is alkalmas az adatbázis.
- Több algoritmus (statisztikai tesztek) végrehajthatóak a rendszeren belül.
- Ad-hoc döntéstámogató rendszerekben csak bizonyos problémák oldhatóak meg, míg az ad-hoc rendszerek mindegyikének elemzési folyamata végrehajtható az UDSS segítségével.
- UDSS optimalizációs problémákat is képes megoldani.
- UDSS vállalati környezetben is használható.

3.3. Új eredmények

1. téziscsoport: Kidolgoztam egy Univerzális Elemző Keretrendszert, amelyet három különböző tudományterületről származó probléma megoldására használtam.

1.1. tézis: Kidolgoztam egy új, flexibilis adatbázis sémát. Az elterjedt adattárház megoldások kevésbé támogatják a későbbi sémamódosításokat. A javasolt adatbázis struktúra bármilyen struktúrájú adat befogadására képes, miközben az egyes összetartozó adatelemek közötti kapcsolatok megőrződnek (metaadatok, dimenziók).

1.2 tézis: Egy olyan Adatintegrációs Modult dolgoztam ki, amely támogatja szétszórt adatelemek integrációját és feltöltését az Univerzális Adatbázisba.

1.3 tézis: Kidolgoztam egy generikus Adatmanipulációs Modult, amely támogatja elemző módszerek rendszerbetörtendő beillesztését, valamint egy flexibilis Megjelenítő réteget, amely támogatja egyedi nézetek definiálását.

1.4 tézis: Az Univerzális Elemző Keretrendszer működését három esettanulmányon keresztül bizonyítottam. Az egységes adatmenedzselés és a flexibilis módszerhasználat tulajdonságok kiterjesztik az ad-hoc döntéstámogató rendszerek képességeit. Teljeskörű kivitelezést végeztem egy erdészeti döntéstámogatási problémát illetően, egy fél-automatizált ionogram elemzést valósítottam meg, valamint egy gazdasági problémára adtam megoldást a keretrendszert használva.

Kapcsolódó publikációk:

Angolul: [B1], [B3], [B6], [B7].

Magyarul: [B2], [B4], [B5].

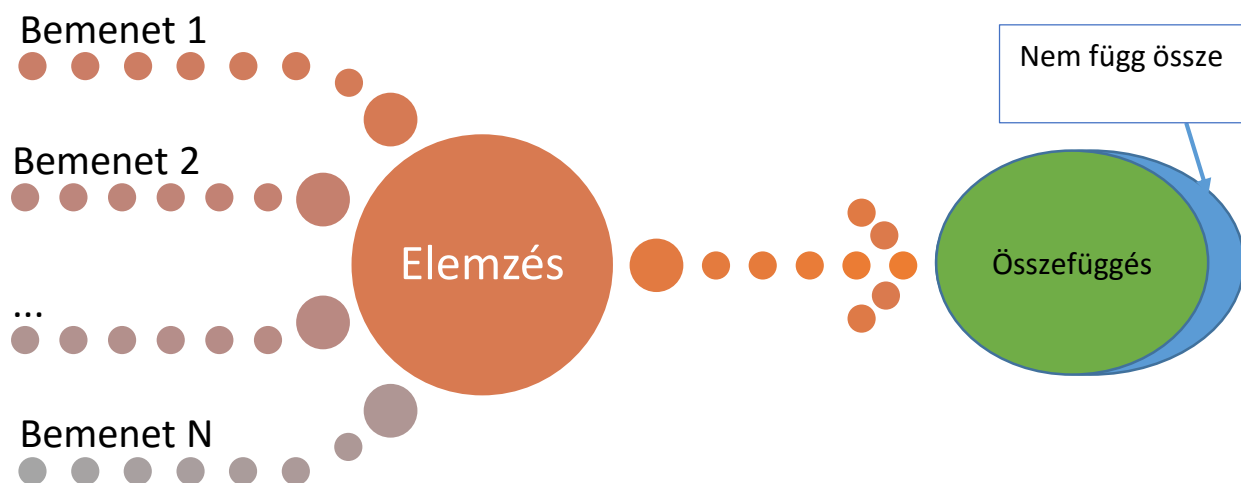
4. Véletlen Összefüggések

A klasszikus megközelítésben adatokat különböző módon mérünk, majd különböző módszerekkel elemezzük azokat. Az eredmények alapján egy döntést hozunk meg, majd ezután ennek megfelelően cselekszünk. Megállapítjuk, hogy két adatsor korrelál egymással, vagy beindítunk egy új termelési gépcsoportot az adott beállítások mellett. Minden döntési folyamatnak van validálási szakasza, ugyanakkor a validálást sokszor csak a döntés meghozatala után tudjuk megtenni. A rossz döntés az előző példánál maradván valótlan korrelációt vagy selejtes termékeket eredményez. De hogyan lehetséges az, hogy adatokon alapuló matematikailag bizonyított elemzések mellett mégis rossz döntéseket hozunk meg? A kérdésre a válaszom egy új elméleti megközelítés, amelyet Véletlen Összefüggéseknek neveztem el. Az elmélet szerint a pontos kutatási módszertanok betartása mellett az eredmények véletlenszerűen is keletkezhetnek, és ez a véletlenszerűség a kutatók előtt is rejtve maradnak. A szakirodalmat tanulmányozva nem találtam olyan módszertant, amely ilyen szempontból figyelembe venné a véletlenszerűséget. Ezért kezdtem el kutatni a Véletlen Összefüggések témakörét.

4.1. Véletlen Összefüggések keretrendszere

4.1.1. Definíció

Általánosságban azon módszert illetve módszerek sorozatát kerestem a kutatás során, amellyel összefüggéseket tudunk találni a különböző változók között úgy, hogy azok véletlenül fűgnek csak össze. Arra vonatkozóan vonunk le következtetéseket, hogy adatsorok [mint valószínűségi változók] a kimutatott [matematikai] kapcsolaton túlmenően valójában nem biztos, hogy összefűgnek.



5. ábra: Véletlen Összefűgések összefoglaló ábrája

Több módszer is létezik, amellyel az eredmények megbízhatóságát lehet vizsgálni. Ilyen például az r^2 vagy a különböző statisztikai tesztek. A Véletlen Összefüggések elmélete nem ezen módszereket helyettesíti. A fő különbség a megbízhatósági elemzések és a Véletlen Összefüggések között a „rossz” eredmények megközelítésének mikéntjében keresendő. Ha az eredmények megbízhatóságának elemzése megfelelő, akkor jobban elhiszük, hogy helyes az eredmény, a keresett összefüggés valóban létezik. A Véletlen Összefüggés ezzel szemben azt feltételezi, hogy bizonyos paraméterek mellett (lásd Paraméterek fejezet) az eredmények, amelyek jó megbízhatósági értékkel rendelkeznek, azok is csak véletlenszerűen jöttek létre. Kiszámoljuk az r^2 -t, kritikus értékeket összehasonlítjuk a próbastatisztikai értékkel, de mindezek ellenére az egész elemzési folyamat a véletlenség eredménye lehet.

A bemeneti adatsorok alapján a fő kérdés az, hogy hogyan kaphatunk egyáltalán más eredményeket. Az 5. ábrán látható, hogy az összes lehetséges eredmény közül az „összefüggés” aránya sokkal nagyobb, mint hogy „nem függ össze” eredményt kapjunk. Akármilyen megbízhatósági elemzést is hajtsunk végre, ez az arány bizonyos paraméterek mellett akkor is jelen van.

4.1.2. Paraméterek

Minden adatforrásnak saját struktúrája van. Az adatok különböző, de előre definiált formája az elemzési módszerek bemenete. Ahhoz, hogy minden adatot elemezni tudjunk Véletlen Összefüggések szempontjából, ehhez egyik részről egységesen kell kezelni az adatokat, másik részről pedig minden elemzést befolyásoló tényezőre fel kell készülni. Például ha az Univerzális Elemző Keretrendszerben egy adott adathalmazt akarunk elemezni regressziós technikákkal, akkor szükségünk van a pontok számára, x és y koordinátájukra és hogy milyen regressziós technikát akarunk végrehajtani (lineáris, kvadratikus, exponenciális, logaritmikus).

Összefoglalva a Véletlen Összefüggések keretrendszer paraméterei:

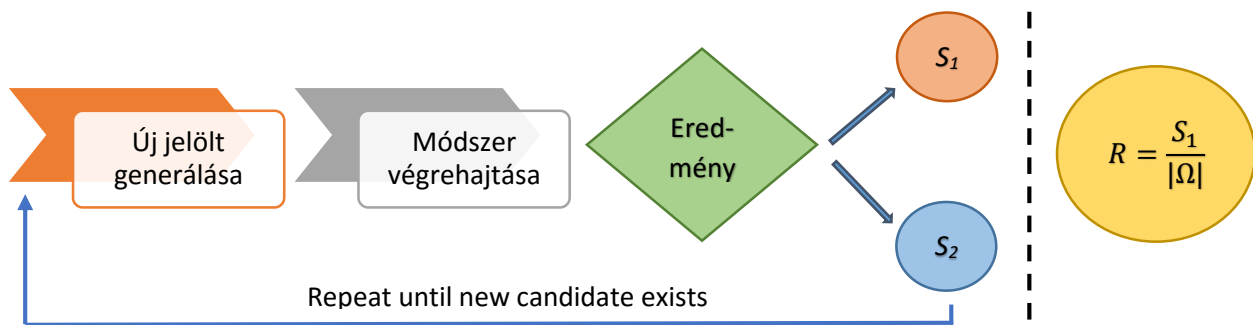
- k , amely az adatsorok száma.
- n , amely az egyes adatsorokban lévő elemek száma.
- r , amely a mért értékek numerikus tartománya.
- t , amely a végrehajtandó elemzési módszerek száma.

4.1.3. Véletlen Összefüggés modelljei és módszerei

A Véletlen összefüggések két modellje a következő:

- (1) A teljes eseménytér kiszámolása (Ω -modell);
- (2) Az ütközések („összefüggést találtunk”) valószínűségének meghatározása (Θ -modell).

Az (1) esetben az összes lehetséges kimenetet legeneráljuk. Ez azt jelenti, hogy az $r(a,b)$ ismeretében az összes lehetséges n -tagú adatsort, amit mérhetünk, előállítjuk. Ehhez az előállításához szükséges az r paraméter, hiszen csak véges r esetén lehet az összes kombinációt előállítani. Ezért is szükséges keretrendszerünkbe az r , mint a tartomány paramétere. Az összes kombinációt előállítjuk, amit a kutatók egyáltalán mérni tudnak. Ezután mindegyik n -tagú adatsor esetén végrehajtjuk azt a módszert, amellyel az adatokat elemezni szeretnénk. Amennyiben az „összefügg” ítélet az eredmény az adott jelöltre vonatkozóan, akkor az S_1 halmaz elemeinek a számát eggyel növelem. Ha ezt az összes lehetséges bemenetre végrehajtom, akkor egy R rátát tudok meghatározni az S_1 és az $|\Omega|$ osztásával. Ez a ráta mutatja meg az adott paraméterek mellett a véletlenszerűség mértékét. Ha például az R ráta értéke 0,99, akkor csak 1%-ban létezik az összes lehetőség közül olyan, ami a „nem függ össze” eredményt adja.



6. ábra: az R ráta számítási folyamata

A második modell esetében egy C rátát számolok ki. Ez a ráta azt mutatja meg, hogy hány adatsort kell mérnem ahhoz, hogy nagy valószínűséggel legalább kettő összefüggjön. A kutatóknak általában megvan a saját hipotézisük, amelyet adatokkal szeretnének alátámasztani. Amennyiben egy elemzés sikertelen („nem függ össze”), akkor másik módon (pl.: egy másik módszerrel, új adatok gyűjtésével) próbálnak meg összefüggéseket találni. A gyakorlatban, ha van egy A adatsorunk, és ez az adatsor nem függ össze a többi mért adatsorral, akkor további adatsorokat mérünk, abból a célból, hogy összefüggést találjunk. Véletlen Összefüggések szempontjából azt a kérdést tettem fel, hogy hány adatsor szükséges az adott paraméterek mellett, hogy biztosan találjunk összefüggést. A C az adatsorok száma, ennyi adatsor esetén nagy valószínűséggel lesz legalább két adatsor, amely mindenképpen összefügg.

A C értékének megfelelően három lehetőséget definiáltam:

- A C nagy szám. Adott paraméterek mellett sok adatsorra van szükség, hogy összefüggést találjunk. Ez a legjobb eset, a Véletlen Összefüggések esélye kicsi.
- A C kielégítő. A véletlenség esélye közepes.
- A C kis szám. A legrosszabb eset. Viszonylag kis számú adatsor esetén magas valószínűséggel mindenképpen találunk összefüggést.

4.1.4. Osztályok

A Véletlen Összefüggések különböző elemzési környezetben különböző okok miatt alakulhatnak ki. Ezért osztályokat definiáltam. Mindegyik osztály egy okot jelent, amely mentén Véletlen Összefüggés kialakulhat.

1. osztály. Ahogy korábban megfogalmaztuk, viszonylag sok tudományos módszer létezik az adatsorok elemzésére vonatkozóan és amennyiben nem találunk összefüggést egy adott módszerrel, akkor addig próbálkozunk más módszerekkel, amíg összefüggést nem találunk. A célunk hogy az adatsorokat elemezve egy adott matematikai módszerrel kapcsolatot találjunk, és a keresést addig folytatjuk, amíg csak lehetséges. Másik oldalról nem definiáljuk [nem tudjuk definiálni], hogy a két adatsor mikor nem függ össze. A módszerek számát még növeli a paraméterezhetőségük változatossága, illetve a hibahatár változatos megválasztása. Ugyanazon adatsor több és több algoritmussal történő elemzésének eredményeképpen nem a tényleges összefüggést találjuk meg, hanem csak véletlenszerűen találunk egy összefüggést. Úgy is fogalmazhatunk, hogy az elemzések számának növelésével biztos lesz legalább egy olyan módszer, amely jó eredményt ad.

2. *osztály*. Az előző osztálynál nagyobb problémát jelent, ha létezik két olyan módszer, amely egymással szemben mutat ki összefüggést. Mindkét módszer matematikailag helyes, mégis egymásnak eltérő ítéleteket vonhatunk le. Mivel az elemzések nagy része abbamarad, amikor már az adott módszerrel megtaláltuk az adott összefüggést, ezért sok esetben nem is derülhet ki, hogy egy másik megközelítésben épp az előző eredmények alapján történő magatartás ellentéte is kimutatható. Inkább az adott módszer pontosabb felparaméterezésének kutatása jellemző. Ezen iránynak két alelete lehetséges: (1) amikor két [vagy több] módszer egymás mellett ad matematikai összefüggést és (2) amikor a két [vagy több] módszer egymásnak ellentétes [egymást kizáró] eredményt ad. Az első eset nagyon jó, hiszen ugyanaz az összefüggés több nézőpontból is helyes, ezzel is erősítve a kapcsolatot a kettő vagy több adatsor között. Ennél rosszabb a második, az egymást kizáró eredmények megléte, hiszen itt nem lehet kialakítani közös ítéletet. Mivel feltételezzük, hogy a mért adatok tulajdonképpen véletlenül is adhatnak összefüggéseket, ezért létezhet kettő [vagy több] olyan módszer, amely ugyanazon adathalmazok esetén mindig ellentétes eredményt mutat. Idetartozik az az eset is, amikor egy adott módszer, valamely Véletlen Összefüggés paraméter változása esetén (pl.: mintaelemszám) egymásnak ellentmondó eredményt produkál.

3. *Osztály*. A harmadik főosztály az adatok növekvő tömegére vonatkozik. Általános az a megfogalmazás, hogy minél több adatunk van, annál pontosabb eredményeket tudunk generálni. Másik oldalról ellentmondásos az a megközelítés, miszerint az adott adatsor egyes része [például az adatgyűjtés kezdetétől a T . időpillanatig] más eredményt ad ugyanazon adatsor egy nagyobb részéhez képest [például az adatgyűjtés kezdetétől a $(T + K)$. időpillanatig beszerzett adatértékek]. Ez azért kritikus, mert nem tudhatjuk, hogy az adatgyűjtésben pontosan hol tartunk, így – bár az elemzés pontos és „jó” eredményt ad adott ideig – a jövőre nézve a felállított modell nem hozza a kívánt pontosságú eredményeket. A jelenségnek két megközelítése lehet értelmezésünkben: (1) az adatsor elemeinek a száma [illetve számának változtatása] önmagában okozza a véletlen összefüggést, valamint (2) rejtett paraméterrel vagy paraméterekkel magyarázható ez az osztály. Ha a rejtett paraméter, amely nem szerepelt a kezdeti tényezők halmazában, tudomásunk nélkül változik meg, akkor a modell összeomlik. A jelenben csak azzal az adathalmazzal dolgozhatunk, amely rendelkezésünkre áll, ezért szükséges olyan kritérium megfogalmazása, amely biztosítja, hogy nem pusztán véletlenszerű összefüggésekről van szó az adott adathalmazra vonatkozóan.

4.1.5. Jelöltek generálása és szimulációs szintek

A dimenziócsökkentő technikák (Space Reducing Techniques, SRT) nagyon nagy paraméterek mellett sem biztosítják a teljes eseménytér kiszámítását valós időben. Ezért különböző szimulációs szinteket határoztam meg, amelyek segítségével két elemzési környezetet elemeztem Véletlen Összefüggések szempontjából.

1. *szint*. Nem az egész eseményteret számolom ki, hanem mintavételezéssel jelölteket generálok az adott k , n és r paraméterek értékeinek megfelelően. A valószínűség definíciója alapján, az így számított R' megközelíti R -t. Ez a legegyszerűbb és leggyorsabb módja, hogy valamilyen megbízhatóságot tudjunk mondani az eredményekre vonatkozóan Véletlen Összefüggések szempontjából. Ugyanakkor a közelítés csak akkor lesz jó, ha a generált jelöltek száma egyre közeledik a teljes eseménytér számosságához.

2. *szint*. Az SRT használata a jelöltek generálásra alkalmas. A négyzetfüggvény miatt az első fázis (jelöltgenerálás) viszonylagosan valós időben kivitelezhető. A gondot a k paraméter okozza, a teljes eseménytér nagy k esetén robbanásszerűen nő. Ezért először előállítom az összes lehetséges jelöltet n -re vonatkozóan, ott is csak azokat tartom meg, amelyek kielégítik az elemző módszer előkövetelményeit (pl.:

normális eloszlás), majd a k paraméter esetében mintavételezek. Ez a szimulációs szint pontosabb R' -t eredményez.

3. szint. Az első fázis n -re vonatkozó műveletét kombinálni lehet az F gyakorisággal. A 2. Szint esetében a szimulált k jelöltek súlya egységesen 1. Ugyanakkor egy teljes jelölt (k és n) nem egyszer, hanem többször, több kombinációban is előfordulhat. Ha például $k = 3$, akkor egy iteráció alatt a 2. Szinten egy eredmény születik, míg 3. Szinten $F_1 * F_2 * F_3$ eredmény. Más szavakkal, amennyiben egy teljes jelölt generálódott, akkor ezen a szinten az összes permutációjának az eredménye is generálódik, amely nem jelent többszámolást, mivel az elemzési módszer eredménye ugyanaz lesz minden esetben és csak a gyakoriságokat kell összeszorozni. Ugyanannyi i iterációt végrehajtva, ezen a szimulációs szinten több eredményünk lesz, ami sokkal jobban megközelíti a teljes eseménytér számosságát. Ezen a szinten R^* -t számolunk ki, amely jobban közelíti az elméleti R -et mint R' .

4.2. ANOVA Véletlen Összefüggésének vizsgálata Ω -modellel (R)

A vizsgálat első fázisában jelöltet generálok, majd utána végrehajtom az ANOVA módszert az adott jelöltön, és végül feljegyzem az eredményt. Először 2. Szintű szimulációval elemeztem a módszert.

A 3. táblázatnak két része van. Egyik részről az elméleti R és az R^* került összehasonlításra. Kisebb paraméterértékeknél van erre lehetőség. Látható, hogy a közelítés viszonylagosan jónak mondható. A szimulációs során 1000 iterációt végeztünk. A táblázat másik része olyan paraméterértékeket tartalmaz, amelyeken hagyományos módszerekkel nem lehet meghatározni R -t. A saját fejlesztésű FUS (Finding Unique Sequences) algoritmus sem képes megfelelően nagy paraméter értékek esetén az összes lehetőséget kiszámolni valós időben, ezért használtuk a 3. Szintű szimulációt. A táblázatban látható, hogy kis paraméter értékek esetén a H_0 bekövetkezési valószínűsége igen magas.

3. táblázat ANOVA eredmények szimulációval

$r(a, b)$	k	n	R	R^*	$r(a, b)$	k	n	R^*
(1, 3)	3	30	0.9523	0.9344	(1, 3)	4	100	0.9151
(1, 3)	3	50	0.9544	0.9737	(1, 3)	7	100	1.09E-9
(1, 3)	5	10	0.9722	0.9629	(1, 3)	10	100	0
(1, 3)	5	15	0.9604	0.9899	(1, 3)	10	500	0
(1, 5)	3	10	0.9774	0.9241	(1, 5)	4	100	0.5889
(1, 5)	4	5	0.9580	0.9782	(1, 5)	5	100	0.0040
(1, 5)	4	9	0.9598	0.9537	(1, 5)	7	100	7.19E-19
(1, 10)	3	5	0.9577	0.9437	(1, 10)	4	10	0.9717
(1, 10)	4	5	0.9565	0.9671	(1, 10)	4	19	0.9601

Megfelelően magas értékeknél viszont épp ellenkezőleg a H_1 bekövetkezési valószínűsége lesz magas, ami ellentmondás, hiszen kis elemhalmazz vizsgálva alig lehet olyan jelölt, amely átlagos eltérés adna, míg nagy paraméter értékek esetén alig lenne olyan jelölt, amely nem mutatna semmilyen eltérést. Ez a mai Big Data környezetre vonatkozóan kritikus, hiszen a standard megközelítés szerint minél több adatalemünk van, annál jobban biztosak lehetünk az eredményekben. Ezzel szemben kutatásom során arra az eredményre jutottam, hogy a Véletlen Összefüggések paramétereinek értékét növelve az ANOVA esetében robbanásszerűen nő a H_1 bekövetkezési valószínűsége. Amennyiben növeljük például a k

értékét, akkor annak a valószínűsége, hogy legalább egy adatsor átlagban eltér a többitől, magas. A pontos k értéket a Θ -modell segítségével lehet meghatározni. Ugyanakkor ez nem változtat azon, hogy az ANOVA érzékeny a Véletlen Összefüggésekre.

4.3. Regressziós technikák elemzése Ω -modellel (R)

Ebben a fejezetben a regressziós technikákat elemzem a Véletlen Összefüggések szempontjából. A standard elemzési módszer egyes lépéseit zárójelbe jelzem. A regressziós technikák matematikai hátterét, az elemzés során használt képleteket a disszertációmban ismertettem [1. lépés]. A regressziós technikák az első osztályba tartoznak [2. lépés]. Ez azt jelenti, hogy amennyiben egyre több és több regressziós technikát használunk az elemzés során, úgy megnöveljük a véletlenség bekövetkezésének valószínűségét [3. lépés]. Ezért a módszerek száma (t) kritikus paraméter [4. lépés]. A k paraméter kihagyható, hiszen két koordinátánk van x és y , ezért $k = 2$ konstans. Az r paraméterből azonban kettő van: r_1 , amely az x koordináta tartományát jelenti [$r_1(a_1, b_1)$] és r_2 , amely az y koordináta tartományát jelenti [$r_2(a_2, b_2)$]. Az első fázis eredményeképpen az összes lehetőség ezért $r_1 * r_2$. Az Ω számítási folyamata k , n és r paramétereken alapul [5. lépés]. A saját fejlesztésű FUS algoritmus csak részben alkalmazható. Az első szintű redukciós tulajdonság itt nem érvényesül, mert a koordináták sorrendje számít. Például az $x' = \{2, 1, 2\}$ és $y' = \{1, 3, 1\}$ nem ugyanazt az r^2 eredményt adja, mint az $x = \{1, 2, 2\}$ és $y = \{1, 1, 3\}$. Ezért az összes jelöltet direktbe kell generálni, amely hosszabb számítási idővel jár. A második szintű redukciós tulajdonság azonban alkalmazható. A számolás lényege, hogy a regressziós technikákat alkalmazzuk és keressük a legjobban illeszkedő egyenest vagy görbét. Akármelyik regresszióval találunk egy megfelelően magas r^2 értéket, úgy eggyel növelem az „összefügg” halmaz elemeinek számát, azaz az 1-sek („összefügg”) számát. Ellenkező esetben a 0-k („nem függ össze”) számát növelem eggyel értelemszerűen. Az elfogadási szint változtatható, én az $r^2 > 0,7$ -es szintet választottam. A regresszió esetében is számolni kell az alkalmazhatósági feltételekkel. Feltételeztem, hogy a függetlenség kritériuma teljesül. A normalitás vizsgálatot D’Agostino-Pearson teszttel, míg a variancia egyenlőséget Bartlett próbával ellenőriztem. A szimulációt lefutása után meghatároztam az R rátát [6. lépés]. Az eredményeket a 4. táblázat mutatja be, ahol lineáris és exponenciális regressziót alkalmaztam, ezért $t = 2$.

4. táblázat: R ráta eredmények $t = 2$ esetén

$t = 2$	$r_1(1,5);r_2(1,3)$	$r_1(1,10);r_2(1,3)$	$r_1(1,3);r_2(1,5)$	$r_1(1,3);r_2(1,10)$
$n = 7$	0.0527	0.0474	0.1453	0.2629
$n = 8$	0.0479	0.0375	0.1348	0.2597
$n = 9$	0.0462	0.0280	0.1334	0.2538

Majd növeltem n értékét 5-től 10-ig, változtattam r_1 és r_2 értékeit és már négy regressziós technikát alkalmaztam, azaz $t = 4$. Az eredményeket az 5. táblázat foglalja össze

5. táblázat: R ráta eredmények $t = 4$ esetén

$t = 4$	$r_1(1,5);r_2(1,3)$	$r_1(1,10);r_2(1,3)$	$r_1(1,3);r_2(1,5)$	$r_1(1,3);r_2(1,10)$
$n = 5$	0.2873	0.3071	0.3122	0.3288
$n = 6$	0.2092	0.2161	0.2530	0.3239
$n = 7$	0.1387	0.1379	0.2204	0.3102
$n = 8$	0.1142	0.1027	0.1947	0.3029
$n = 9$	0.1057	0.0796	0.1894	0.2927

A 4. és 5. táblázatot összehasonlítva arra a következtetésre jutottam, hogy a plusz két regressziós technika alkalmazásával – néha akár a duplájára is – növelni tudtam az „összefügg” valószínűségét. Ez azt jelenti, hogy a regressziós technikák esetében a t paraméter növeli az R rátát. Az $r_1(1,3);r_2(1,10)$ 0,3 körüli értéke stabilnak mondható. Ez azt jelenti, hogy a paraméter értékek megváltoztatása után is ugyanannyi az „összefügg” és „nem függ össze” aránya. Másik oldalról a 0,5-ös ráta nem megfelelő feltétel regresszió esetében, mivel ebben az esetben az „összefügg” és a „nem függ össze” aránya nem más, mint egy 50-50 százalékos érmefeldobás valószínűsége. Ahhoz, hogy valamilyen összefüggést összefüggésnek minősítsünk ennél szigorúbb feltételnek kell megfelelni. Ezért a 0,3-as arány ilyen szempontból is megfelelő. További regressziós elemzések eredményeit láthatjuk a 6. táblázatban.

6. táblázat: További regressziós eredmények

$r_1(a_1, b_1); r_2(a_2, b_2)$	n	R
$(1, 3);(1, 3)$	5	0.3465
$(1, 3);(1, 3)$	10	0.1087
$(1, 5);(1, 5)$	5	0.5332
$(1, 5);(1, 5)$	8	0.2491
$(1, 5);(1, 5)$	9	0.2196
$(1, 4);(1, 6)$	5	0.4153
$(1, 4);(1, 6)$	8	0.2406
$(1, 4);(1, 6)$	9	0.2248
$(1, 6);(1, 4);$	5	0.5419
$(1, 6);(1, 4);$	8	0.2472
$(1, 6);(1, 4);$	9	0.2147

Amennyiben n -t növeljük, R csökken. Ha további R csökkentést feltételezünk, és figyelembe vesszük a 30 elemszámú statisztikai ökölszabályt, akkor annak az esélye, hogy $r^2 > 0,7$ szintű „összefüggést” találjunk alacsony. Ha elfogadjuk azt, hogy az „összefüggés” valószínűsége kisebb kell, hogy legyen (nem 50-50), akkor megállapíthatjuk, hogy a regresszió nem nagyon érzékeny a Véletlen Összefüggésekre.

4.4. Új eredmények

2. téziscsoport: A döntéshozatal támogatására kidolgoztam egy keretrendszert, amelyet Véletlen Összefüggéseknek neveztem el. Ezt felhasználva elemzési eredményeinket megadott szempontrendszerek szerint tudjuk validálni, segítve ezáltal a hatékony döntéshozatalt.

2.1 tézis: Megadtam a Véletlen Összefüggések pontos definícióját, miszerint a módszertanilag helyes eredmények ellenére a vizsgált adatsorok mégsem függenek össze. Definiáltam a Véletlen Összefüggések négy paraméterét és három osztályát, amelyek segítségével meghatározható az eredmény véletlen faktora.

2.2 tézis: Két alapvető módszer definiáltam, melyek közül az egyik a teljes eseménytér kiszámolása a másik pedig az ütközések keresése. Az első alapgondolata, hogy az összes lehetséges adatkombináció alapján határozzuk meg a véletlen faktort. A másik módszer arra a kérdésre ad választ, hogy hány adatsort kell vizsgálni ahhoz, hogy magas valószínűséggel kapcsolatot találjunk.

2.3 Alkalmaztam a Véletlen Összefüggések rendszerét az ANOVA statisztikai teszt véletlenre való érzékenységének elemzésére. Megállapítottam, hogy az ANOVA érzékeny a Véletlen Összefüggésekre.

2.4 Alkalmaztam a Véletlen Összefüggések rendszerét a regressziós technikák véletlenre való érzékenységének elemzésére. Megállapítottam, hogy a regressziós technikák kevésbé érzékenyek a Véletlen Összefüggésekre.

Kapcsolódó publikációk:

Angolul: [B9], [B10], [B11].

Magyarul: [B8].

Hivatkozások

- [1] J. A. Khan, "Research methodology," APH Publishing Corporation, New Delhi, 2008
- [2] J. Kuada, "Research Methodology: A Project Guide for University Students," Samfundslitteratur, Frederiksberg, 2012
- [3] P. Lake, H. B. Benestad, B. R. Olsen, "Research Methodology in the Medical and Biological Sciences," Academic Press, London, 2007
- [4] A. Mohapatra, P. Mohapatra, "Research methodology," Partridge Publishing, India, 2014
- [5] G. D. Jackson, N.A. Moltshaniwskyj, "Spatial and temporal variation in growth rates and maturity in the Indo-Pacific squid *Sepioteuthis lessoniana* (Cephalopoda: Loliginidae)," *Marine Biology*, vol. 140, pp747–754, 2002
- [6] G. T. Pecl, G. D. Jackson, "The potential impacts of climate change on inshore squid: biology, ecology and fisheries," *Reviews in Fish Biology and Fisheries*, vol. 18, pp 373–385, 2008
- [7] E. S. Zavaleta, B. D. Thomas, N. R. Chiariello, Gregory P. Asner, M. Rebecca Shaw, Christopher B. Field, "Plants reverse warming effect on ecosystem water balance," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp9892–9893, 2003
- [8] W. Liu, Z. Zhang, S. Wan, "Predominant role of water in regulating soil and microbial respiration and their responses to climate change in a semiarid grassland," *Global Change Biology*, vol. 15, pp184–195, 2009
- [9] J. A. Church, N. J. White, "A 20th century acceleration in global sea-level rise," *Geophysical Research Letters*, vol. 33, pp1–4, 2006
- [10] J.R. Houston, R.G. Dean, "Sea-Level Acceleration Based on U.S. Tide Gauges and Extensions of Previous Global-Gauge Analyses," *Journal of Coastal Research*, vol. 27, 409–417, 2011
- [11] P. K. Aggarwal, R. K. Mall, "Climate Change and Rice Yields in Diverse Agro Environment of India. II. Effect of Uncertainties in Scenarios and Crop Models on Impact assessment," *Climatic Change*, vol. 52, pp331–343, 2002
- [12] J. R. Welch, J. R. Vincent, M. Auffhammer, P. F. Moya, A. Dobermann, D. Dawe, "Rice yields in tropical/subtropical Asia exhibit large but opposing sensitivities to minimum and maximum temperatures," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp14562–14567, 2010
- [13] L. Hooper, C. Bartlett, G. D. Smith, S. Ebrahim, "Systematic review of long term effects of advice to reduce dietary salt in adults," *British Medical Journal*, vol. 325, pp628–632, 2002
- [14] S. Pljesa, "The impact of Hypertension in Progression of Chronic Renal Failure," *Bantao Journal*, vol. 1, pp71-75, 2003
- [15] Climate Change 2007, Impacts, Adaptation, vulnerability, report, 2007
- [16] H. Nkurunziza, J. Pilz, "Impact of increased temperature on malaria transmission in Burundi," *International Journal of Global Warming*, vol. 3, pp78–87, 2011

- [17] P. Martens, R.S. Kovats, S. Nijhof, P. de Vries, M.T.J. Livermore, D.J. Bradley, J. Cox, A.J. McMichael, "Climate change and future populations at risk of malaria," *Global Environmental Change*, vol. 9, pp89–107, 1999
- [18] P. W. Gething, D. L. Smith, A. P. Patil, A. J. Tatem, R. W. Snow, S. I. Hay, "Climate change and the global malaria recession," *Nature*, vol. 465, pp342–345, 2010
- [19] H. J. Fowlera, M. Ekstrom, "Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes," *International Journal of Climatology*, vol. 29, pp385–416, 2009
- [20] E. J. Burke, R. H. J. Perry, S. J. Brown, "An extreme value analysis of UK drought and projections of change in the future," *Journal of Hydrology*, vol. 388, pp131–143, 2010
- [21] I. M. Held, T. L. Delworth, J. Lu, K. L. Findell, T. R. Knutson, "Simulation of Sahel drought in the 20th and 21st centuries," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp1152–1153, 2006
- [22] R. J. Haarsma, F. M. Selten, S. L. Weber, M. Kliphuis, "Sahel rainfall variability and response to greenhouse warming," *Geophysical Research Letters*, vol. 32, pp1–4, 2005
- [23] A. Giannini, "Mechanisms of Climate Change in the Semiarid African Sahel: The Local View," *Journal of Climate*, vol. 23, pp743–756, 2010
- [24] S. M. Crimmins, S. Z. Dobrowski, J. A. Greenberg, J. T. Abatzoglou, A. R. Mynsberge, "Changes in Climatic Water Balance Drive Downhill Shifts in Plant Species' Optimum Elevations," *Science*, vol. 331, pp324–327, 2011
- [25] J. Grace, F. Berninger, L. Nagy, "Impacts of Climate Change on the Tree Line," *Annals of Botany*, vol. 90, pp537–544, 2002
- [26] T. A. Dueck, R. de Visser, H. Poorter, S. Persijn, A. Gorissen, W. de Visser, A. Schapendonk, J. Verhagen, J. Snel, F. J. M. Harren, A. K. Y. Ngai, F. I. Verstappen, H. Bouwmeester, L. A. C. J. Voesenek, A. van der Werf, "No evidence for substantial aerobic methane emission by terrestrial plants: a ¹³C-labelling approach," *New Phytologist*, vol. 175, pp29–35, 2007
- [27] F. Keppler, J. T. G. Hamilton, M. Braß, T. Röckmann, "Methane emissions from terrestrial plants under aerobic conditions," *Nature*, vol. 439, pp187–191, 2006
- [28] L. Siliang, L. Ronggao, L. Yang, "Spatial and temporal variation of global LAI during 1981–2006," *Journal of Geographical Sciences*, vol. 20, pp323–332, 2010
- [29] G. P. Asner, J. M. O. Scurlock, J. A. Hicke, "Global synthesis of leaf area index observations: implications for ecological and remote sensing studies," *Global Ecology and Biogeography*, vol. 12, pp191–205, 2003
- [30] C. Jaramillo, D. Ochoa, L. Contreras, M. Pagani, H. Carvajal-Ortiz, L. M. Pratt, S. Krishnan, A. Cardona, M. Romero, L. Quiroz, G. Rodriguez, M. J. Rueda, F. de la Parra, S. Morón, W. Green, G. Bayona, C. Montes, O. Quintero, R. Ramirez, G. Mora, S. Schouten, H. Bermudez, R. Navarrete, F. Parra, M. Alvarán, J. Osorno, J. L. Crowley, V. Valencia, J. Vervoort, "Effects of Rapid Global Warming at the Paleocene-Eocene Boundary on Neotropical Vegetation," *Science*, vol. 330, pp957–961, 2010
- [31] L. F. Salazar, C. A. Nobre, M. D. Oyama, "Climate change consequences on the biome distribution in tropical South America," *Geophysical Research Letters*, vol. 34, pp1–6, 2007
- [32] M. Hulme, R. Doherty, T. Ngara, M. New, D. Lister, "African climate change: 1900–2100," *Climate Research*, vol. 17, pp145–168, 2001
- [33] A. P. Williams, C. Funk, "A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa," *Climate Dynamics*, vol. 37, pp2417–2435, 2011
- [34] M.D. Flannigan, Y. Bergeron, O. Engelmark, B.M. Wotton, "Future wildfire in circumboreal forests in relation to global warming," *Journal of Vegetation Science*, vol. 9, pp469–476, 1998
- [35] E. S. Kasischke, N. L. Christensen, B. J. Stocks, "Fire, Global Warming, and the Carbon Balance of Boreal Forests," *Ecological Applications*, vol. 5, pp437–451, 1995

- [36] M. E. Visser, A. C. Perdeck, J. H. Van Balen, C. Both, "Climate change leads to decreasing bird migration distances," *Global Change Biology*, vol. 15, pp1859–1865, 2009
- [37] N. Doswald, S. G. Willis, Y. C. Collingham, D. J. Pain, R. E. Green, B. Huntley, "Potential impacts of climatic change on the breeding and non-breeding ranges and migration distance of European *Sylvia* warblers," *Journal of Biogeography*, vol. 36, pp1194–1208, 2009
- [38] F. Pulido, P. Berthold, "Current selection for lower migratory activity will drive the evolution of residency in a migratory bird population," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp 7341–7346, 2010
- [39] A. R. Huete, K. Didan, Y. E. Shimabukuro, P. Ratana, S. R. Saleska, L. R. Huttyra, W. Yang, R. R. Nemani, R. Myneni, "Amazon rainforests green-up with sunlight in dry season," *Geophysical Research Letters*, vol. 33, pp33–39, 2006
- [40] A. Samanta, S. Ganguly, H. Hashimoto, S. Devadiga, E. Vermote, Y. Knyazikhin, R. R. Nemani, R. B. Myneni, "Amazon forests did not green-up during the 2005 drought," *Geophysical Research Letters*, vol. 37, pp1–5, 2010
- [41] D. T. Shindell, R. L. Miller, G. A. Schmidt, L. Pandolfo, "Simulation of recent northern winter climate trends by greenhouse-gas forcing," *Nature*, vol. 399, pp452–455, 1999
- [42] V. Petoukhov, V. A. Semenov, "A link between reduced Barents-Kara sea ice and cold winter extremes over northern continents," *Journal of Geophysical Research*, vol. 115, D21111, 2010
- [43] P. Knippertz, U. Ulbrich, P. Speth, "Changing cyclones and surface wind speeds over the North Atlantic and Europe in a transient GHG experiment," *Climate Research*, vol. 15, pp109–122, 2000
- [44] R. Vautard, J. Cattiaux, P. Yiou, J.-N. Thépaut, P. Ciais, "Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness," *Nature Geoscience*, vol. 3, pp756–761, 2010
- [45] I. Bogardi, I. Matyasovzky, "Estimating daily wind speed under climate change," *Solar Energy*, vol. 57, pp239–248, 1996
- [46] M. Rebetz, R. Lugon, P.-A. Baeriswyl, "Climatic Change and Debris Flows in High Mountain Regions: The Case Study of the Ritigraben Torrent (Swiss Alps)," *Climatic Change*, vol. 36, pp371–389, 1997
- [47] M. Stoffel, M. Beniston, "On the incidence of debris flows from the early Little Ice Age to a future greenhouse climate: A case study from the Swiss Alps," *Geophysical Research Letters*, vol. 33, L16404, 2006
- [48] M. Stoffel, M. Bollschweiler, M. Beniston, "Rainfall characteristics for periglacial debris flows in the Swiss Alps: past incidences–potential future evolutions," *Climatic Change*, vol. 105, pp263–280, 2011
- [49] B. Nosek et al. (Open Science Collaboration), "Estimating the reproducibility of psychological science," *Science*, vol. 28, aac4716, 2015
- [50] A. De Meyer, R. Estrella, P. Jacxsens, J. Deckers, A. Van Rompaey, J. Van Orshoven, "A conceptual framework and its software implementation to generate spatial decision support systems for land use planning," *Land Use Policy*, vol. 35, pp.271–282, 2013

Publikációk

Univerzális Elemzési Keretrendszer

[B1] Jos M.F. Van Orshoven, Vincent Kint, Anja Wijffels, René Estrella, **Gergely Bencsik**, Pablo Vanegas, Bart Muys, Dirk Cattrysse, Stefaan Dondeyne, “Upgrading Geographic Information Systems to Spatio-Temporal Decision Support Systems”, *Mathematical and Computational Forestry & Natural-Resource Sciences*, vol. 3, pp36–41, 2011

[B2] **Gergely Bencsik**, Attila Gludovátz, László Jereb, “Integrált informatikai elemző keretrendszer alkalmazása a magyar felsőoktatásban”, Proc. of Informatika a felsőoktatásban konferencia 2011, pp1040–1047, Debrecen, Hungary, 2011 (*In Hungarian*)

[B3] **Gergely Bencsik**, Attila Gludovátz, László Jereb, “Adaptation of analysis framework to industry related economic problems”, Proc. of The Impact of Urbanization, Industrial and Agricultural Technologies on the Natural Environment: International Scientific Conference on Sustainable Development and Ecological Footprint, pp1–6, Sopron, Hungary, 2012

[B4] **Gergely Bencsik**, Attila Gludovátz, “Univerzális elemző keretrendszer gazdasági alkalmazásának lehetőségei különös tekintettel a termelővállalatokra”, Proc. of International Symposium on Business Information Systems (OGIK 2012), pp36–37, Győr, Hungary, 2012 (*In Hungarian*)

[B5] **Gergely Bencsik**, Attila Gludovátz, László Jereb, “Decision support framework with wood industrial application”, Proc. of 8th International PhD & DLA Symposium, pp154, Pécs, Hungary, 2012

[B6] **Gergely Bencsik**, Attila Gludovátz, “Adaptation of a universal decision support system in forestry”, Proc. of Implementation of DSS tools into the forestry practice, *Technical University Zvolen*, pp37–49, 2013

[B7] **Gergely Bencsik**, László Baczárdi, “Towards to decision support generalization: the Universal Decision Support System Concept”, Proc. of IEEE 19th International Conference on Intelligent Engineering Systems (INES), pp277–282, Bratislava, Slovakia 2015

Véletlen Összefüggések

[B8] **Gergely Bencsik**, László Baczárdi, “Statisztikai adatok közötti véletlen összefüggések tanulmányozása”, Proc. of Informatika a felsőoktatásban konferencia 2014, pp403–412, Debrecen, Hungary, 2014 (*In Hungarian*)

[B9] **Gergely Bencsik**, László Baczárdi, “Effects of Random Correlation on ANOVA and Regression”, Proc. of the 9th International Conference on Knowledge, Information and Creativity Support Systems, pp396–401, Lemesos, Cyprus, 2014

[B10] **Gergely Bencsik**, László Baczárdi, “New Methodology to Analyze the Random Impact Level of Mathematically Proved Results”, Proc. of 15th IEEE International Symposium on Computational Intelligence and Informatics (CINTI), pp33–38, Budapest, Hungary, 2014

[B11] **Gergely Bencsik**, László Baczárdi, “Novel methods for analyzing random effects on ANOVA and regression techniques”, (to be published in Springer Advances in Intelligent Systems and Computing, ISSN 2194-5357)

Egyéb publikációk:

- [B12] Attila Gludovátz, **Gergely Bencsik**, “Egy felsőoktatási képzés Balanced Scorecard alapú mintarendszer működésének demonstrálása”, Proc. of 5th International Symposium on Business Information Systems, pp88, Győr, Hungary, 2007 (*In Hungarian*)
- [B13] **Gergely Bencsik**, Attila Gludovátz, László Bacsárdi, “Tudásmenedzsment módszerek faipari alkalmazása”, Proc. of. Inno Lignum konferencia, Sopron, Hungary, 2010 (*In Hungarian*)
- [B14] **Gergely Bencsik**, “Analyzing the Adaptability Condition of Decision Support Systems and Data mining in Forestry”, Short Time Scientific Mission report, COST FP0804, Leuven, Belgium, 2010
- [B15] Márton Edelényi, **Gergely Bencsik**, Attila Gludovátz, “Adaptation possibilities of knowledge management tools in higher education”, Proc. of. Szellemi tőke, mint versenyelőny avagy A tudásmenedzsment szerepe a versenyképességben, pp883–897, Komárno, Slovakia, 2010
- [B16] **Gergely Bencsik**, Attila Gludovátz, “Experience with universal data analyses”, Proc. of Forest Management Decision Support Systems (FORSSYS) Conference, COSTFP0804, pp53–54, Umea, Sweden, 2013
- [B17] **Gergely Bencsik**, “Virtual Environment to simulate business processes related to ERP systems”, Presentation on Microsoft Dynamics Convergence 2014 conference, Barcelona, Spain, 2014
- [B18] Gergely Pieler, **Gergely Bencsik**, “Relevant and related area extraction from binary images”, Proc. of 11th International Symposium on Business Information Systems, Budapest, Hungary, 2014
- [B19] Péter Kiss, **Gergely Bencsik**, “Virtual Economic Environment”, Proc. of 11th International Symposium on Business Information Systems, Budapest, Hungary, 2014
- [B20] Péter Kiss, **Gergely Bencsik**, László Bacsárdi, “From ERP trainings to business: a new approach of simulations in economics”, Proc. of 12th International Symposium on Business Information Systems, pp49–50, Veszprém, Hungary, 2015
- [B21] Attila Gludovátz, **Gergely Bencsik**, László Bacsárdi, “IT challenges of a production system”, Proc. of 12th International Symposium on Business Information Systems, pp31, Veszprém, Hungary, 2015