

**Doktori (PhD) értekezés**

**Nyugat-magyarországi Egyetem**

**Simonyi Károly Műszaki, Faanyagtudományi és Művészeti Kar**

**Cziráki József Faanyagtudomány és Technológiák Doktori Iskola**

**Vezető: Prof. Dr. Tolvaj László egyetemi tanár**

**Doktori program: Informatika a faiparban**

**Programvezető: Dr. Jereb László**

**Tudományág: anyagtudomány és technológiák**

**Idősorok elemzési lehetőségeinek kiterjesztése és alkalmazhatósága erdészeti, faipari  
döntéstámogatásban**

**Készítette: Pödör Zoltán**

**Témavezető: Dr. Jereb László**

**Sopron**

**2014**

**Idősorok elemzési lehetőségeinek kiterjesztése és alkalmazhatósága erdészeti, faipari  
döntéstámogatásban**

**Értekezés doktori (PhD) fokozat elnyerése érdekében**

**a Nyugat-Magyarországi Egyetem Cziráki József Faanyagtudomány és Technológiák  
Doktori Iskolája**

**Informatika a faiparban programja**

**Írta:**

**Pödör Zoltán**

**Készült a Nyugat-Magyarországi Egyetem Cziráki József Doktori Iskola**

**Informatika a faiparban programja keretében**

**Témavezető: Dr. Jereb László**

**Elfogadásra javaslom (igen / nem)**

**(aláírás)**

**A jelölt a doktori szigorlaton ..... % -ot ért el,**

**Sopron/Mosonmagyaróvár .....**

**.....**

**a Szigorlati Bizottság elnöke**

**Az értekezést bírálóként elfogadásra javaslom (igen /nem)**

**Első bíráló (Dr. ....) igen /nem**

**(aláírás)**

**Második bíráló (Dr. ....) igen /nem**

**(aláírás)**

**(Esetleg harmadik bíráló (Dr. ....) igen /nem**

**(aláírás)**

**A jelölt az értekezés nyilvános vitáján.....% - ot ért el**

**Sopron/Mosonmagyaróvár,**

**.....**

**a Bírálóbizottság elnöke**

**A doktori (PhD) oklevél minősítése.....**

**.....**

**Az EDHT elnöke**

# 1 Tartalomjegyzék

1	Módszertani - irodalmi áttekintés .....	7
1.1	Döntéstámogatás .....	7
1.1.1	DSS rendszerek .....	7
1.1.2	Az erdészeti döntéstámogató rendszerek áttekintése.....	10
1.2	Idősorok fogalma, idősorok közti összefüggés-vizsgálatok módszerei .....	14
1.2.1	Főkomponens analízis és klaszterelemzés .....	16
1.2.2	Korrelációelemzés és regressziószámítás.....	18
1.2.3	Válaszfüggvény-elemzés.....	20
1.2.4	Összefüggések stabilitását javító eljárások .....	22
1.2.5	Mozgó időintervallumok vizsgálata, evolúciós módszer .....	23
1.3	Idősorok és elemzési módszereik erdészeti alkalmazásai.....	26
1.3.1	Vizsgálati körülmények.....	27
1.3.2	Elemzési módszerek alkalmazásai .....	29
2	CReMIT módszer .....	34
2.1	CReMIT alpmódszer.....	36
2.1.1	Ablakok definiálása az idősor elemei felett.....	36
2.1.2	Transzformált idősorok létrehozása.....	37
2.1.3	A CReMIT módszer mint magasabb absztrakciós szint .....	38
2.2	CReMIT kiterjesztés több változóra.....	39
2.2.1	CReMIT kiterjesztése a függő paraméterre.....	39
2.2.2	Kiterjesztés több független paraméterre .....	41
2.2.3	A többváltozós lineáris regresszió alapjai .....	42
2.3	A CReMIT módszer alkalmazása.....	48
2.3.1	Egyváltozós alkalmazás – példák.....	48
2.3.2	Többváltozós kiterjesztések – példák.....	55
3	A CReMIT alkalmazásának támogatása.....	61
3.1	Az elemző folyamat .....	61
3.1.1	Az adatelőkészítő modul .....	62
3.1.2	Az elemzési modul.....	63
3.2	Töréspont keresés idősorokban .....	64
3.2.1	Töréspontok keresésének módszerei.....	65
3.2.2	Töréspont keresés t-próba alkalmazásával és a töréspontok jelentősége .....	70
3.3	Növekedési görbék.....	76

3.3.1	Növekedési görbék .....	77
3.3.2	Görbeillesztés folyamata, módszerei .....	85
3.3.3	Növekedési görbe illesztése - eredmények.....	88
4	Alkalmazások.....	96
5	Összefoglalás .....	98
5.1	Következtetések .....	99
5.2	Tovább lépési lehetőségek.....	100
6	Tézisek .....	102
7	Köszönetnyilvánítás.....	103
8	Irodalomjegyzék .....	104

## **Kivonat**

### **Idősorok elemzési lehetőségeinek kiterjesztése és alkalmazhatósága erdészeti, faipari döntéstámogatásban**

Az erdészeti döntéstámogatás feladata, hogy különböző szakmai kérdésekben tudományosan és szakmailag megalapozott döntéseket hozzunk. Ehhez elengedhetetlen a rendelkezésre álló adatsorok minél mélyebb feldolgozása, az összefüggések feltárása, megismerése. Szakirodalmi áttekintésünk alapján meghatároztuk azokat a statisztikai, matematikai módszereket, melyeket az erdészetben, alapvetően a fanövekedés-klíma adatokat leíró adatsorok közötti kapcsolatok vizsgálatában jellemzően alkalmaznak. Ezen tapasztalatok alapján kidolgoztuk a CReMIT (Cyclic Reverse Moving Interval Techniques) módszert, mely a periodikus idősorok szisztematikus kiterjesztésével képes a vizsgálatok körét az alkalmazott elemzési módszerektől függetlenül kiszélesíteni. A módszer elvi, matematikai alapjait egy idősor vektor reprezentációján mutattuk be, a CReMIT módszert több változóra is kiterjesztettük, illetve a többváltozós elemzéseket meg is valósítottuk. Az eljárást egy egységes struktúrát biztosító elemzési folyamatba építettük be, az adatelőkészítő és az elemző modulok mellé. A kialakított elemző folyamat előkészítő részének két speciális eleme a töréspontok keresése idősorokban és az adatsorokra illeszthető görbék. Az eljárást több valós probléma kapcsán, illetve kutatás-fejlesztési projektek keretében alkalmaztuk fák növekedése, egészségi állapota, mortalitása, valamint lepkefogási adatok vizsgálatában. A dolgozat egyértelmű célja az alkalmas módszer kidolgozása és ismertetése volt, a bemutatásra kerülő adathalmazok elemzése kapcsán ezért az alkalmazhatóság demonstrálását tekintettük elsődlegesnek és nem tekintettük magunkat kompetensnek erdészeti eredmények megfogalmazásában.

Az összefüggés-vizsgálatok esetében fontos kérdés az összefüggések időbeli stabilitása, annak vizsgálata, hogy vannak-e az idősorokban, s ha igen, hogyan mutathatóak ki olyan pontok, ahol statisztikailag is igazolhatóan ugrásszerű változások történtek. Bemutatjuk a töréspontok kimutatására szolgáló eljárásokat és az ezek közül kiválasztottat alkalmaztuk meteorológiai adatsorokra. A CReMIT módszert felhasználó elemzési eredményeken keresztül azt is igazoljuk, hogy ezek a hirtelen változások mennyire komoly hatással vannak az összefüggés-vizsgálatok eredményeire. A kérdéskörnek a feltételezett klímaváltozás különleges jelentőséget ad. Az éven belüli, heti-kétheti gyakoriságú növekedési adatokra megfelelő növekedési görbét illesztve képesek vagyunk a tényleges éves növekedési adatokon túl további, a növekedést jellemző paramétereket objektív módon definiálni és azokat az elemzésekben felhasználni.

Vizsgálatokat végeztünk az általunk felhasznált adatsorokra illeszthető növekedési görbékre vonatkozóan, kiválasztottuk a legkedvezőbbnek tekinthetőt, az illesztett görbe alapján további növekedési paramétereket definiáltunk és azokat a CReMIT módszer alapján részletesen elemeztük.

## **Abstract**

### **The Extension of Time Series Analysis Opportunities and its Applicability in Decision Support of Forestry and Wood Sciences**

The decision support systems related to various professional fields help make scientifically based professional decisions. Forestry gives rise to a whole range of tasks and problems, in which the time sequence of collected data holds significance. Derived from the periodicity of the basic data, the CReMIT (Cyclic Reverse Moving Interval Techniques) method, and an analytical process, with three modules was developed to extend the range of possible analytical processes. The first module in the process prepares the time series to be studied also in different ways, such as detecting break points and fitting the right curves to the data series. The varied applications definitely underline the general approach of the method, and highlight its wide-scale applicability. The procedures developed can also be applied in any other fields where the analysis of relationships between periodic time series is of fundamental importance.

## **1 Módszertani - irodalmi áttekintés**

### **1.1 Döntéstámogatás**

#### **1.1.1 DSS rendszerek**

A döntéstámogató rendszerek (DSS), illetve magának a döntéstámogatás fogalmának definiálása rendkívül nehéz feladat, hiszen gyakorlatilag minden olyan – akár egyszerű, akár komplex – rendszer, mely segít a szakmai döntések meghozatalában, döntéstámogató rendszernek tekinthető. Ez lehet akár egy egyszerű Excel tábla, akár egy összetett célszoftver. Az az egyszerű definíció, mely szerint a döntéstámogató rendszer segíti, megalapozottá teszi bizonyos döntéseinket, ugyan kezdetben megfelelő lehetett, de jelenleg már annyira általános, hogy nehezen értelmezhető és nem igazán megfelelő.

Ma már a definíció szerves részét képezi az interaktivitás, azaz a döntéstámogató rendszer egy olyan interaktív, számítógép alapú rendszer mely adatbázisok és modellek felhasználásával segíti a döntéshozókat az adott területet jellemző, tipikusan nem jól strukturált problémák megoldásában. (Nem jól strukturáltak tekintjük azt a problémát, melynek nem ismerjük összes

megoldási alternatíváját, azok értékeit és egymáshoz viszonyított preferenciáit). Kezdetben a döntéstámogató rendszerek egyedi célú, kifejezetten egy-egy jól strukturált probléma megoldására létrehozott célrendszerek voltak (Wikström és mtsai, 2011). Később, az adattárházak és az OLAP adatbázis-kezelők, az adatbányászat térnyerésével azonban a döntéstámogató rendszerek elvesztették célalkalmazás jellegüket és mára már egy általános célú döntés-előkészítési, döntéstámogatási eszközként tekinthetünk rájuk. Egy döntéstámogató rendszer alapvetően három fő komponenst tartalmaz: adatbázisok, a rájuk épülő modellezési réteg és végül a megjelenítésért felelős réteg (Power, 2002).

Az adatok rendelkezésre állhatnak a döntéstámogató rendszerhez tartozó saját adatbázisban (ez ma még nagyon sokszor egyedi PC-ket jelent) vagy ideális esetben a döntéstámogató rendszer adattárházában. Mindenképpen lényeges momentum, hogy a tartalmilag és szerkezetileg is megfelelően felépített adatbázis a kulcsa nem csak az adatbányászatnak, hanem az OLAP és relációs adatbázis alapú döntéstámogató rendszereknek is, mivel az adatokat ezekből tudjuk kigyűjteni, lekérni.

A kommunikációs réteg feladata, hogy biztosítsa a felhasználó és a rendszer közti kapcsolat lehetőségét. Ez vonatkozik az adatok be- és kivételére egyaránt, gyakorlatilag ez a réteg biztosítja a rendszer interaktivitását. Mivel a DSS rendszerek felhasználói jellemzően nem informatikai szakemberek, így a felhasználhatóság szempontjából fontos, hogy egy barátságos, könnyen kezelhető felhasználói felülettel rendelkezzen a rendszer. A kommunikációs alrendszer részét képezik a hagyományos értelemben vett hardveres periféria eszközök, illetve az adatbevitelt, a párbeszédet, lekérdezéseket, modellek megalkotását lehetővé tevő különböző szoftvereszközök is.

A modellező réteg feladata, hogy a megfelelő adatbázisban rendelkezésre álló adatokból hasznos információkat nyerjünk ki. Ezek jelenthetnek egyszerű, akár Excelben is megvalósítható függvényeket (összeg, átlag, minimum, maximum, stb.), vagy összetettebb matematikai, statisztikai, illetve adatbányászati eljárásokat, modelleket. A modell mindig egy valós, létező probléma absztrakciója, mely lehetővé teszi, hogy szimuláljuk egy adott döntés meghozatalának következményeit anélkül, hogy annak tényleges negatív hatásaival a valóságban egyelőre számolnunk kellene. A modellek számtalan előnnyel rendelkeznek, például:

- kevésbé költséges és sok esetben kevésbé időigényes, mint a tényleges, valós kísérletek végrehajtása,
- a jó modellek képesek a jövőre vonatkozó előrejelzésekre is,
- a modellezés segíti a tényleges fizikai folyamatok pontosabb megismerését és megértését.

Ugyanakkor nem szabad elfeledkeznünk bizonyos hátrányokról sem, mint például a hiba lehetősége, vagy a modellek megalkotásához szükséges matematikai ismeretek szükségessége.

A modellek különböző szempontok szerint csoportosíthatóak, így például az időbeli viselkedés alapján beszélhetünk statikus és dinamikus modellekről. Vannak modellek, melyek valószínűségi változókat használnak, ezeket sztochasztikus vagy valószínűségi, míg a többit determinisztikus modellnek nevezzük. Funkcióik alapján megkülönböztethetünk elemző modelleket, melyek egy-egy döntési alternatíva kimenetelét szimulálják, de magát a döntést a szakértő felhasználóra bízják. Az optimalizáló modellek adott problémára meghatározzák a lehetséges megoldások halmazából kiválasztott legoptimálisabb megoldást. A DSS rendszerekben alkalmazott modellek sokféle feladat megoldására lehetnek alkalmasak, azonban jellemzően négy fő feladatkörbe szokás azokat csoportosítani:

- a „mi van, ha” jellegű elemzések: a vizsgálatba bevont változók, paraméterek értékeit, a köztük fennálló összefüggéseket vizsgáljuk,
- érzékenységvizsgálatok: az előző eset egy speciális típusa, amikor egy változó értékét változtatjuk és vizsgáljuk, hogy ez a változás hogyan hat a többi változóra,
- célkereső elemzés: az előző két eset ellentettjének tekinthető abban az értelemben, hogy itt egy kiemelt változónak adunk egy előre definiált célértékét és a többi paraméter értékét úgy kell meghatározni, hogy ezt az előre definiált célértéket elérjük vagy legalábbis adott pontossággal megközelítsük,
- optimalizáló elemzések: az előző eset kiterjesztésének tekinthetőek abban az értelemben, hogy azt a célt fogalmazzuk meg, hogy a bizonyos kiválasztott változók által elért érték optimális legyen.

A döntéstámogató rendszereket lehet csoportosítani a működésük, a felhasznált adatok, illetve a felhasználónak nyújtott támogatás alapján is:



- modell alapú: különböző, már létező alternatívákat kínál fel a megfelelő döntési modell összeállításához,
- adat alapú: itt adatok, általában idősorok kezelésével, elemzésével támogatja a döntési folyamatot,
- dokumentum alapú: különböző dokumentumok feldolgozásával, az azokban történő keresési lehetőség felkínálásával, megvalósításával támogatja a döntési folyamatokat,
- tudás alapú: a vizsgált problémát leíró tudást felhasználva nyújt segítséget,
- kommunikáció alapú: a hálózati technika és kommunikáció kínálta lehetőségeket használja fel, például a csoportos döntéshozatal során.

A DSS rendszerek száma óriási, csak az erdészet területét tekintve is több tucatra rúg az ismertebb és elterjedtebbnek tekinthető, és vélhetően több százra a kisebb programok száma. A továbbiakban kizárólag az erdészetben alkalmazott döntéstámogatási rendszerekkel foglalkozunk. Megmutatjuk néhány rendszer alapján a jellemző feladatokat, kérdéseket, majd felvillantjuk az ezen területen alkalmazott DSS rendszerek széles körét és egy rövid áttekintést nyújtunk róluk.

### ***1.1.2 Az erdészeti döntéstámogató rendszerek áttekintése***

Az erdészeti tudomány területén is rengeteg DSS rendszer létezik és működik. Ezek között vannak általánosabb és specifikusabb céllal létrehozottak. Éppen ezek nagy száma miatt a teljes körű és részletes áttekintésre nem vállalkozhatunk, azonban a szakma által is számon tartott, fontosabb eszközöket röviden körüljárjuk.

Packalen és mtsai (2013) áttekintették a FORSYS wiki rendszerében található erdészeti döntéstámogató rendszereket, összesen 23 ország 62 DSS rendszerével foglalkoztak. Figyelembe véve ezt a nagy – és még korántsem teljes – számot a rendszerek alapvető szempontok szerinti jellemzésére és összehasonlítására törekedtek. A könnyebb áttekinthetőség és értelmezhetőség miatt mindezt táblázatos formában tették meg. Vizsgálták a szoftvereket funkcionalitásuk, képességeik szerint az alábbi szempontokat figyelembe véve: térbeli és nem térbeli adatelemzés, impakt analízis (egy adott változás hatásainak elemzése), kockázat-elemzés, dinamikus erdő és tájkép szimulációk, optimalizációs képesség, stb. Megállapították, hogy a szoftverek egyik fele nem alkalmas, a másik fele képes térbeli adatelemzésre is. Jelentős részük képes erdő szimulációkra, ugyanakkor a többi, előbb felsorolt funkciót csak a szoftverek kis része képes megvalósítani.

Korábban már említettük, hogy egy DSS rendszer alapvetően három fő alkotóelemet tartalmaz, ezért ezek alapján is elemezték a 62 vizsgált rendszert. Megállapították, hogy 14 kivétellel a többség egyedi PC-ken fut, nem kapcsolódik nagyobb adattárházakhoz, hanem a gépen található saját adatbázisokból dolgozik és ezekbe egyelőre nincs is beépítve ennek a közvetlen lehetősége. A felhasználóval való kapcsolat, a neki nyújtott információk tekintetében a rendszerek zöme passzívnak tekinthető, azaz anélkül próbál segíteni, hogy világos és egyértelmű megoldási, döntési javaslatokat adna. Csak nagyon kicsi részükre (ugyancsak 14) jellemző, hogy konkrét megoldásokat, vagy döntési javaslatokat nyújtana. Végül vizsgálták a DSS rendszerek által nyújtott támogatás módjait, amiket már mi is felsoroltunk a DSS rendszerek ismertetésének végén. Megállapították, hogy a rendszerek egy nagy része tudás-, egy kisebb része pedig modell, valamint adat alapúnak tekinthető.

Ezek az eszközök, ahogy már korábban mi is említettük, általában nem informatikai szakembereknek készülnek, így fontos, hogy felhasználóbarát felületet, elérhetőséget biztosítsanak hozzá. A vizsgált eszközök majdnem mindegyike teljesítette (47 a 62-ből) ezt, azaz rendelkezik grafikus interfésszel. Azok a rendszerek pedig, amelyek ilyen felülettel nem rendelkeznek olyan moduloknak vagy klienseknek tekinthetők, amelyek részkomponensként egy-egy nagyobb rendszerbe integrálhatóak.

Reynolds és mtsai (2008) 10 általuk kiválasztott és fontosnak vélt rendszert tekintettek át és különböző szempontok (mint például a szoftver elsődleges célja, a kiértékelés alapja, tartalmaz-e mesterséges intelligenciát, alkalmas-e jövőbeni előrejelzésre, stb.) alapján vizsgálták azokat. A vizsgált erdészeti DSS rendszerek az AFFOREST, DSD, EMDS, ESC, FORESTAR, ForestGALES, LMS, NED, SADfLOR és a Woodstock voltak. A kiválasztás alapja az volt, hogy régóta (legalább 10 éve) fejlesztés alatt álló európai, ázsiai és észak-amerikai rendszerek is szerepeljenek az áttekintésben, illetve, hogy a kiválasztott rendszerek ezen eszközök sokszínűségét és sokféleségét is reprezentálják.

Több eszköz (EMDS, LMS, SADfLOR) is foglalkozik erdészeti tájképtervezési, azaz az erdészeti útvonalak, építmények, menedzsment kérdések problémáival, míg a ForestGALES alapvető célja a szél és viharok hatásainak elemzése és vizsgálata.

A vizsgált rendszerek közül hat is tartalmaz növekedési modelleket, amelyből látszik, hogy ez az erdészet egyik fontos kérdése. Viszont az, hogy a növekedési probléma kérdését milyen paraméterekből, jellemzőkből vezetik le, nagyon sokféle alapon nyugodhat. Lehetnek ezek meteorológiai, környezeti jellemzők, mesterséges beavatkozások hatásainak vizsgálatai, erdőtelepítési, tájképtervezési kérdések. Az AFFOREST rendszert alapvetően a környezeti

hatások és az erdősítés, erdőtervezés közti relációk vizsgálatára hozták létre, ehhez pedig fontos, hogy ismerjük a különböző környezeti komponensek és a vizsgált fafajok kapcsolatrendszerét.

A DSD rendszer létrehozásának egyik alapvető célja az volt, hogy a vizsgált projektterületeken mely fajok, fafajcsoportok kerülhetnek perifériára a klímaváltozás hatásainak figyelembevételével. Három fő szempontot tekintettek az elemzések során: a feldolgozható fa mennyisége, a természet sokféleségének megőrzése, valamint az adott terület produktivitásának fenntartása vagy fejlesztése. Az ESC rendszer egyik alapvető célja, hogy becsléseket adjon a hektáronkénti törzsnövekedés mértékére. Fő befolyásoló tényezőként a hőmérsékletet tekintik, az aszályosság mértéke, a talajnedvesség és az óceántól való távolság (kontinentalitás) mellett.

A FORESTSTAR program egyik fő modulja az erdészeti kitermelést segítő modul, mely alapvetően két fő részből áll: (i) Az egyik rész feladata, hogy a területek különböző szempontok szerinti összehasonlítása alapján kiválassza a területről a kitermelés szempontjából optimális részletet. (ii) A második rész a terület kiválasztása után meghatározza a kitermelés optimális módját, ebbe beleértve a sorrendet is.

A fenti alkalmazások jól mutatják, hogy az erdészet és a faipar mennyire szorosan összefüggenek egymással, hiszen a faipari feldolgozás szempontjából fontos a nagy mennyiségű és jó minőségű alapanyag. A faipar lehetőségeit sokszor már a telepítés, majd az azt követő erdőgazdálkodás határozza meg. A kérdéskör fontos részét jelenti a fák növekedése, hiszen egyértelmű cél, hogy egy adott területen adott körülmények között a későbbi felhasználás, feldolgozás számára optimális mennyiségű és minőségű növedéket érjünk el.

A fa felhasználása nagyon sokrétű, ebbe beletartozik például a bútorgyártás, a fűrészáru-, a falemez-, a parketta-, az épületasztalos-ipari termék- és a tároló fatermékek gyártása. Magyarországon az éves szinten kitermelt fa mennyisége KSH adatok alapján 2012-ben 7 732 000 m<sup>3</sup> volt, ami, bár kissé kevesebb, mint a 2011-es év adata (8 080 000m<sup>3</sup>), azonban több, mint 2004 óta a 2011-et megelőző bármely más évben volt.

A KSH adatok azt mutatják, hogy Magyarországon az élőfakészlet 2004 óta folyamatosan emelkedő tendenciát mutat, hasonlóan a kitermelt fa mennyiségéhez. A teljes élőfaállomány 366 252 000 m<sup>3</sup> volt 2012-ben. A terület gazdasági fontosságára utal, hogy szintén KSH adatok szerint a fa és faáru, valamint a bútor és bútorelem export forgalma 2012 első háromnegyed évében több mint 350 milliárd forint volt, ami körülbelül 50 milliárd forintos bővülés a korábbi

év azonos időszakához képest (Milei, 2013). A faipar és a fafeldolgozás így gazdaságilag kiemelt fontosságú kérdés Magyarországon, aminek alapja az erdőgazdálkodás.

Az erdőgazdálkodás, és ezen keresztül a faipar egyik kritikus kérdése a fák növekedése (alapvetően átmérő irányú növekedést értve alatta) és a meteorológiai paraméterek kapcsolatainak vizsgálata, amelyet Reynolds és mtsai (2008) korábban is említett munkája, illetve a vizsgált DSD, ESC, AFFOREST, FORESTAR döntéstámogató rendszerek sokféle szempontú növekedés-klíma kapcsolat vizsgálati lehetőségei és a kapott eredmények is alátámasztanak. Számos cikk, vizsgálat foglalkozik ezen kapcsolatok feltárásával és elemzésével.

Az elkövetkezendő időszak csapadék és hőmérséklet viszonyaira vonatkozóan a klímaszcenáriók alapján vannak megállapítások. Magyarország vonatkozásában például Pieczka és mtsai (2011) a PRECIS regionális klímamodell eredményeit mutatják be három különböző scenárió vonatkozásában a 2071-2090-es időszakra vonatkoztatva az 1961-90-es bázisidőszakhoz képest. Mindhárom forgatókönyv 95%-os szignifikancia szinten melegedést jósol a kárpát-medencei területre különösen a nyári hőmérsékletek vonatkozásában. Emellett a csapadék- és hőmérsékleteloszlások átrendeződése is várható: gyakoribbak lesznek a nyári aszályok, ugyanakkor a téli csapadék mennyiségének növekedés várható. Gálos és mtsai (2012) a REMO regionális klímamodell (Jacob 2001, Jacob és mtsai 2007) alapján a 21. század utolsó harmadára 3-3,5°C-os emelkedést jósolnak a nyári időszakra; a nyári csapadékösszeg csökkenése pedig akár 30-35% is lehet az 1961-1990-es bázisidőszakhoz képest. A klímaváltozás hatásaira történő felkészülése fontosságára hívja fel a figyelmet Mátyás (2010) és Mátyás és mtsai (2011) is cikkeikben.

A faipar erdőgazdálkodástól, az erdők egészségi állapotának és növekedésének klímaviszonyoktól való függése egyértelműen indokolja, hogy olyan módszereket keressünk, amelyek lehetővé teszik a meteorológiai jellemzők változásának hatáselemzését. Ezekben a kutatásokban a felhasznált adatsorok időbelisége fontos tényező, azaz idősorok vizsgálatáról, azok közötti kapcsolatok kereséséről van szó. A fejezet hátralévő részében ezért először bemutatjuk az idősorokkal kapcsolatos, e területen érdekes és fontos alapfogalmakat, valamint a kapcsolódó elemzési módszereket. Ezt követően áttekintést adunk az erdészeti szakirodalom fanövekedés-klímaviszonyok kapcsolatával foglalkozó szelétéről, bemutatjuk a vizsgálatok körülményeit, és az azokban alkalmazott vizsgálati módszereket.

## 1.2 Idősorok fogalma, idősorok közti összefüggés-vizsgálatok módszerei

Számtalan tudományterületre jellemző, hogy a gyűjtött adatok időbeliséggel ellátottak, azaz fontos tényező az adatok időbelisége, ilyenkor beszélhetünk idősorokról. Az idősorok vizsgálata, a közöttük fennálló kapcsolatok elemzése a statisztika és az adatbányászat egy nagyon fontos, kiemelt területe. Az idősorokkal kapcsolatban két fontos kutatási terület határozható meg, az egyik az idősorok modellezése, azok belső struktúrájának megismerése, a másik pedig az idősorok közötti kapcsolatok feltárása (Tusnádi és Ziermann, 1986). Mindkét eljárás alkalmas lehet előrejelzések, jövőre vonatkozó jóslások előállítására.

Az idősorok modellezésén belül alapvetően két jól definiált irány különíthető el, a determinisztikus és a sztochasztikus modellezés. Előbbi esetben az idősort jellemzően 3 (4) komponensre bontják fel  $Y = T + P + A$  alakra (Abonyi, 2006), ahol

- *Tendencia (T)*: ez az idősor fő vonala, jellemzője.
- *Szezonálitás (P)*: egy rövid periodicitású (legfeljebb éves hosszúságú) összetevő. Például minden évben nyáron a fagyalt eladások mennyisége jelentősen megnövekszik az év többi időszakához képest.
- *Véletlen hatások (A)*: ezek az előre nem jelezhető, véletlen elemek. Például az utóbbi évek gazdasági válságának kitörése. Ezt szokás hibatagnak is nevezni.

Némely modellben megkülönböztetnek még egy negyedik komponenset, ez a *ciklikusság (C)*. Ez a szezonális komponenssel ellentétben egy hosszabb (akár 5-10 év) ciklusú folyamat, azonban a legtöbb modellben ezt beolvasztják a tendencia komponensbe. A determinisztikus modell alkalmazása során az alapötlet, hogy az idősort a három komponensnek megfelelően bontjuk fel, trend, periodicitás és véletlen elemekre. Ezt követően ezen komponenseket elemezzük és segítségével tudjuk például a jövőbeni értékeket előre jelezni. Az egyes komponensek elkülönítésére, definiálására számtalan technika létezik, azok részletezésétől azonban itt eltekintünk. A véletlen itt csupán olyan értelemben fordul elő, hogy a determinisztikus modell általában nem képes pontosan leírni a valóságot, mindig vannak olyan tényezők, amelyek a modell szempontjából véletlennek tekinthetők, így nem modellezhetőek.

A sztochasztikus modellezés esetében az idősort egy sztochasztikus folyamatnak fogjuk fel, mely függ a korábbi elemektől, a véletlen hatásoktól és a korábbi hibáktól. Azaz itt a véletlent a folyamat szerves részének tekintjük és nem csak egy felesleges rossznak. Így a modell itt is minden egyes pontban eltérhet a tényleges adattól, azonban úgy tekintünk erre az eltérésre, hibára, mint ami az idősor további elemeinek előállításában fontos szerepet tölt be. Ellentétben

a determinisztikus modellel itt nem bontjuk komponenseire az idősort, hanem a maga teljes egységében kezeljük és próbáljuk meg az idősor elemei közti belső kapcsolatokat feltárni (például AR, MA, ARMA, ARIMA, stb. modellek). Az autoregresszív (AR) folyamatok lényege, hogy a folyamat (legalábbis részben) a saját múltjára épülő lineáris regresszióként írható fel. A *moving average* (MA) folyamatban gyakorlatilag minden egyes elmozdulás egy vagy több elmozdulás átlagával jellemezhető. A folyamatok rendje utal arra, hogy visszamenőleg hány tényező értéke határozza meg az aktuális elem értékét. Az idősorok vizsgálata, a determinisztikus, sztochasztikus modellezések önmagukban is egy-egy komoly kutatási területet jelentenek, ezért jelen dolgozatban ezek részleteire a továbbiakban nem térünk ki.

Már a determinisztikus modellezés kapcsán is felmerül a regressziós technika alkalmazása a trendfolyamat feltárásában (Abonyi, 2006). Ebben az esetben a regressziós technikát analitikus trendszámításnak szokták nevezni. Az itt alkalmazott regresszió esetén a vezető indikátor az idő összetevő, amire egy egyszerű egyváltozós regressziót alkalmazunk. A cél, hogy az idősor adataira egy olyan kompakt alakban felírható  $f(t)$  függvényt illesszünk, amely a mérési adatokat valamilyen értelemben (pl. legkisebb négyzetek módszere) a legjobban leírja. A felírt  $f(t)$  függvény típusa alapján beszélhetünk lineáris, polinomiális, illetve egyéb nemlineáris regresszióról (exponenciális, logaritmusos, stb.).

Ha idősorainkat megfelelő módon össze tudjuk hasonlítani egymással, újabb fontos összefüggésekhez, információkhoz juthatunk. Csoportokba rendezhetjük, osztályozhatjuk azokat, eltéréseket kereshetünk, különböző szabályokat tárhatunk fel. Így ilyenkor az alapfeladat az, hogy hogyan definiáljuk, illetve mérjük két adatsor hasonlóságát. Erre az egyik lehetőség, ha definiáljuk az idősorok távolságát, illetve egy másik lehetőséget jelent az idősorok közti korreláció mérése. Az aktuális (erdészeti) kutatásokat tekintve a legtöbb vizsgálat az utóbbi módszer mellett döntött. A korrelációelemzés előnye a távolság alapú modellel szemben, hogy nincs szükség komolyabb előkészítő lépésekre (például az értékkészletek rendezésére) és különböző értéktartományú adatsorok összevetésére is alkalmas.

A továbbiakban az idősorok közti összefüggések vizsgálatának azon módszereit mutatjuk be, melyeket az erdészetben jellemzően a növekedésvizsgálatokkal kapcsolatos kutatásokban alkalmaznak. Ezek ismertetése előtt azonban röviden bemutatunk két olyan speciális módszert, melyeket felhasználunk egyes elemzési technikák részeként, illetve az adatsorok előkészítésében is alkalmazhatóak.

### 1.2.1 Főkomponens analízis és klaszterelemzés

A főkomponens elemzés (Principal Component Analysis – PCA) a matematikai statisztika egy jól ismert eszköze (Joliffe, 2002), mely  $p$ , esetleg egymással korreláló  $x_1, x_2, \dots, x_p$  változók olyan  $z_1, z_2, \dots, z_q$ -val jelölt lineáris kombinációit határozza meg, ahol a kapott  $z_1, z_2, \dots, z_q$  főkomponensek ( $q \leq p$ ) egymással már nem korrelálnak és mégis jelentős információvesztés nélkül képesek leírni az eredeti változókat. A főkomponenseket szokás látens, nem megfigyelt változóknak is nevezni, ahol  $z_1$  képviseli az eredeti adathalmaz varianciájának legnagyobb,  $z_2$  a második legnagyobb részét, és így tovább. Ideális esetben az eredeti adatok varianciája adekvát módon leírható néhány olyan  $z$  változóval, melyek varianciája nem elhanyagolható és ezek fogják az adatok „dimenzióit” mérni.

Az eljárás módját szolgálta arra, hogy az adatok mögött rejlő kevesebb, eleve nem korreláló feltételezett változókat megtaláljuk. Szemléletesen megfogalmazva a főkomponens analízis egyenértékű a koordinátarendszer olyan elforgatásával, amely azt eredményezi, hogy a tengelyek rendre az adathalmaz legnagyobb szórásainak irányába állnak be. A módszer további előnye a lehetséges adatredukció.

Fontos hangsúlyozni, hogy a PCA nem mindig működik abban az értelemben, hogy az eredeti, nagyszámú változókat kisebb számú változókká alakítjuk. Sőt, ha az eredeti változók nem korrelálnak egymással, akkor az elemzés nem is szolgálta eredményt. A legjobb eredményt akkor kaphatjuk, ha az eredeti változók erősen korrelálnak egymással - akár pozitív, akár negatív értelemben. Ebben az esetben könnyen elképzelhető, hogy 20-30 eredeti változót adekvátan reprezentálhat 2-3 főkomponens. Ha pedig ez teljesül, akkor a fontosabb főkomponensek (melyek varianciája elég nagy) lesznek csupán érdekesek, hiszen ezek fogják az adatok „dimenzióit” mérni.

A kapott főkomponensek (nem megfigyelt, látens változók) értelmezése azonban nehézkes lehet, mert azok az eredeti, megfigyelésen alapuló változók lineáris kombináció, így közvetlen értelemmel nem bírnak. A közvetlen magyarázat előállítására általában a faktorok (standardizált főkomponensek) és az eredeti változók közti korrelációs mátrixot, a faktorstruktúrát használják.

A csoportosítás másik lehetséges módja az adatbányászat egy jól ismert eszköze, a klaszteranalízis (Han és Kamber, 2006, Abonyi, 2006), amely egy adathalmaz pontjainak hasonlóság szerinti csoportosítását jelenti. Szokás felügyelet nélküli tanulásnak is nevezni, mert az osztályozással ellentétben itt nem rendelkezünk előzetes információkkal a kialakítandó

csoportokról. A klaszterezés célja, hogy az elemeknek egy olyan partícióját adja, amelyben a közös klaszterbe kerülő elempárok lényegesen hasonlóbba egymáshoz, mint azok a pontpárok, melyek két különböző csoportba sorolódtak. A klaszteranalízis kiindulópontja az elemek közötti hasonlóság vagy távolság fogalom, ami sokféle módon definiálható. Ezzel kapcsolatban általában rendelkezünk előzetes információkkal, amelyek alapján kiszámíthatjuk ezeket a hasonlóságokat vagy távolságokat.

A klaszteranalízis több szempontból is igen hasznos eljárás. Először is, gyakran szükség van arra, hogy a „valódi” csoportokat határozzuk meg. Például a különböző elméleteknek megfelelően, többféle módon csoportosíthatjuk a vásárlói magatartásformákat, a klaszteranalízis egy objektív módszert kínál a csoportok kialakításához. Másodszor, a klaszteranalízist adatredukció céljából is használhatjuk. Az adatredukció fontosságáról a főkomponens analízis kapcsán már volt szó. Harmadrészt a klaszteranalízis segítségével kapott, esetleg meglepő klaszter-csoportosulások új fényt vethetnek az eddig vizsgált kapcsolatok rendszerére.

A klaszterezési módszerek két alapvető csoportba sorolhatóak:

- hierarchikus módszerek: legközelebbi társ vagy centroid módszer,
- a  $K$ -közép módszer.

A hierarchikus módszereknél nem kell előzetesen ismernünk a létrehozandó klaszterek számát, ebben különféle grafikonok segítenek majd bennünket. A  $K$ -közép módszernél ezzel szemben már kiinduláskor adott a klaszterek száma, a mi feladatunk csak a megfigyelések besorolása. A másik fontos különbség, hogy egy hierarchikus módszer általában időigényesebb, mint egy  $K$ -közép klaszterezés, amelyet emiatt gyakran neveznek gyors klaszterezésnek is. A két klaszterező módszer lényege röviden a következőkben összegezhető.

### **Hierarchikus klaszterezés**

A hierarchikus eljárások az egyes személyek, objektumok, esetek közötti távolság meghatározásával kezdődnek. A csoportok, klaszterek kialakítása történhet összevonáson vagy felosztáson alapuló módszerekkel. Az összevonó módszerek abból indulnak ki, hogy minden egyes elem egy önálló csoportot alkot, majd fokozatosan vonják össze az egyelemes, illetve kisebb csoportokat a nagyobb csoportokba. Ezzel szemben a lebontó módszerben kezdetben az összes elem egyetlen csoportba tartozik, és ezt a csoportot osztjuk fel kettőre, majd egyre több csoportra. Mindkét eljárás addig ismétlődik, míg egy adott megállási feltételt el nem érünk.



## A $K$ -közép módszer

A  $K$ -közép algoritmus minden egyes elemet ahhoz a klaszterhez sorol, amelyiknek a középpontja a legközelebb esik az adott elemhez.

Az algoritmus lépései a következők (MacQueen, 1967):

- meghatározza a klaszterek számát ( $k$ ),
- véletlenszerűen létrehoz  $k$  számú klasztert, és meghatározza minden klaszter közepét, vagy azonnal létrehoz  $k$  véletlenszerű klaszter középpontot,
- minden egyes pontot abba a klaszterbe sorol, amelynek középpontjához a legközelebb helyezkedik el,
- kiszámolja az új klaszter középpontokat,
- addig ismétli az előző két lépést (iterál), amíg valamilyen konvergencia kritérium nem teljesül (általában az, hogy a besorolás nem változik).

Az algoritmus legnagyobb előnye az egyszerűsége és a sebessége, ami lehetővé teszi alkalmazását nagy adattömbökön is. Hátránya viszont, hogy nem ugyanazt az eredményt adja különböző futtatások után, mert a klaszterezés eredményét befolyásolja a kezdeti véletlen alapú besorolás.

### 1.2.2 Korrelációelemzés és regressziószámítás

A statisztikai alkalmazások gyakori problémáját képezik azok a vizsgálatok, melyek során azt elemezzük, hogy egy vagy több független változó milyen hatással van a függő változóra, milyen erős a kapcsolat közöttük, illetve hogyan írható le függvényszerűen ez a reláció (Spiegel, 1995). A kapcsolat elemzésnek az első fajtája a korreláció-, az utóbbi a regressziószámítás. A két technika egymással szorosan összefügg, egymást kiegészítik.

#### Pearson-féle korrelációs együttható

Két változó között felírt lineáris függvénykapcsolatokat számszerűen a korrelációs együttható ( $r$ ) és a determinációs együttható ( $R^2$ ) jellemezheti, amely paraméterek az esetleges kapcsolatok irányát ( $r$ ) és erejét is mérik. A korrelációszámításnál fontos szem előtt tartani, hogy a számítás a populációból vett mintákkal történik, de az eredmény az egész populációra érvényesítendő. Az  $r$  jelentőségének megalapozottsága egy  $n - 2$  szabadságfokú  $t$ -statisztika felhasználásával (ahol  $n$  a minta elemszáma) ellenőrizhető.

Fontos megemlíteni, hogy – bár sok kutatásban erre nem fektetnek hangsúlyt – a lineáris korrelációanalízis alkalmazhatóságának komoly feltételei vannak, úgymint az adatsorok normalitása, a vizsgált adatsorok mintáinak véletlensége és függetlensége, a kiugró (outlier) elemek vizsgálata. E feltételek teljesülését az elemzések megkezdése előtt ellenőrizni kell.

A korrelációs számítás segítségével általában a változók közti lineáris kapcsolatot vizsgáljuk, azonban az alapmódszer általánosításával mód van nemlineáris kapcsolatok elemzésére is. A vizsgálatok jelentős része mégis megelégszik a lineáris korrelációelemzés módszerével annak egyszerűsége, illetve a nemlineáris vizsgálat viszonylagos bonyolultsága miatt. Utóbbi esetben két lehetőség áll rendelkezésre:

1. megfelelő transzformációkkal (pl. logaritmizálás) lineárisra tesszük a kapcsolatot és az ismertetett eljárást használjuk,
2. nemlineáris regressziós görbét illesztünk a vizsgált változók által meghatározott ponthalmazra és a korrelációs együtthatóval gyakorlatilag azt jelezzük, hogy mennyire pontosan illeszkednek ezek a pontok a regressziós görbére, azaz a regresszió pontosságát értékeljük. Ennek nehézségét – már egy változó esetében is – az okozza, hogy nehéz meghatározni az illesztendő görbét.

A korreláció- és regresszióanalízis kiterjeszhető több változóra is, ahol szintén megkülönböztethetünk lineáris és nemlineáris eseteket. A többváltozós regresszió esetében az illeszkedés jóságának mérésére általában a korrigált determinációs együtthatót ( $\hat{R}^2$ ) szokás használni, mert ez figyelembe veszi a bevont független paraméterek számát is. A témakörrel a 2.2.3. fejezet részletesebben is foglalkozik.

### **Rangkorreláció**

A lineáris (paraméteres) korrelációanalízis akkor végezhető el – az egyéb feltételek teljesülése mellett – ha a vizsgált adatok eloszlásáról előzetes információkkal rendelkezünk. Amennyiben ilyen információnk nincs vagy a vizsgált adatsorok eloszlása nagyon ferde, illetve outlier értékek szerepelnek benne, akkor nemparaméteres korrelációanalízist lehet alkalmazni a két vizsgált paraméter vonatkozásában. A nemparaméteres korrelációanalízis az eredeti adatsorainkból csak az adatok nagyságrendi viszonyait veszi figyelembe, de a konkrét értékeket nem.

Az utóbbi korrelációt rangkorrelációnak is hívják (Spiegel, 1995). A módszer az egyváltozós lineáris korrelációs együttható speciális esetének tekinthető. A lineáris korreláció gyakorlatilag azt méri, hogy a két adatsorra felírható regressziós egyenes mennyire pontos, azonban ez a

módszer csak numerikus adatsorok esetén alkalmazható és érzékeny a kilógó adatokra. A rangkorrelációs együtthatók ezzel szemben azt mérik, hogy a két adatsor együtt változik-e, de nincsenek tekintettel azok arányaira és alkalmazhatóak például ordinális típusú adatsorok esetén is (alma íze és színezettsége közti kapcsolat vizsgálata). Példaként a (0,1); (10,100); (101,500); (102, 2000) számpárok sorozata esetén a rangkorrelációk teljes egyezést mutatnak, hiszen mindkét sorozat nő, míg a lineáris korrelációs érték csak  $r = 0.456$ , vagyis a számpárok viszonylag távol esnek a regressziós egyenestől.

A rangkorreláció mérésére jellemzően kétfajta módszert szoktak alkalmazni:

- Spearman-féle rangkorreláció,
- Kendall-féle rangkorreláció.

Mindkét módszer alkalmas az adatsorok úgynevezett együttfutásának vizsgálatára. A kettő között a különbség az idősorok azonos indexű  $x_i, y_i$  elempárjai közti rangszámok (az idősor elemeinek nagyság szerint növekvő sorba rendezése esetén az adott indexű elem sorrendbeli pozíciója) különbségének eltérő felhasználásában rejlik. A Spearman-féle rangkorreláció esetében  $r_s = 1 - \frac{6 \sum_{i=0}^n d_i^2}{n^3 - n}$ , ahol  $d_i$  az idősorok  $x_i, y_i$  elempárjai rangszámainak különbsége. Míg a Kendall-féle korrelációs együttható,  $r_K = 1 - \frac{4Q}{n(n-1)}$ , ahol  $Q$  a két idősor rangszámoszlopainak megfelelő ellenkező irányú eltéréspárok darabszáma.

Tekintsük példaként az  $x$ : 11, 13, 15, 12, 14 és  $y$ : 13, 15, 14, 11, 12 idősorokat. Az adatsorok elemeinek rangszámai  $rang_x$ : 1, 3, 5, 2, 4,  $rang_y$ : 3, 5, 4, 1, 2, így  $r_s = 1 - \frac{6 \cdot 14}{20} = 0,3$ . A Kendall-féle korrelációs együtthatóhoz a rangszámoszlopok minden lehetséges elempárja alapján képzett eltérések az  $x$  vektorra: +, +, +, +, +, -, +, -, -, +, az  $y$  vektorra +, +, -, -, -, -, -, -, -, +. Például az  $x$  vektor (1,5) elempárja természetes sorrendben van, + jelet kap, míg a (3,5) nem természetes sorrend, ezért - jelet kap. Az ellentétes irányú eltéréspárok száma,  $Q = 4$ , a korrelációs együttható értéke pedig  $r_K = 1 - \frac{4 \cdot 4}{20} = 0,2$ .

### 1.2.3 Válaszfüggvény-elemzés

Az egyszerű lineáris korrelációelemzés módszere nagyon sokszor nem elégséges komplex rendszerek mélyebb összefüggéseinek megértéséhez, valamint nehézséget okozhat a független változók túl nagy száma is és a köztük fennálló korrelációs kapcsolatok (multikollinearitás). A multikollinearitás problémájának kezelésére több statisztikai módszer is ismert, ezek közül pedig a válaszfüggvény-elemzés (RF, Response Function, Fritts, 1976; Biondi és Waikul, 2004) az, amit kifejezetten a növekedés-klíma kapcsolatok vizsgálatához fejlesztettek.

A meteorológiai paraméterekre mint független változókra alkalmazott PCA segítségével egymással már nem korreláló, új magyarázó változókat definiálunk. Az eredeti függő és a

kapott látens változókra pedig többváltozós lineáris regressziót hajtunk végre. Az RF gyakorlatilag az így kapott regressziós egyenlet, ezért tekinthetjük a módszert a főkomponens-regresszió egy speciális esetének is.

Az RF-elemzés fő célja, hogy a lehető legzártabb módon írja le a környezeti paraméterek és a növekedés közti statisztikai kapcsolatokat. Megmutatja, hogy melyek azok a környezeti paraméterek, amelyek hatással vannak a növekedésre, továbbá meghatározza a feltárt kapcsolatok erősségét és irányát.

A módszer lényege a normalizált környezeti elemekre végrehajtott PCA alkalmazásában rejlik, ami egyrészt átláthatóbbá teszi az elemzést a független változók számának jelentős csökkentésével, másrészt a kapott főkomponensek növekedéssel való összevetése miatt alkalmas a több, illetve különböző típusú független változók együttes hatásainak vizsgálatára is. Az előállított főkomponensek olyan, az eredeti független változókból megfelelő módon előállított már korrelálatlan látens változók, melyek jól reprezentálják az eredeti adathalmazt, így az előállított főkomponensekre mint új független változókra alkalmazzuk a többváltozós regressziót. Mivel ezek az új változók korrelálatlanok, így felesleges a többlépéses (például a stepwise-regression) regresszió alkalmazása.

Az RF függvényben használt főkomponensek együtthatói a megfelelő lineáris transzformációval visszaalakíthatóak az eredeti független paraméterekre jellemző regressziós együtthatókra. Biondi (2004) alapján a folyamat formálisan az alábbi módon írható le:

- Jelölje  $X$  és  $Y$  a már normalizált független változók mátrixát, illetve a szintén normalizált függő változó vektorát.
- Számítsuk ki az  $X$ -hez tartozó normalizált sajátvektorokat tartalmazó  $A$  négyzetes mátrixot, amiből kiválasztjuk a lényegeseket, így kapjuk  $A'$ -t.
- $X$  és  $A'$  felhasználásával előállítjuk a főkomponenseket tartalmazó  $Z$  mátrixot,  $Z = X \cdot A'$ .
- $Z$  és  $Y$  közt többváltozós lineáris korrelációanalízist alkalmazva előállítható a  $Z \cdot K = Y$  regressziós egyenlet, illetve annak megoldását tartalmazó  $K$  együttható vektor.
- Végül  $K$  ismeretében már előállíthatóak azok a  $(B = A \cdot K)$  válaszfüggvény együtthatók, melyek az eredeti változók közti kapcsolatokat jellemzik.

A válaszfüggvény-elemzés hátránya, hogy alapvetően a főkomponensekre és nem közvetlenül az eredeti magyarázó változókra ad eredményt, inkább csak meghatározza a releváns (pl. klimatikus) paraméterek természetét valamint megmutatja a kapcsolat erejét és irányát. Az RF-

elemzés korlátai és hátrányai már Fritts előtt is ismertek voltak és több kutató is felhívta rá a figyelmet (Brett, 1982, Norton, 1983), mégis számtalan kutató használta és használja azóta is ezt a speciális módszertant ezen a kutatási területen.

#### **1.2.4 Összefüggések stabilitását javító eljárások**

Az elemzési eredmények megbízhatóságának és stabilitásának javítására több módszer is ismert. A módszerek mindegyikének alapja, hogy a meglévő mintából további (rész-, al- stb.) mintákat állítanak elő. Az igazi fejlődés itt a számítástechnikának volt köszönhető, mivel ezek a módszerek nagyon számításigényesek:

- véletlen részminták, csoportok,
- kiegyenlített félminták,
- jackknife,
- bootstrap-módszerek.

E módszerek alapelve úgy is megfogalmazható, hogy az „észt helyettesítsük erővel”, azaz az analitikus formulákat nagy tömegű számítással pótoljuk. A számításigényesség az egyre erősebb gépek birtokában már nem okoz nehézséget és a felsorolt eljárások lényegében csak abban különböznek egymástól, hogy miként generáljuk az egyes rész- vagy almintákat.

A továbbiakban két olyan statisztikai eljárást ismertetünk – Efron (1979) tanulmánya alapján – , melyek alkalmasak az összefüggések eredményei megbízhatóságának és stabilitásának javítására.

A jackknife az egyik legelterjedtebb, kezdetleges formájában már az 1950-es években használt technika, amelyet eredendően becslőfüggvények torzításának csökkentésére alakítottak ki, ám mára a varianciabecslés egyik kedvelt módszerévé vált. A módszer lényege, hogy minden egyes lépésben az eredeti mintából elhagy egy vagy több elemet (1-törléses jackknife, vagy  $d$ -törléses jackknife), így képezve a másodlagos mintákat. Az 1-törléses jackknife eljárás iteratív módon egyesével végiglépked az adatsor egyes elemein és kihagy egyet-egyet közülük. A  $d$ -törléses jackknife  $\binom{n}{d}$  darab mintát állít elő, úgy hogy minden egyes lépésben az eredeti minta  $n$  eleméből  $d$  elemet elhagy. A jackknife során előállított másodlagos minták hossza – a törölt elemek számától függően – rövidebb, mint az eredeti. A módszer minden újabb mintára kiszámítja a vizsgált statisztikát, végül az egyes iterációkra kapott statisztikákból meghatározható egy átlag vagy egy konfidencia intervallum. Ily módon becsülhető a statisztika változékonysága az adott másodlagos mintákra számolt statisztikák szórásából. Különösen

hasznos lehet ez az eljárás, ha az adatsorban kilógó értékek vannak, vagy az eloszlás széles szórású (Ramachandran és Tsokos, 2009). Gond lehet azonban az, hogy vannak olyan változók (például decilisek, kvantilisek) amelyek esetében, feltehetően a mintából adódó becslés torzítotttsága miatt, nem ad kielégítő eredményt.

A bootstrap-módszer kifejlesztése (Efron, 1979) nemcsak a SE kiszámítását tette lehetővé olyan esetekben, amikor a korábbi technikákat nem lehetett használni, hanem ráirányította a figyelmet arra, hogy a számítógépek kapacitásának növekedésével a replikációs eljárások elméletileg is új megközelítéseket adnak: az ismételt mintavétel lehetővé tette az analitikus formulák helyett számítógépes módszerek alkalmazását.

A bootstrap-módszert leginkább akkor használják, amikor a statisztika eloszlása ismeretlen vagy a normalitás feltételei nem teljesülnek. Ezekben az esetekben például a korrelációs együtthatók szignifikanciájának vizsgálata is bizonytalanná válik (Biondi, 1997). A módszer az eredeti mintából – pszeudo-véletlen számok felhasználásával – ismétléses mintavételezést alkalmazva képez másodlagos mintákat úgy, hogy a másodlagos minták hossza rendszerint megegyezik az eredeti minta hosszával. A visszatevéses mintavételnek köszönhetően egy-egy elem akár többször is előfordulhat a másodlagos mintákban, illetve lehet olyan elem, amely egyszer sem szerepel azokban. A bootstrap nem csak arra használható, hogy a becslőfüggvények standard hibáját becsüljük, hanem egyben alkalmas konfidencia intervallumok becslésére, illetve statisztikai hipotézisek tesztelésére is.

A bootstrap módszer abban is különbözik a jackknife-től, hogy míg az utóbbi esetén az eredeti adatsor méretétől függ az előállított részminták száma, addig a bootstrap-nél ez az érték elméletileg korlátlan. A pszeudo-véletlen számok alkalmazása miatt a bootstrap ugyanazon adatsoron való újabb futtatása más eredményt szolgáltat, míg a jackknife esetén egy megismételt eljárás biztosan ugyanazt az eredményt adja. A bootstrap eljárás hatékonysága függ az eredeti minta reprezentativitásától (Ramachandran és Tsokos, 2009).

### ***1.2.5 Mozgó időintervallumok vizsgálata, evolúciós módszer***

Sok esetben nem használják fel a teljes rendelkezésre álló idősort az elemzés egy-egy lépésében. Ennek egyrészt oka lehet az eljárás pontosságának mérése tanuló és validáló halmazok alkalmazásával, másrészt megfelelő hosszúságú adatsorok esetén elemezhető a paraméterek közti kapcsolatok időbeli változása is.

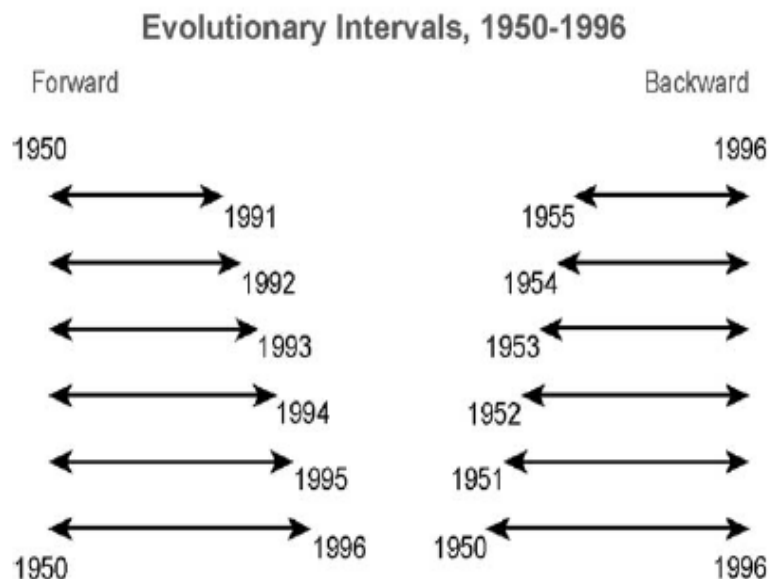
Az előre- és a visszalépéses evolúciós módszerek, valamint a mozgó intervallumok használhatók az utóbbi problémák vizsgálatára (Biondi, 2000, Biondi és Waikul, 2004). Alkalmazhatóságuk feltétele, hogy egy adott lépésben a ténylegesen vizsgált adatsorok hossza

( $H$ ) kevesebb legyen, mint az eredeti adathossz 80%-a, illetve, hogy  $H$  legalább kétszerese legyen a magyarázó változók számának.

Az evolúciós technika minden egyes lépésben egy-egy időegységgel (pl. évvel) növeli a kiinduló adatsor hosszát (1. ábra). Definiálunk egy  $L$  hosszúságú bázis intervallumot, melynek kezdőpontját rögzítjük és minden lépésben egy-egy évvel növeljük az aktuálisan vizsgált intervallum hosszát.

1. előre evolúció: első lépésében a vizsgált intervallum kezdőpontja az időben legkorábbi adat, majd minden egyes lépésben, időben előre haladva növeli a vizsgált adatsor hosszát,
2. hátra evolúció: esetében a vizsgált intervallum kezdőpontja az időben legkésőbbi adat, majd minden egyes lépésben, időben visszafelé haladva növeli a vizsgált adatsor hosszát.

Az eljárást addig ismételjük, míg az adatsorok végére nem érünk.



1. ábra – Evolúciós technika (Biondi és Waikul, 2004)

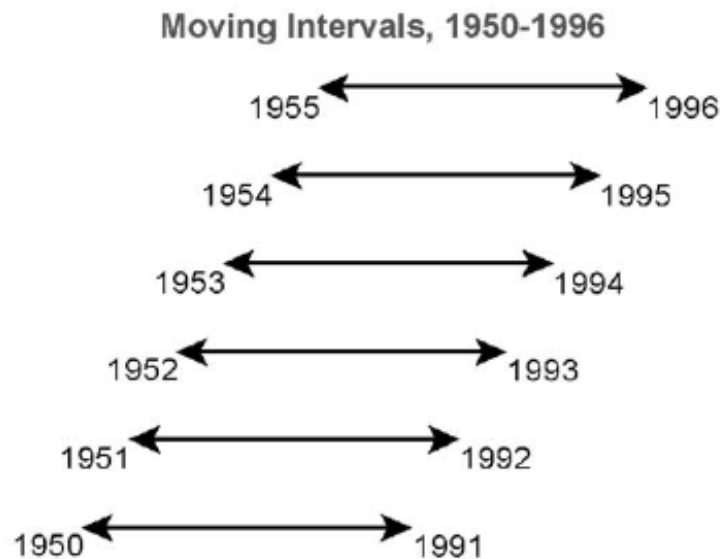
Tegyük fel az egyszerűség kedvéért, hogy két adatsorunk van, melyek hossza  $n$ :  $x_1, x_2, \dots, x_n$  és  $y_1, y_2, \dots, y_n$ . Definiáljuk  $L$ -et a korábban már leírt feltételek alapján, azaz  $L < n * 0.8$  és  $L > 2 * pred$ , ahol  $pred$  jelöli a magyarázó változók számát.

Ekkor a létrehozott intervallumok az alábbiak:

1. táblázat – Evolúciós intervallumok

<b>Előre evolúció</b>				
<b>1.</b>	<b>2.</b>	<b>...</b>	<b><math>n - L</math>.</b>	<b><math>n - L + 1</math>.</b>
$x_1, x_2, \dots, x_L$	$x_1, x_2, \dots, x_{L+1}$	...	$x_1, x_2, \dots, x_{n-1}$	$x_1, x_2, \dots, x_n$
$y_1, y_2, \dots, y_L$	$y_1, y_2, \dots, y_{L+1}$	...	$y_1, y_2, \dots, y_{n-1}$	$y_1, y_2, \dots, y_n$
<b>Hátra evolúció</b>				
<b>1.</b>	<b>2.</b>	<b>...</b>	<b><math>n - L</math>.</b>	<b><math>n - L + 1</math>.</b>
$x_{n-L+1}, x_{n-L+2}, \dots, x_n$	$x_{n-L}, x_{n-L+1}, \dots, x_n$	...	$x_2, x_3, \dots, x_n$	$x_1, x_2, \dots, x_n$
$y_{n-L+1}, y_{n-L+2}, \dots, y_n$	$y_{n-L}, y_{n-L+1}, \dots, y_n$	...	$y_2, y_3, \dots, y_n$	$y_1, y_2, \dots, y_n$

Mozgó intervallumok esetén rögzíteni kell a vizsgált intervallum hosszát, melynek kezdőpontja az időben legkorábbi adatpont, majd minden egyes lépésben időben egy-egy évvel előbbre ugrik az intervallummal (2. ábra).



2. ábra – Mozgó intervallumos technika (Biondi és Waikul, 2004)

Azonos feltételek mellett az egyes intervallumok:

2. táblázat – Mozgó intervallumok

<b>1.</b>	<b>2.</b>	<b>...</b>	<b><math>n - L</math>.</b>	<b><math>n - L + 1</math>.</b>
$x_1, x_2, \dots, x_L$	$x_2, x_3, \dots, x_{L+1}$	...	$x_{n-L}, x_{n-L+1}, \dots, x_{n-1}$	$x_{n-L+1}, x_{n-L+2}, \dots, x_n$
$y_1, y_2, \dots, y_L$	$y_2, y_3, \dots, y_{L+1}$	...	$y_{n-L}, y_{n-L+1}, \dots, y_{n-1}$	$y_{n-L+1}, y_{n-L+2}, \dots, y_n$

Mindhárom technika esetében minden egyes lépésre kiszámítjuk a korrelációs együtthatókat, így az eljárások végére korrelációs együtthatók sora áll rendelkezésre, mely alkalmas arra, hogy a vizsgált változók közti kapcsolatok hosszútávú időbeli változását vizsgáljuk.



### **1.3 Idősorok és elemzési módszereik erdészeti alkalmazásai**

Kutatásaink alapvető célja az idősorok közti kapcsolatok elemzése, feltárása elsősorban olyan erdészeti jellegű vizsgálatokban, mint a fanövekedés és a klimatikus jellemzők kapcsolatai. A fák növekedési folyamata (akár átmérő, akár magassági növekedésről beszélünk) régóta az erdészeti kutatások fontos területe. Ma a kutatási téma aktualitását – tágabban értelmezve – a klímaváltozás vizsgálata adja. A környezeti tényezők fák növekedésére gyakorolt hatása közvetlenül nem szolgáltat érveket a klímaváltozás mellett vagy épp ellene, de mindenképpen segít helyére tenni azoknak a lehetséges (prognosztizált) változásoknak a hatásait, amelyek a szaksajtóban és a közmédiában is nagy nyilvánosságot kapnak.

Figyelembe véve, hogy a konkrét elemzési feladatok Magyarországon gyűjtött adatokhoz köthetőek a szakirodalmat is elsősorban európai, vizsgálatainkkal hasonlóságokat mutató (fajok, földrajzi adottságok, klimatikus viszonyok) cikkek áttekintésével dolgoztuk fel (Manninger és mtsai, 2011a; Edelényi és mtsai, 2011a).

A vizsgált kutatási terület rendkívül bőséges és széleskörű szakirodalommal rendelkezik, sok szakterületet érint, maguk a környezeti tényezők is számosak lehetnek. A növekedés fogalma is elég tág, hiszen a fák növekedését is többféle módon jellemezhetjük. Éppen ezért bizonyos megszorításokat alkalmazunk a továbbiakban a szakirodalmi áttekintés vonatkozásában.

Alapvetően azokkal a kutatásokkal foglalkoztunk, amelyek a növekedést az átmérő változásaként definiálják függetlenül attól, hogy ezt az adatot közvetlenül mérik (átmérő-, illetve kerületmérés) vagy származtatják (évgyűrűelemzés). A sokféle környezeti tényező közül a léghőmérsékletet és a csapadékot, tehát a két legáltalánosabban használt meteorológiai tényezőt emeltük ki. Ennek oka alapvetően a gyakorlatiasságban keresendő: ez a két környezeti paraméter az, ami általában hozzáférhető adott területre és adott időszakra. Világos, hogy a növekedés sok egyéb környezeti és nem környezeti paraméter által is definiált, a bemutatott és általunk kifejlesztett módszerek alkalmasak is ezek kezelésére. Külön figyelmet fordítottunk az olyan publikációkra, amelyekben ezen tényezők valamilyen időszaki és/vagy késleltetett hatását vizsgálták.

Mivel a disszertáció központi eleme egy újfajta elemzési módszer és alkalmazási lehetőségeinek bemutatása, ezért a szakirodalmi áttekintés kapcsán is alapvetően az elemzési módszertan eszközeire fektetjük a hangsúlyt. Azonban a vizsgálatok pontosabb megismeréséhez röviden összefoglaljuk az egyes vizsgálatok körülményeit, céljait is.

A növekedés, mint erdészeti paraméter és a környezeti tényezők kapcsolatrendszerének vizsgálata az adatok jellege miatt idősorok közti relációk keresését jelenti. A függő változó kutatásainkban a növekedési adatsor, míg a független változók a különböző környezeti paraméterek (jellemzően a két alapvető klimatikus jellemző, a csapadék és a hőmérséklet). Az áttekintett szakirodalom alapján körülhatároltuk azokat az elemzési módszereket, amelyeket akár éves, akár éven belüli növekedés esetén a fenti összefüggés-elemzésekben alkalmaztak. Az elemzési módszertan szempontjából gyakorlatilag nincs jelentősége, hogy konkrétan milyen paramétereket használunk függő, illetve független változónak.

Az idősorok közötti kapcsolatok megismerése, az ezen alapuló és a jövőre vonatkozó előrejelzések készítése lényeges elem a döntéstámogató rendszerekben, hiszen sok esetben a jelenben kell meghoznunk olyan döntéseket, melyek eredménye a jövőben, akár 20-30 év, vagy még több idő távlatában fog realizálódni. Gondoljunk például a már említett klímaszcenáriókra.

### ***1.3.1 Vizsgálati körülmények***

Ahogy korábban már említettük, a szakirodalmi áttekintés elsősorban, de nem kizárólag európai kutatásokra összpontosít. Az áttekintett szakirodalomban a vizsgálatok egész Európa területére kiterjednek. Az elemzések többnyire egy földrajzi értelemben szűkebben vett régió mintaterületeire vonatkoznak (Carrer és Urbinati, 2001; Lebourgeois, 2004; Manninger, 2004; Bouriaud és mtsai, 2005; van der Werf, 2007; Wilczyński és Podlaski, 2007; Pichler és Oberhuber, 2007; Čufar és mtsai, 2008; Oberhuber és mtsai, 2008; Feliksik és Wilczyński, 2009; Gutiérrez és mtsai, 2011; Maxime és Hendrik, 2011), és az eredményeket erre a kiemelt területre vonatkozóan fogalmazzák meg. Egy adott országon belül több régiót vizsgált Szabados (2002), Tuovinen (2005), Manninger (2008), Martín-Benito és mtsai (2008), Szabados (2008), Bogino és mtsai (2009), Novák és mtsai (2010), a kapott eredmények közti kapcsolatokat is vizsgálták, illetve általános érvényű összefüggéseket próbáltak megfogalmazni.

Vannak átfogó, nagy mintaterületekre kiterjedő vizsgálatok is. Briffa és mtsai (2002) például az északi félgömb 387 fenyővel borított mintaterületét (köztük észak- és dél-európaiakat) vették figyelembe kutatásukban. Pärn (2003) három északkelet-észtsországi ipari terület közelében lévő, légszennyezett erdő növekedését elemezte. Piovesan és mtsai (2005) az olasz-félsziget bükköseiből kialakított hálózat évgyűrű-elemzési adatait tanulmányozták. Franciaországban Lebourgeois és mtsai (2005) a RENECOFOR hálózat 15 idős, különböző klíma- és talajviszonyokkal jellemezhető bükkös állományát vizsgálták. Savva és mtsai (2006) a Tatra lengyelországi részén 10 különböző tengerszint feletti magasságú terület mintáit használták. Mäkinen és mtsai (2000), valamint Büntgen és mtsai (2006) egy országon

(Finnországon, illetve Svájcban) belül szintén több régió adatait használták tanulmányukban. Nagyobb földrajzi régióra, Közép- és Észak-Európára vonatkozó megállapításaikat ismertették Mäkinen és mtsai 2002-ben és 2003-ban, míg Büntgen és mtsai (2007) a Tátra lengyel és szlovák részéről származó adatsorokat vizsgálták. A cél minden esetben az volt, hogy a rendelkezésre álló, nagy területet lefedő adatok felhasználásával egyrészt általános érvényű állításokat fogalmazzanak meg, illetve a területek közti különbségeket is azonosítsák.

A kutatásokban a hőmérsékleti és csapadékadatok legtöbb esetben a közeli meteorológiai állomások adatsoraiból, ritkábban modellekből vagy helyben mért adatokból számolták, illetve származtatták. A tanulmányozott cikkekkel kapcsolatban elmondható, hogy a környezeti tényezők és a növekedés kapcsolatának elemzésében a meteorológiai paramétereket leginkább a havi átlagos léghőmérséklet és a havi szabad területi csapadékösszeg képviseli. Az ettől való eltérés – néhány kivételtől eltekintve – a bevont paraméterek számának növelését jelenti. Ennek oka vélhetően az, ami miatt elemzéseinkben mi is erre a két paraméterre szűkítettük a vizsgálati kört: a csapadék és hőmérséklet adatok azok, melyek általánosan, bármely területre elérhetőek, így alkalmasak az eredmények kiterjesztésére.

A növekedéshez felhasznált adatokat alapvetően két módszerrel szokták meghatározni. A legelterjedtebb eljárás az évgyűrűelemzés, melynek mind a terepi, mind a feldolgozási része jól kidolgozott. Az elemzésekhez leggyakrabban a fából növedékcspot vesznek, ritkábban a kidöntött fákból vágott korongot használják fel. Egyes kutatásokban az évgyűrűszélességen túl a korai és kései pászta szélességét (Lebourgeois és mtsai, 2004; Tuovinen, 2005; Lebourgeois és mtsai, 2005; Savva és mtsai, 2006; van der Werf és mtsai, 2006; Pichler és Oberhuber, 2007; Martín-Benito és mtsai, 2008), olykor sűrűségét (Briffa és mtsai, 2002; Büntgen és mtsai, 2007) is meghatározták.

A növekedési adatok másik forrása a mellmagassági átmérő, illetve a kerület mérése. Ezek a módszerek a mérések gyakoriságától és az alkalmazott mérőeszközöktől függően az átmérő éven belüli kismértékű változását (növekedés és zsugorodás) is képesek kimutatni, automata mérőeszközök esetén akár a napon belülieket is (Manninger, 2004; Bouriaud és mtsai, 2005).

Kézi és automata kerületmérő szalagokat egyidejűleg alkalmazott Manninger (2004). A heti gyakorisággal mért kézi mérőeszközökkel határozta meg az éves kerületnövekedést a vegetációs időszak előtt lemért kezdőkerület arányában, míg az automata szalagok órás adatai alapján értékelte a környezeti tényezők napi hatását.

Bouriaud és mtsai (2005) a törzsön különböző magasságban rögzített, az átmérő változását sugárirányban mérő automata mérőeszközöket (az ún. pont-dendrométereket) alkalmazták. Az órás adatokból konvertálták az éves növekedési görbét, illetve határozták meg az éven belüli vizsgálatok számára az 5 napos növekedési értékeket.

Gutiérrez és mtsai (2011) kézi kerületmérő szalagokkal vizsgálták a növekedést. Az átlagos mintavételezési gyakoriság  $28 \pm 1$  nap volt, de tavasszal és ősszel sűrűbben ( $17 \pm 1$  nap), míg nyáron és télen ritkábban ( $39 \pm 2$  nap) mértek. A kapott értékeket korrigálták, szűrték, majd havi átmérőváltozást számoltak belőlük. A kerületmérésből számolt évgűrűszélességeket mintafákból vett korongok adataival hasonlították össze. Novák és mtsai (2010) a növekedést a fiatalabb állományokban az átmérő mérésével, az idősebb állományokban évgűrűelemzéssel határozták meg.

Néhány cikkben (Manninger, 2004; van der Werf és mtsai, 2007; Gutiérrez és mtsai, 2011) az éven belüli növekedést vizsgálták a meteorológiai paraméterekkel összefüggésben, azonban az áttekintett cikkek többségében éves növekedési adatokat és jellemzően havi bontású meteorológiai paramétereket használtak.

### ***1.3.2 Elemzési módszerek alkalmazásai***

Az idősorok vizsgálata az erdőszetben jellemzően a determinisztikus megközelítést jelenti. Azonban a sztochasztikus módszert is gyakran alkalmazzák az évgűrű adatok vizsgálatához, az abban fellépő anomáliák szűréséhez. Druckenbrod (2004) cikkében megfogalmazza, hogy az évgűrűk idősorokon alapuló elemzése jó lehetőség a fa növekedésében fellépő zavarok feltárásában. Hasonló vizsgálatokat végez dolgozatában Monserud (1986), aki szintén idősorok segítségével elemzi az évgűrűket. Dolgozatában azt vizsgálja, hogy melyik ARIMA modell a legmegfelelőbb a modellezésre.

Az erdőszeti elemzések egyik alapvető, fontos feladata a növekedéssel szignifikáns kapcsolatot mutató havi (időszaki) környezeti paraméterek meghatározása. Az adatsorok időbelisége releváns tényező, így az alpmódszer jellemzően a lineáris korrelációelemzés. Ez önmagában azonban még nem feltétlenül elégséges egy ilyen összetett rendszer vizsgálatára, így a ténylegesen alkalmazott módszerek jellemzően ennek az alaptechnikának valamiféle továbbfejlesztéseit jelentik.

Az adatsorok előkészítésére, illetve a válaszfüggvény elemzési eljárás részeként a kutatók gyakran PCA és klaszteranalízis módszereket alkalmaztak. Mivel a növekedési adatok mérése jellemzően egyedenként történik, sok esetben több száz vagy több ezer egyed bevonásával

történnék, értelmezhető eredmények gyakorlatilag csak az adatsorok csoportosításával állíthatók elő. Ha nem áll rendelkezésre egy előre definiált, kötött csoportosítási szempont, akkor valamilyen módon nekünk kell csoportosítani a növekedési adatsorokat, és egységes mutatót képezni. Erre így az egyik lehetőség a főkomponens elemzés, amelyet Mäkinen és mtsai (2000), Briffa és mtsai (2002), Mäkinen és mtsai (2002), Macias és mtsai (2006), Büntgen és mtsai (2007), Oberhuber és mtsai (2008), Novák és mtsai (2010) használtak.

Több erdészeti kutatásban is alkalmaztak klaszteranalízist az évgyűrű adatsorok csoportosítására (Piovesan és mtsai 2005). Feliksik és Wilczyński (2009) az erdei-, a luc-, a duglász-, a szitkaluc- és jegenyefenyő növekedését klaszterelemzéssel vizsgálták, míg a különböző évgyűrű adatsorok közti hasonlóságokat, illetve különbözőségeket meghatározó faktorok kijelölésére főkomponens analízist alkalmaztak.

Mindkét bemutatott módszer alkalmas a növekedési adatsorok csoportosítására és a vizsgált változók mennyiségének csökkentésére, hiszen a hasonló növekedést produkáló egyedek így egy csoportba kerülnek, és képezhető például az erős, közepes és gyenge növekedést adó egyedek halmaza. A módszer lehetőséget teremt annak a vizsgálatára is, hogy a különböző növekedési erélyű csoportok másként reagálnak-e a környezeti paraméterekre.

A növekedés mint függő változó és külön-külön az egyes havi környezeti paraméterek közti lineáris korreláció meghatározásának alapmódszere a szimpla egyváltozós lineáris korrelációanalízis. Több cikkben is alkalmaztak az éves növekedés és a havi bontású csapadék és/vagy hőmérséklet adatok közti kapcsolatokat vizsgálatában lineáris korrelációanalízist (Briffa és mtsai, 2002; Mäkinen és mtsai, 2003; Wilczyński és Podlaski, 2007; Manninger, 2008; Szabados, 2008; Feliksik és Wilczyński, 2009; Novák és mtsai, 2010; Petráš és Mecko, 2011). Manninger (2008) a különböző csapadék adatok, valamint a csapadék és növekedés adatok kapcsolatának vizsgálatában használt lineáris regresszióanalízist. A cikkek között azonban viszonylag kevés olyan található, amely a lineáris korrelációelemzést csak önmagában alkalmazta, mert az alkalmazott vizsgálatok általában tartalmaztak más típusú elemzéseket is mellette. Így például Savva és mtsai (2006), illetve Rybniček és mtsai (2009) a tengerszint feletti magasság szerepét vizsgálták a kapcsolatokban. Gutiérrez és mtsai (2011) az összefüggés vizsgálatok mellett részletesen elemezték az évgyűrű adatsorokat. Hasonló vizsgálatot végzett Maxime és Hendrik (2011) is. Pichler és Oberhuber (2007) lineáris korrelációanalízist alkalmazott (a válaszfüggvény-analízis mellett) a 2003-as meteorológiailag extrém év növekedésre gyakorolt hatásainak elemzésében.

Rangkorreláció-analízist alkalmazva az adatsorok együttfutását vizsgálta Csókáné (2002) fák évgyűri adatai és a csapadék adatsorok kapcsolatában. Bilodeau-Gauthier és mtsai (2011) Spearman-féle rangkorrelációt alkalmaztak a különböző talajtípusokon élő fák növekedési adatainak összehasonlítására. Briffa és mtsai (2008) különböző évgyűri adatsorok összehasonlítására használt Kendall-féle rangkorrelációt. Jansen és mtsai (2012) különböző korú egyedek növedékadatait vetették össze Spearman-féle rangkorrelációt alkalmazva és evvel a kor szerepét vizsgálva.

Többváltozós lineáris regresszióelemzést alkalmazott Bouriaud és mtsai (2005) a növekedés mint függő és a pentádokra bontott csapadék, talajvízhiány, átlaghőmérséklet és az erdősűrűség mint független változók között. Tuovinen (2005) az évgyűrik szélessége, a korai és kései pászta, valamint a havi csapadékösszeg és átlaghőmérsékletek közti kapcsolatokat vizsgálta.

Válaszfüggvény elemzést alkalmazott Pärn (2003) azt vizsgálva, hogy van-e különbség a szennyezett és nem szennyezett területek között a talált kapcsolatok vonatkozásában és azt tapasztalta, hogy a légszennyezéssel sújtott helyen a csapadéknak egyre erősebb hatása van a növekedésre. Lindholm és mtsai (2000), Dittmar és mtsai (2003) és Bijak (2010) RF-analízist alkalmaztak az évgyűri adatok által definiált növekedés és a havi csapadékösszegek és átlaghőmérsékletek kapcsolatának vizsgálatában. Más cikkek általában az RF-elemzés kissé módosított, bővített verzióit használták, hogy javítsák a módszer és az együtthatók megbízhatóságát.

Az utóbbi évek szakirodalmából arra lehet következtetni, hogy az összefüggések stabilitásának javítására szolgáló, korábban bemutatott eljárások közül az erdészeti jellegű vizsgálatokban a leggyakrabban alkalmazott módszer a bootstrap eljárás.

Az ilyen kutatásokban használt korrelációs és regressziós vizsgálatokra alkalmazott bootstrap eljárás alapja, hogy párhuzamosan teszteli a korrelációs, regressziós együtthatókat, illetve az RF stabilitását. Az ilyen módon javított korreláció- és RF-elemzés rendkívül széles körben használt eljárás a növekedés-klíma kapcsolatrendszer vizsgálatában. Több kutató is felhasználta a bootstrap eljárást a korreláció- és válaszfüggvény-elemzésükben (Piovesan és mtsai, 2005; Martín-Benito és mtsai, 2008; Bogino és mtsai, 2009; Michelot és mtsai 2012), míg Carrer és Urbinati (2001) a két módszer hatékonyságát a neurális hálókkal hasonlította össze. Van der Werf és mtsai (2006), valamint Čufar és mtsai (2008), Kern és Popa (2007) csak a korrelációelemzésükhöz, míg néhányan (Lebourgeois és mtsai, 2005; Pichler és Oberhuber, 2007; Oberhuber és mtsai, 2008) csak a válaszfüggvény-elemzésükben használták a módszert. Tuovinen (2005) a regressziós elemzésének javításához használta a bootstrap-et.

A módszer alkalmas annak pontosabb meghatározására, hogy a különböző típusú klímáparaméterek közül melyiknek van erősebb hatása a növekedésre és kijelöli a növekedés szempontjából releváns időintervallumokat is.

A mozgó intervallumok és az evolúciós technika módszerét alkalmazva kiszélesíthető az elemzések köre, hiszen vizsgálhatóvá válik az összefüggések időbelisége, az állandóság, illetve az esetleges változások. Büntgen és mtsai (2006), Carrer és Urbinati (2006), Andreau és mtsai (2007), Oberhuber és mtsai (2008), Rybníček és mtsai (2010), Zhang és mtsai (2008, 2011) vizsgálták a növekedés és a klíma összefüggéseit a korábbi időszakokhoz képest. Jellemzően egy legalább 30 évnyi időintervallumot mozgattak a vizsgált időszakban, az alkalmazott elemzési módszer pedig vagy korreláció-, vagy válaszfüggvény-analízis volt. Macias és mtsai (2006) mozgó intervallumokat és evolúciós technikát is alkalmaztak válaszfüggvény-analízishez. Biondi (1997) mozgó intervallumos és evolúciós technikán alapuló válaszfüggvény-elemzés alkalmazási lehetőségeit vizsgálta évgyűri adatokon. Wilczyński és Podlaski (2007) vadgesztenyefák növekedési adataira korreláció- és RF-elemzést végeztek. A rendelkezésre álló adatsorok hossza (1932-2003) lehetővé tette mozgó intervallumok alkalmazását is, valamint a 70 éves intervallum három, egymást fedő részintervallumra való felosztásával vizsgálták, hogy a fák egyes életkoraiban mennyire különbözőek a növekedési és klímaadat kapcsolatok.

Az összefüggések időbeli stabilitása, illetve esetleges változása különösen fontos kérdés például a klímakutatásban, hiszen általában elfogadott tény, hogy a klímaváltozás folyamata az utóbbi évtizedekben gyorsult fel igazán. Így egyáltalán nem biztos, hogy a 20. század elejére jellemző kapcsolatok változatlan formában kimutathatóak a 20. század végének, vagy a 21. század elejének adataira is.

Az elemzéseket az erdészeti szakirodalomban alapvetően a lineáris megközelítés jellemzi, holott a tényleges kapcsolatokat valószínűsíthetően nem a lineáris összefüggések határozzák meg a legoptimálisabban. Előbbi okairól már szoltunk röviden a nemlineáris regresszió kapcsán. A nemlineáris összefüggések definiálásának összetettségére láthatunk példát a 4. fejezetben, a növekedési görbék illesztése kapcsán.

A szakirodalmi áttekintés alapján körülhatároltuk azon elemzési módszerek körét, melyeket a különböző kutatásokban a fák növekedése és a klimatikus komponensek közti kapcsolatok vizsgálatában jellemzően alkalmaznak. Megjegyzendő, hogy ugyanezen módszerek más, hasonló jellegű paraméterek (akár függő, akár független változó vonatkozásában) közti kapcsolatokat vizsgálatában is használhatóak, illetve használják is azokat. A vizsgálatokban

fontos szerepe van az időben eltolt és különböző időszakok hatáselemzésének. A bemutatott módszerek erre csak részben (lásd mozgó intervallumok és evolúciós módszerek), vagy bizonyos előfeltevések mellett alkalmasak. Ezért kidolgoztunk egy olyan általános módszert, mely egyrészt magában foglalja a bemutatott ablakos módszerek lehetőségeit és azokon túlmutatva, a periodikus idősorok szisztematikus kiterjesztésén alapuló bővítést hordoz magában.



## 2 CReMIT módszer

Megállapítottuk, hogy a kutatásokban sokszor nem a teljes rendelkezésre álló idősort használják fel az elemzés, az összefüggés vizsgálat egy-egy lépésében, hanem annak csak egy-egy megfelelően definiált ablaka által meghatározott szeletét, hiszen alkalmas hosszúságú adatsorok esetén így a kapcsolatok időbeli változása az előre- és a visszalépéses evolúciós technika, valamint mozgó intervallumok alkalmazásával vizsgálhatóvá válik.

Az ablakos technikák alkalmazása sok egyéb statisztikai eljárásban is felmerül; gondoljunk csak például az egyszerű mozgóátlagolásra, vagy az események felfedésére idősorokban (Pelech-Pilichowski és Duda, 2010), vagy éppen az idősorok bányászatában is alkalmazott szegmentálási eljárásokra (Fu, 2011; Keogh és mtsai., 1993). A szimpla mozgóátlag esetében előre definiált szélességű ablakot mozgatunk végig az adatsoron, és az ablakban levő elemek számtani közepét képezzük. Az idősorokban fellelhető események szegmentálása során az aktuálisan vizsgált ablak kezdőpontja rögzített, és a szélességét változtatjuk egy bizonyos megfigyelt indikátor változásáig. Ezen eljárások közös jellemzője, hogy vagy csak a vizsgált ablak szélessége, vagy csak az ablak kezdőpontja változik az egyes iterációs lépésekben.

Az ablakos technikák nemcsak a teljes vizsgálandó adatsor intervallumokra bontása során alkalmazhatóak, hanem periodikus idősorok esetében ennél speciálisabb értelemben is használhatóak. A természetes környezetünk megfigyelésén alapuló vizsgálatokban felhasznált idősorokhoz számos esetben rendelhető periódus. Gondoljunk a hőmérséklet, csapadék, légnyomás, fanövekedés, szaporulat, stb. adatokra, melyekhez például éves ciklusokat tudunk definiálni. Az ezekhez köthető vizsgálatok sok esetben igénylik, hogy időben korábbi, akár több ciklusra visszamenő, illetve egy-egy perióduson belül különböző szélességű, aggregált adatokat tartalmazó ablakok is képezhetőek legyenek, ezek összefüggéseit is vizsgáljuk. Az erdészeti kutatásoknak is sokszor fontos eleme, hogy korábbi periódusok (évek) hatásait is vizsgáljuk egy-egy paraméter, például a növekedés vonatkozásában.

Kutatásaink során alapvetően erdészeti jellegű idősorokhoz köthető problémákat vizsgáltunk, azon belül is elsősorban a klímát jellemző paraméterek erdészeti változókra (növekedés, egészségi állapot, fogási adatok, stb.) gyakorolt hatásait. Ablakos technikák ezekben a kutatásokban is nagy számmal találhatóak. A fák növekedése és a környezeti paraméterek kapcsolatainak vizsgálatában több esetben alkalmaznak mozgó ablakos (Büntgen és mtsai, 2006; Carrer és Urbinati, 2006; Andreau és mtsai, 2007; Oberhuber és mtsai, 2008; Rybníček és mtsai, 2010; Zhang és mtsai, 2008; 2011) és evolúciós technikát (Biondi, 1997; Macias és mtsai, 2006) is a klimatikus változók hatásainak kimutatására, illetve használnak speciálisan

képzett időszakai adatokat is. Carrer és Urbinatti (2001) felhasználták a február-március-áprilisi és május-június-júliusi időszakok átlaghőmérsékletét és csapadékösszegét is. Briffa és mtsai (2002) az előző év júniusától adott év szeptemberéig felhasznált havi adatokat és előző év októberétől adott év szeptemberéig, illetve márciusáig, adott év április-szeptember, május-augusztus és június-augusztus időszakokat vizsgálták. Szabados (2002) adott év tavaszának és nyarának 2-3 hónapos időszakai adatait képezve vizsgálta a csapadék és növekedés kapcsolatát. Az osztrák Alpokban végzett elemzéseik során Pichler és Oberhuber (2007) háromhavi időszakai változókat alkalmaztak a havi átlaghőmérsékletek átlagolásával és a havi csapadékmennyiségek összegzésével. Friedrichs és mtsai (2009) a havi meteorológiai adatok alapján képezték még a március-május, április-szeptember és június-augusztusi átlaghőmérséklet és csapadékösszeg adatokat. Novak és mtsai (2010) a havi adatok vizsgálata mellett képezték a február-júniusi, november-júniusi és április-júniusi csapadékösszegeket, valamint az április-augusztus, április-szeptember és április-júniusi átlaghőmérsékleteket. Majd ebben a sorrendben a megfelelő csapadékösszegek és átlaghőmérséklet párok hányadosait is elemezték. Rybníček és mtsai (2009) pedig előző és adott év július-szeptember, április-szeptember időszakainak csapadékösszegeit vonták be még a vizsgálatba a havi adatok mellett. Schwarzwelber és mtsai (2011) a szimpla havi adatok mellett (előző év májusától adott év októberéig) az alábbi speciális időszakokat képezték mind a csapadék, mind a hőmérséklet vonatkozásában: március-május, június-augusztus, április-szeptember és az éves adatok. A felsorolt esetekben a kutatók előre definiáltak olyan az alapadatokból képzett időszakai adatokat, melyek hatását vizsgálják. Fontos leszögeznünk, hogy ezen ezen időszakok kiemelése vízforgalmi és szervesanyag-képzési ismereteken, olykor hipotéziseken alapul, és a szerzők nem törekedtek az összes lehetőség vizsgálatára.

Az ismertetett módszereket alapul véve, azok előnyös tulajdonságait felhasználva egy szisztematikus ablakos megközelítést hoztunk létre (Edelényi és mtsai, 2011a; 2011c, Pödör és mtsai, 2014b). A megoldás ötvözi a mozgó intervallumok és az evolúciós technika lényegét, az ablakok szisztematikus mozgatását és a minden egyes lépéshez tartozó, különböző szélességű ablakok szisztematikus képzését. Így a korábbiaknál egy sokkal általánosabb módon definiált ablak alapú technikát alkalmazhatunk a periodikus idősorok vizsgálatára. Az eljárás egy közös folyamatban biztosítja az ablakok eltolási és szélességi értékeinek kombinálását, jelentősen kiszélesítve ezzel az elemzési lehetőségek körét. A módszert alapvetően, de nem kizárólag – a fent már említett – periodikus idősorokat használó erdészeti jellegű problémák megoldásához fejlesztettük ki.

Az eljárás önmagában is felhasználható már létező rendszerekbe implementálva, beépítve de felhasználását most egy olyan elemzési folyamatba foglalva mutatjuk be, mely tartalmazza az adatok előkészítését, az adattranzformációt és az elemzési eljárás modulokat. Ezek a lehetőségekhez mérten függetlenek és alakíthatóak, ezáltal is növelve és szélesítve az elemzések spektrumát. A transzformációs modul (CReMIT: Cyclic Reverse Moving Intervals Techniques) az alapadatok fent említett ablakos technikán alapuló, szisztematikus transzformációinak felhasználásával szélesíti ki a vizsgálatba bevonható változók körét. A módszer kidolgozását inspirálták és jelentősen segítették Manninger Miklóstól származó problémafelvetések és a Vele folytatott szakmai megbeszélések is.

## 2.1 CReMIT alpmódszer

Legyen adott egy tetszőleges idősor, melynek természetes periódusát jelölje  $P$  (például havi adatok esetében  $P = 12$ , heti adatok esetében  $P = 52$  és így tovább). Az idősor elemeit az  $x$  vektorban tároljuk, ennek első eleme legyen az időben legfrissebb elem és ennek megfelelően vegyük az adatok egy természetes sorszámozását:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

### 2.1.1 Ablakok definiálása az idősor elemei felett

Az idősor periodicitását figyelembe véve ennek a periodicitásnak megfelelően ismétlődő, adott szélességű és eltolású ablakok definiálhatóak a vektor elemei felett. Jelölje  $K$  ( $1 \leq K \leq P$ ) az aktuálisan alkalmazott vizsgálat kezdőpontját, ez a vektor  $K$ . eleme. Ehhez a kezdőindexhez képest definiáljuk az idősorra éppen alkalmazott ablak eltolásának ( $i$ ) és szélességének ( $j$ ) értékeit. Az eltolás és az ablak szélesség minimális értéke egyaránt 0 lehet. A 0 eltolás érték jelenti, hogy a  $K$  kezdőponthoz képest az ablakot nem toltuk el, azaz az első ablak első elemének indexe éppen  $K$ . Az ablak szélesség 0 értéke pedig azt takarja, hogy gyakorlatilag egy egységnyi szélességű ablakot nyitunk, azaz az első ablak végső eleme is éppen a  $K$  indexű elem lesz. Felhasználva az idősor periodicitását az így definiált ablakok periódusonként ismétlődnek és a paramétereiktől, valamint az adatsor hosszától függő maximális ciklusszámhoz ( $M CN$ ) hozhatóak létre. Ez az  $M CN$  érték határozza meg az adott paraméterezés mellett létrehozott ablakok számát és ezen keresztül a létrehozandó transzformált vektor dimenzióját.

Az aktuális  $K$ ,  $i$  és  $j$  értékek mellett létrehozott ablakok kezdő és végponti indexeinek általános alakja  $[K + i + l * P ; K + i + j + l * P]$ ,  $0 \leq l \leq M CN - 1$  alakban adható meg. Ezen

paraméterezés mellett az ablakok határait definiáló indexértékek tárolására két temporális vektort hozunk létre az eljárás minden egyes iterációs lépésében. Jelölje ezeket a vektorokat  $index_b$  és  $index_e$ :

$$index_b = \begin{pmatrix} K + i + 0 * P \\ K + i + 1 * P \\ \vdots \\ K + i + (MCN - 1) * P \end{pmatrix}$$

$$index_e = \begin{pmatrix} K + i + j + 0 * P \\ K + i + j + 1 * P \\ \vdots \\ K + i + j + (MCN - 1) * P \end{pmatrix}$$

Így az  $x$  vektor elemei felett az aktuális  $K$ ,  $i$  és  $j$  paraméterek mellett  $MCN$  darab ablak kerül definiálásra.

### 2.1.2 Transzformált idősorok létrehozása

Az ablakokban szereplő elemek felhasználásával az eredeti adatsorból származtatható, másodlagos idősorokat lehet előállítani. Ezen származtatott idősorok egyes elemei a  $K$  kezdőponthoz képest időben eltolt ( $i$ ) és adott szélességű ( $j$ ) ablakok elemeiből állíthatóak elő. Az egyes ablakok elemeire a feladattól függően előre definiált  $TR$  transzformációs függvényt alkalmazzuk, és így állítjuk elő az aktuális paraméterezéshez tartozó transzformált  $tr_{x_{K,i,j}}$  vektort.

$$tr_{x_{K,i,j}} = \begin{pmatrix} TR(index_b[1]; index_e[1]) \\ TR(index_b[2]; index_e[2]) \\ \dots \\ TR(index_b[MCN]; index_e[MCN]) \end{pmatrix}$$

A  $TR$  transzformációs függvény lehet egyszerű, elemi függvény, mint például szimpla átlag, összeg, minimum, maximum. Azonban tetszőleges, az előzőeknél jóval bonyolultabb, például nemlineáris függvények is alkalmazhatóak.

A módszer teljességét az a szisztematikus megközelítés biztosítja, melynek lényege, hogy a feladattól függően előre definiált  $I$  maximális eltolás érték ( $0 \leq i \leq I$ ),  $J$  maximális ablakszélesség ( $0 \leq j \leq J$ ) és  $K$  ( $1 \leq K \leq P$ ) kezdőpont mellett az összes lehetséges  $tr_{x_{K,i,j}}$  transzformált vektort előállítjuk az eljárás során. Formálisan ez azt jelenti, hogy a CReMIT módszer az eredeti idősorból  $I * J$  darab transzformált idősort állít elő. Ezek az új idősorok bevonhatóak az aktuális elemzési folyamatba, jelentősen kiszélesítve a vizsgálati lehetőségek körét.

Az egyes iterációs lépésekben fontos az aktuálisan létrehozható maximális ciklusszám értékének meghatározása, azaz a transzformált vektor dimenziója. Ezt az  $x$  vektorra vonatkozóan ( $MCN_x$ ) egy adott iterációban az aktuálisan vizsgált első ablak utolsó (időben legrégebbi) eleme, a  $(K + i + j)$ . elem és a  $P$  periódus értéke határozza meg az alábbi módon:

$$MCN_x = \left\lceil \frac{n-(K+i+j)}{P} \right\rceil + 1, \text{ ahol } \lceil \cdot \rceil \text{ az egészrész függvény.}$$

A CReMIT eljárás eredményeképpen az eredeti  $x$  vektorból adott  $K, I, J, P$  paraméterek és  $TR$  transzformációs függvény mellett szisztematikus módon az összes lehetséges transzformált adatsort létrehozunk. Majd ezeket a másodlagos adatsorokat az aktuális elemzésekben megfelelő módon felhasználva tesszük teljessé és az eddigieknél komplexebbé a vizsgálatot.

### **2.1.3 A CReMIT módszer mint magasabb absztrakciós szint**

A bemutatott CReMIT transzformációs eljárás a korábban ismertetett ablakos technikák (evolúciós és mozgó intervallumos módszerek, 1.2.5. fejezet) általánosításaként, kiterjesztéseként fogható fel. Módszertanilag egyesíti azok előnyös jellemzőit, ezáltal gyakorlatilag felülről majorálja azokat. Képes egyrészt az evolúciós technikát jellemző szisztematikus ablakszélesség, másrészt a mozgóintervallumokat meghatározó szisztematikus kezdőpont változtatásra. Előnye ezekkel a módszerekkel szemben éppen abban rejlik, hogy a két elvet egyesítve, egyszerre tudja teljes körű módon kezelni a kombinált kezdőpontú és szélességű ablakokat.

A CReMIT a megfelelő paraméterek kiválasztása után speciálisan éppen ezeket az alapttechnikákat hajta végre. Amennyiben az eltolás mértékét,  $i$ -t tekintjük rögzítettnek és az ablakszélességet változtatjuk ( $0 \leq j \leq J$ ), akkor az evolúciós technika elvét alkalmazzuk: adott kezdőpont mellett minden iterációs lépésben az ablak szélességét eggyel növeljük. Ha az ablakszélességet tekintjük rögzítettnek egy adott  $j$  értékkel és az eltolás mértékét változtatjuk ( $0 \leq i \leq I$ ), akkor a mozgóintervallumok módszerét valósítjuk meg: az ablakszélesség rögzített és az ablakot minden egyes iterációban egy-egy lépéssel előbbre mozgatjuk. A paraméterek  $i = 0$  és  $j = 0$  választása mellett speciálisan a két idősor időben aktuális,  $K$ . elemének kapcsolatát vizsgálhatjuk, míg amennyiben  $P = 1$ , akkor adott eltolás és ablakszélesség értékek mellett éppen az eredeti evolúciós és mozgóintervallumos technika kerül megvalósításra nem periódusos idősoron. Az összes többi paraméterezés mellett olyan esetek kerülnek megvalósításra, melyek előállítására a fenti módszerek önmagukban nem alkalmasak. Így módszertanilag a CReMIT egy magasabb absztrakciós szintre emeli az

evolúciós és mozgó intervallumos technikák nyújtotta lehetőségeket, megőrizve és egyúttal tovább bővítve azok előnyös jellemzőit.

A CReMIT transzformációs eljárást periodikus idősorok szisztematikus kiterjesztésére hoztuk létre az idősorok közti kapcsolatok teljesebb, mélyebb vizsgálatára, például a független változóra történő alkalmazással. Az eljárást, annak elvi lényegét és matematikai háttérét egy vektoron mutattuk be, melynek elemei időbélyeggel ellátottak. A módszer így megfelelő implementáció mellett alkalmazható, beépíthető bármely már létező rendszerbe annak egy moduljaként.

## 2.2 CReMIT kiterjesztés több változóra

A CReMIT alaplódszer egyszerűen kiterjeszthető akár a függő, akár több független változóra is. Bemutatjuk a kiterjesztés elméleti alapjait, a kapott paraméterekre alkalmazott többváltozós lineáris regresszió megvalósítását, ennek felhasználási és alkalmazási lehetőségeit.

### 2.2.1 CReMIT kiterjesztése a függő paraméterre

A módszert függő változóra történő kiterjesztésével képesek lehetünk a független paraméter(ek) által a függő paraméter különböző időszakaira gyakorolt hatásainak elemzésére. Gondoljunk például a havi (vagy akár még sűrűbb) gyakoriságú növekedési adatokból

képezhető különböző időszaki adatok vizsgálatára. Jelölje  $y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$  a független változót,

például a havi, heti bontású növekedési idősort. Ez ugyanúgy időbélyeggel ellátott adatsort

jelöl, mint a módszer definiálása során használt  $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$  vektor mint független paraméter. A

két adatsor időben nem biztos, hogy teljes fedésben van egymással, ezért a folyamat adatelőkészítő fázisának egyik feladata, hogy az alkalmazott elemzési feladatban definiált eltolás és ablakszélesség értékektől függően fedésbe hozza egymással azokat. Figyelemmel kell lennünk arra is, hogy a két adatsor periodicitása sem feltétlenül azonos; gondoljunk arra, hogy például a növekedési adatok általában heti, havi, vagy éves bontásúak, ugyanakkor a meteorológiai adatok felbontása még ennél finomabb, akár napi szintű, vagy hőmérséklet adatok esetében akár órás szintű is lehet. Mivel a módszer célja, hogy a kidolgozott eljárás minél általánosabb és egyéb területeken is felhasználható legyen, így törekedtünk a minél univerzálisabb megközelítésre az elméleti háttér kidolgozása során.

A két változóra történő kiterjesztés kapcsán még viszonylag könnyen kezelhető az a helyzet, hogy a két adatsorra eltérő kezdőpont, eltolás, ablakszélesség paramétereit és transzformációs függvényt alkalmazzunk. Értelemszerűen ez az áttekintendő és feldolgozandó kimenet méretét is jelentősen megnöveli az alapszámhoz képest.

Jelölje a továbbiakban  $K_x, I_x, J_x, P_x$  az  $x$ -szel jelölt független, és  $K_y, I_y, J_y, P_y$  az  $y$ -nal jelölt függő változóra alkalmazott paramétereit, ahol  $0 \leq i_x \leq I_x$  és  $0 \leq i_y \leq I_y$ , valamint  $0 \leq j_x \leq J_x$  és  $0 \leq j_y \leq J_y$  az aktuális iterációs lépésekben alkalmazott eltolás és ablakszélesség értékek. Továbbá  $TR_x$  és  $TR_y$  a két változóra alkalmazott transzformációs függvényeket és  $MCN_x, MCN_y$  a két változó esetében az aktuális paraméterezés mellett definiálható maximális ciklusszámot jelöli. A két változóra – az alapszámhoz hasonlóan – definiálni kell két-két temporális vektort az ablakok kezdő és végindexeinek tárolására, legyenek ezek  $index_{x,b}, index_{x,e}$  és  $index_{y,b}, index_{y,e}$ :

$$index_{x,b} = \begin{pmatrix} K_x + i_x + 0 * P_x \\ K_x + i_x + 1 * P_x \\ \vdots \\ K_x + i_x + (MCN_x - 1) * P_x \end{pmatrix},$$

$$index_{y,b} = \begin{pmatrix} K_y + i_y + 0 * P_y \\ K_y + i_y + 1 * P_y \\ \vdots \\ K_y + i_y + (MCN_y - 1) * P_y \end{pmatrix},$$

valamint a végindexeket tároló vektorok:

$$index_{x,e} = \begin{pmatrix} K_x + i_x + j_x + 0 * P_x \\ K_x + i_x + j_x + 1 * P_x \\ \vdots \\ K_x + i_x + j_x + (MCN_x - 1) * P_x \end{pmatrix},$$

$$index_{y,e} = \begin{pmatrix} K_y + i_y + j_y + 0 * P_y \\ K_y + i_y + j_y + 1 * P_y \\ \vdots \\ K_y + i_y + j_y + (MCN_y - 1) * P_y \end{pmatrix}$$

A temporális vektorokban tárolt indexek által definiált ablakok elemeire alkalmazva a megfelelő transzformációs függvényeket állítjuk elő az aktuális paraméterekhez tartozó transzformált vektorokat:

$$tr\_x_{K_x, i_x, j_x} = \begin{pmatrix} TR(index_{x,b}[1]; index_{x,e}[1]) \\ TR(index_{x,b}[2]; index_{x,e}[2]) \\ \dots \\ TR(index_{x,b}[MCN_x]; index_{x,e}[MCN_x]) \end{pmatrix}$$

$$tr\_y_{K_y, i_y, j_y} = \begin{pmatrix} TR(index_{y,b}[1]; index_{y,e}[1]) \\ TR(index_{y,b}[2]; index_{y,e}[2]) \\ \dots \\ TR(index_{y,b}[MCN_y]; index_{y,e}[MCN_y]) \end{pmatrix}$$

Fontos ugyanakkor megjegyezni, hogy a tényleges elemzési feladatok végrehajtásához az aktuális lépésben alkalmazott maximális ciklusszámot a két vektor esetében össze kell hangolni  $MCN = \min\{MCN_x; MCN_y\}$  módon, amennyiben az alkalmazott elemzési eljárás mint például a korrelációanalízis megköveteli.

Az előállított transzformált vektorok száma így  $X_{db} = I_x * J_x$ , illetve  $Y_{db} = I_y * J_y$ . A CReMIT módszert jellemző szisztematikus megvalósítás mellett ez azt jelenti, hogy a létrehozott összes  $tr\_y$  vektort összevetjük a  $tr\_x$  vektorokkal, ami így  $X_{db} * Y_{db}$  számú eredménysort fog jelenteni. A fenti paraméterezés mellett kapott transzformált adatokra alkalmazott elemzési feladat végrehajtása után kapott kimenet egy 4 dimenziós mátrixként képzelhető el a két vektorra alkalmazott eltolás és ablakszélesség paraméterek által definiálva. A kimenet egyszerűbb áttekinthetősége érdekében célszerűbb ezt például a függő változóra alkalmazott paraméterek által definiált számú kimenet mátrix előállításával megragadni és ezeket külön-külön vizsgálni. Azaz minden  $0 \leq i_y \leq I_y$  és  $0 \leq j_y \leq J_y$  paraméterpárhoz tartozóan állítunk elő egy az alapló módszer definiálása során bemutatott (3. táblázat) táblázatot, ami így  $Y_{db}$  darab kétdimenziós kimenet mátrixot jelent.

### 2.2.2 Kiterjesztés több független paraméterre

A több független változóra történő kiterjesztés során legyenek  $x_1 = \begin{pmatrix} x_{11} \\ \vdots \\ x_{1n1} \end{pmatrix}$ ,  $x_2 = \begin{pmatrix} x_{21} \\ \vdots \\ x_{2n2} \end{pmatrix}$ , ...,

$x_w = \begin{pmatrix} x_{w1} \\ \vdots \\ x_{wnw} \end{pmatrix}$  a magyarázó változók, ahol  $w$  jelöli a vektorok számát. Az általánosság

megkötése nélkül és a gyakorlati szempontok figyelembevételével élünk avval a feltevéssel, hogy az adatelőkészítő folyamat során ezek a vektorok az alkalmazott paramétereknek és elemző eljárásnak megfelelő időszakokat fednek le. A legegyszerűbb esetben feltételezhetjük, hogy a vizsgálatba bevont független és a CReMIT módszerben



alkalmazott paraméterek periodicitása megegyező, azaz ugyanolyan eltolás és ablakszélesség értékeket alkalmazunk. Ekkor az  $MCN$  érték és az ablakok indexeit tároló két temporális vektor definiálása egyszerű, hiszen minden vektorra azonos módon definiálhatóak, hasonlóan, mint az alapmódszer esetében.

Az elemzési feladat adott iterációs lépésében  $MCN$  értékének definiálásához az összes alkalmazott vektort figyelembe kell venni amennyiben ezek periodicitása eltérő (havi, heti, napi adatok), vagy különböző  $K_t, I_t, J_t, 1 \leq t \leq w$  értékeket akarunk definiálni hozzájuk:

$$MCN = \min\{MCN_{x_1}; MCN_{x_2}; \dots; MCN_{x_w}\},$$

ahol  $w$  jelöli a vizsgálatba bevont változók számát,  $MCN_{x_m}$  ( $1 \leq m \leq w$ ) pedig az  $x_m$   $1 \leq m \leq w$  vektorhoz definiálható maximális ciklusszámot. Továbbá minden egyes változóra külön-külön kell/lehet definiálni az ablakok kezdő- és végindexét tároló temporális vektorokat ( $w$  darab vektorpár), és a  $TR$  transzformációs függvény is definiálható akár vektoronként eltérő módon.

Ugyanakkor látni kell, hogy az ilyen irányú kiterjesztés – különösen, ha vektoronként eltérő paraméterezést és transzformációs függvényt használunk – jelentős módon megnöveli a kimenetként kapott transzformált vektorok számát, és ezen keresztül az elemzési feladat végrehajtása után kapott áttekintendő eredményhalmaz méretét is. Gondoljunk arra, hogy a kimenet ebben az esetben egy  $2w$  dimenziós mátrixként képzelhető el az egyes vektorokra alkalmazott  $I_t, J_t, 1 \leq t \leq w$  paramétereknek megfelelően.

A módszer több változóra történő kiterjesztésének alapvető célja a többváltozós elemzések megfelelő támogatása, továbbá a vizsgálatba bevonható változók körének az alapmódszerre is jellemző szisztematikus megközelítésen alapuló kiterjesztése, amely így bővíti és teljessé teszi az elemzési lehetőségeket. Fontos látni, hogy a kapott változókra alkalmazott elemzési módszertan gyakorlatilag itt is független a transzformációs modultól. A vizsgálatunkban lépésenkénti többváltozós lineáris regressziót használtunk, azonban természetesen a korábban bemutatott elemzési módszerek bármelyike is alkalmazható.

### **2.2.3 A többváltozós lineáris regresszió alapjai**

A függő paraméterek jellemzően nemcsak egy, hanem általában több magyarázó változóval hozhatók összefüggésbe. Pontosabb összefüggéseket és eredményeket kaphatunk, ha a vizsgálatokba a ténylegesen hatással bíró változók mindegyike bevonható. A többváltozós regresszióanalízis éppen ezt célozza meg. Természetesen ennek is lehet lineáris és nemlineáris

esete, látni kell azonban, hogy a nemlineáris eset már egy magyarázó változó esetében sem triviális probléma. Ekkor ugyanis el kell tudnunk dönteni, hogy milyen jellegű (nemlineáris) kapcsolat áll fent a függő és az egyetlen független változó között. Ez gyakran csak szemléletesen, az adathalmaz szemrevételezése alapján tehető meg. A megfelelő függvény kiválasztása után pedig vagy ezzel a függvénnyel, vagy pedig ennek linearizált változatával tudunk dolgozni. Több magyarázó változó esetében a probléma még nehezebbé válik, hiszen három magyarázó változó mellett már szemléletesen elképzelni sem könnyű az illesztendő függvényt. Ez az oka annak, hogy a többváltozós elemzések jellemzően lineárisak.

A regressziószámítás lehetővé teszi, hogy lineáris kapcsolatot állítsunk fel egy függő és több független változó között, felépítve rájuk egy lineáris modellt:

$$y = b + a_1x_1 + a_2x_2 + \dots + a_nx_n,$$

ahol  $y$  a függő,  $x_1, x_2, \dots, x_n$  a független változók,  $b, a_1, a_2, \dots, a_n$  pedig a regressziós együtthatók. Hasonlóan az egyváltozós esethez a felépített lineáris modell esetében is mérhető a kapcsolat jósága a többszörös determinációs együttható segítségével. Ez a paraméter azonban csak azt mutatja meg, hogy a vizsgált változók között milyen erősségű lineáris kapcsolat áll fent de arra nem kapunk választ, hogy ebben a lineáris kapcsolatban az egyes független változók milyen súllyal vesznek részt. Ennek meghatározására szolgál a parciális korreláció, amely két változó kapcsolatát úgy vizsgálja, hogy a többi változó hatását konstansnak tekinti.

A parciális korrelációs együttható annyiban különbözik a páronkénti együtthatótól, hogy számításánál a többi változótól nem tekintünk el, de hatásukat kiküszöböljük. Az így kapott parciális korrelációs együttható az mutatja meg, hogy milyen szoros a kapcsolat valamelyik kiválasztott tényezőváltozó és a függő változó között, ha a többi tényezőváltozó hatását mind a vizsgált tényezőváltozóból, mind az eredményváltozóból kiszűrjük. A vizsgálathoz az aktuális modellben szereplő összes változóra vonatkozóan felírjuk a korrelációs mátrixot (az  $y$  függő és  $p$  darab független változóra):

$$R_m = \begin{bmatrix} 1 & r_{y1} & r_{y2} & \dots & r_{yp} \\ r_{1y} & 1 & r_{12} & \dots & r_{1p} \\ r_{2y} & r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{py} & r_{p1} & r_{p2} & \dots & 1 \end{bmatrix}$$

Az  $R_m$  korrelációs mátrix első sora és első oszlopa az eredményváltozó és az egyes tényezőváltozók közötti kapcsolat szorosságát mérő lineáris korrelációs együtthatókat tartalmazza, a mátrix többi eleme pedig a tényezőváltozók egymás közötti korrelációját méri.

Az  $Rm$  mátrix számos információt tartalmaz a kapcsolat természetére vonatkozóan. Megállapítható, hogy melyek azok a magyarázó változók, amelyek a legszorosabb összefüggésben vannak a függő változóval és mely magyarázó változók között van számottevő korrelációs összefüggés. Ez utóbbi a multikollinearitás veszélyére hívja fel a figyelmet.

A parciális korrelációs együttható pozitív korrelációnál pozitív, negatív korrelációnál negatív előjelű lesz, abszolút értéke 0 és 1 között helyezkedik el. A sokváltozós modellben általánosan a korrelációs mátrix inverze alapján határozhatjuk meg a parciális korrelációs együtthatókat. A korrelációs mátrix inverze:

$$Rm_{inv} = \begin{bmatrix} q_{yy} & q_{y1} & \dots & q_{yj} & \dots & q_{yp} \\ q_{1y} & q_{11} & \dots & q_{1j} & \dots & q_{1p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{jy} & q_{j1} & \dots & q_{jj} & \dots & q_{jp} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q_{py} & q_{p1} & \dots & q_{pj} & \dots & q_{pp} \end{bmatrix}$$

A parciális korrelációs együtthatókat az inverz mátrixból a következő összefüggés szerint számolhatjuk ki:

$$r_{yj.1,2,\dots,(j-1),(j+1),\dots,p} = \frac{-q_{yj}}{\sqrt{q_{yy}q_{jj}}}$$

A parciális korrelációs együttható az  $y$  és az  $x_j$  változók kapcsolatának szorosságát méri, miután a többi  $(p - 1)$  magyarázó változó hatását mindkét változóból kiszűrtük. A parciális korrelációs együttható négyzetét parciális determinációs együtthatónak nevezzük. A parciális determinációs együttható arra ad választ, hogy az  $x_j$  magyarázó változó mekkora hányadot képes megmagyarázni az  $y$  függő változó varianciájának azon részéből, amelyekre az  $x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_p$  változók nem adnak választ.

A modellbe bevont független változók számának növelésével a modellt jellemző determinációs együttható értéke minden egyes lépésben biztosan nem romlik, általában javul is valamennyit. Ez azt a tévképzetet keltheti bennünk, hogy a legjobb modell a legtöbb változó bevonásával érhető el. Ezzel szemben az optimális modell előállításához meg kell határoznunk azon változók minimális körét, melyek érdemi, statisztikailag is mérhető hatást fejtenek ki a függő változóra és egy-egy újabb változó bevonásával szignifikánsan javítjuk a modellt.

Egy egyszerűnek tűnő megközelítés, hogy minden potenciális független változó és az eredményváltozó között számítsuk ki a korrelációs együtthatót (ez könnyű feladat), majd ezek közül a meghatározott számú legerősebbet vonjuk be a többváltozós modellbe. Ennek előnye a

könnyű megvalósíthatóság, hátránya viszont, hogy az így kapott modell a gyakorlatban általában nem a legoptimálisabb megoldást adja, azaz más változók bevonásával hasonló jóságú modell kevesebb változóval, vagy ugyanannyi változó bevonásával jobb modell is felépíthető.

Multikollinearitás alatt a magyarázó változók között létező (lineáris) kapcsolatot értjük, ami a gyakorlatban sokszor megfigyelhető jelenség. Ez jelentősen befolyásolhatja a modellezés folyamatát. Felesleges a modellbe két olyan független változót is bevonni, melyek között a lineáris korrációs mértéke magas, hiszen ezek gyakorlatilag hasonló kapcsolatot mutatnak a függő változóval. Ugyanakkor a függő változóval való kapcsolatuk ereje – éppen a fennálló multikollinearitás miatt – nagyon hasonló, így ha egyik bevonásra kerül a kialakítandó modellbe, akkor várhatóan a másik is.

Ezért olyan megoldásra kell törekedni, mely során a modellbe csak a minimális számú, 0-tól szignifikánsan eltérő együtthatójú, egymással minél kevésbé összefüggő magyarázó változó kerüljön bevonásra, mégpedig úgy, hogy a kapott modell még megfelelő biztonsággal írja le a vizsgált folyamatot. Összegezve, a modellépítés során optimális egyensúlyra törekszünk a gazdaságosság és a jó közelítés között.

Alapvető kérdés ezért a fenti feltételeknek megfelelő paraméterek körének kijelölése. A feladat megoldására különböző technikák léteznek: az egyik ismert és elterjedt módszertant a lépésenkénti regressziós technikák jelentik. A lépésenkénti regressziós technikáknak alapvetően három típusát szokták megkülönböztetni:

- forward selection,
- backward elimination,
- stepwise regression.

A módszerek mindegyikének alapötlete, hogy egyesével vizsgáljuk a lehetséges változókat és egyenként döntjük el, hogy az adott változóra szükség van-e az épülő modellben. Annak eldöntésére, hogy egy változó beépítése a modellbe szignifikáns javulást hoz-e az eddig korábbi állapothoz képest,  $F$ -próbát használunk. Annak vizsgálatára, hogy egy beépítendő változó együtthatója a modellben szignifikánsan eltér-e 0-tól  $t$ -próbát alkalmazhatunk. Nyilvánvalóan teljesen felesleges egy olyan – egyébként javulást hozó – változót bevonni a modellbe, melynek együtthatója gyakorlatilag 0.

A három említett lépésenkénti módszerben a megközelítés irányában van különbség.

- A forward selection során egyesével vesszük a lehetséges magyarázó változókat, és döntjük el, hogy beépítésre kerüljön-e vagy sem. Így a modell kezdetben egyetlen független változót sem tartalmaz, majd minden egyes iterációban egy-egy elemmel bővíthet ez a halmaz, bevonva azt a változót, mely a legerősebb kapcsolatot mutatja a függő változóval.
- A backward selection ennek éppen az ellentettje. A kezdő lépésben minden lehetséges független változót bevonunk a modellbe, majd az egyes iterációs lépésekben egyesével hagyjuk el azokat a változókat, melyek a legkevésbé gyakorolnak hatást a függő változóra.
- Stepwise módszer pedig a fenti két eljárás ötvözete. Egy-egy iterációs lépésben bevonunk egy új változót, mely szignifikáns javulást okoz a modellben, majd vizsgáljuk, hogy a már bevont változók közül el tudunk-e hagyni úgy, hogy az ne okozzon statisztikailag mérhető romlást a modell jóságát tekintve.

A forward selection eljárás részletei, lépései (Montgomery és mtsai, 2012):

Legyenek  $x_1, x_2, \dots, x_n$  a magyarázó és  $y$  a függő változó.

1. A szimpla determinációs együttható alapján meghatározzuk azt a független paramétert, mely a legerősebb kapcsolatot mutatja a függő változóval, legyen ez  $x_i$ . Eldöntjük, hogy ez a változó szignifikáns kapcsolatban van-e a vizsgált függő változóval, azaz érdemes-e bevonni a modellbe. Erre az alábbi elméleti  $F$ -próba lehet alkalmas:

$$F = \frac{(r_{y.1,2,\dots,p}^2 - r_{y.1,2,\dots,p-1}^2)}{\frac{1 - r_{y.1,2,\dots,p}^2}{n - p - 1}}$$

A gyakorlatban azonban ezen próba nehézkes számíthatósága miatt az alábbi  $F$ -próbát szokás alkalmazni:

$$F = \frac{\widehat{\beta}_i^2}{Var(\beta_i)}$$

Ekkor a próbafüggvény a vizsgált változó négyzete és varianciájának hányadosaként áll elő. Ha már az első, legerősebb változó sem mutat statisztikailag értékelhető kapcsolatot  $y$ -nal, akkor vége az eljárásnak, egyébként  $x_i$ -t bevonjuk a modellbe, így adódik egy kezdeti egyváltozós modell, jelöljük ezt *modell*-gyel. Az első változó bevonását követi a tényleges iterációs művelet sor.

2. Az eddig bevonásra nem került változók,  $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$  mindegyikére számítjuk a parciális korrelációs együtthatókat, és azt az elemet tekintjük jelöltnek, mely esetében a parciális korrelációs együttható négyzete maximális lesz, legyen ez  $x_j$ , a hozzá tartozó modell pedig *model2*. A jelölt kapcsán először megvizsgáljuk *F*-próbával, hogy az annak bevonásával kapott újabb modell, a *model2* statisztikai értelemben javult-e *modell1*-hez képest. Amennyiben nem, vége az eljárásnak. Ha igen akkor a jelölt változót tovább vizsgáljuk.

3. A jelölt bevonásával kapott *model2* paraméterei együtthatóinak 0-tól való eltérését *t*-próbával vizsgáljuk. Ha a regressziós együttható értéke nem különbözik nullától, akkor az új (*j*-edik) változót elhagyjuk a modellből, s visszatérünk a 2. lépésnél meghatározott feladatokra a fennmaradó változókra vonatkozóan. Ha a parciális regressziós együtthatók értéke szignifikánsan különbözik nullától, akkor a változót bevonjuk a modellbe, *model2* elkészült és továbblépünk.

4. Ezek után gyakorlatilag a 2. és 3. lépéseket ismételjük amíg lehetséges, azaz a még nem vizsgált magyarázó változók közül kiválasztjuk a legnagyobb parciális korrelációs együttható négyzettel bírót, ez lesz a jelölt. Megvizsgáljuk *F*-próbával, hogy ennek bevonása mérhető javulást okoz-e a modellben, ha nem, akkor leállunk, ha igen, akkor *t*-próbával ellenőrizzük az együtthatókat a 3. pontnak megfelelően. Így a folyamat végén az adott értelemben optimális regressziós függvény áll elő a megfelelő paraméterekkel és együtthatókkal.

Az illesztés, modellezés során az illeszkedés jóságának mérésére leggyakrabban használt mutató az  $R^2$  determinációs együttható. Az együtthatónak azonban van egy komoly hibája, amire a korábbiakban már utaltunk is: ha a már meglévő változók mellé egy újabb változót építünk be a modellbe, akkor  $R^2$  értéke mindig legalább annyi marad, mint volt, de a gyakorlatban általában növekszik is. Így, ha csak az  $R^2$  kritériumot alkalmazzuk a modellek illeszkedési pontosságának mérésére, akkor az lenne a legjobb modell, amibe minden lehetséges magyarázó változót beépítünk. Ugyanakkor a változók számának növelésével megnő a multikollinearitás veszélye, aminek következtében jellemző módon nőnek a paraméterbecslések hibái, valamint a regresszió értéktelenné válhat. Másrészt a túl sok magyarázó változó csökkenti a szabadságfokot (a megfigyelések és a becsülni kívánt paraméterek számának különbségét), s ezáltal nem engedi meg a becslés statisztikai tulajdonságainak érvényesülését. Ezért az  $R^2$  mellett olyan mutatók alkalmazása is célszerű, amelyek figyelembe veszik a becslés során a bevont változók számát is, és ezáltal a kevés számú paramétert tartalmazó modelleket – még ha kevésbé tűnnek is pontosnak, mint a több paramétert tartalmazó társaik – versenyképesé teszik a több változót, illetve paramétert

tartalmazó modellekkel. A legegyszerűbb ilyen mutató a Theil-féle, szabadságfokkal korrigált determinációs együttható, amelynek alakja a következő:

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2)$$

A mutató értéke a paraméterek számának növekedésével csökkenhet, és csökken is akkor, ha a rendszerbe utolsónak bevont változó csak kis befolyással bír az eredményváltozóra. Negatívumként meg kell említeni, hogy nagyon speciális esetekben negatív értéket is felvehet. A mutató a különböző modellek összehasonlításában kiemelkedő fontosságú, így nem véletlen, hogy a különböző statisztikai szoftverek (R, Statistica, SPSS) kiterjedten alkalmazzák. Az  $\hat{R}^2$  mutató mellett számos más, hasonló célt szolgáló, eltérő elméleti alapokon nyugvó mutató létezik, melyek közül az Akaike információs kritériumon alapulóról ejtünk még néhány szót.

Akaike a mintában meglévő információ felhasználásának maximalizálását tűzte ki célul, és mutatója – amely az AIC rövidítéssel vált ismertté – olyan konstrukciójú, hogy ennek minimalizálása a maximális információ-felhasználású modellhez vezet. A regressziószámításban alkalmazott leggyakoribb formája:

$$AIC = \frac{SSE}{n} e^{\frac{2k}{n}} \rightarrow MIN$$

alakú. Ez a mutató előnyben részesíti a jó illeszkedésű, ugyanakkor bünteti a nagyszámú változót tartalmazó modelleket. Mivel a mutató az SSE-re épít, kis értékei jelzik a jó modellt (Burnham és Anderson, 2002).

A CReMIT módszer eredményeit felhasználó forward selection alapú többváltozós regressziós technika alkalmazására két olyan példát is bemutatunk (2.3.2. fejezet), melyek mindegyike az alap CReMIT módszer kapcsán bemutatásra kerülő elemzési feladatok (2.3.1. fejezet) többváltozós kiterjesztését jelenti.

## **2.3 A CReMIT módszer alkalmazása**

A továbbiakban bemutatjuk, hogy a kidolgozott CReMIT alapmódszer hogyan ültethető át a gyakorlatba annak akár egy, akár több változóra történő alkalmazásával. A 2.3. fejezetben bemutatott példák a CReMIT módszer működését és felhasználhatóságát hivatottak bemutatni.

### **2.3.1 Egyváltozós alkalmazás – példák**

A transzformációs eljárás működését teljesen általánosan egy  $x$  vektor kapcsán mutattuk be. Ebből fakadóan egy természetes alkalmazási lehetősége, ha egyetlen idősort veszünk, mint

független változót és erre alkalmazzuk a CReMIT eljárást, majd a statikus függő változót mellé téve végzünk összefüggés vizsgálatokat. A megoldás lehetővé teszi, hogy a vizsgált független változó különböző hosszúságú, és időben eltoltsz szakaszainak a függő paraméterre gyakorolt hatásait vizsgálhassuk.

Az irodalmi áttekintés alapján egyértelműen látható, hogy az erdőszetben kiemelt fontosságú feladat a fák növekedése és a klimatikus paraméterek közti kapcsolatok vizsgálata. Tapasztaltuk, hogy utóbbira a havi csapadékösszeg és átlaghőmérséklet a jellemző alapadat a korábban már leírt okok miatt. Az elemzések jelentős része – ahogy láttuk – megelégszik a szimpla havi adatok alkalmazásával és csak viszonylag kevés olyan kutatást találtunk ahol ezekből képzett speciális időszakokat is alkalmaztak. Utóbbiakra is inkább a hipotetikus megközelítés volt a jellemző, azaz feltételezték, hogy bizonyos időszakoknak kiemelt szerepe lehet és csak ezeket képezték és vonták be a vizsgálatba. Ahogy az általunk végzett elemzések eredményei is mutatják módszerünk alkalmazásának egyik lényegi eleme, hogy a szakmai felhasználó által definiált határok között az összes lehetséges származtatott adatsort előállítja és bevonja a vizsgálatba.

A CReMIT módszert több kutatásban is sikeresen alkalmaztuk, többek között fák növekedése (Edelényi és mtsai, 2011a; 2011c; Führer és mtsai, 2012; Manninger és mtsai, 2011b), egészségi állapota, levélvesztés, koronaelhálás (Jereb és mtsai, 2013a), mortalitás (Jereb és mtsai, 2013b) és lepkefogási adatok (Csóka és mtsai, 2012a; 2012b; 2013; Pödör és mtsai, 2013a; 2013b), mint függő változók kapcsolatának vizsgálatában. Ezek közül kettőt mutatunk be részletesen a következőkben. Az egyik a CReMIT módszer kifejlesztését megalapozó fanövekedés-klíma, a másik pedig lepkefogási adatok és klíma kapcsolatok vizsgálatára vonatkozik. A példák azt hivatottak bemutatni, hogy a módszer független a létrehozásában meghatározó problémától, minden egyéb, hasonló jellegű feladatra, idősorra alkalmazható. Jelen dolgozatban nem célunk a kapott eredmények erdőszeti értékelése, csak a módszer alkalmazhatóságának bemutatása. A megfogalmazott eredmények az együttműködésben részt vevő erdőszeti szakemberrel együtt működve születtek, azok már vagy publikálásra kerültek, vagy jelenleg publikálás alatt állnak. A CReMIT módszer lényegi eleme az időben eltoltsz, késleltetett és időben eltérő hosszúságú időszakok vizsgálata és ezek bevonása az elemzési folyamatokba. A bemutatott példák kapcsán a továbbiakban az alábbi általános jelölések kerülnek bevezetésre:  $H$  a hőmérséklet,  $CS$  a csapadék adatok vizsgálatára utal,  $pp$  a kettővel,  $p$  az egyvel korábbi és  $a$  az aktuális év adatait mutatja. A jelölések végén szereplő számérték pedig a hónap sorszámára utal, így például  $H.pp11$  kettővel korábbi év novemberének hőmérséklet adatát jelenti, míg  $H.p10-a2$  előző év



októberétől adott év februárig tartó időszak átlaghőmérsékletét. A különböző megbízhatósági szinteken statisztikailag szignifikáns kapcsolatok jelölésére pedig a statisztikai szoftverek kapcsán (például R) megszokott jelöléseket alkalmazzuk, azaz: . -  $\alpha = 0,1$ ; \* -  $\alpha = 0,05$ ; \*\* -  $\alpha = 0,01$ , végül \*\*\* -  $\alpha = 0,001$ .

### **Fanövekedés adatok és a klíma kapcsolata – eredmények illusztrálása**

Egy mintaterület 1985-2007 közti éves terület növekedés adatait vetettük össze az OMSZ-tól származó és az adott koordinátára interpolált havi alap klimatikus paraméterekkel (csapadékösszeg és átlaghőmérséklet). Az 1998-as mérési adat technikai okok miatt hiányzik, amit az elemzések elvégzése során is figyelembe vettünk (Führer és mtsai, 2012).

Az állomány szerkezetének megfelelően kiválasztottak 9 törzset, melyek paramétereinek átlaga jellemző az egész állomány átlagára és Liming-szalagok segítségével 22 éven át (1998 kivételével 1985-től 2007-ig) hetente mérték a kiválasztott faegyedek kerületének változását. A feldolgozás során 2 törzset, amelyek alászorult helyzetben voltak, kihagytuk az értékelésből, hiszen azoknál a növekedést az időjáráson kívül a szociális helyzetük is nagymértékben befolyásolta. A hét törzs mérési adataiból számítottuk ki aztán az éven belüli növekedési időszakokra és az egész évre vonatkozó átlagos körlap-növekményeket. A CReMIT által előállított másodlagos (meteorológiai) adatok és a fák éves körlap-növekedése között lineáris korrelációelemzéssel kerestük a kapcsolatokat, melyek szignifikanciáját  $t$ -próbával ellenőriztük.

Führer Ernő szakmai iránymutatása alapján, amikor összefüggést kerestünk a fák körlap-növekedése és az időjárás között, három év hőmérséklet- és csapadékadait vontuk be a vizsgálatokba és legfeljebb 4 hónap szélességű ablakokat vizsgáltunk. Az értékelés lineáris korrelációanalízissel történt egyrészt havi, másrészt időszaki (több, szomszédos hónap csapadékösszege és átlaghőmérséklete) komponensekre (CReMIT) alapozva. Ez utóbbiak képzésénél figyelembe vettük a növekedési szakaszokat, vagyis a teljes eredményhalmazból igyekeztünk a végső eredményekben szakmai szempontból is értelmezhető, a növekedést nagy valószínűséggel szignifikáns módon befolyásoló időszakokat kiválasztani. Az elemzés első lépéseként az egyes időjárási tényezők havi és időszaki komponenseinek évenkénti körlap-növedékre gyakorolt hatását vizsgáltuk, melyek szignifikáns kapcsolatait rögzíti a 3. táblázat. A korrelációs együtthatók ( $r$ ) segítségével látható a hatások iránya és mértéke is.

3. táblázat – A növekedéssel szignifikánsan korreláló meteorológiai jellemzők

havi komponensek	H.pp10	H.p6	H.p11	H.a6	CS.pp6	Cs.pp11	CS.p5	CS.a4	CS.a6
<i>r</i>	-0,41.	-0,42.	-0,54*	-0,53*	0,5*	0,48*	0,44*	0,45*	0,45*
időszaki komponensek	H.pp10-pp11	H.p4-p6	H.p10-p11	H.a6-a7	CS.pp7-pp9	CS.pp10-pp11	CS.p5-p6	CS.a2-a4	CS.a4-a6
<i>r</i>	-0,44*	-0,45*	-0,65*	-0,43*	-0,47*	0,47*	0,44*	0,47*	0,42.

A havi hőmérséklet-komponensek esetében 90%-os szinten szignifikáns hatású a két évvel korábbi október (H.pp10) és az egy évvel korábbi június (H.p6), míg 95%-os szinten az egy évvel korábbi november (H.p11) és a tárgyévi június (H.a6). Az egyértelmű negatív előjel azt jelzi, hogy a vizsgált állomány hőmérsékleti viszonyai magasabbak, mint a faj számára optimális hőmérsékleti tartomány. A havi csapadék-komponensek közül a körlap-növedéket 95%-os szinten szignifikánsan befolyásolja a két évvel korábbi júniusi (CS.pp6) és novemberi (CS.pp11), az egy évvel korábbi májusi (CS.p5), valamint a tárgyévi áprilisi (CS.a4) és júniusi (Cs.a6) adat. A korrelációs együtthatók kivétel nélkül pozitív előjelűek, azaz a nagyobb mennyiségű csapadék kedvez a növekedésnek. A szignifikánsan ható havi komponensek olyan időszakokban fordulnak elő, amikor fontos fiziológiai folyamatok játszódnak le az év során (rügyképződés, lombhullás, szervesanyag-képződés).

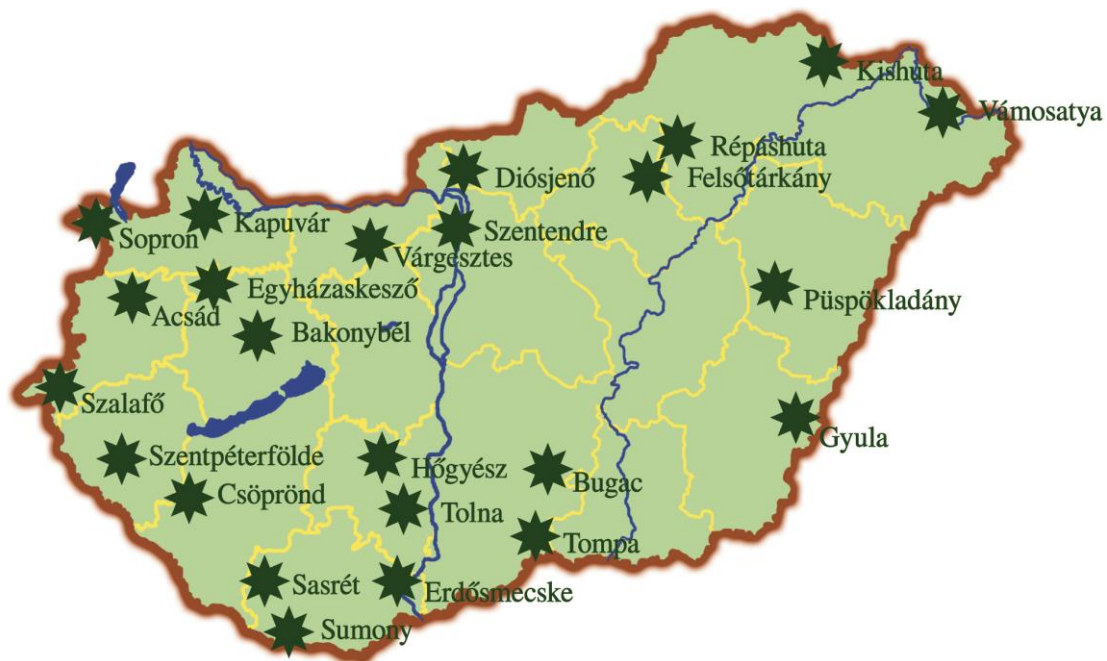
A fenti vizsgálatokat elvégeztük időszaki komponensekre is. Időszakoknak a legfeljebb négy, egymást követő hónapot tekintettük. Hőmérséklet esetén átlagot, csapadékokra összeget képeztünk. Megjegyzendő, hogy a havi komponensek esetében nagyobb szerepet játszó hónapok az időszaki komponensekben is megjelennek. Így a hőmérséklet adatok esetében a kettővel korábbi év ősze (H.pp10-pp11), az egyvel korábbi év tavasza-nyara (H.p4-p6) és ősze (H.p10-p11), valamint adott év nyara (H.a6-a7) mutat egységesen negatív előjelű kapcsolatokat. A csapadék adatok esetében jellemzően pozitív előjelű kapcsolatok adódtak, kivétel ez alól a kettővel korábbi év nyara és ősze (CS.pp7-pp9). A kapott releváns havi és időszaki adatokat felhasználtuk többváltozós modellek képzésére is, ennek módját és eredményeit a 2.3.2. fejezetben mutatjuk be részletesen.

### Nagylepkefajok populációdinamikája klímafüggése – eredmények illusztrálása

A tapasztalatok alapján a szakértők tudják, hogy sok nagylepke, rovarfaj élete, szaporodása erősen függ a környezeti hatásoktól, különösen az általunk is vizsgált két fontos paramétertől, a csapadéktól és hőmérséklettől. Ezen kapcsolatok pontos, fajspecifikus feltérképezése, a releváns időszakok meghatározása és a jövőre történő kivetítése, előrejelzések készítése

ugyanakkor még sok további feladatot jelent. Az Erdészeti Fénycsapda Hálózat (3. ábra) Magyarországon az 1960-as évek eleje óta működik, ami európai viszonylatban is kivételesnek számít. Olyan hosszúságú és minőségű adatsorokkal rendelkezünk ezért, mellyel kevés más ország dicsekedhet. 24 vizsgálati hely éves szinten összesített fogási adatait vontuk elemzés alá, összevetve azokat a havi csapadékösszeg és átlaghőmérséklet adatokkal a CReMIT módszert alkalmazva.

Fontos látni, hogy módszertanilag itt ugyanaz a feladat, mint a fák növekedése esetében volt, csak éppen a függő változó jelenleg nem az éves növekedési adatokat, hanem az éves fogási adatokat takarja. A meteorológiai adatok az OMSZ-tól származó, a fogási helyek koordinátáihoz rendelt adatokat jelentenek.



3. ábra – Erdészeti Fénycsapda Hálózat fogási helyei (Hirka és mtsai, 2011)

A vizsgálatok célja az egyes fajok éves összesített fogási adatai alapján kapott populációdinamika klímfüggésének vizsgálata. A kapott eredmények alkalmasak a fontos, fajra jellemző meteorológiai időszakok kijelölésére és ezt felhasználva akár a jövőre vonatkozó előrejelzések elkészítésére. Figyelembe véve, hogy a disszertációban nem célunk közvetlen erdészeti eredmények megfogalmazása, két kiválasztott faj kapcsán mutatjuk be az elemzés menetét és adunk példát a kapott eredményekre, ezeket azonban szakmai szempontból nem vizsgáljuk tovább.

Az egyik kiválasztott faj a gyapottok bagolylepke (*Helicoverpa armigera*), mely egy észak-afrikai, dél-európai vándorlepke faj. A kutatók a globális felmelegedés vonzataként hazánkban újonnan megjelenő rovarkártevők közé sorolják többek között a gyapottok-bagolylepkét is (Szeőke és Csóka, 2012). Az 1990-es évek előtt szinte alig fordult elő Magyarországon ez a faj (ezt a fogási adatok is alátámasztják), azonban az elmúlt 20 évben egyre nagyobb számban és területen jelent meg és kártevése is egyre erőteljesebbé vált.

A meteorológiai adatokat a CReMIT eljárás felhasználásával vetettük össze az éves fogási adatokkal lineáris korrelációanalízist alkalmazva. Szignifikancia szűréssel meghatároztuk a faj szempontjából relevánsnak tekinthető időszakokat mind a csapadék, mind a hőmérséklet adatok tekintetében. A CReMIT módszert Csóka György szakmai iránymutatásai alapján a következő paraméterezéssel alkalmaztuk:  $K = 3$ ,  $I = 21$ ,  $J = 5$ ,  $P = 12$ . Ez azt jelenti, hogy a kezdőpont október, onnan maximum előző év áprilisáig léptünk vissza és legfeljebb 6 hónap szélességű ablakokat vizsgáltunk. A könnyebb áttekinthetőség érdekében csak azokat az időszakokat jelenítettük meg (4. táblázat), melyek a vizsgált fogási helyek (12 darab) legalább felénél statisztikailag igazolható kapcsolatot mutattak (Pödör és mtsai, 2013a; 2013b).

4. táblázat – Gyapottok bagolylepke fogási adatainak klímafüggése

helyek	csapadék (alfa=0.05)								hőmérséklet (alfa=0.01)															
	CS.a2-a4	CS.a2-a6	CS.a3-a4	CS.a3-a6	CS.a3-a8	CS.a3-a9	CS.a6	CS.a6-a9	H.p2-p5	H.p2-p7	H.p3-p5	H.p3-p7	H.a3-a8	H.a4-a6	H.a4-a7	H.a4-a8	H.a5	H.a5-a6	H.a5-a7	H.a5-a8	H.a6	H.a6-a7	H.a6-a8	
FELSŐTÁRKÁNY									0.49	0.49	0.53	0.49												
TOLNA			-0.45					-0.39	0.49	0.50	0.47	0.42	0.49	0.55	0.59	0.60	0.63	0.75	0.79	0.74	0.69	0.68	0.65	
TOMPA									0.58	0.66	0.57	0.65	0.56	0.57	0.57	0.56		0.53	0.56	0.54	0.53	0.50	0.50	
VÁRGESZTES	-0.45		-0.43						0.52	0.59	0.48	0.57	0.47	0.52	0.55	0.57	0.56	0.71	0.71	0.69	0.70	0.59	0.60	
ERDŐSMECSKE	-0.45	-0.40	-0.52	-0.43	-0.40	-0.41	-0.39	-0.39					0.50	0.58	0.56	0.59	0.64	0.78	0.76	0.73	0.74	0.61	0.64	
SUMONY	-0.58	-0.47	-0.61	-0.48	-0.50	-0.46	-0.42	-0.40					0.50	0.55	0.55	0.61	0.57	0.70	0.71	0.73	0.66	0.60	0.67	
KAPUVÁR											0.59	0.48												
BUGAC									0.68	0.72	0.62	0.64							0.46					
SASRÉT	-0.47	-0.44	-0.44	-0.39	-0.41	-0.42		-0.42	0.54	0.46			0.49	0.55	0.55	0.61	0.53	0.66	0.66	0.70	0.63	0.56	0.64	
HÖGYÉSZ	-0.43	-0.42	-0.46	-0.52	-0.58	-0.66	-0.46	-0.65					0.69	0.74	0.69	0.78	0.63	0.67	0.65	0.74	0.56	0.48	0.67	
BAKONYBÉL	-0.48	-0.43	-0.52	-0.44	-0.42	-0.44	-0.42	-0.40					0.56	0.53	0.54	0.65	0.60	0.76	0.75	0.80	0.76	0.63	0.77	
PÜSPÖKLADÁNY	-0.55	-0.60	-0.68	-0.68	-0.65	-0.64	-0.60	-0.60								0.56	0.73	0.76	0.71	0.72	0.61		0.64	

A csapadék adatok esetében negatív előjelű korrelációs együtthatók adódtak, és azok is az adott év tavaszi és nyári időszakaira. Ez matematikai értelemben azt jelenti, hogy a nagyobb mennyiségű csapadék visszaveti a fajt, míg a kevesebb csapadék inkább segíti azt. A hőmérséklet adatok vonatkozásában azt látjuk, hogy az előjelek egységesen pozitívak, azaz a magasabb hőmérséklet a nagyobb fogási adatokat és ezen keresztül vélhetően a nagyobb számú megjelenést predesztinálja. Ez tapasztalható az előző és adott évi releváns időszakokra is. Különösen adott év tavaszának végére és nyár elejére adódtak erős pozitív kapcsolatok.

A kapott eredmények alapján látható, hogy a faj az enyhe telek, az aszályosodás, és a nyári melegek egyik nagy nyertesének tekinthető, mindezek a körülmények segítették az

előfordulások gyakoriságának ezt a komoly növekedését. Ez teljes mértékben összehang Szeőke és Csóka (2012) által megfogalmazottakkal.

A másik példában a tölgy búcsújáró lepke (*Thaumetopoea processionea* L.) három nyugat-magyarországi erdészeti fénycsapda (Acsád, Sopron, Szalafő) éves fogási adatait (mint függő változókat) vetettük össze lineáris korrelációanalízist alkalmazva havi csapadékösszeg és átlaghőmérséklet adatokkal. Vizsgáltuk a tárgyévi, illetve a megelőző évek időjárásának hatásait is (Csóka és mtsai, 2012a; 2012b; 2013). A vizsgálat feltételei, paraméterei egyeznek az előző példában ismertetettekkel. A vizsgálatokat az 1998-2012 közti 15 éves időszakra végeztük az egységesség kedvéért. A faj kiválasztásának oka, hogy hazánkban már régóta ismertek erdei kártételei, de erdővédelmi jelentőségénél fontosabbak potenciális humán-egészségügyi hatásai. Hernyószőrei ugyanis erősen csalánzó hatásúak, súlyos bőrgyulladást okozhatnak.

A CReMIT módszert alkalmazva meghatároztuk azokat az időszakokat, melyek a vizsgált faj és a három elemzésbe bevont fogási hely kapcsán egységesen statisztikailag szignifikáns kapcsolatokat mutatnak.

5. táblázat – Tölgy búcsújáró lepke fogási adatainak hőmérséklet függése

<b>hőmérséklet adatok</b>				
<i>mettől</i>	<i>meddig</i>	<i>Sopron</i>	<i>Szalafő</i>	<i>Acsád</i>
<i>H.a05</i>	<i>H.a07</i>	0,66*	0,64*	0,66*
<i>H.a05</i>	<i>H.a08</i>	0,59*	0,55*	0,63*
<i>H.a06</i>	<i>H.a07</i>	0,59*	0,62*	0,57*

Az 5. táblázatban szereplő eredmények alapján egyértelműen megfogalmazható, hogy adott év május-július időszakának átlaghőmérséklete pozitív hatással van a faj fogási adataira és így valószínűsíthetően a populáció méretre is, ami szakmailag is könnyen magyarázható kapcsolat.

6. táblázat – Tölgy búcsújáró lepke fogási adatainak csapadék függése

csapadék adatok				
<i>mettől</i>	<i>meddig</i>	<i>Sopron</i>	<i>Szalafő</i>	<i>Acsúd</i>
<i>CS.p10</i>	<i>CS.a01</i>	-0,39	-0,51*	-0,54*
<i>CS.p10</i>	<i>CS.a02</i>	-0,45.	-0,53*	-0,37
<i>CS.p11</i>	<i>CS.a01</i>	-0,43	-0,45.	-0,30
<i>CS.p11</i>	<i>CS.a02</i>	-0,50	-0,49.	-0,14
<i>CS.a05</i>	<i>CS.a07</i>	-0.51*	-0,42	-0,40
<i>CS.a05</i>	<i>CS.a08</i>	-0,40	-0,45.	-0,18
<i>CS.a06</i>	<i>CS.a07</i>	-0,42	-0,32	-0,45.

A csapadék adatok szinte kivétel nélkül minden esetben (az itt feltüntetetteken kívül is) negatív hatással bírnak a fogási adatokra (6. táblázat), azaz a nagyobb csapadék adatok inkább visszavetik a faj fogási adatait, míg az aszályosság inkább növeli azt.

Ez utóbbit összevetve a hőmérséklet adatok kapcsán tapasztaltakkal azt a megállapítást fogalmazhatjuk meg, hogy az aszályosság mértéke – különösen a május, június és július hónapokban – pozitív hatással van a faj népségére, illetve az előző év telének aszályossága is pozitívan hat a fajra. Tehát a növekvő gyakoriságú és mértékű aszályos időszakok bekövetkezése esetén a faj további terjeszkedésével és kárterületének további növekedésével kell számolnunk.

Mindkét bemutatott példa jól mutatja, hogy a CReMIT módszer szisztematikus megközelítését alkalmazva meg tudjuk határozni azokat az időszakokat, melyek szignifikáns kapcsolatot mutatnak a függő változóval, ami lehet akár növekmény, akár lepkefogási adat (vagy természetesen egyéb más adat is). Nincs szükség komoly előzetes feltevésekre, csak a paraméterek értékeit kell megfelelően definiálni. Az eljárás minden lehetséges időszaki adatot előállít és elvégzi az előre definiált elemzési feladatokat is. A szakértő feladata aztán, hogy ezen matematikai-statisztikai összefüggésekből kiválassza azokat, melyek szakmailag is értelmesek és megfelelően magyarázhatóak.

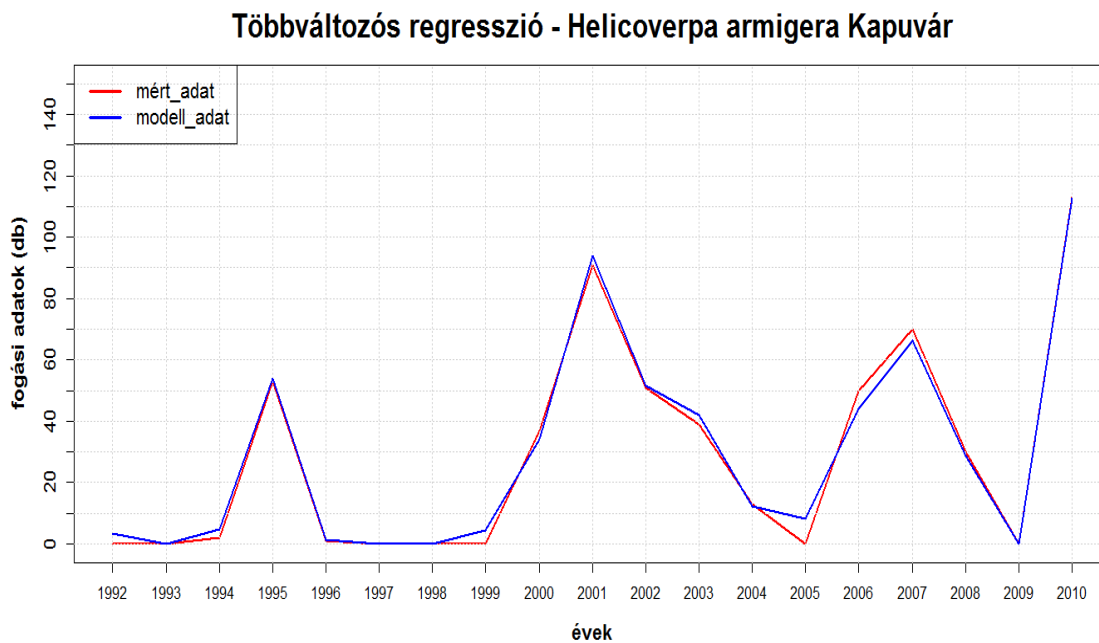
### 2.3.2 Többváltozós kiterjesztések – példák

A többváltozós elemzések egyik lehetősége az alap CReMIT módszer által a független változóra létrehozott transzformált vektorok felhasználása a modellépítés során. Ez lehetővé teszi, hogy a szisztematikusság elve alapján létrehozott összes lehetséges transzformált adat (időszaki adatok) közül azokat a legmegfelelőbbeket vonhassunk be a regressziós modellbe, melyre egyébként nem lenne lehetőségünk.

## Nagylepke fogási adatok vizsgálata

A nagylepke fogási adatokkal kapcsolatban végzett elemzések során már a CReMIT módszert felhasználva előállítottuk mind a hőmérséklet, mind a csapadék adatok esetében adott paraméterezés mellett az összes transzformált adatsort. Ezen adatokat használtuk fel a többváltozós modellek felépítése során a forward stepwise módszert alkalmazva.

Az eljárást minden korábban már említett vizsgált fajra és minden lehetséges fogási helyre lefuttattuk. A kapott eredményekre példákat itt is a gyapottok bagolylepke kapcsán mutatunk be (Pödör és mtsai., 2013b), ahol az összes független paraméterek száma 266. Két fogási helyet választottunk ki a módszer alkalmazhatóságának, az eredmények és a lehetséges felhasználások illusztrálására, Kapuvárt és Sasrétet. A modellépítésben használt  $F$ - és  $t$ -próbákat  $\alpha = 0.05$  megbízhatósági szinten alkalmaztuk.



4. ábra – Gyapottok bagolylepke kapuvári fogási adatai és a modell adatok összevetése

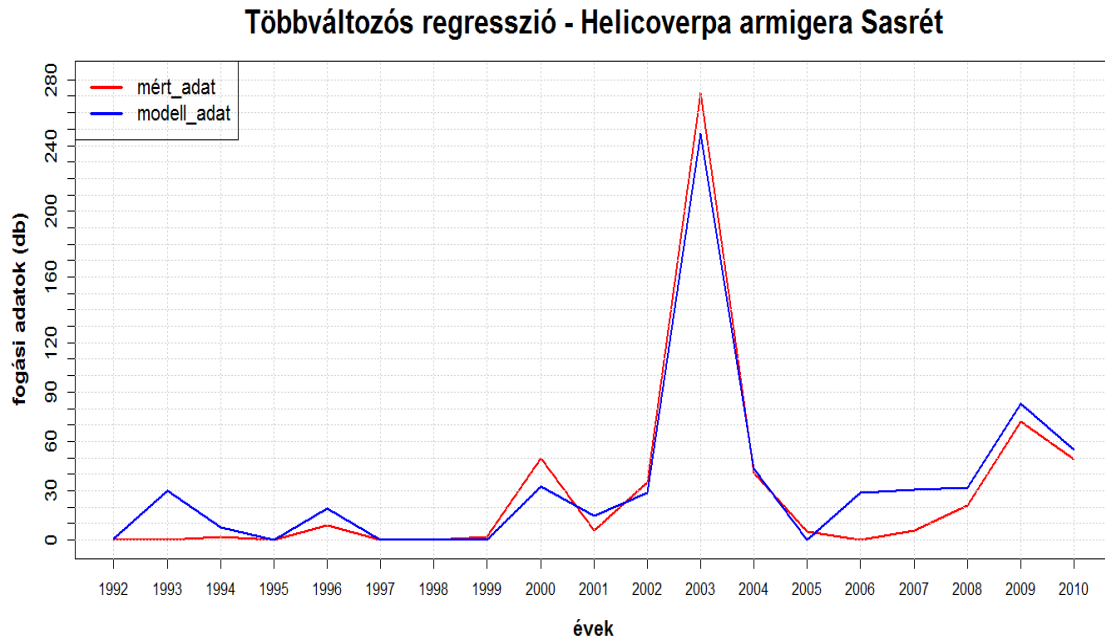
Az eljárás alapján a modellbe bevont független változók köre (a modellbe történt beépítés sorrendjében):

- előző év szeptember-novemberi átlaghőmérséklet (H.p9-p11)
- előző év április-májusi átlaghőmérséklet (H.p4-p5)
- adott év április-szeptemberi csapadékösszeg (CS.a4-a9)
- adott év júniusi csapadékösszeg (CS.a6)
- előző év október-decemberi átlaghőmérséklet (H.p10-p12)

- előző év április-augusztusi átlaghőmérséklet (H.p4-p8)

Az előállított regressziós modell pedig az alábbi:

$$y = -505,64 + 41,4476 * H.p9 - p11 + 25,88 * H.p4 - p5 + 0,18 * CS.a4 - a9 - 0,29 * CS.a6 - 16,76 * H.p10 - p12 - 12,70 * H.p4 - p8, \text{ és } \hat{R}^2 = 0,92 \text{ és } R^2 = 0,95.$$



5. ábra – Gyapottok bagolylepke sasréti fogási adatai és a modell adatok összevetése

Az eljárás alapján a modellbe bevont független változók köre (a modellbe történt beépítés sorrendjében):

- adott év május-augusztusi átlaghőmérséklet (H.a5-a8)
- adott év január-februári átlaghőmérséklet (H.a1-a2)
- előző év május-júniusi átlaghőmérséklet (H.p5-p6)
- előző év szeptember-decemberi csapadékösszeg (CS.p9-p12)
- Az előállított regressziós modell pedig az alábbi:

$$y = -1156,71 + 50,11 * H.a5 - a8 - 15,81 * H.a1 - a2 + 15,85 * H.p5 - p6 - 0,17 * CS.p9 - p12, \text{ és } \hat{R}^2 = 0,89 \text{ és } R^2 = 0,91.$$

### Fanövekedés adatok vizsgálata

Az alap CReMIT módszer alkalmazása során már bemutatott mintaterület növekedés adatok havi szinten is rendelkezésre állnak április-október vonatkozásában 1985-2007 között (1998



kivételével). Ez lehetőséget ad arra, hogy ezen növekedés adatokra, mint függő változóra is kiterjesszük a CReMIT módszert speciálisan a havi adatok éven belüli összegzéseivel. Így a korábban már vizsgált éves növekedés (április-októberi növekedésösszeg) mellett vizsgálni tudjuk az egyes hónapok és a különböző havi növekedésösszegek klimatikus komponensekkel való összefüggéseit. Az elvégzett vizsgálatokban a növekedés adatokat csak adott éven belül összegeztük, ez 28 különböző időszakot jelent beleértve a szimpla havi adatokat is. Végül csak a szakmailag is könnyebben értelmezhető időszakokat vizsgáltuk meg részletesen: áprilisi (a kezdeti növekedés), május-augusztusi (főnövekedési időszak), szeptember-október (a befejező növekedés) és természetesen az április-októberi (teljes növekedési) időszak. Utóbbival kapcsolatos egyváltozós eredményeket már bemutattuk, most ezt terjesztjük ki többváltozós elemzés alkalmazásával.

Egyrészt szimpla egyváltozós korrelációanalízissel összevetettük a különböző növekedési adatokat és a CReMIT módszerrel képezett csapadék és hőmérséklet adatsorokat. Ezzel gyakorlatilag  $28 \times 2$  kétdimenziós kimenet táblázatot kaptunk a 28 vizsgált növekedési időszakra és a két környezeti paraméterre vonatkozóan. Az összefüggés elemzéseket itt is ugyanolyan paraméterezéssel futtattuk le, mint korábban, azaz kettővel korábbi év áprilisától adott év októberéig, legfeljebb 4 hónap széles ablakkal.

Másrészt a lepkefogási adatok kapcsán elvégzett többváltozós elemzésekhez hasonló elemzéseket végeztünk itt is a növekedési adatok vonatkozásában külön csak a csapadék, csak a hőmérséklet és a kettő paraméter együttes felhasználásával is. Az eredmények közül az április-októberi, teljes növekedési időszak vonatkozásában mutatjuk be a módszert. A korábbi többváltozós elemzésekhez képest az volt a különbség, hogy itt nem pusztán a stepwise módszert alkalmaztuk a modell előállításánál, mert szakmai szempontokat is figyelembe kívántunk venni a modellbe bevonandó környezeti paraméterek kapcsán. Ezért a CReMIT módszerrel előállított szignifikáns meteorológiai komponensek felhasználásával előállítottuk az összes, matematikailag lehetséges többváltozós klímaindex-modellt (továbbiakban *CI*: Climate Index) mind a hőmérséklet, mind a csapadék és a kettő additív vonatkozásában havi, időszakos és együttes szinten is (ügyelve a multikollinearitás problémájára). A kapott regressziós egyenletekből a statisztikailag szignifikánsak  $\alpha = 0,05$ ) közül kiválogattuk a szakmailag is megfelelőeket. A továbbiakban az alábbi rövidítéseket alkalmazzuk *CI* (*Climate Index*) vonatkozásában: *T* a hőmérséklet, *P* a csapadék, *m* (monthly) a havi, *s* (seasonal) az időszakos komponensekre utal.

A releváns szignifikáns paraméterek figyelembevételével előállítottuk azokat a regressziós klímaindexeket (7. táblázat), amelyek a vizsgált 22 év átlagában leginkább befolyásolják a fák éves körlap-növekedését.

7. táblázat – Az egyes szempontok szerint kiválasztott modellek

modell	$\hat{R}^2$	$R^2$
$CI_{Tm} = -1,42 * H.p11 - 1,86 * H.a6 + 51,38$	0,43	0,49
$CI_{Pm} = 0,11 * CS.pp11 + 0,07 * CS.p5 + 0,07 * CS.a4 + 0,03 * CS.a6 - 1,03$	0,64	0,71
$CI_{TPm} = 0,09 * CS.pp11 + 0,06 * CS.p5 - 0,76 * H.p11 + 0,05 * CS.a4 - 1 * H.a6 + 22,71$	0,71	0,78
$CI_{Ts} = -2,24 * H.pp10 - pp11 - 1,26 * H.p4 - p6 - 2,91 * H.p10 - p11 - 0,05 * H.a6 - a7 + 63,52$	0,54	0,63
$CI_{Ps} = 0,06 * CS.pp10 - pp11 + 0,04 * Cs.p5 - p6 + 0,02 * CS.a4 - a6 - 2,49$	0,46	0,54
$CI_{TPs} = -0,03 * CS.pp7 - pp9 - 2,23 * H.pp10 - pp11 - 2,23 * H.pp10 - pp1 - 2,97 * H.p10 - p11 + 0,04 * CS.a2 - a4 - 0,01 * CS.a4 - a6 - 0,41 * H.a6 - a7 + 57,86$	0,66	0,75
$CI_{TPsm} = 1,09 * H.pp10 + 0,1 * CS.pp11 - 1,21 * H.p4 - p6 + 0,03 * CS.p5 - p6 - 1,32 * H.p10 - p11 + 0,06 * CS.a4 - 0,01 * CS.a6 - 1,1 * H.a6 + 34,86$	0,72	0,82

A táblázatban a vizsgálatok alapját képező korrigált determinációs együttható mellett a szimpla determinációs együtthatót is feltüntettük. Látható, hogy a független változók számát is figyelembe vevő korrigált determinációs együttható mindenütt alacsonyabb értékeket ad, mint a szimpla.

A havi komponensekből képzett  $CI_{Tm}$  és  $CI_{Pm}$  indexek alapján megállapítható, hogy a csapadék ( $\hat{R}^2 = 0,64$ ) egyértelműen erősebb hatással van a növedékre, mint a hőmérséklet ( $\hat{R}^2 = 0,43$ ). A kiválasztott együttes (hőmérséklet-csapadék) havi klímaindex modellre ( $CI_{TPm}$ ) kapott  $\hat{R}^2 = 0,71$  érték határozott összefüggést mutat.

Az időszaki additív hőmérsékleti klímaindexre ( $CI_{Ts}$ ) számított korrigált determinációs együttható értéke 0,54, ami egy kissé nagyobb, mint a havi komponensekre kapott érték. Az időszaki csapadék komponensekkel képzett additív indexre ( $CI_{Ps}$ ) számított korrigált determinációs együttható értéke pedig 0,46, ez viszont egy kicsivel alacsonyabb, mint a havi csapadék komponensek vonatkozásában. Vagyis az időszaki klímaindex esetében a hőmérsékletnek van erősebb hatása. Amennyiben az időszaki hőmérséklet és csapadék

paraméterek additív hatását ( $CI_{TPs}$ ) tekintjük akkor nem adódott szorosabb összefüggés ( $\hat{R}^2 = 0,66$ ), mint a havi komponensek esetében.

A havi és az időszaki komponensek összevonásával kapott modellre ( $CI_{TP}$ ) számított korrigált determinációs együttható értéke 0,72, ami határozott összefüggést mutat. A mért és a modellel számított évenkénti körlap-növedékek hasonló karakterisztikájú futásait mutatja a 6. ábra. Jól látható szemléletesen is, hogy a modell adatok jól követik az eredeti mérési adatokat mindhárom esetben.



6. ábra: A körlapnövedékek mért és a modell alapján számított értékeinek futása ( $n=22$ ,  $\alpha = 0,01$ )

A CReMIT módszert alkalmazva szisztematikus módon ki tudjuk terjeszteni a modellbe bevonható független paraméterek körét, majd ezekre az adatsorokból a többváltozós regressziós modellek segítségével egyrészt meghatározhatóak azok, az egymással nem korreláló független paraméterek (időszaki adatok), amelyek statisztikailag értékelhető hatással bírnak a kiválasztott függő változóra (például lepkefogási adatok, fanövekedés adatok). Másrészt a felállított regressziós egyenlet alkalmas lehet a függő változó jövőbeni értékeinek becslésére is a

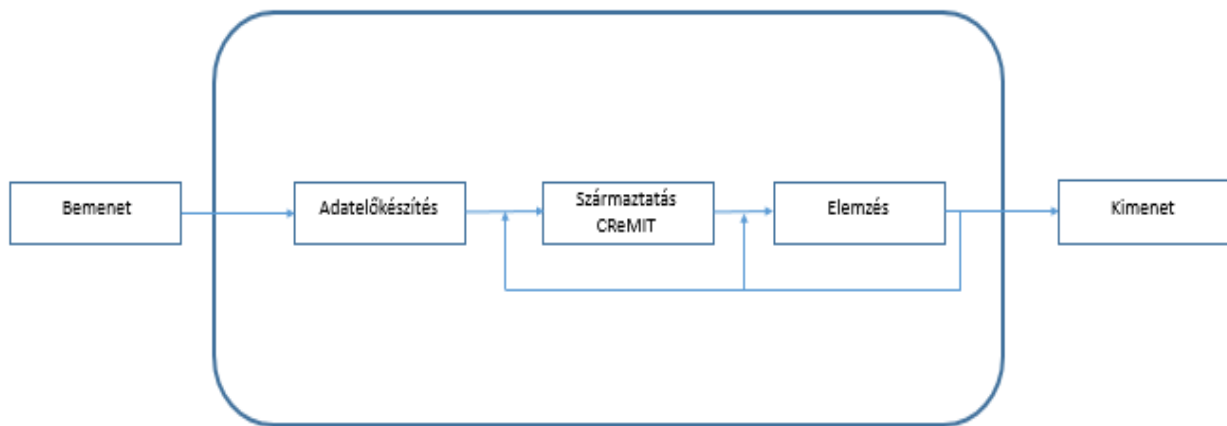
felhasznált környezeti paraméterek klímaszenáriók által jósolt értékeinek figyelembe vétele mellett.

A CReMIT módszer által generált másodlagos paraméterek, illetve a módszer több változóra történő párhuzamos alkalmazása jelentősen kiszélesíti a többváltozós modellekbe bevonható magyarázó változók körét. Így pontosabban meghatározható azon független paraméterek köre, melyek együttesen is statisztikailag igazolható hatással bírnak, jól magyarázzák a függő változót. Javul a modellek pontossága és gyakorlati alkalmazhatóságuk a kapcsolatok leírásában és a jövőbeni értékek előrejelzésében. Ez fontos feladat az erdőszetben is, hiszen segít annak a meghatározásában, hogy egy-egy terület jövőbeni klimatikus körülményeinek ismeretében milyen növekedés adatokkal számolhatunk az adott faj kapcsán. A lepkefajok alapvető klimatikus komponensekkel történő összefüggés vizsgálatai, az összefüggések feltárásai pedig segítenek abban, hogy akár a gazdasági, akár az egészségügyi kártevésekre is jobban felkészülhessünk.

### **3 A CReMIT alkalmazásának támogatása**

#### **3.1 Az elemző folyamat**

A CReMIT módszer kapcsán bemutatott példák alapján látható, hogy a különböző elemzési feladatok eltérő adatokat, feldolgozási szempontokat jelentenek. Fontos azonban, hogy az elemzési feladatok különböző lépései egy egységet alkossanak. A CReMIT módszer akár egy, akár többváltozós felhasználásának gyakorlati megvalósításához az eljárást egy szintén általunk kidolgozott elemző folyamatba illesztettük (7. ábra). A létrehozott folyamat további célja, hogy a kidolgozott transzformációs módszer más rendszerektől függetlenül is alkalmazható legyen elemzési feladatok elvégzésére. A dolgozatban már bemutatott és bemutatásra kerülő vizsgálatok, elemzések és azok eredményei ezen folyamatot felhasználva kerültek végrehajtásra, előállításra. A továbbiakban ismertetésre kerül az elemző folyamatot, azonban annak CReMIT moduljára már nem térünk ki.



7. ábra – Az elemző folyamat

A transzformációs eljárás köré épített folyamat lehetővé teszi, hogy a teljes előkészítési-transzformálási-elemzési folyamat egységként legyen kezelhető, ugyanakkor a moduláris felépítés nagyfokú rugalmasságot is biztosít a vizsgálatba bevont változók, azok transzformáltjai, valamint a különböző elemzési módszerek tekintetében. Az első fő modul az adatelőkészítés, amely a nyers adatokat alkalmassá teszi arra, hogy bekerülhessenek a transzformációs modulba.

A második modul a CReMIT transzformációs modul, amely az eljárás lényegét és újdonságát tartalmazza. Ez a blokk felelős az előzőekben bemutatott ablakos transzformáción alapuló másodlagos adatsorok szisztematikus előállításáért. A harmadik modul a transzformáció során előállított adatsorokat fogadja, és végrehajtja az előre definiált elemzési eljárást. Az egymásra épülő, ám egymástól bizonyos szintig mégis független modulrendszer jelentős rugalmasságot és bővíthetőséget biztosít mind az előkészítés, mind a transzformálás, mind az elemzési módszerek tekintetében. Az R szoftverkörnyezetben megvalósított modulok könnyen illeszthetőek egyéb rendszerekhez, mint például a RapidMiner ingyenes adatbányász szoftverhez, vagy a Statistica szoftvercsomaghoz.

### 3.1.1 Az adatelőkészítő modul

Az adatelőkészítő modul kapcsán nem csupán a hagyományos értelemben vett adatelőkészítésre kell gondolnunk (adattisztítás, adathiányok kezelése, stb.), hanem a nyers adatok idő vagy egyéb szempontok szerinti összevonására illetve felbontására – például esemény alapú észlelések azonos időegységekké konvertálására – is. Ez az adatelőkészítés teszi lehetővé, hogy az adatokat megfelelő periódusokba szervezhessük. Ez különösen a

meteorológiai adatok esetében lehet fontos feladat, hiszen például a csapadékadatok gyakran eseményhez kötöten vannak megadva, ezekből heti, havi adatokat kell előállítani a vizsgálatokhoz.

Felmerülhetnek ugyanakkor olyan problémák is, ahol a függő, illetve a független változókra alkalmazott CReMIT alapú transzformációs megközelítés nem nyújthat teljes körű kiterjeszhetőséget. Az előkészítő modul részeként ezért két olyan technika is bemutatásra és alkalmazásra kerül, melyek speciálisan a függő, illetve a független változók irányából tesznek lehetővé további kiterjesztéseket. Így bővíthető az előkészítő modulból a CReMIT modulnak átadható alapadatok, és áttételesen az elemzési lehetőségek köre is. A kutatásainkban független paraméterként felhasznált meteorológiai adatokra a töréspont keresés (3.2. fejezet), míg a függő paraméterként felhasznált fák éven belüli növekedési adataira a függvényillesztés (3.3. fejezet) eljárások kerültek be az előkészítő folyamatba.

Mindkét technika alkalmat ad a felhasznált adatok még pontosabb és részletesebb megismerésére, a bennük rejlő lehetőségek mélyebb feltárására az elemzési folyamatokban. Az egyes eljárások részletei és az eredmények a megfelelő fejezetekben kerülnek tárgyalásra.

### **3.1.2 Az elemzési modul**

A CReMIT modul által előállított transzformált vektorok az elemző modulba kerülnek. Az elemzések a felhasználó által előre definiált módon kerülnek végrehajtásra, és alapvetően az idősorok közötti kapcsolatok vizsgálatát jelentik, közöttük leggyakrabban a korreláció-, és regresszióanalízist. Fontos megjegyezni, hogy az elemző modul gyakorlatilag független a transzformációs modul működésétől. Azaz a CReMIT modul által előállított adatsorokra bármely megfelelő elemző eljárás alkalmazható az egyszerű korreláció- és regresszióanalízistől kezdve a PCA-n és egyéb többváltozós elemzéseken keresztül.

Az alap CReMIT módszer tekintetében (a vizsgált  $x$  vektorra vonatkozóan) az elemzések végrehajtásának eredményeként a folyamat az elemző modulban a vizsgált változóra alkalmazott  $I$  és  $J$  paraméterek értékétől függő 2-dimenziós mátrix-kimenetet állít elő (8. táblázat). A sorok az aktuális eltolás, az oszlopok az ablakszélességet jelölik, az egyes cellák pedig az adott  $i$  eltolás és  $j$  ablakszélesség mellett kapott elemzési eredményeket tartalmazzák. Egy-egy cella tartalma lehet egy egyszerű korrelációs együttható, vagy bonyolultabb adatstruktúra is a vizsgálat jellegétől függően.

8. táblázat – Kimenet mátrix

		<b>J</b>					
		<i>0</i>	<i>1</i>	...	<i>j</i>	...	<i>J</i>
<b>I</b>	<i>0</i>						
	<i>1</i>						
	...						
	<i>i</i>						
	...						
	<i>I</i>						

Amennyiben a transzformációs eljárást a függő, vagy több független változóra is kiterjesztjük ennek megfelelően a kimenet mátrix dimenziója is természetesen növekszik.

Az elemző folyamat és a CReMIT módszer megvalósítása a szabad hozzáférésű R szoftverkörnyezetben történt meg (R 2.15.2 verzió). A jelenleg implementált elemi *TR* függvények az átlag, összeg, minimum és maximum, az alapértelmezett elemző eljárás pedig a lineáris korrelációanalízis. A felhasznált *K, I, J* paraméterek értéke az aktuális feladattól függ.

Az alábbiakban bemutatjuk az elemző folyamat adatelőkészítő moduljában alkalmazott két speciális előkészítő technikát, azok elméleti hátterét és jelentőségét az elemzési folyamatokban.

### 3.2 Töréspont keresés idősorokban

A megfelelő hosszúságú idősorok esetében fennáll annak a lehetősége, hogy a teljes adatsor nem feltétlenül homogén, lehetnek komoly, ugrásszerű változások benne. Ezért az ilyen idősorok közötti kapcsolatok sem tekinthetők minden esetben a teljes vizsgálati időszak vonatkozásában stabilnak. Különösen igaz ez a természettel, környezetünkkel kapcsolatos vizsgálatokban felhasznált adatokra, hiszen mi magunk is tapasztalhatjuk, hogy időjárásunkban, klímánkban változások zajlanak le. A klimatikus komponensek közül leggyakrabban a hőmérsékletváltozás különböző lehetőségeit szokták vizsgálni, mert ez egy olyan elem, aminek változásait a hétköznapi emberek is közvetlen módon érzékelhetik. A Meteorológiai Világszervezet 2000. évi állásfoglalása (WMO, 2000) többek között megállapította, hogy a Föld felszíni átlaghőmérséklete mintegy 0,7°C-kal emelkedett az 1900-as évek eleje óta, valamint, hogy az 1980-2000 közti időszakban a hőmérséklet kimutathatóan gyorsabban emelkedett, mint akármely másik azt megelőző 20 éves időszakban. A legfrissebb, nyilvánosan hozzáférhető 2011-es évi állásfoglalás (WMO, 2011) szerint az abban az évben mért 14°C-os globális átlaghőmérséklet 0,3-0,5°C-kal haladja meg az 1961-1990-es

bázisidőszak átlagát, illetve a legutolsó tízéves periódus (2002-2011) átlaga közel  $0,5^{\circ}\text{C}$ -kal haladja meg a bázisidőszakét. Természetesen ezek a Föld egészére vonatkozó átlagos megállapítások, azaz az egyes régiókban ettől fel- és lefele való eltérések is tapasztalhatóak. Például az évi középhőmérsékletek akár  $5^{\circ}\text{C}$ -kal is meghaladták a normált Oroszország északi-sarkvidéki partjainál,  $3^{\circ}\text{C}$ -kal, vagy még annál is többel emelkedtek a normál fölé mind Oroszország, mind pedig Kanada északi részének jelentős területe felett.

Az idősorok közti összefüggések elemzése kapcsán így felmerül a kérdés, hogy valóban korrekt eredményt generálunk-e akkor, ha a teljes vizsgálati időszakot egy egységnek tekintjük, vagy érdemes azt részekre osztva is megvizsgálni. Ez a kérdésfelvetés a meteorológiai adatsorok vizsgálata során különösen indokolt. Az idősorok felbontása során törekednünk kell arra, hogy ez a tagolás – ellenőrzött módon ugyan –, de minél objektívebb és egységes legyen, azaz egy meghatározott, minden változóra egyformán alkalmazható elv mentén történjen.

Különösen a természeti jellemzők megfigyelése kapcsán igaz, hogy ezek a változások nem egyik napról a másikra következnek be, hanem egy hosszabb távú folyamat eredményeképpen. Egy ilyen hosszabb távú folyamat kapcsán is lehetőségünk van ugyanakkor olyan pontok kijelölésére, melyek mentén viszonylag élesen kettéválasztható az adatsorunk. Ez az eljárás a töréspont keresés és ezeket a hirtelen, ugrásszerű változásokat töréspontoknak (Molnár és Izsák, 2011) nevezzük az idősorokban. Megfelelő hosszúságú idősorok esetében tehát vizsgálható, hogy az idősor által lefedett időszakban statisztikailag kimutatható-e hirtelen, törésszerű szignifikáns változás az adatokban.

### **3.2.1 Töréspontok keresésének módszerei**

Töréspontok kimutatására többféle módszer is ismert a statisztikában (Molnár, 2003), mint például

- a szórások minimalizálásának módszere,
- az anomáliák kumulatív összegzése,
- a Pettitt-féle nemparaméteres próba,
- a jel-zaj arány vizsgálata,
- a részátlagok összevetése Student-féle  $t$ -próbával.

A lehetőségek mindegyike módszertanilag könnyen illeszthető a CReMIT módszerhez. A továbbiakban röviden bemutatjuk az egyes eljárások lényegét.



## Részátlagok összevetése Student-féle $t$ -próbával

A módszer lényege (Péczeley, 1979), hogy a rendelkezésre álló adatsort egy-egy adott év mentén kettévontjuk és  $t$ -próbával vizsgáljuk, hogy az így kapott két időszak részátlagai között kimutatható-e statisztikailag is igazolható, szignifikáns eltérés. Fontos megjegyezni, hogy e módszer alkalmazhatóságának feltétele, hogy a vizsgált adatsorok normális eloszlásúak legyenek, amit például a Kolmogorov-Szmirnov, vagy a Shapiro-Wilk teszttel vizsgálhatunk meg. Amennyiben az adatsorok nem tekinthetők normálisnak, akkor például nemparaméteres próbát (Pettitt-próba), vagy más módszert kell alkalmazni a töréspont keresésben.

Az átlagok összevetésére a próba során generált  $t$ -érték alkalmazható,  $t = \frac{\bar{x}_1 - \bar{x}_2}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ , ahol  $\bar{x}_1$  és

$\bar{x}_2$  a két minta átlagát,  $n_1$  és  $n_2$  azok elemszámát,  $S$  pedig a két minta közös szórását jelenti. Utóbbi a két részminta szórásának ( $\sigma_1$  és  $\sigma_2$ ) ismeretében az alábbi módon könnyen számítható:

$$S = \sqrt{\frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}}.$$

Az időszakok felbontása során figyelemmel kell lenni arra, hogy a kapott részidőszakok hossza egy előre definiált minimum értéket elérjen a statisztikailag korrekt összevethetőség miatt. Ezen minimumérték meghatározása részben függ a rendelkezésre álló adatsor hosszától, részben pedig statisztikai értelemben is megfelelő kell, hogy legyen. Minden egyes potenciális töréspont által meghatározott két részzakaszhoz kiszámítjuk a Student-féle  $t$ -próbához tartozó  $t$ -értéket, a választópontot pedig az első lehetséges értéktől egy-egy évvel mozgatjuk előre az utolsó lehetséges értékig. Az így kapott  $|t|$  értékek sorozatából a maximális mutatja azt a pontot, évet, amely mentén az idősort kettéválasztva, a két részsorozat átlaga között maximális az eltérés mértéke. Amennyiben ez az eltérés statisztikailag kimutathatóan szignifikáns is az adott megbízhatósági szinten, akkor mondhatjuk, hogy ez a pont egy töréspont a vizsgált adatsor tekintetében. Azt a pontot tekintjük ekkor töréspontnak (legalábbis statisztikai értelemben), amely által definiált két részidőszak átlaga között a maximális az eltérés és az statisztikailag is igazolható adott szignifikancia szint mellett.

Ahogy korábban is említettük már, az idősorban tapasztalható változások gyakran egy hosszú távú folyamat eredményei. E folyamatok kimutatásának egy lehetséges és gyakran alkalmazott módja a trendvizsgálat. Amennyiben a vizsgált idősorban statisztikailag alátámasztva trendszerű és töréspontszerű változás is kimutatható, akkor feltétlenül kérdés, hogy hogyan tudjuk meghatározni, hogy a kimutatott változás inkább egy folyamat következménye, vagy valóban egy hirtelen, ugrásszerű változás történt-e a kijelölt évben.

A kérdés eldöntéséhez össze kell hasonlítani a vizsgált folyamatot leíró lineáris trend és a kapott töréspont jellemzőit, pontosabban az eltérés négyzetösszegeket (SS). Illesszünk az idősorunkra regressziós egyenest, legyen ez  $x_{trend} = mt + b$  alakú trendvonal. Határozzuk meg a trendegyenes adott pontokhoz tartozó helyettesítési értékeit, jelöljük ezeket  $x_{trend,i} = mt_i + b$ -vel,  $1 \leq i \leq n$  az  $n$  adatot tartalmazó adatsor esetében. Képezzük ezek eltérését az adott pont tényleges  $x_i$  értékeitől, majd számítsuk az eltérés négyzetösszeget,

$$SS_{trend} = \sum_{i=1}^n (x_i - x_{trend,i})^2.$$

Adott ekkor továbbá az idősorban kimutatott töréspont által meghatározott két részidőszak és azok átlagai ( $\bar{x}_1$  és  $\bar{x}_2$ ). Az első részidőszak elemei  $x_1, x_2, \dots, x_k$ , míg a másodiké  $x_{k+1}, x_{k+2}, \dots, x_n$ . Állítsuk elő a két részidőszakra külön-külön az egyes tényleges mérési adatok és a részidőszak átlagának különbségeit és képezzük ebből az eltérés négyzetösszegeket:

$$SS_{tp} = \sum_{i=1}^k (x_i - \bar{x}_1)^2 + \sum_{j=k+1}^n (x_j - \bar{x}_2)^2.$$

Az illeszkedések jóságának, pontosságának összehasonlítására képezzük a két eltérés négyzetösszeg hányadosát:  $h = \frac{SS_{trend}}{SS_{tp}}$ -t.

A kapott  $h$  hányados értéke alapján meghatározható, hogy melyik eljárás ad pontosabb eredményt, hiszen ha  $h > 1$ , akkor a töréspontok által meghatározott átlagok, ha  $h < 1$ , akkor a teljes adatsorra illesztett trendvonal illeszkedik jobban a vizsgált adatsorra.

### Szórások összevetésének módszere

Az eljárás lényege, hogy a potenciális töréspontok mindegyikére vonatkozóan meghatározzuk az általuk definiált két-két részidőszak közös szórásnégyzetét a korábban már ismertetett képlet alapján,  $S^2 = \frac{(n_1-1)\sigma_1^2 + (n_2-1)\sigma_2^2}{n_1+n_2-2}$ . Ezt követően kiszámítjuk a teljes, rendelkezésre álló adatsor szórásnégyzet értékét is,  $\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$  és ezt vetjük össze  $\chi^2$ -próba segítségével az aktuálisan vizsgált két részidőszak közös szórásával. Az alkalmazott statisztika

$$\chi^2 = (n-1) \frac{\sigma^2}{S^2}$$

alakú, ahol  $\sigma^2$  a teljes, rendelkezésre álló időszak,  $S^2$  pedig a két részidőszak közös szórásnégyzete,  $n$  pedig a minta elemszáma. Az így definiált próbastatisztika megközelítőleg  $n-1$  szabadsági fokú  $\chi^2$  eloszlást követ (Dévényi és Gulyás, 1988)

Azt a pontot tekinthetjük a jelöltek közül tényleges töréspontnak, ahol a  $\chi^2$ -érték maximális (lokálisan), mert ezekben a pontokban tér el a két részidőszak közös szórása lefele a teljes időszak szórásához képest. Természetesen ezt is adott szignifikancia szint mellett vizsgálva kaphatjuk meg a statisztikai értelemben is definiálható töréspontot.

Molnár (2003) összevetette a két módszertant és megállapította, hogy a gyakorlatban az egyes idősorok különböző szakaszokra való tagolásával kapott  $\chi^2$ -értékek időbeli futásának karakterisztikája erős hasonlóságot mutat a megfelelő idősor  $t$ -próbás vizsgálatánál kapott  $t$ -értékekével. Még a maximumok időpontjai is megegyeznek, azonban megadott szinten szignifikáns maximum jóval kevesebb van, mint az előző vizsgálatnál.

### **Anomáliák kumulatív összegzése**

Általánosan megfogalmazva azt mondhatjuk, hogy anomáliának tekinthetjük azt, ami az átlagostól eltér. Természetesen nem mindegy, hogy egy adat mennyire tér el az átlagtól, például outlier, kiugró adatok definiálásának egy jól bevált módja az átlagtól legalább az interkvartilis terjedelem (IQR) másfélszeresével való eltérés mérték.

Meteorológiai adatsorok esetében – általános anomáliavizsgálat kapcsán – is az anomáliának az átlagtól való eltérését tekintjük, ezt azonban jelentősen befolyásolhatja a klimatikus idősorokban tapasztalható havi, évszakos trendek és ciklusok jelenléte. Utóbbiakat megfelelő technikákkal ki kell küszöbölni az adatsorokból az anomáliák definiálása előtt, illetve az éves bontás mellett a vizsgálatokat havi szinten is célszerű elvégezni.

Általános értelemben az  $x_i$  ( $1 \leq i \leq n$ ) adathoz tartozó anomália az alábbi egyszerű módon definiálható:

$$a_i = x_i - \bar{x}, 1 \leq i \leq n,$$

ahol  $n$  az adatsor hossza,  $\bar{x}$  pedig a teljes adatsor átlaga.

Ezek alapján minden potenciális töréspontra meghatározzuk az anomáliák kumulált értékét:

$$A_{szum} = \sum_{j=1}^k a_j, k < n.$$

Azt a pontot tekinthetjük töréspontnak, ahol ez az összeg extrém nagy vagy éppen extrém kicsi. A módszernek két nagy hátránya is van, az egyik hogy az idősorok részletesebb és bonyolultabb előkészítést igényelnek, mint a korábbi két módszer esetében, illetve nem teljesen definiált az sem, hogy a kumulált anomáliák esetében mit tekintünk extrém nagyoknak, illetve extrém kicsi értéknek.

## Pettitt-féle nemparaméteres próba

Amennyiben a vizsgált adatsorok eloszlása szignifikánsan eltér a normálistól, a részidőszakok átlagának eltérésén alapuló módszer nem alkalmazható. Ekkor jöhet számításba annak egy nemparaméteres változata, a Pettitt-próba. Ennek lényege, hogy minden potenciális töréspont esetében kiszámítjuk az

$$X_k = 2R_k - k(n - 1)$$

próbastatisztika értékét, ahol  $R_k = \sum_{j=1}^k r_j$  és  $r_j$  az  $x_j$  elem – rangkorellációk esetéhez hasonlóan definiált – rangszáma. A teljes minta elemszáma  $n$ , és  $k = 1, 2, \dots, n$ . A töréspont az a  $k$  pont lehet, ahol az  $X_k$  értéke maximális vagy minimális, attól függően, hogy a tapasztalt ugrásszerű változás fel- vagy lefele irányul-e.

Annak eldöntésére, hogy a módszer által adott  $X_k$  érték valóban tényleges, statisztikai értelemben is kimutatható töréspontot határoz-e meg ki kell számítani az annak valószínűségét adó, az  $X_k$  értékhez tartozó szignifikanciaszintet. Ezt az alábbi összefüggés szolgáltatja (Mares és Mares, 1994):

$$\alpha = 2e^{-\frac{6(X_{extrem})^2}{n^3+n^2}},$$

ahol  $x_{extrem}$  az  $X_k$  extrém értékéhez tartozó elemszámnak felel meg. Ha az  $\alpha$  értéke kisebb volt, mint 0,05, az  $x_{extrem}$  értékhez tartozó pont által elválasztott részek eltérésének valószínűsége meghaladta a 95%-ot, azaz szignifikáns a felfedett töréspont.

## Jel-zaj arány vizsgálat

Ahogy az idősorok bemutatása során már szó volt róla a determinisztikus modellezés esetében feltételezünk az idősor vonatkozásában egyfajta állandóságot, ez a tartósnak tekinthető tendencia, a szezonális és a ciklikusság, míg a véletlent valóban véletlennek és ilyen értelemben egy zavaró tényezőnek, zajnak tekintjük. Ez a fajta megközelítés – szemben a sztochasztikus modellezéssel – ezért inkább leíró jellegű, a hosszabb távú hatások feltérképezésben erős, a véletlennel kevésbé foglalkozik.

A statisztikai szakirodalomban a hosszú távú, potenciálisan előre jelezhető változásokat jelnek (ami mérhető, tervezhető), míg a természetes változékonyságot zajnak tekintjük. Töréspont kereshető ezen két komponens idősorban történő előfordulási arányainak vizsgálatával is. A jel–zaj arány becslésére többféle módszer létezik, köztük például az alábbi, Yamamoto által ajánlott és alkalmazott (Mares és Mares, 1994):

$$A_{jel/zaj} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma_1 + \sigma_2},$$

ahol  $\bar{x}_1$ ,  $\bar{x}_2$  és  $\sigma_1$  és  $\sigma_2$  a két vizsgált részidőszak átlagai és szórásai. Töréspontnak a jel–zaj arány maximumát produkáló választópontot tekinthetjük, vagyis ahol  $A_{jel/zaj}$  a maximális értéket veszi fel. A töréspontot a gyakorlatban akkor tartjuk reálisnak, ha  $A_{jel/zaj} \geq 0,5$  (ez inkább egy empirikus értéknek tekinthető).

A töréspontbecslés különféle módszereit alkalmazva felmerülhet a kérdés, hogy mennyire eltérő ezek érzékenysége akár a teljes idősor, akár annak különböző szakaszai tekintetében. Ezt részletesen vizsgálta dolgozatában Molnár (2003). Különböző meteorológiai (köztük hőmérséklet) adatsorokon történő futtatások, vizsgálatok alapján megállapította, hogy az anomáliák kumulatív összegzése és a Pettitt-próba kiemeli az adatsor középső szakaszán jelentkező töréspontokat, míg a széleken kevésbé érzékeny. A jel–zaj arány esetében éppen ennek az ellenkezője figyelhető meg. A többi alkalmazott próba érzékenységében nem talált szakaszfüggőséget.

Figyelembe véve ezeket a megállapításokat és azt, hogy e dolgozatnak nem konkrét témája a töréspont kereső módszerek részletes vizsgálata és összehasonlító elemzése, az általunk elvégzett vizsgálatokban a  $t$ -próbán alapuló töréspont kereső módszert alkalmaztuk. Megjegyezzük, hogy a vizsgált hőmérséklet és csapadék adatsorok is teljesítették a normalitásra vonatkozó feltételeket a Kolmogorov-Szmirnov tesztet elvégezve.

### 3.2.2 Töréspont keresés $t$ -próba alkalmazásával és a töréspontok jelentősége

A megfelelő hosszúságú meteorológiai adatsorokban a töréspontok vizsgálatának fontosságát a korábban már említett mintaterület meteorológiai és növekedési adatsor kapcsán mutatjuk be. Rávilágítunk, hogy egy ilyen töréspontnak milyen fontos szerepe lehet az összefüggés vizsgálatokban, mert azok erősségét, de sokszor irányát is befolyásolhatja.

A meteorológiai adatok 1961-2007 vonatkozásában állnak rendelkezésre. Hasonlóan, ahogy Molnár és Izsák (2011), itt is 5 évben definiáltuk a minimális részidőszak hosszát. Az éves átlaghőmérsékletek vonatkozásában a töréspont 1998-ra adódott (9. táblázat), mely valóban egy ugrásszerű, hirtelen változás statisztikailag az alábbi értékekkel:

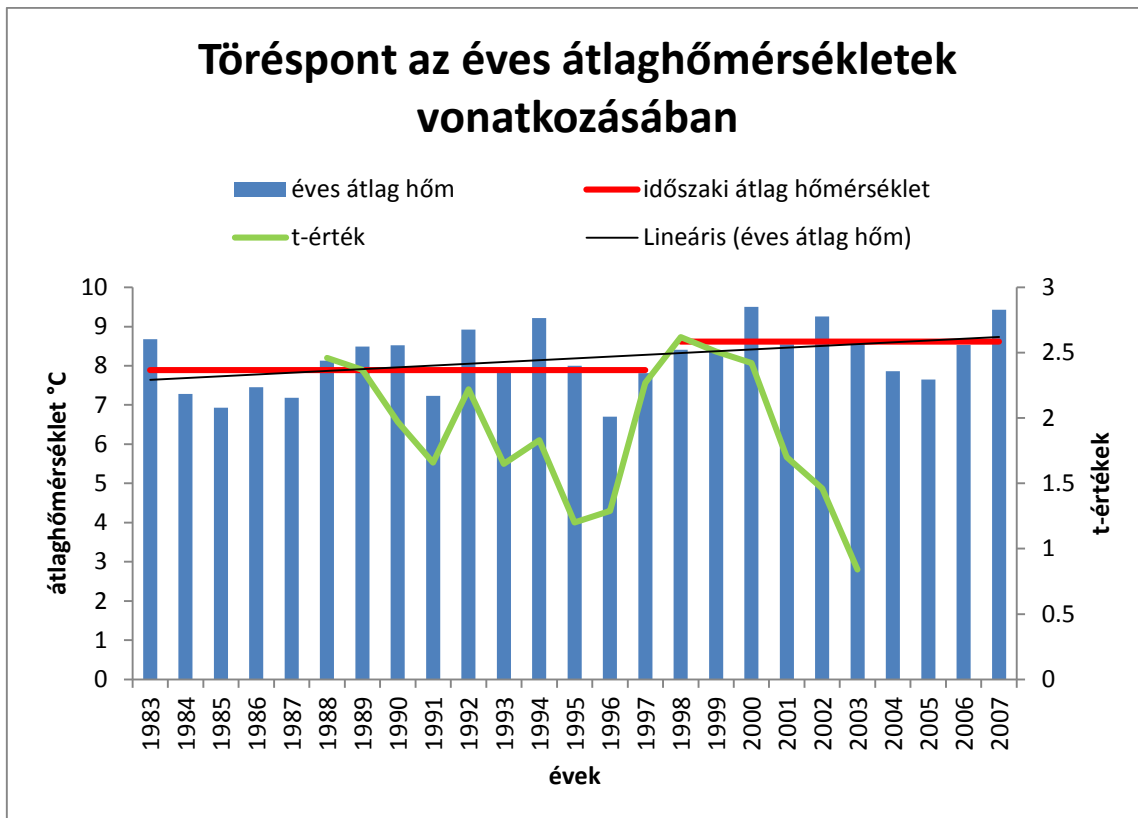
9. táblázat – Töréspont a mintaterület hőmérséklet adatsorban

Éves átlaghőmérséklet			
	átlag előtte (°C)	átlag mögötte (°C)	eltérés (°C)
1998	8,01	9,05	1,04

Az éves csapadékösszegek vonatkozásában nem kaptunk statisztikailag igazolható töréspontot ezzel a módszerrel. A kapott eredmények teljes összhangban vannak Molnár és Izsák (2011) által a Beregszászi Meteorológiai Állomás éves adataival végzett vizsgálatainak kapcsán kapott eredményekkel.

Ahogy a bevezető szakaszban már említettük, a meteorológiai adatokban található töréspontok felfedése az összefüggés-vizsgálatokban is fontos lehet. A mintaterületre a növekedés adatok (körlapterület növekedés) rendelkezésre állása 1985-2007 közti, azonban az 1998-as év hiányzik. Így külön megvizsgáltuk az ezen időszakhoz köthető meteorológiai adatokat is, azaz a havi csapadékösszegek és átlaghőmérsékletek felhasználásával kerestünk töréspontot az 1983-2007-es időszakban mind a havi, mind az éves, mind a naptári évszakok vonatkozásában. A vizsgált szakaszok minimális hosszát itt is 5 évben definiáltuk, így az első lehetséges töréspont 1988, az utolsó pedig 2003 lett. E vizsgálatunkban azt a pontot tekintjük statisztikai értelemben is definiált töréspontnak, ahol 95%-os szinten szignifikáns eltérés mutatható ki az aktuálisan vizsgált szakaszátlagok között. A havi és időszaki adatok vizsgálatának oka, hogy ezekkel az adatokkal keressük összefüggéseket az éves növekedés adatok tekintetében. Illetve az 1983-as kezdés oka, hogy az összefüggés vizsgálatokban a CReMIT módszerrel adott év növekedését adott, eggyel és kettővel korábbi év meteorológiai adataival vetettük össze.

Az éves átlaghőmérséklet és csapadékösszeg vonatkozásában a fenti feltételek mellett az alábbi eredményeket kaptuk. A csapadék adatsor esetében a maximális  $t$ -érték (1,46) 2000-ben adódott, azonban  $p = 0,16$  érték mellett, így ezt nem tekinthetjük szignifikáns eltérésnek az általunk kijelölt 95%-os szinten. Az éves átlaghőmérséklet adatok tekintetében (8. ábra) a legnagyobb  $t$ -érték – ugyanúgy, mint a hosszabb adatsor vonatkozásában – 1998-ra adódott 2,62-es értékkel, illetve a hozzá tartozó  $p$ -érték 0,02, ami alapján ez az év két, 95%-os szinten szignifikánsan eltérő részidőszakra bontja az idősort: 1983-1997 és 1998-2007 felbontásban. A kapott két részátlag  $7,89^{\circ}\text{C}$  és  $8,62^{\circ}\text{C}$ , azaz az eltérés  $0,73^{\circ}\text{C}$  a két időszak között.



8. ábra – Töréspont az éves átlaghőmérséklet vonatkozásában

A hőmérséklet adatokra a teljes időszak vonatkozásában illesztett regressziós egyenes egyenlete  $y = 0,0455x + 7,594$  és  $R^2 = 0,18$ , ez utóbbi érték 95%-os szinten szignifikáns illeszkedést mutat.

Ez azt jelenti, hogy a hőmérséklet adatok tekintetében az elvégzett vizsgálatok mind a lineáris trend, mind a töréspont vonatkozásában 95%-os szinten szignifikáns eredményeket hoztak. Annak meghatározására, hogy a hőmérséklet adatokban bekövetkezett változásokat melyik módszer írja le pontosabban (folyamatos, trendszerű változás, vagy ugrásszerű, a töréspont által definiált változás) a már ismertetett eltérések négyzetösszegein ( $SS$ ) alapuló módszert alkalmaztuk.

A mi esetünkben  $SS_r = 12,11$  és  $SS_{tp} = 10,97$  értékek adódtak. Azaz a hőmérséklet adatokban tapasztalható változás inkább hirtelen, ugrás jellegű, mintsem egy folyamatos emelkedő trend eredménye. Mindez felveti azt a kérdést, hogy a teljes időszak vizsgálata helyett/mellett nem lenne-e célszerű a két részüszakot külön-külön is vizsgálni.

A havi és évszakos szinten elvégzett vizsgálatok esetében karakterisztikáját tekintve hasonló eredmények adódtak, mint a teljes év vonatkozásában. A hőmérséklet adatsorok tekintetében sok statisztikailag értelmezhető eltérés adódott (10. táblázat). A tavaszi időszak adataira teljesen

hasonló eredmény adódott, mint a teljes év tekintetében, azaz az 1983-1998-as időszak átlaghőmérséklete, 7,89°C szignifikánsan ( $\alpha = 0,05$ ) alacsonyabb, mint az 1999-2007 közti időszaké, ami 8,99 °C volt. A nyári időszak tekintetében ugyanezen évre szintén adódik egy több mint 0,8°C-os eltérés, azonban ez a vizsgált szinten nem szignifikáns. Egy ennél nagyobb és már statisztikailag is értékelhető eltérés adódott az 1992-es évre, több mint 1°C eltéréssel az 1992-t követő időszak javára. Megjegyezzük, hogy az éves átlaghőmérsékletek tekintetében is kijött a szignifikáns eltérés az 1992-es évre, azonban ennél nagyobb eltérés adódott 1998-ra.

Havi adatok vonatkozásában az április a tavasszal, június pedig a nyári időszakokkal fedésbe hozható eltéréseket mutat.

10. táblázat – Töréspontok a havi, időszaki hőmérséklet adatokban

<b>átlaghőmérsékletek (1983-2007)</b>				
	<i>eltérés (év)</i>	<i>átlag előtte (°C)</i>	<i>átlag mögötte (°C)</i>	<i>eltérés (°C)</i>
<i>éves átlag</i>	1998	7,89	8,62	0,73
<i>tavas</i>	1998	7,89	8,92	1,03
<i>nyár</i>	1992	17,11	18,18	1,07
<i>április</i>	1998	7,85	9,08	1,23
<i>június</i>	1992	14,90	16,71	1,81
<i>november</i>	2000	2,09	3,91	1,81

A csapadék adatsorok esetében az éves csapadékösszegek vonatkozásában itt sem adódott statisztikailag kimutatható töréspont, ugyanakkor havi és évszaki szinten már igen (11. táblázat). A tavaszi időszakra 2000-ben adódott egy ugrásszerű töréspont, ami azt mutatja, hogy a tavasszal lehullott csapadék átlagos mennyisége között közel 40 mm különbség van az 1983-1999-es időszak javára. A havi szinten vizsgált adatsorok esetében május, június és július hónapokra kaptunk még statisztikailag kimutatható eltéréseket, májusban 2000-re és júniusra 1998-ra hasonló eredménnyel, mint amit a teljes tavaszi időszak vonatkozásában is tapasztaltunk, míg ezzel szemben júliusra 1996 vonatkozásában egy a tavasszal előjelében ellentétes eredmény adódott közel 30 mm eltéréssel.



11. táblázat – Töréspontok a havi, időszaki csapadék adatokban

<b>csapadékösszegek (1983-2007)</b>				
	<i>eltérés (év)</i>	<i>átlag előtte (mm)</i>	<i>átlag mögötte (mm)</i>	<i>eltérés (mm)</i>
<i>tavas</i>	2000	212	174	38
<i>május</i>	2000	101	66	35
<i>június</i>	1998	106	71	35
<i>július</i>	1996	65	98	33
<i>október</i>	1990	36	61	24

Már ez a viszonylag rövid adatsor is azt mutatja, hogy akár egy 25 éves adatsorban is felfedhetőek olyan ugrásszerű változások, melyek a növekedési adatokkal való kapcsolatokban is hozhatnak érzékelhető változásokat. Ennek bemutatására a következő vizsgálatot végeztük el: a területre jellemző átlagos növekedési adatsort (1985-2007, 1998 kivételével) vetettük össze a két környezeti paraméterrel (havi csapadékösszegek és átlaghőmérsékletek) a CReMIT módszert alkalmazva. Mindezt elvégeztük a teljes lehetséges időszak, 1985-2007 és a hőmérséklet adatsorban az éves adatok vizsgálata alapján kapott töréspont mentén ketté bontva az adatsort, azaz egy 1985-1997 és 1999-2007-es időszakra nézve.

Az elemzéseket a kettővel korábbi év áprilisától adott év októberéig tartó havi adatok bevonásával végeztük el, legfeljebb 4 hónap szélességű ablakok létrehozásával. Fontos, hogy nem célunk erdészeti eredmények, összefüggések megfogalmazása, csupán arra próbálunk matematikai, statisztikai oldalról rávilágítani, hogy az idősorokban tapasztalható hirtelen változások, töréspontok mennyire befolyásolhatják az ezen idősort felhasználó elemzések eredményeit. Az alábbi megállapítások fogalmazhatóak meg felsorolás jelleggel annak tekintetében, hogy mennyiben változtak meg, vagy éppen nem változtak meg az összefüggések a teljes és a két részidőszak tekintetében.

#### **Csapadék adatok tekintetében:**

1. A teljes és a második részidőszakra kapott összefüggések (ahol mindkettő szignifikáns) jellemzően azonos előjelűek, azonban a második részidőszak összefüggései sokkal erősebbek. Pl. CS.a4-a6..
2. A két részidőszak – különösen az adott év vonatkozásában – sokszor ellentétes előjelű, de szignifikáns összefüggéseket mutat. Pl. CS.a6-a7, CS.a5-a7, CS.a5-a8. Hasonló tapasztalható a szimpla hónapok esetében is. Ennek oka nyilván a nyári hónapok csapadék adatainak szignifikáns eltéréseiben keresendő.

3. A havi adatok kapcsán is tapasztalható – különösen adott év nyara tekintetében -, hogy a két részidőszak kapcsolatainak iránya is ellentéte, mint pl. CS.a5, CS.a6, CS.a7.

12. táblázat – Összefüggések a részidőszakokban, csapadék

<b>csapadék adatok</b>				
<i>mettol</i>	<i>meddig</i>	<i>1985-2007</i>	<i>1985-1997</i>	<i>1999-2007</i>
<i>CS.pp10</i>	<i>CS.p1</i>		0,61.	-0,6.
<i>CS.a4</i>	<i>CS.a6</i>	0,39.		0,63.
<i>CS.a5</i>	<i>CS.a8</i>		-0,48.	0,77.
<i>CS.a6</i>	<i>CS.a7</i>		-0,68.	0,69.

### Hőmérséklet adatok tekintetében:

1. Az első részidőszakra sokkal kevesebb szignifikáns kapcsolat adódott, mint a másodikra, illetve a teljes időszakra vonatkozó kapcsolatok közül sok a két részidőszak tekintetében nem mutatható ki.
2. A második részidőszak nyarára adódtak olyan összefüggések, melyek sem a teljes, sem az első részidőszakra nem adódtak, mint pl.: H.a7-a8, H.a9.
3. Az adott év nyári hónapjai 1999-2007 között erős, de nem szignifikáns negatív kapcsolatot mutatnak a növekedéssel (augusztus 90%-on szignifikáns), ami a korábbi részidőszakban egyáltalán nem volt jellemző. A korábbi és a kettővel korábbi év második részidőszakra kapott összefüggési hasonló képet mutatnak, mint ami a teljes időszakot is jellemzi (H.pp10, H.p11). Ugyanakkor érdekes, hogy 1985-1997 között a kapcsolatok sokszor még előjelükben sem egyeznek a második részidőszakkal, pl. H.a5, H.a7.

13. táblázat – Összefüggés a részidőszakokban, hőmérséklet

<b>hőmérséklet adatok</b>				
<i>mettől</i>	<i>meddig</i>	<i>1985-2007</i>	<i>1985-1997</i>	<i>1999-2007</i>
<i>H.p8</i>	<i>H.p11</i>	-0,53.		-0,69.
<i>H.p10</i>	<i>H.p11</i>	-0,65.		-0,73.
<i>H.p5</i>	<i>H.p6</i>	-0,43.		
<i>H.a4</i>	<i>H.a5</i>		0,65.	
<i>H.a6</i>	<i>H.a8</i>	-0,41.		-0,62.
<i>H.a7</i>	<i>H.a8</i>			-0,69.

A bemutatott eredmények egyrészt arra hívják fel a figyelmet, hogy az alkalmazott elemzési módszer tekintetében (lineáris korrelációanalízis) a független paraméterekben bekövetkező változások komoly hatással bírnak az összefüggés vizsgálatok kimenetére. Azt is látható, hogy a hosszabb vizsgálati időszakra vonatkozó összefüggések, kapcsolatok nem feltétlenül állnak fent annak egy-egy rövidebb részintervallumára, sőt akár az összefüggések irányát tekintve is bekövetkezhetnek változások. Érdekes tapasztalat továbbá, hogy az összefüggések időben is változnak, aminek felismerése és kezelése annak érdekében fontos kérdés, hogy valóban reális, a valósághoz jobban közelítő eredményeket fogalmazhassunk meg. Az időbeli változások vizsgálatára a mozgó intervallumok, evolúciós technika mellett az általunk létrehozott eljárás is alkalmas és fontos eszköz.

### **3.3 Növekedési görbék**

Amennyiben a növekedési adatok mérési sűrűsége lehetővé teszi, akkor képesek lehetünk az éves növekedési adatsorokra egy megfelelő növekedési görbét illesztve azt folytonossá tenni. Valamint az illesztett görbe paramétereiből akár közvetlen, akár közvetett módon egyéb jellemzők generálására, mint például érdemi növekedési folyamat kezdete, befejezése, meredeksége. E paraméterek felhasználásával, vagy a görbék további jellemzőinek meghatározásával pedig más, a görbére jellemző tényezők is definiálhatóak. Először megvizsgáljuk a növekedési folyamatok kapcsán szóba jöhető függvények körét, majd megfelelő szempontok alapján kiválasztjuk az általunk alkalmazottat.

Számos olyan természeti és ember által generált mesterséges jelenséget ismerünk, melynek időbeli alakulása kezdetben olyan exponenciális görbére utal mely a határtalan növekedés érzetét keltheti, azonban egy bizonyos szint fölé (szaturációs küszöb, telítődési szint) nem emelkedhet gyakorlati korlátok miatt. Ilyen folyamatra példa egy tumor növekedése, vagy bizonyos ökológiai folyamatok (pl. egy növény növekedési fejlődése). Ezek a korlátos exponenciális növekedési folyamatok a hagyományos exponenciális függvényekkel nem írhatóak le utóbbiak határtalan növekedése miatt. Ugyanakkor bizonyos feladatok esetében a folyamatnak van egy lecsengése is, amikor egy exponenciális jellegű csökkenő folyamatot láthatunk a kezdeti emelkedés és a szaturációs szint elérése után. Az elmúlt több mint 150 évben számtalan megoldás született ilyen problémák kezelésére.

### 3.3.1 Növekedési görbék

Zeide (1993), valamint Sit és Poulin-Costello (1994), tanulmányai alapján, a gyakorlati szempontokat is figyelembe véve a vizsgált függvényeket három csoportba szokás sorolni az inflexiós pontjaik száma alapján:

- inflexiós pont nélküliek,
- két inflexiós ponttal rendelkezők,
- egy inflexiós ponttal rendelkezők.

A függvények nagy száma miatt nem vállalkozunk a teljes körű áttekintésre, azonban igyekszünk a nevezetesebb lehetőségeket felsorolni és a kiválasztott, általunk is alkalmazott függvényeket kissé részletesebben is bemutatni (Kehl és Sipos, 2009). A fenti sorrendiség oka, hogy először általánosan jellemezzük az adott függvény osztályokat, majd külön kitérünk a mi vizsgálataink szempontjából kiemelt szerepet játszó osztályra, az egy inflexiós ponttal rendelkező görbékre. Emellett a függvények megadásában is lehetnek kisebb-nagyobb eltérések a különböző irodalmakban, Mi minden esetben igyekeztünk az általános formát bemutatni.

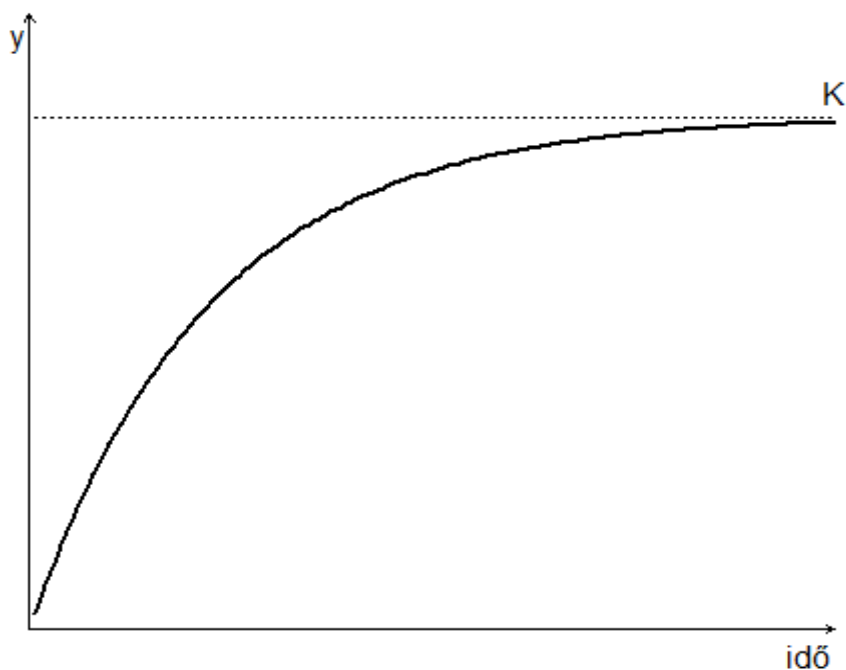
#### Az inflexiós ponttal nem rendelkező függvények

Ezek teljes értelmezési tartományukon konkáv módon viselkednek,  $K$  minden esetben a telítődési szintet jelöli.

14. táblázat – Inflexiós ponttan nem rendelkező növekedési függvények

Függvény neve	formula
Egyszerűen modifikált exponenciális függvény	$y(t) = K - ae^{-rt}$
Mitscherlich	$y(t) = K(1 - e^{-rt})$
Bertalanffy	$y(t) = K(1 - be^{-rt})$
Törnquist 1.	$y(t) = \frac{Kt}{t + a}$
Törnquist 2.	$y(t) = \frac{K(t + a)}{t + b}$
Levakovic	$y(t) = K \left( \frac{t^d}{b + t^d} \right)^c$

A fenti függvények alapkarakterisztikája általános módon a 10. ábrán látható görbével jellemezhető:



9. ábra – Inflexiós ponttal nem rendelkező görbék általános alakja

Az általunk vizsgált növekedési folyamatok leírására (fák növekedése) ezek a görbék nem alkalmasak éppen az inflexiós pontjuk hiányából fakadó korlátok miatt. Így ezek részletesebb vizsgálatával a továbbiakban nem foglalkozunk.

### A két inflexiós ponttal rendelkező növekedési-csökkenési görbék

Az életgörbe trendfüggvény alapvetően egy termék életgörbe alakulását mutatja, nevét is innen kapta, mivel a termék piaci forgalmának (volumenének) alakulását ábrázolja az idő függvényében. Két jellemző függvényt szoktak ebbe a csoportba sorolni:

$$\text{Haustein: } y(t) = \frac{K}{e^{\omega^2(t-r)^2}}$$

$$\text{Hubbert: } y(t) = \frac{bue^{b(t-r)}}{(1+e^{b(t-r)})^2}.$$

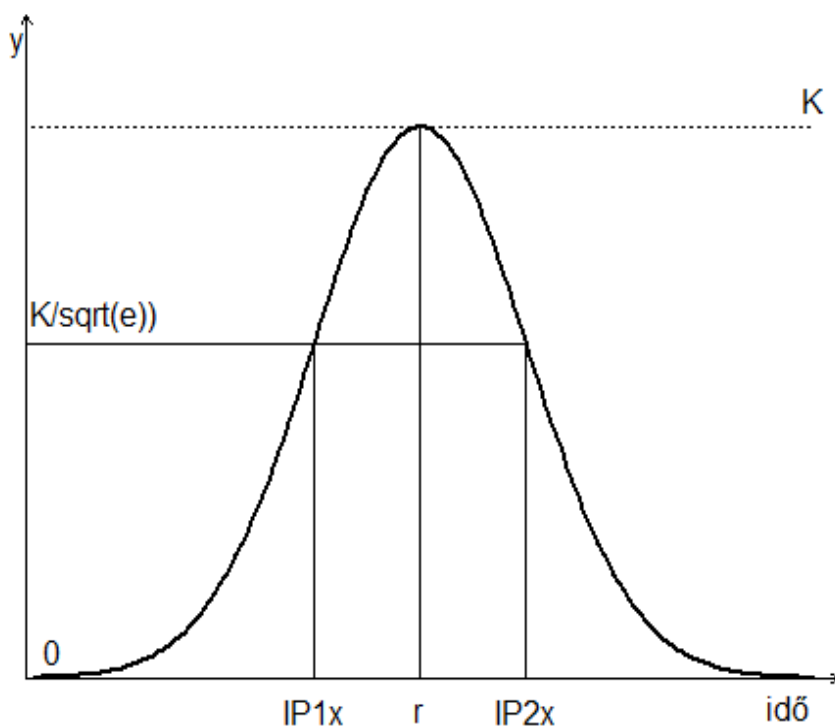
A két függvény alapkarakterisztikája hasonló, ezért csak az egyiket mutatjuk be kissé részletesebben. A Haustein függvény:  $y(t) = Ke^{-\omega^2(t-r)^2}$  paramétereinek gazdasági értelmezése jellemző módon az alábbi szokott lenni:

$K$ : az éves termelési értékesítési volumenek várható maximuma

$\omega$ : a görbe alakját, az inflexiós pontok helyét meghatározó alakparaméter

$r$ : a maximális értékesítés várható időpontja

az inflexiós pont koordinátái ( $IP1x, IP2x$ ):  $\left(r \pm \frac{1}{\sqrt{2}\omega}, \frac{K}{\sqrt{e}}\right)$



10. ábra – Két inflexiós ponttal rendelkező Hausteine életgörbe

Az élet- vagy az inflexiós ponttal nem rendelkező görbék a fák éven belüli növekedési folyamatának függvényesített leírására nem alkalmasak, ezért a tovább részletezésüktől eltekintünk. Az elkövetkezőkben a feladatra alkalmas 1 inflexiós ponttal rendelkező görbékkel foglalkozunk részletesebben.

#### **Az egy inflexiós ponttal rendelkező telítődési függvények**

Ezek azok a görbék, melyek már alkalmasak lehetnek az általunk is vizsgált, fák éves növekedési folyamatainak leírására (Edelényi és mtsai, 2011b). Az egy inflexiós ponttal rendelkező telítődési függvények jellemzője, hogy a rövid, lassú kezdeti növekedés után egy intenzívebb szakasz következik, ami után a lassú, befejező növekedés egy telítődési ponthoz tart. A dinamikus növekedési folyamat a görbe inflexiós pontján fordul át a lassabb növekedésbe. A fák éven belüli növekedésmenete ilyen, „S” alakot leíró, ún. telítődési (szigmoid) függvényekkel általában jól közelíthető.

A 15. táblázat felsorolását követően csak azokat a függvényeket jellemezzük részletesebben, melyek a szakirodalom alapján gyakran alkalmazottak fák éves növekedési folyamatainak leírásában. A továbbiakban a következő általános jelöléseket alkalmazzuk a függvények kapcsán:  $t$  jelöli az időt,  $K$  a telítődési szintet,  $c$  a növekedési sebességet,  $m$  pedig az inflexiós pont pozícióját határozza meg.

15. táblázat – Az egy inflexiós ponttal rendelkező növekedési függvények

Függvény neve	formula
logisztikus függvény	$y(t) = \frac{K}{1 + e^{-c(t-m)}} = \frac{Ke^{ct}}{e^{cm} + e^{ct}}$
késleltetett logisztikus függvény	$y(t) = \frac{K}{1 + \left(\frac{T}{t}\right)^a}$
négyzetesen logisztikus függvény	$y(t) = \left(\frac{K}{1 + be^{-ct}}\right)^2$
Gompertz függvény	$y(t) = Kb^{ct}$ $y(t) = Ke^{-be^{-ct}}$
A 63%-os telítődési függvény:	$y(t) = A - \frac{K}{e^{\left(\frac{t}{T}\right)^a}}$
Johnson függvény:	$y(t) = e^{K\frac{b}{c+t}}$
Az általánosított Richards-féle függvény:	$y(t) = \frac{K}{(1 + ve^{-c(t-m)})^{\frac{1}{v}}}$
Korf függvény	$y(t) = Ke^{-bt^{-c}}$
Sloboda függvény	$y(t) = Ke^{-be^{-ct^d}}$
Verhulst	$y(t) = \frac{K}{1+e^{-c(t-m)}} = \frac{Ke^{ct}}{e^{cm}+e^{ct}}$
Weibull	$y(t) = K(1 - e^{-bt^c})$
Pearl – Reed	$y(t) = \frac{K}{be^{-ct} + 1}$
Chapman – Richards	$y(t) = K(1 - e^{-bt})^c$

Ezeket a növekedési görbéket sok kutatási területen alkalmazzák, így többek között az erdőszetben is, különböző folyamatok modellezésére és egyéb, az illetett görbe alapján definiálható paraméterek, értékek meghatározására.

Az erdőszeti szakirodalom alapján gyakran alkalmazott 3 függvénnyel, a logisztikus ( $y_L$ ), a Gompertz- ( $y_G$ ) és a Richards-függvénnyel ( $y_R$ ) foglalkozunk a továbbiakban. Mindegyik függvénynek definiáljuk az eredeti, matematikai alakját, majd a konkrét problémához konvertált alakját és az ehhez tartozó, már speciálisan biológiai értelmezéssel is bíró paraméterezését (Zwietering és mtsai, 1990).

## A logisztikus függvény

A korlátos növekedési folyamatok leírására az első értékesebb megoldást Francois Verhulst belga matematikus publikálta. Abból indult ki, hogy minden folyamat rendelkezik egy rá jellemző  $K$  telítődési szinttel, amit beépített modelljébe. 1845-ben publikált második cikkében már logisztikus névvel hivatkozott rá, ami általánosan elfogadott elnevezéssé vált, s az alábbi képlet írja le:

$$y(t) = \frac{K}{1 + e^{-c(t-m)}}$$

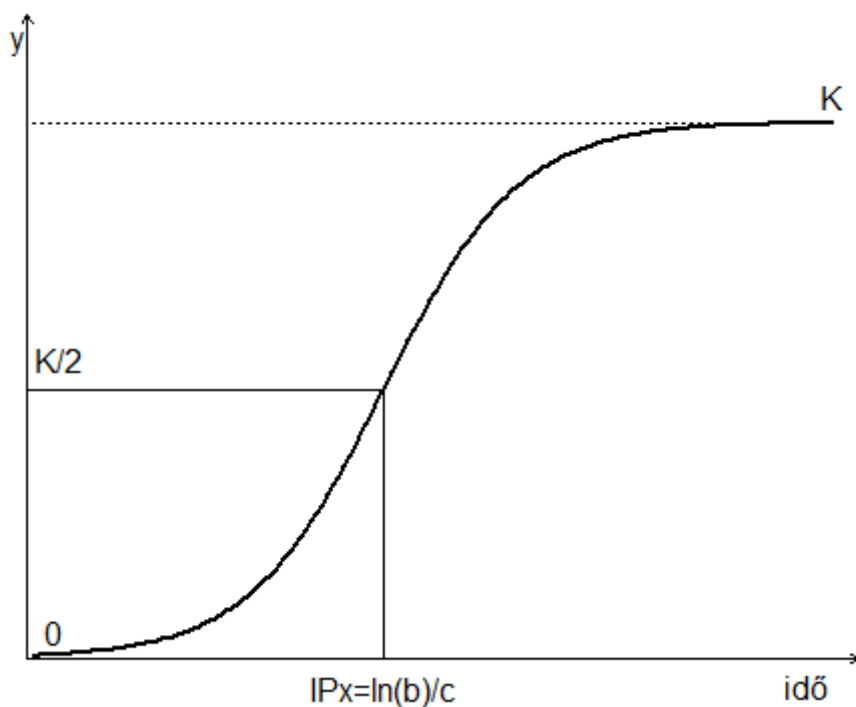
A görbe inflexiós pontja,  $IP = \left(m, \frac{K}{2}\right)$  független a paraméterektől, továbbá az ebben a pontban felvett függvényérték mindig  $\frac{K}{2}$ . A logisztikus függvény egyszerű, paraméterei könnyen értelmezhetőek. Ha  $c > 0$  logisztikus növekedés, ha  $c < 0$ , akkor logisztikus csökkenésről van szó.

A függvény szimmetrikus az inflexiós pontra, ahol a görbe a telítettségi szint felét  $\left(\frac{K}{2}\right)$  éri el. Napjainkban logisztikus trendfüggvény névvel az egyik legelterjedtebben alkalmazott függvény a Pearl–Reed-féle logisztikus függvény:

$$y(t) = \frac{K}{1 + be^{-ct}} = \frac{Ke^{ct}}{e^{ct} + b}$$

A  $c$  paraméter értelmezése hasonló, mint az eredeti, Verhulst által definiált változatban, a függvény inflexiós pontjának koordinátái pedig:  $IP = \left(\frac{\ln b}{c}, \frac{K}{2}\right)$ . Könnyen be lehet látni, hogy a Verhulst és a Pearl-Reed függvények gyakorlatilag azonosnak tekinthetőek.





11. ábra – Logisztikus függvény

Megjegyzendő, hogy sok esetben a vizsgált folyamatok nem ennyire szabályosak, szimmetrikusak, éppen ez jelentheti ezen függvény gyakorlati alkalmazásának egyik komoly korlátját.

### **Gompertz függvény**

Több kutató próbálkozott a növekedési folyamatok más, kevésbé szimmetrikus módon történő függvényesítésével. Időben ezek közül Gompertz tekinthető az elsőnek, aki az alábbi csökkenő jellegű általános exponenciális függvényből indult ki:

$$y(t) = Ke^{-ct}.$$

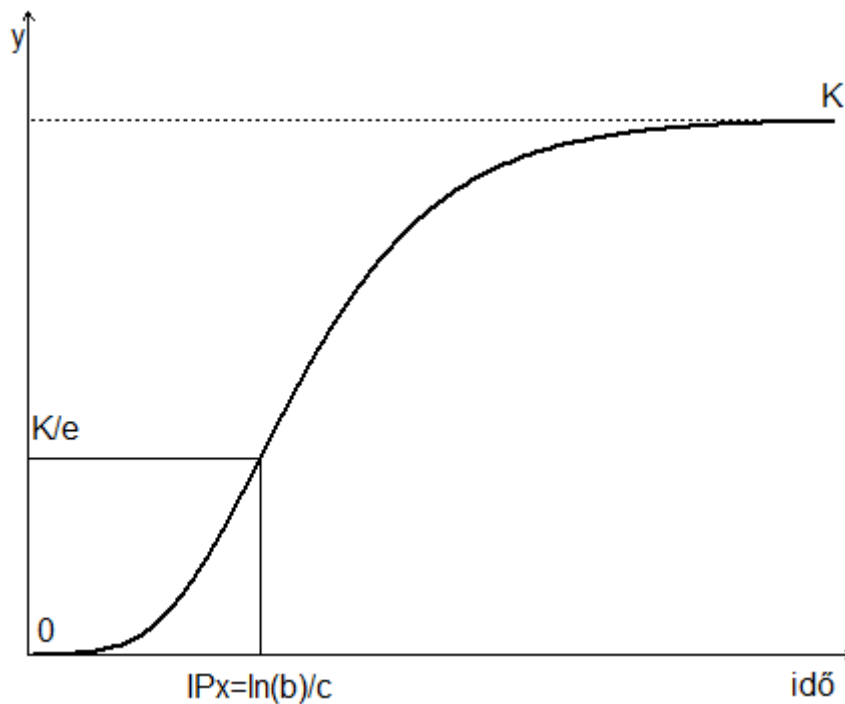
1825-ben írt tanulmányában azt vizsgálta, hogy a  $c$  paraméter változása akár exponenciális jellegű is lehet, így ennek megfelelően az alapfüggvényt az alábbi módon javította:

$$y(t) = Ke^{-be^{-ct}}$$

Az inflexiós pont koordinátái pedig:  $IP = \left(\frac{\ln b}{c}, \frac{K}{e}\right)$  alakban definiálhatóak.

Ez a forma tekinthető a nemzetközileg leginkább elterjedt és elfogadott Gompertz formának. A logisztikus függvényhez képest a fenti formula egy nem szimmetrikus S-görbét határoz meg, melynek az inflexiós pontja előrébb van és mintegy a telítődési szint harmadánál következik

be. A Gompertz-függvény a logisztikus függvénynél meredekebben emelkedik a korai, fejlődési szakaszban és így hamarabb éri el a telítettségi szint ( $K$ ) közeli állapotot.



12. ábra – Gompertz függvény

### Richards- és az általánosított Richards-féle függvény

Egy másik lehetséges megoldás a Richards által kidolgozott függvény, mely a Verhulst-féle logisztikus függvény kiegészítése egy negyedik,  $v$  paraméterrel, az inflexiós pontban aszimmetrikussá téve a görbét:

$$y(t) = \frac{K}{(1 + ve^{-c(t-m)})^{\frac{1}{v}}}$$

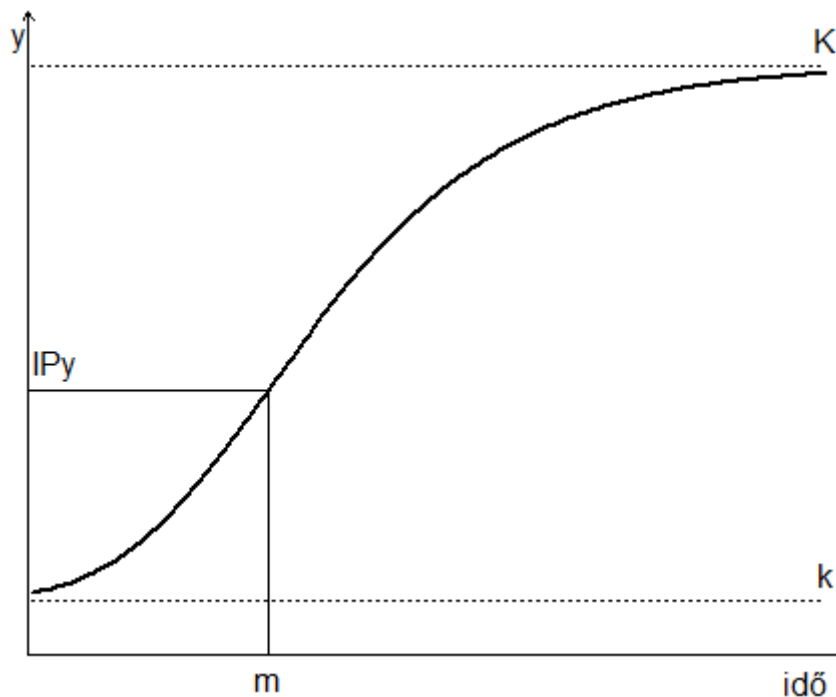
Az újonnan bevezetett  $v > 0$  paraméter szabályozza az inflexiós pontban felvett függvényértéket, azaz az inflexiós pont koordinátái:  $IP = \left(m, \frac{K}{(1+v)^{\frac{1}{v}}}\right)$ .  $c > 0$  továbbra is a növekedési sebességet jellemző, míg  $m > 0$  az inflexiós pont helyzetét definiáló paraméter.

Az általánosított Richards-féle logisztikus trendfüggvénynek a szakirodalom alapján jelenleg általánosan elfogadott formulájában kezelték már azt a kezdeti feltételezést is, hogy a függvény nem rendelkezik alsó aszimptotával. Ezzel egy olyan öt paraméterrel rendelkező függvényt definiáltak, mely egy olyan S-alakú görbét határoz meg, melynek van alsó korlátja ( $k$ ) és a szokásos telítődési szintje:

$$y(t) = k + \frac{K - k}{(1 + ve^{-c(t-m)})^{\frac{1}{v}}}$$

Az inflexiós pont koordinátái:  $IP = \left(m, k + \frac{K-k}{(1+v)^{\frac{1}{v}}}\right)$  alakban adhatóak meg. Ugyanakkor könnyen belátható, hogy a paraméterek megfelelő megválasztása mellett bizonyos növekedési függvények a Richards függvény speciális esetének tekinthetőek. Azaz valóban tekinthető ez a fajta függvény egy több esetet is magában foglaló általánosításnak.

- ha  $k = 0$  és  $v = 1$ , akkor az általánosított Richards-féle trendfüggvény logisztikus függvénné alakítható,
- ha  $k = 0$ ,  $m = 0$  és  $v \rightarrow 0$ , akkor a Richards-függvény Gompertz-függvénné alakul át,
- ha  $k = 0$ ,  $m = 0$  és  $v = -1$  a Richards-függvény Mitscherlich-trendfüggvénné alakul.



13. ábra – Richards függvény

A telítődési függvényeknek általában három jellegzetes, biológiai jelentéssel is bíró pontja van: az intenzív növekedés kezdete ( $\lambda$ ), a növekedés meredeksége ( $\mu$ ), valamint a telítődési érték ( $A$ ). A Richards-függvény inflexiós pontjában felvett függvényértéket a  $v$  paraméter értéke határozza meg. Az eredetileg és a korábbi definíciókban is alkalmazott függvényparamétereknek pusztán matematikai jelentése van, azokból közvetlen módon nem,

vagy csak nehezen határozhatóak meg az említett gyakorlati jelentőséggel is bíró paraméterek. Ezért Zwietering és mtsai (1990) a növekedési folyamat jellemzőit  $(\lambda, \mu, A, \nu)$  kifejezte az alapmodell paramétereinek függvényeként, majd visszahelyettesítette azokat az eredeti képletbe.

$$y_L = \frac{K}{1 + e^{b-ct}} \quad \rightarrow \quad \frac{A}{1 + e^{(2+(\lambda-t)^4\mu/A)}}$$

$$y_G = Ke^{-e^{(b-ct)}} \quad \rightarrow \quad Ae^{-e^{(1+(\lambda-t)\mu e/A)}}$$

$$y_R = K(1 + \nu e^{k(\tau-t)})^{-1/\nu} \quad \rightarrow \quad A(1 + \nu e^{(1+\nu)e^{\mu/A(1+\nu)(1+1/\nu)(\lambda-t)}})^{-1/\nu}$$

Az így kapott új formulákat alkalmazva már közvetlenül kinyerhetjük az illesztést követően a görbét, és így azon keresztül az azt jellemző paramétereket.

### 3.3.2 Görbeillesztés folyamata, módszerei

Az illesztési folyamat során a feladat az adott függvény paramétereinek meghatározása úgy, hogy adott értelemben (pl. legkisebb négyzetek elve) a lehető legjobb illeszkedést érjük el. Az illesztés során vizsgált függvények egy adott értelemben legjobb paramétereit keressük, ami matematikai értelemben egy többváltozós függvény lokális (lehetőség szerint globális) szélsőértékének a meghatározása, amely általában egy nemlineáris egyenletrendszer megoldására vezet az alábbihoz hasonló alakban:

$$\begin{aligned} f_1(x) &= 0 \\ f_2(x) &= 0 \\ &\vdots \\ f_m(x) &= 0 \end{aligned}$$

Amennyiben  $f(x) = (f_1(x), \dots, f_m(x))^T$ , akkor  $\|f(x)\|^2 = \sum_{i=1}^m (f_i(x))^2 \rightarrow \min$  szélsőérték feladat megoldását keressük. Direkt módon elérhető megoldás a legtöbb esetben nem létezik, numerikusan meghatározható közelítő megoldások kereshetőek, melyek jellemzője, hogy a megoldást véges sok lépésben csak adott hibahatár mellett képesek előállítani. Csak közelítő megoldások állíthatóak elő, amelyek azonban természetesen a gyakorlatban tökéletesen elegendőek. A megoldások jelentős része ún. lokális módszer, ami azt jelenti, hogy az iterációs folyamathoz előre szükséges definiálni a megfelelő kezdőértékeket. A módszerek általában – többek között az általunk ismertettek is – lokális szélsőértékek meghatározását teszik lehetővé. Három általánosan elterjedt eljárást mutatunk be erre röviden: a Gauss-Newton-, a Gradiens (lejtő módszer)- és a Levenberg-Marquardt-módszert (Veress, 2007). Mindegyik olyan iteratív eljárásként fogható fel, mely a megfelelően megválasztott kezdeti paraméter

értékekből kiindulva minden egyes lépésben egyre jobban közelíti az optimumot. Nem célunk a módszerek komoly elméleti háttérének részletes ismertetése, csupán a gyakorlati felhasználásra koncentrálunk.

### **Gauss-Newton-módszer**

Az egyváltozós függvények kapcsán jól ismert Newton-eljárás általánosítása többváltozós esetre. A módszer gyakorlatilag egy iteratív közelítő eljárás, mely az optimumnak mindig csak egy közelítését adja, illetve azt nem is feltétlenül éri el. Minden egyes iterációs lépésben azt várjuk, hogy a tényleges optimumnak egy jobb közelítését kapjuk. A módszerhez szükséges az optimum egy megfelelő kezdeti sejtése, azaz egy jól megválasztott kezdőérték az iterációhoz. Az eljárás az eredetileg megoldandó nemlineáris legkisebb négyzetes problémát – Taylor sorok alkalmazásával – visszavezeti a lineáris legkisebb négyzetes problémára. Így az egyes iterációs lépésekben a megfelelő módon választott  $x_0$  kezdőérték mellett már lineáris egyenletrendszereket kell megoldani.

Minden egyes iterációs lépésben az adott  $x_i$  közelítés mellett linearizálni akarjuk a minimalizálandó célfüggvényt az alábbi módon:

$$f(x) = f(x_i) + J_f(x_i)(x^* - x_i),$$

ahol  $J$  a Jacobi mátrix. Ezt felírva az alábbi alakban:

$$f(x_i) = A_i x - b_i \text{ alakban, ahol } A_i = J_f(x_i) \text{ és } b_i = J_f(x_i)x_i - f(x_i)$$

az eredeti nemlineáris probléma egy lineáris közelítését kapjuk, majd a közelítést felhasználva a többváltozós Gauss-Newton iterációs formula az alábbi már egyszerűsített módon írható fel:

$$x_{i+1} = x_i - \left( J_f(x_i)^T J_f(x_i) \right)^{-1} J_f(x_i)^T f(x_i)$$

A gyakorlatban általában a fenti egyenletben szereplő mátrix inverzet nem közvetlen invertálással számoljuk, hanem  $x_{i+1} = x_i + \delta_i$  helyettesítést alkalmazva egy újabb nemlineáris egyenletrendszert oldunk meg.

## Gradiens-módszer

Egy megfelelően megválasztott kezdeti paraméterezésből kiindulva az egyes paraméterek értékét iteratív módon úgy határozza meg a módszer, hogy veszi egy  $G : \mathbb{R}^m \rightarrow \mathbb{R}$  többváltozós, nemlineáris veszteség-függvény (pl. a legkisebb négyzetek elve alapján) gradiens vektorát.

$$\text{grad}(G(x)) = \left[ \frac{\partial G}{\partial x_1}(x); \dots; \frac{\partial G}{\partial x_m}(x) \right]^T$$

A gradiens vektor geometriailag a  $G$  függvény, mint az  $m + 1$  dimenziós  $\mathbb{R}^{m+1}$  térben értelmezett felület adott  $x$  pontjába húzható érintők közül a legmeredekebb irányába mutat. A módszer kiszámolja az aktuális paraméterezés mellett azt az értéket, amely meghatározza, hogy milyen irányba kell a paramétereket mozgatni, majd egy  $\alpha$  szorzótényezőt alkalmazva súlyozza a kapott értékeket. A javított paramétereket az előző lépés paraméterei és az előbbi súlyozott értékek különbsége határozza meg. Az egyes paramétereket a módszer egy-egy iterációs lépésben párhuzamosan állítja elő. Amennyiben a gradiens értéke már „elég” kicsi elértük a szélsőértéket adott pontosság mellett.

A  $G$  függvény görbületét, vagyis konvex, illetve konkáv voltát (s ezáltal a szélsőérték típusát tekintve) egy  $x^* \in \mathbb{R}^m$  pontban a Hesse mátrix mutatja meg, melynek elemeit a  $[H_G]_{ij} = G''_{x_i x_j}$ . A  $G$  függvénynek (lokális) minimuma van az  $x^* \in \mathbb{R}^m$  pontban ha  $\text{grad}(G(x^*)) = \bar{0}$  és  $H_G(x^*) > 0$ , azaz a Hesse mátrix pozitív definit.

Így a gradiens módszer algoritmusának alaplépései a következők adott  $x_0$  kezdőérték, veszteségfüggvény,  $\varepsilon$  hibakorlát és  $\delta$  lépésköz mellett:

1.  $i := 0$  a kezdő iterációs érték és  $x_i = x_0$ ,
2. számítsuk ki a veszteségfüggvény  $x_i$  pontbeli  $\text{grad}(G(x_i))$  vektorát,
3. ha a gradiens vektor már elég „kicsi”, azaz az  $\varepsilon$  hibahatáron belül van, akkor vége az eljárásnak, megtaláltuk a minimumot,
4. egyébként lépünk egyet a negatív gradiens irányába, azaz  $x_{i+1} = x_i - \text{grad}(G(x_i))\delta$ , ahol  $\delta$  egy előre definiált lépésköz érték,
5. növeljük eggyel az iterációs számlálót,  $i := i + 1$  és folytassuk a 2. ponttal.

### Levenberg-Marquardt-eljárás

Ez ötvözi az előző két módszert. Minden egyes iterációs lépésben egy  $\gamma$  paraméter értékének változtatásával gyakorlatilag a Gauss-Newton- és a Gradiens-eljárások között váltani tud. Mindig az az eljárás kerül végrehajtásra, amely az adott pillanatban kedvezőbbnek tűnik, ugyanis ha a közelítés gyorsan csökken, akkor kis  $\gamma$  értékekkel a Gauss-Newton felé, míg lassú csökkenés esetén nagy  $\gamma$  értékekkel a gradiens módszer felé terelődik az eljárás.

Formálisan az eljárás egy előre definiált  $p$  vektor által adott irány mentén keresi a minimumot. Így az eljárás elején meg kell adnunk egy kezdő  $p$  vektort, általában ez a  $p = (1; 1; \dots; 1)^T$  vektor. Minden egyes iterációban a  $p$  irányt próbáljuk javítani egy  $p + q$  vektorral, hogy

$f_i(p + q) = f(p) + J_f(q)$ , ahol  $J_f(q)$  a Jacobi determináns

linearizációs formulát alkalmazzuk. Tudjuk, hogy egy függvény ott lehet minimális, ahol a parciális derivált függvények értéke 0, így adódik  $q$ -ra, hogy

$$\left( J_{f_i}^T(x_i) J_{f_i}(x_i) \right) q = -J_{f_i}^T(x_i) f_i(x_i)$$

amiből  $q$  értéke egy egyszerű mátrix invertálás és szorzás után előállítható. A Levenberg-Marquardt-eljárás ezen egyenletet helyettesíti annak egy skálázott alakjával:

$(J^T(x_i)J(x_i) + \lambda_i I)q_i = -J^T(x_i)F(x_i)$ , ahol  $\lambda_i \in R$ ,  $I$  pedig az egységmátrix.

A  $\lambda$  skalárok bevezetésével sikerült elérni, hogy a módszer tetszőlegesen mozoghat a Gauss-Newton és a gradiens módszerek között.

Az illeszkedés jóságának mérésére ugyanazok a mutatók használhatóak, mint a többváltozós modellek esetében bemutatott lehetőségek, ugyanakkor érdemes megjegyezni, hogy a gyakorlatban (különösen az erdészeti elemzések kapcsán) a determinációs ( $R^2$ ) és korrigált determinációs együtthatót ( $\hat{R}^2$ ) szokták általában használni.

#### 3.3.3 Növekedési görbe illesztése - eredmények

Az egy inflexiós ponttal rendelkező növekedési görbéket számtalan esetben alkalmazzák erdészeti kutatásokban. Rennols (1995) Chapman-Richards függvényt használt a fák magassági növekedési adatainak függvényesítésére. Liao és mtsai (2003) négy függvényt, Mitscherlich, Logisztikus, Gompertz és Korf alkalmaztak a simafenyő éves magassági és átmérő irányú növekedési adataira, majd az illesztett görbe alapján különböző éves paramétereket származtattak, mint például éves abszolút, relatív és átlagos növekedés, miközben a növekedési folyamatot három részre osztották az illesztett görbe inflexiós pontja alapján. Deslauriers és

Morin (2005) balzsamfenyők sejtképzésének elemzéséhez Gompertz-függvényt használt. Sánchez-González és mtsai (2005) Spanyolországban domináns tölgy egyedek magassági és átmérő növekedési adataira illesztettek Lundqvist-Korf, McDill-Amateis és Richards görbéket. Azt vizsgálták, hogy a 3 lehetséges függvény közül melyik alkalmazható a legjobban a magasság, illetve az átmérő növekedés leírására. McCarthy és Weetman (2006) kor- és méreteloszlásokat vizsgáltak boreális erdőkben, és az újulat kor-átmérő, valamint kor-magasság viszonyát elforgatott Chapman-Richards görbével közelítették. Petritan és mtsai (2009) egy idős bükkös felújulási szintjének éves átmérő- és hajtáshossz-növekedését vizsgálták a megvilágítás függvényében logisztikus, a Michaelisz-Menten-, a Gompertz- és a Chapman-Richards-függvények felhasználásával.

Többen használtak a fák éven belüli növekedési adataira illesztett görbéket, illetve azokból származtatott egyéb adatokat. Thorp és mtsai (1995) avokádó fák növekedési adataira illesztettek logisztikus függvényt két egymást követő szezonban, összehasonlítva azokat egymással. Vizsgálták, hogy különböző időszakokban a teljes növekedés mekkora aránya realizálódott. Bouriaud és mtsai (2005) lucfenyő növekedésmenetére illesztettek Gompertz-függvényt, a kapott görbét az adott mérethez tartozó dátum pontos definiálására használták fel. Čufar és mtsai (2008) bükkök évgűrű-kialakulását, időbeli növekedését vizsgálták Gompertz-függvény illesztésével. Wipfler és mtsai (2009) bükk és erdei fenyő éven belüli növekedésmenetének leírására használtak egy speciálisan átalakított Weibull-függvényt. Deslauriers és mtsai (2009) nemesnyár fajták átmérő- és hossznövekedési adataira illesztettek Gompertz-függvényt, és ezt felhasználva számolták ki az átlagos éves növekedés mértékét. Seo és mtsai (2011) erdeifenyők növekedésmenetét elemezték Gompertz-függvénnyel, miközben a függvényt az adatok szűrésére is felhasználták. Jezik és mtsai (2011) szintén a Gompertz-függvényt alkalmazták bükkfák átmérő növekedésének jellemzésére. A függvény első deriváltjának segítségével vizsgálták a növekedés változásának sebességét. Mindkét esetben szoros illeszkedést állapítottak meg.

Saját kutatásinkban az éven belüli növekedési adatokra illesztett görbék vizsgálata kapcsán először arra kerestük a választ, hogy a korább bemutatott és a szakirodalom alapján is relevánsnak tekinthető három görbe (logisztikus, Gompertz és Richards) közül melyiket használjuk általánosan az általunk vizsgált területek és faegyedek esetében a további kutatásokhoz (Edelényi és mtsai., 2011b; 2012; Pödör és mtsai., 2014a).

A görbe kiválasztásához szükséges vizsgálatokat az Erdészeti Tudományos Intézet által működtetett intenzív erdészeti monitoring mátrai bükkösének 2002-2011 közötti adatait



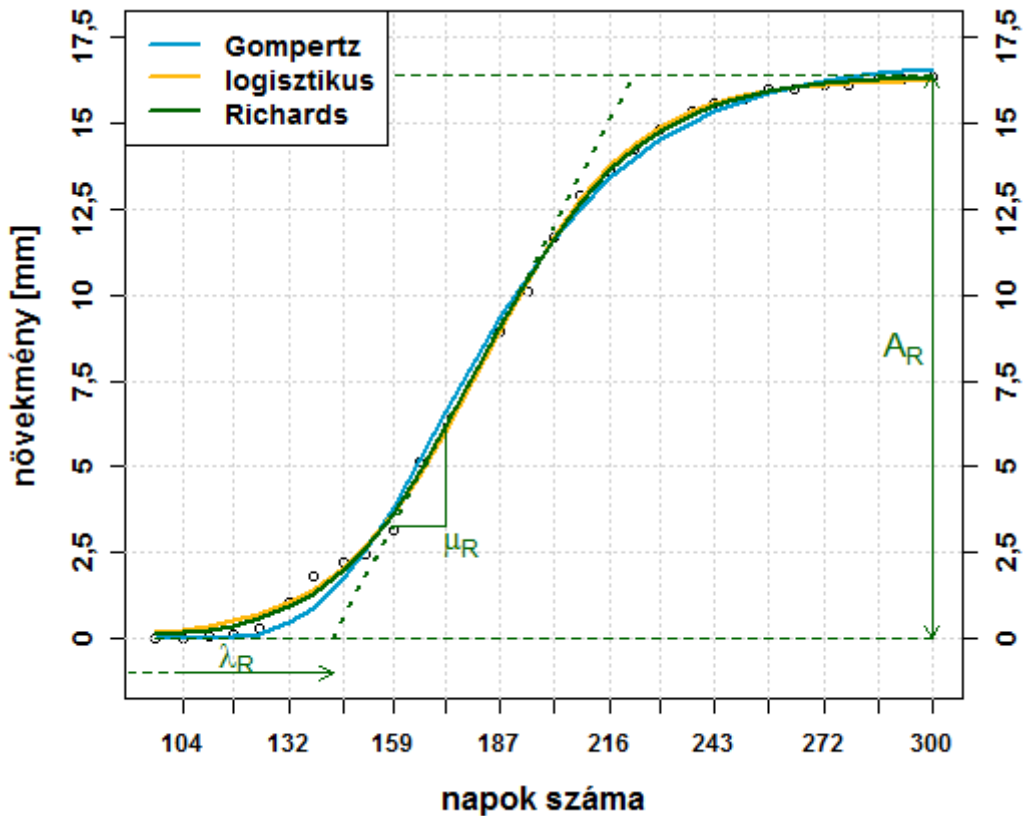
felhasználva végeztük el Manninger Miklós szakmai iránymutatásai mellett. A 100 éves bükkös kerületmérő szalaggal felszerelt kimagasló vagy uralkodó szintű fái közül a hat legjobb növekedést mutató fát választottuk ki. A heti gyakoriságú mérések eszköze a mellmagasságban felszerelt és Járó Zoltán által módosított Liming-féle szalag volt. Az éves adatsorból a vegetációs időszakot magába foglaló április 1. és október 31. közötti időszak adatait vontuk be az elemzésbe, hogy kizárólag a növekedéssel, a szervesanyag-képzéssel összefüggő méretváltozásokat vizsgáljuk.

A bükk fafaj mellett azért döntöttünk, mert a legtöbb fafajtól eltérően éves növekedésmenete elég egyenletes, nincsenek benne nagy ugrások, törések, tehát függvény illesztésére jó lehetőséget teremt.

A mintaadatokra az illesztést az R szoftver (*R Development Core Team, 2012*) grofit csomagjával végeztük el. A grofit csomag magába foglalja a fenti függvényeket, a nemlineáris legkisebb négyzetes (Gauss-Newton) illesztést alkalmazza, és a kezdeti értékeket automatikusan határozza meg (Kahm és mtsai, 2010).

A modellek rangsorolása az Akaike módosított információs kritériumon alapuló súly ( $w$ ) szerint történt. Az R szoftverben a  $w$  értékét az  $AIC_{modavg}$  bővítménnyel számoltuk (Mazerolle, 2012). Az illeszkedés jóságát a kiválasztott modellek esetében a korrigált determinációs együtthatóval ( $\hat{R}^2$ ) jellemeztük, ami a  $w$ -hez hasonlóan figyelembe veszi a paraméterek számát is.

A kapott paraméterértékek elemzésekor megállapítható, hogy a növekedés kezdetére pár nap különbség jön ki: a Gompertz-függvény szerint indul be a növekedés a legkorábban ( $\lambda = 141$ ), míg a logisztikus függvény szerint a legkésőbb ( $\lambda = 145$ ). A növekedés mértékét jelző paraméter ( $A$ ) esetében a legnagyobb értéket a Gompertz-, a legkisebbet a logisztikus függvény adja. A legnagyobb és a legkisebb érték között a különbség mindössze 0,57 mm, ami nem túl jelentős, alig 4%-nyi eltérés. A növekedés meredeksége ( $\mu$ ) mindhárom modell esetében gyakorlatilag azonosnak tekinthető.



14. ábra – Görbeillesztés, példa

A hat fa összes növekedésmentére vonatkozó számítások eredményeit a 16-18. táblázatokban foglaljuk össze. A 16. táblázatban évek (2002-2011), a 17. táblázatban fák szerinti bontásban adjuk meg az adatokat, míg a 18. táblázat összevontan tartalmazza az összes esetre vonatkozó eredményeket.

A legjobbnak nevezett modell kiválasztása – ahogyan azt már korábban említettük – az Akaike módosított információs kritériumon alapuló súly ( $w$ ) alapján történt. A modellek jóságát és illeszkedését jellemző paraméterek ( $w$  és  $\hat{R}^2$ ) átlagos értékén az adott modell teljes mintaszámra (hat fára vagy 10 évre) vonatkozó értékeinek átlagát értjük, tehát nem csak azoknak az átlagát, mikor az adott modell a legjobb illeszkedést adta. Ezek alapján vizsgálható, hogy:

- az egyes években a három modell közül melyik hány faegyednél mutatta a legjobb illeszkedést (évek szerinti bontás)
- az egyes faegyedek esetében a 10 év alatt melyik görbe hány alkalommal bizonyult jobbnak a többinél (fák szerinti bontás).

A táblázatokban a legjobb modell sorban látható, hogy adott évben/adott fára hány alkalommal bizonyult a legjobbnak az aktuális görbe.

16. táblázat – Modellillesztési eredmények évek szerinti bontásban

évek alapján (n=6)		2002	2003	2004	2005	2006	2007	2008*	2009	2010	2011
<b>G</b>	<b>legjobb modell</b>	0	3	0	2	1	3	3	3	4	4
	<i>w</i> átlag	0,06	0,53	0,26	0,34	0,19	0,44	0,37	0,45	0,49	0,56
	$\hat{R}^2$ átlag	0,9958	0,9956	0,9978	0,9979	0,9918	0,9970	0,9986	0,9966	0,9978	0,9978
<b>L</b>	<b>legjobb modell</b>	0	0	0	0	3	0	0	0	0	0
	<i>w</i> átlag	0,00	0,00	0,00	0,01	0,36	0,03	0,00	0,04	0,04	0,00
	$\hat{R}^2$ átlag	0,9880	0,9899	0,9967	0,9967	0,9929	0,9953	0,9953	0,9942	0,9959	0,9960
<b>R</b>	<b>legjobb modell</b>	6	3	6	4	2	3	4	3	2	2
	<i>w</i> átlag	0,94	0,47	0,74	0,65	0,45	0,53	0,63	0,51	0,47	0,44
	$\hat{R}^2$ átlag	0,9968	0,9958	0,9983	0,9983	0,9933	0,9973	0,9987	0,9968	0,9982	0,9980

\* 2008-ban a 26-os fára a Gompertz és a Richards azonos súlyú

A 16. táblázat adatai alapján megállapítható, hogy 2002-ben mind a hat fánál a Richards-függvény volt a legjobb modell. Ugyanakkor 2009-ben 3-3 fára a Gompertz-, illetve a Richards-függvény adta a legjobb eredményeket, majd az azt követő két évben már a Gompertz-függvény bizonyult a legjobbnak. A vizsgált 10 év vonatkozásában az tapasztaltuk, hogy kezdetben a Richards-függvény volt a legjobb, majd ez folyamatosan eltolódott a Gompertz-függvény irányába. Látható az is, hogy csupán egyetlen olyan év volt (2006), amikor a logisztikus függvény jobb eredményt adott a másik két modellhez képest. A vizsgált 10 évben négy alkalommal a Richards-, két esetben a Gompertz-, egyszer a logisztikus függvény mutatta a legjobb illeszkedést a fák többségére. Három évben pedig a Gompertz- és Richards-függvény is azonos számú (3-3) fára bizonyult a legjobbnak. Az Akaike-súly átlagát figyelembe véve az évek többségében (7 esetben) a Richards-függvény volt a legjobb, míg a többi esetben a Gompertz. A logisztikus modell egyik évben sem volt jobb, mint a másik kettő.

A fák szerinti bontásban (17. táblázat) a Richards négyszer (4, 19, 22 és 26-os fák), a Gompertz két esetben (2, 21-es fák) tekinthető a legjobbnak. Az Akaike-súlyokból képzett átlagok szempontjából egyértelműen a Richards-függvény adja a legjobb eredményt.

17. táblázat – Modellillesztési eredmények fák szerinti bontásban

fák alapján ( $n=10$ )		2	4	19	21	22	26*
<b>G</b>	<b>legjobb modell</b>	6	2	4	6	2	3
	$w$ átlag	0,41	0,26	0,48	0,50	0,16	0,41
	$\hat{R}^2$ átlag	0,9973	0,9946	0,9970	0,9972	0,9974	0,9967
<b>L</b>	<b>legjobb modell</b>	1	1	0	0	1	0
	$w$ átlag	0,07	0,07	0,00	0,01	0,11	0,03
	$\hat{R}^2$ átlag	0,9950	0,9926	0,9919	0,9935	0,9975	0,9940
<b>R</b>	<b>legjobb modell</b>	3	7	6	4	7	8
	$w$ átlag	0,52	0,67	0,52	0,50	0,73	0,56
	$\hat{R}^2$ átlag	0,9977	0,9951	0,9973	0,9974	0,9985	0,9969

\* 2008-ban a 26-os fára a Gompertz és a Richards azonos súlyú

Az elvégzett 60 illesztés összevont eredményei alapján (18. táblázat) megfogalmazható, hogy a három vizsgált függvény közül a logisztikus adta a leggyengébb eredményeket, a Gompertz-23, míg a Richards-függvény 35 esetben bizonyult a legjobbnak. Ugyanakkor az illeszkedés jóságát tekintve a logisztikus függvény is alkalmas lehet az ilyen jellegű feladatra.

18. táblázat – Összevont modellillesztési eredmények

összes együtt ( $n=60$ ) *	<b>G</b>	<b>L</b>	<b>R</b>
<b>legjobb modell</b>	23	3	35
$w$ átlag	0,37	0,05	0,58
$\hat{R}^2$ átlag	0,9967	0,9941	0,9971

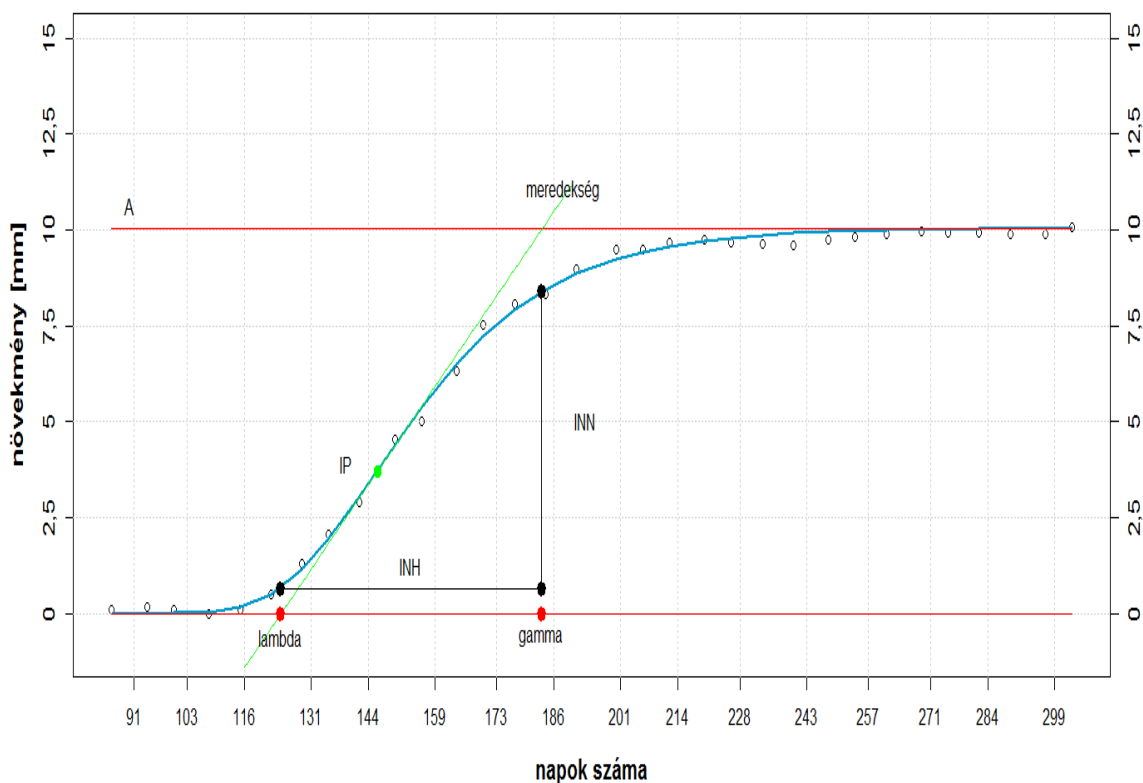
\* 2008-ban a 26-os fára a Gompertz és a Richards azonos súlyú

Megállapítottuk, hogy a korrigált determinációs együttható alapján a három módszer között nincs számottevő különbség, bármelyik alkalmas lehet az illesztési feladatok elvégzésére. Ugyanakkor az Akaike információs (AIC) kritériumon alapuló vizsgálatok azt mutatták, hogy a logisztikus függvény minden esetben alulmaradt a másik két görbével szemben, melyek egymáshoz képest hasonló eredményeket mutattak. A további elemzések elvégzéséhez a Gompertz görbe alkalmazása mellett döntöttünk (Pödör és mtsai., 2014a) egyrészt figyelembe véve a korábbi vizsgálataink során kapott eredményeket és az áttekintett szakirodalmat. Másrészt a Gompertz görbe három paramétere már megfelelő rugalmasságot biztosít vizsgálatainkhoz, és ezek a paraméterek jól értelmezhetőek és társítható hozzájuk megfelelő biológiai jelentés is (Zwietering és mtsai, 1990).

A függvények illesztésével a célunk egyrészt, hogy az egyébként diszkrét adatsorokat folytonossá tegyük. Ez lehetőséget nyújt arra, hogy a hiányzó értékeket pótolhassunk, illetve olyan időpontokra is generálhassunk növekmény adatot, amikor valódi mérési adat nem áll rendelkezésre. Az illesztett görbe alkalmas az adatsorokban fellépő zajok szűrésére, kezelésére és megkönnyíti a kiugró adatok ellenőrzését, minősítését, illetve ha szükséges, akkor javítását. Másrészt az illesztett görbék alkalmasak a növekedést jellemző egyéb paraméterek objektív módon történő definiálására. Ezek között vannak olyanok, melyek az illesztett görbe paraméterei alapján közvetlen módon kaphatók meg, mint például az intenzív növekedés kezdete ( $\lambda$ ) napokban megadva, a növekedés meredeksége ( $\mu$ ) és a növekedés mértéke ( $A$ ).

Ugyanakkor ezen paraméterek és a görbék felhasználásával további jellemzők is könnyen előállíthatóak. Ilyen lehet például a ténylegesen növekedésnek tekinthető folyamat vége, amit  $\gamma = \frac{A}{\mu}$  értékkel tudunk definiálni, felhasználva, hogy az illesztett görbe esetében mind az intenzív növekedés kezdete, mind annak vége az inflexiós pontba húzott érintő egyenes által definiált. Továbbá megadható az intenzívnek tekinthető növekedés hossza  $INH$ , ami  $INH = \gamma - \lambda$  értékkel definiált. Utóbbi két adat szintén napokban van megadva. A Gompertz görbe tulajdonságait felhasználva könnyen előállíthatóak az inflexiós pont koordinátái, hiszen ennek helye az eredeti Gompertz függvény definíció esetében  $x = \frac{b}{c}$ , ahol  $c = \frac{\mu e}{K}$  és  $b = \frac{\mu e}{K} \lambda + 1$  módon határozható meg, az inflexiós pont értéke pedig az  $A$  paraméter által egyértelműen definiált  $\frac{A}{e}$  értékkel. Az inflexiós pont helye napokban, értéke a növekedésnek megfelelő mértékegységgel adott. Meghatározható a  $\lambda$  és  $\gamma$  pontok által definiált időszakhoz köthető intenzív növekedés nagysága:  $INN = y_G(\gamma) - y_G(\lambda)$ .

## Növekedési paraméterek



15. ábra – Származtatott növekedési paraméterek

Felhasználva a kapott jellemzőket további paraméterek képezhetők, mint például az abszolút adatok (napok száma, mm) alapján képzett különböző arányszámok: intenzív növekedés aránya a teljes növekedéshez képest, intenzív növekedés hosszának aránya a teljes növekedési időszakhoz képest, stb.

Az így előállított paraméterek előnye, hogy a pusztán matematikai szempontok figyelembe vétele mellett illesztett görbékből teljesen egységes és objektív módon állíthatóak elő. Természetes, hogy ez a folyamat legfeljebb félautomatizáltnak tekinthető abban az értelemben, hogy a kapott görbéket, eredményeket erdészeti szakmai szempontból ellenőrizni kell az elemzésekbe történő bevonás előtt. Ugyanakkor ezen teljesen objektív módon definiált növekedést jellemző paraméterek tovább szélesíthetik az elemzések körét és segíthetnek a növekedés-klíma kapcsolatok feltárásában.

## 4 Alkalmazások

A dolgozatban bemutatott CReMIT módszer és az azt magában foglaló elemzési folyamat, illetve annak előkészítő lépései sok – részben a dolgozatban már érintett – erdészeti feladat kapcsán kerültek felhasználásra. Az erdészeti jellegű eredmények bemutatása nélkül rövid áttekintést adunk ezekről a feladatokról, azok céljairól, a felhasznált adatokról és a konkrét elemzési folyamatról.

A dolgozatban bemutatott elemző folyamat és a CReMIT módszer kidolgozását alapvetően a fanövekedés és a klimatikus paraméterek közötti kapcsolatok vizsgálatának kiszélesítése, mélyítése vetette fel. Ennek során több együttműködés keretében végeztünk és végzünk jelenleg is vizsgálatokat. Ezek mindegyikében a felhasznált meteorológiai adatok havi csapadékösszegek és középhőmérséklet átlagok voltak, az alkalmazott összefüggés vizsgálati eljárás pedig szignifikancia vizsgálattal egybekötött egy- és többváltozós korrelációelemzés.

Manninger Miklóssal az Intenzív Monitoring két területe adatainak vizsgálatával foglalkozunk. Az egyik területen ez kilenc, a másikon tíz egyed tizenöt évnyi éven belüli növekedési adatsorainak vizsgálatát jelenti. Az Intenzív Monitoring területeken végzett gyakori (heti, kétheti) mérések lehetővé teszik, hogy az éves növekedés mértékén túl további, a növekedést jellemző paraméterek is előállíthatóak legyenek, mint például az intenzív növekedés mértéke (mm), az intenzív és teljes növekedés aránya, a teljes és intenzív növekedés hossza (nap) és ezek aránya, a teljes és intenzív növekedés kezdete, vége (nap) és ezek különböző arányai. A paraméterek értékeinek definiálását a mért adatok alapján Manninger Miklós végezte el. A CReMIT módszer alkalmazásával vizsgáltuk ezen éves növekedési jellemzők kapcsolatát az előző év áprilisától adott év augusztusáig bezárólag képzett különböző időszaki környezeti jellemzőkkel (Edelényi és mtsai., 2011c; Manninger és mtsai., 2011b).

Ezekben a vizsgálatokban alkalmazott paraméterek előállítása a mérési adatok alapján a szakember kijelölései alapján történt. Amennyiben a diszkrét növekedési adatsorokra megfelelő növekedési függvény illeszthető, akkor vizsgálni lehet ezen folyamat legalább részleges automatizálásnak lehetőségét és a paraméterek objektív, bár pusztán matematikai szempontok szerinti előállíthatóságát. Természetesen a kapott eredmények szakmai ellenőrzése ekkor is fontos és szükséges feladat. Ez irányba mindkét fent említett terület adatait felhasználva történtek vizsgálatok, ahol a felhasznált növekedési adatsorokból kivettük a gyenge növekedésű egyedeket. Első lépésben a vizsgálatok célja annak meghatározása volt, hogy melyik növekedési görbe lehet alkalmas az illesztési feladatokra. Az illesztésekkel kapcsolatos

vizsgálatok és bizonyos eredmények (Edelényi és mtsai., 2011b; 2012; Pödör és mtsai., 2014a) a dolgozat 3.3. fejezetében jelentek meg.

Ugyancsak a növekedés és a klíma komponensek közötti kapcsolatokat vizsgáltuk Führer Ernővel egy másik terület adatain, mellyel kapcsolatos eredmények a dolgozat 2.3. fejezetében jelentek meg (Führer és mtsai., 2012). Az éves növekedési adatokat vetettük össze az adott, az előző és a kettővel korábbi év meteorológiai adataival legfeljebb négy hónapnyi szélességű meteorológiai adatokat képezve a CReMIT módszerrel. Ebben az esetben egy és többváltozós vizsgálatok egyaránt történtek. Az egyváltozós vizsgálatok segítségével meghatároztuk a releváns időszakokat mind a csapadék, mind a hőmérséklet adatok tekintetében. Majd részben ezek felhasználásával a többváltozós elemzésekben az éves növekedés modellezése volt a cél a megfelelő klimatikus jellemzők kiválasztása mellett.

Mindkét fent említett esetben éven belüli növekedési adatokról, illetve azokból képzett éves növekedési adatokról és egyéb paraméterek vizsgálatáról volt szó. A növekedési adatok azonban sok esetben évgyűrű adatok alapján állnak rendelkezésre, ahol az éves növekedés mértéke a vizsgált alapparaméter. Ilyen jellegű adatokat vizsgáltunk az ERTI-ből Szabados Ildikóval, illetve a NÉBIH-ből, Kolozs Lászlóval és Solti Györggyel végzett közös munkáinkban. Több mint 50 faegyed évgyűrű adatokból származtatott növekedési indexét vetettük össze előző év áprilisától adott év októberéig terjedő időszak 1-2-3 havi időszaki csapadék és hőmérséklet adataival az elemző folyamatot felhasználva. Jelenleg is folyamatban vannak ilyen jellegű vizsgálatok 107 faegyedre felhasználva az elemzési folyamat előkészítő lépésének görbeillesztési funkcióját. A megkapott évgyűrű adatokra Hegershoff függvényt illesztve képezzük a növekedési indexet és ezt, mint függő változót vonjuk a fentiekhez hasonló vizsgálatok alá. A havi környezeti paraméterek mellett speciális erdészeti indexeket (FAI, PADI) is használunk, mint független változókat.

A CReMIT módszer kialakítása során fontos szempont volt az univerzalitásra való törekvés, azaz a módszer ne csak az alapproblémának tekinthető fanövekedés, klíma, hanem egyéb függő és akár független paraméterekre is alkalmazható legyen. Erre mutat példát a dolgozat 2.3. fejezetében már bemutatott nagylepke fogási adatok vizsgálata (Csóka és mtsai., 2012a; 2012b; 2013; Pödör és mtsai., 2013a; 2013b). A fogási adatokat Csóka Györggyel közös munkáinkban dolgoztuk, dolgozzuk fel. Az elvégzett egy- és többváltozós elemzések célja a releváns csapadék és hőmérsékleti időszakok meghatározása volt. 24 fogási hely, több mint 20 faj több mint 50 éves adatsorait vontuk elemzés alá. Az éves összesített fogási adatoknak adott és előző év környezeti jellemzőktől való függését vizsgáltuk fogási helyenként és fajonként legfeljebb



6 hónapnyi szélességű ablakokat képezve a CReMIT modulban. Az eredmények könnyebb értelmezhetősége miatt a kapott szignifikáns összefüggéseket egy-egy fajra fogási helyenként összesítettük és azokat az időszaki adatokat tekintettük fontosnak, ahol a fogási helyek legalább adott százalékában szignifikáns kapcsolat adódott.

A dolgozatban csak említésre került, azonban komoly vizsgálatok folytak a fák egészségi állapota (levélvesztés, koronaelhalás) és mortalitás adatok vonatkozásában. Az együttműködő partner a NÉBIH (Kolozs László, Solti György) volt. A feladat ebben az esetben az éves és egyed szinten rendelkezésre álló levélvesztés, koronaelhalás és mortalitás adatoknak az erdőszetben alkalmazott klíma indexekkel való összefüggés vizsgálata volt. Ez a FAI és PADI indexek mellett további 32 index vizsgálatát jelentette. Vizsgálat alá vonásra került adott év, előző, az azt megelőző évek és ezek kombinációnak hatása a függő változókra. A nagyságrendileg 33000 faegyed egyedi szinten történő elemzése nehezen áttekinthető méretű kimenetet adott volna, ezért a szakemberek által kijelölt 13 fafajcsoportra bontva történtek meg az említett elemzések.

Az erdőszeti mérések egy speciális csoportját alkotják a korszaki növedék adatok. Ez azt jelenti, hogy a növekedési adatok nem éves, hanem különböző hosszúságú időszaki bontásban állnak rendelkezésre. Az alapfeladat ebben az esetben is a havi szinten rendelkezésre álló környezeti paraméterek és a növekedés kapcsolatának vizsgálata. A rendelkezésre álló növekedési adatok felbontásából fakadóan itt némileg eltérő technikák alkalmazására lesz szükség. Ezek vizsgálata és a megfelelő elemzési technikák kidolgozása jelenleg is folyamatban van.

## **5 Összefoglalás**

Világunkra jellemző, hogy különösen a számítógépek, informatika rohamos fejlődése mellett jelentős méretű adathalmazok generálódnak. Ezek önmagukban, hasznosítás nélkül azonban csak haszontalan adattemetőknek tekinthetőek. Fontos, hogy a puszta adatokból kinyerjük a hasznos információkat, a tudást majd ezt fel is használjuk például a különböző döntéseink meghozatala során. A döntéstámogató rendszerek célja éppen ez: az adott szakterület adatait feldolgozva, azokból az értékes összefüggéseket, információkat kinyerve tudományosan megalapozott és alátámasztott döntéseket hozhassunk. Nem kivétel ez alól természetesen az erdőszet és a faipar sem. Ha csak a jelen dolgozat egyik fontos kérdését jelentő fanövekmény-klíma kapcsolatok problémájára gondolunk is, látható, hogy sok helyen folynak ilyen jellegű mérések (Intenzív monitoring területeken év közbeni mérések, évgyűri adatok felvételezése,

korszakos növekmény adatok, stb.) Ezek vizsgálata alapvetően idősorok elemzését, a köztük fennálló kapcsolatok feltárását jelenti.

## **5.1 Következtetések**

A dolgozatban széleskörű szakirodalmi áttekintés alapján feltérképezésre került azon jellemző statisztikai módszerek köre, melyeket a fák növekménye és a klimatikus paraméterek közti kapcsolatok vizsgálatában alkalmaznak. Az elemzési módszereket áttekintve, azok hátrányos tulajdonságait figyelembe véve és előnyös tulajdonságaikat felhasználva kidolgozásra került egy olyan szisztematikus transzformációs módszer (CReMIT) periodikus idősorokra, mely az alap idősorból létrehozott összes lehetséges származtatott adatsor létrehozásával és elemzésbe történő bevonásával szélesíti ki a vizsgálatok körét. Ugyanakkor a létrehozását megalapozó módszerek majorációjának tekinthető abban az értelemben, hogy azokat magában foglalja, amellett, hogy ki is bővíti. A módszer több erdészeti jellegű feladat kapcsán alkalmazásra került már; két különböző példát mutattunk ezzel kapcsolatban illusztrálendő az alapló módszer működését és alkalmazhatóságát. Az alap CReMIT több változóra, akár a függő, akár több magyarázó változóra is kiterjeszhető. Bemutatásra került ennek elméleti háttere és gyakorlati alkalmazhatósága fanövekmény és lepkefogási adatok kapcsán. A két viszonylag eltérő példa azt is hivatott szemléltetni, hogy ugyan a módszer alapvetően fanövekmény adatok vizsgálata kapcsán került kifejlesztésre, azonban bármely más jellegű idősor kapcsán is alkalmazható akár az erdészetben, faiparban, akár azon kívül is.

Az elemzési feladatok egységes kezelhetősége céljából kialakított elemző folyamat központi eleme a CReMIT modul, azonban az előkészítő modulon keresztül ennek szerves részét képezi két speciális előkészítő technika, a töréspont keresés, valamint a függvényillesztés. Vizsgáltuk annak fontosságát, hogy az idősorok közti kapcsolatok nem mindig tekinthetőek időben stabilnak. Ez különösen igaz lehet a meteorológiai paramétereket (csapadék, hőmérséklet) használó kutatásokban, hiszen sokan foglalkoznak az ezekben tapasztalható komoly változásokkal. Ennek fényében fontos vizsgálni, hogy ezekben az adatsorokban statisztikailag is kimutatható-e változás, erre lehet alkalmas a töréspont keresés módszere. A feltárt töréspont(ok) mentén érdemes lehet az adatsorokat és így az elvégzendő vizsgálatokat is megbontani és külön-külön is elvégezni. Ismertetésre kerülnek a töréspont keresésének lehetséges módszerei és annak jelentősége egy konkrét fanövekmény-klíma elemzési feladat kapcsán, bemutattva, hogy a két részidőszak közt milyen komoly eltérések lehetnek a kimutatható összefüggések tekintetében a CReMIT módszer alkalmazásával.

A növekmény adatok lehetnek heti-kétheti rendszerességgel mért éven belüli növekmény adatok is. A nyers adatpontok alapján sokszor nehéz feladat az éves növekedés mértékén túlmutató, egyéb növekedési jellemzők objektív meghatározása, illetve két különböző mérési időpont közti növekmény adatok kezelése, definiálása. A megfelelő növekedési görbe kiválasztásával és illesztésével megfelelő választ adhatunk erre a kihívásra. Megmutattuk, hogy vizsgálataink alapján több lehetséges alternatíva is szóba jöhet az illesztendő növekedési görbe kapcsán, illetve mi miért a Gompertz görbét alkalmaztuk. Az alkalmazott görbével folytonossá tehetőek az egyébként diszkrét adatsorok és az illesztett görbék, illetve azok paramétereinek felhasználásával objektív módon állíthatóak elő a növekedést jellemző egyéb paraméterek. Utóbbiak elemzésbe történő bevonásával szélesíthető az elvégezhető vizsgálatok köre, különösen ha a dolgozatban bemutatott transzformációs modul alkalmazása mellett használjuk fel ezeket az objektív paramétereket.

## **5.2 Továbblépési lehetőségek**

Fontos hangsúlyozni, hogy a kialakított elemzési folyamat és az abban alkalmazott technikák alapvető célja jelen szakaszban elsősorban az erdészeti terület szakembereinek támogatása a számukra értékes szakmai eredmények előállításában. A bemutatott alap CReMIT módszernek már a dolgozatban is bemutatásra került néhány kiterjesztési lehetősége a több változóra történő alkalmazása kapcsán, miközben a transzformációs függvények körének bővítése egy komoly továbblépési pontot jelent. Az eddig elvégzett és bemutatott elemzési feladatok kapcsán az alkalmazott  $TR$  függvény hőmérséklet adatok esetén szimpla átlag, csapadék adatok esetében pedig összeg volt. A módszer kialakítás során szem előtt tartott univerzalitás lehetőséget teremt arra, hogy akár egyéb tudományterületeken, akár ipari környezetben is ki lehessen használni az eljárás nyújtotta előnyöket. Ezek az új területek részben ki is jelölhetik a továbblépési irányokat, melyek közül három már most is látszik.

Az egyik továbblépés irányt a függvények körének ezen elemi függvényeken túlmutató kiterjesztése jelenti. Ez lehet speciális, nemlineáris függvények alkalmazása, vagy valamilyen súlyfüggvényé (például exponenciális, hogy az időben régebbi elemek hatása exponenciálisan csökkenjen, vagy akár egyéb, speciális szakmai szempontokat figyelembe vevő súlyokat is definiálhatunk). A súlytényezők alkalmazásának egy speciális esete lehet az ablakok elemeire egyszerű bináris súlyok alkalmazása, ami lehetővé teszi nem folytonos ablakok definiálását is a módszerben. Ez utat nyithat az erdészetben is gyakran alkalmazott indexek (FAI, Führer és mtsai, 2011 és PADI, Pálfai és Herceg, 2011) irányába, illetve magában foglalhatja az indexek komponenseinek, illetve magának az egész indexnek a képzési lehetőségét is. További irányt

jelöl ki a transzformációs függvények mellett alkalmazott feltétel függvények bevonása, ami lehetővé teszi, hogy adott ablakban szereplő elemek közül csak azokra alkalmazzuk a kijelölt *TR* transzformációs függvényt, melyek a feltétel(eke)t kielégítik.

Az elemző modul kapcsán rengeteg továbblépési, fejlesztési pont látszik akár a többváltozós elemzések körének bővítése, akár az alkalmazott elemzési technikák nemlineáris irányba történő fejlesztés kapcsán. Ehhez jó alapot nyújtanak a különböző adatbányászati technikák, mint például a neurális hálók alkalmazásának vizsgálata.

Célunk, hogy folytassuk az eddigi együttműködések, segíthessük az értékes szakmai eredmények előállítását a módszer fejlesztésével és felhasználásával. Ugyanakkor a módszer ipari körülmények között történő alkalmazhatóságának feltételei is látszódnak, hiszen egy gyártási folyamatra tekinthetünk úgy is, mint olyan egymást követő lépések sorozatára, ahol az előző lépések bármelyike – akár külön-külön, akár együttesen – hatással lehet az adott lépésre.

## 6 Tézisek

Az alapadatok periodicitásának felhasználásával kidolgozásra került a CReMIT (Cyclic Reverse Moving Interval Techniques) módszer (Edelényi és mtsai., 2011a; 2011c; Pödör és Jereb, 2013c; Pödör és mtsai., 2014b), amely egy független változó idősoron értelmezve, szabadon rögzíthető paraméterhatárok között különböző szélességű és eltolású ablakokat állít elő (2.1.1. fejezet).

Az ablakok felett olyan transzformált idősorok értelmezhetőek, amelyek alkalmasak a kapcsolatok korábbinál szélesebb körű, szisztematikus feltárására (Edelényi és mtsai., 2011a, 2011c; Pödör és Jereb, 2013c; Pödör és mtsai., 2014b) (2.1.2. fejezet).

Az eljárás egy magasabb absztrakciós szintre emeli az evolúciós és mozgó intervallumos technikák nyújtotta lehetőségeket (Pödör és mtsai., 2014b) (2.1.3. fejezet).

Kidolgozásra és bemutatásra került a CReMIT alapszámítás több változóra történő kiterjesztése. A kiterjesztés magában foglalja az alapszámításnak akár a függő, akár több független paraméterre történő alkalmazását, még inkább kiszélesítve az elemzésbe bevonható változók körét és ezen keresztül az elemzési lehetőségeket (Pödör és mtsai., 2014b) (2.2. fejezet).

A CReMIT módszer alkalmazásával egy általános elemzési folyamat egy saját keretrendszerbe foglalva került megvalósításra. Az adatelőkészítő, CReMIT és elemző modulokból álló rendszer lehetővé teszi a teljes elemzési folyamat végrehajtását. (Edelényi és mtsai., 2011a; 2011c; Pödör és Jereb, 2013c; Pödör és mtsai., 2014b) (3.1. fejezet).

Az előkészítő modul részét képezi az a töréspont keresési funkció, amely módot ad arra, hogy a bemeneti adatsorokban statisztikailag igazolható töréspontok kerüljenek meghatározásra, és ez a tény figyelembe vehető az összefüggés vizsgálatokban (3.2. fejezet).

Az előkészítő modulnak ugyancsak részét képezi a növekedési adatokra illeszthető függvények. Mintaadatokon a logisztikus, Gompertz és Richards függvények illeszkedése került összehasonlításra, és megállapítható volt, hogy az illesztések eredményei és praktikus szempontok figyelembevétele mellett a Gompertz függvény a legkedvezőbb a növekedési folyamat leírására (Edelényi és mtsai., 2011b, 2012; Pödör és mtsai., 2014a) (3.3. fejezet).

## **7 Köszönetnyilvánítás**

Köszönetemet szeretném kifejezni elsősorban témavezetőmnek, Dr. Jereb Lászlónak, aki mindenre kiterjedő figyelmével folyamatosan segítette, irányította kutatómunkámat. Köszönöm kitartó ösztönzését, támogató tanácsait, javaslatait, melyekkel munkám elkészültéhez nagyban hozzájárult.

Köszönettel tartozom az ERTI több munkatársának is, elsősorban Manninger Miklósnak, akivel elkezdett közös munkánk alapozta meg a kutatásokat. Értékes tanácsaival és szakmai javaslatival folyamatosan segítette munkámat és jelentősen elősegítette a módszer fejlesztését is. Dr. Führer Ernőt és Dr. Csóka Györgyöt is külön köszönet illeti a szakmai segítségnyújtásért, amit a velük folytatott közös munka során folyamatosan kaptam, az eredmények értelmezésében, feldolgozásban nyújtott szerepükért, amivel szintén hozzájárultak a sikeres kutatásokhoz.

Szeretném megköszönni volt doktorandusztársamnak, Edelényi Mártonnak is a közös munkát, akinek komoly szerepe volt a módszerek fejlesztésében. Rengeteg segítséget jelentett, hogy a felmerült problémákat, kérdéseket meg tudtuk beszélni és így hosszú időn keresztül folytatott közös munkával tudtunk a kutatási feladatokban előre lépni. Valamint köszönöm a munkahelyi vitabírálóinak, Dr Charaf Hassannak és Dr. Bejó Lászlónak értékes javaslatait és tanácsait, melyek nagyban hozzájárultak a végleges forma és tartalom kialakulásához

A kutatások anyagi támogatásában nagy szerepe volt a TÁMOP-4.2.2.C–11/1/KONV-2012-0015 számú (Föld-rendszer) projektnek.

## 8 Irodalomjegyzék

- Abonyi, J. (2006): Adatbányászat a hatékonyság eszköze, Computerbooks, Budapest 2006, pp.400.
- Andreau, L., Gutierrez, E., Macias, M., Ribas, M., Bosch, O., Camarero, J. (2007): Climate increases regional tree-growth variability in Iberian pine forests. *Global Change Biology* (2007) 13, 1–12.
- Bijak, SZ. (2010): Tree-ring chronology of silver fir and its dependence on climate of the Kaszubskie lakeland (Northern Poland). *GEOCHRONOMETRIA* 35 (2010), 91-94.
- Bilodeau-Gauthier, S., Daniel Houle, D., Gagnon, C., Côté, B., Messier C. (2011): Assessment of sugar maple tree growth in relation to the partitioning of elements in xylem along a soil acidity gradient. *Forest Ecology and Management* 261 (2011), 95–104.
- Biondi, F. (1997): Evolutionary and moving response functions in dendroclimatology. *Dendrochronologia* 15, 139-150.
- Biondi, F. (2000): Are climate–tree growth relationships changing in north-central Idaho? *Arctic, Antarctic, and Alpine Research* 32, 111–116.
- Biondi, F., Waikul, K. (2004): DENDROCLIM2002: A C++ program for statistical calibration of climate signals in tree-ring chronologies. *Computers & Geosciences* 30 (2004), 303–311.
- Bogino, S.; Nieto, M. J. F. and Bravo, F. (2009): Climate Effect on Radial Growth of *Pinus sylvestris* at Its Southern and Western Distribution Limits. *Silva Fennica* 43 (4), The Finnish Society of Forest Science, The Finnish Forest Research Institute, 609-623.
- Bouriaud, O.; Leban, J.-M. and Bert, D.; Deleuze, C. (2005): Intra-annual variations in climate influence growth and wood density of Norway spruce. *Tree Physiology* 25, 651–660.
- Brett, D.W. 1982. Comment in climate from Tree-rings. Cambridge University Press. pp. 223.
- Briffa, K. R.; Osborn, J. T.; Schweingruber, H. F.; Jones, D. P.; Shiyatov, G. S. and Vaganov, A. E. (2002): Tree-ring width and density data around the Northern Hemisphere: Part 1, local and regional climate signals. *The Holocene* 12,6. SAGE Publications, 737–757.
- Briffa, K. R., Shishov, V. V., Melvin, T. M., Vaganov, E. A., Grudd, H., Hantemirov, R. M., Eronen, M., Naurzbaev, M. M. (2008): Trends in recent temperature and radial tree growth

- spanning 2000 years across northwest Eurasia. *Phil. Trans. R. Soc. B* (2008) 363, 2271–2284.
- Burnham, K. P. és Anderson, D. R. (2002): *Model Selection and Multi-Model Inference*. Springer, 2. kiadás, pp. 485.
- Büntgen, U.; Frank C. D.; Schmidhalter, M.; Neuwirth, B.; Seifert, M. and Esper, J. (2006): Growth/climate response shift in a long subalpine spruce chronology. *Trees* 20, 99–110.
- Büntgen, U.; Frank, C. D.; Kaczka, J. R.; Verstege, A.; Zwijacz-Kozica, T. and Esper, J. (2007): Growth responses to climate in a multi-species tree-ring network in the Western Carpathian Tatra Mountains, Poland and Slovakia. *Tree Physiology* 27, 689–702.
- Carrer, M., Urbinati, C. (2001): Assessing climate-growth relationships: a comparative study between linear and non-linear methods. *Dendrochronologia* 19 (1), 57–65.
- Čufar, K.; Prislan, P.; de Luis, M. and Gričar, K. (2008): Tree-ring variation, wood formation and phenology of beech (*Fagus sylvatica*) from a representative site in Slovenia, SE Central Europe. *Trees* 22, 749–758.
- Csóka, Gy., Pödör, Z., Hirka, A., Szócs L., (2012a): Az időjárási tényezők hatása a tölgy bűcsújáró lepke (*Thaumetopoea processionea* L.) populációinak fluktuációjára. Magyar Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.
- Csóka, Gy., Pödör, Z., Hirka, A., Führer, E. and Szócs, L. (2012b). Influence of weather conditions on population fluctuations of the oak processionary moth (*Thaumetopoea processionea* L.) in Hungary. Joint IUFRO 7.03.10 – “Methodology of forest insect and disease survey” and IUFRO WP 7.03.06 – “Integrated management of forest defoliating insects” Working Party Meeting, Palanga, 10–14 September 2012.
- Csóka, Gy., Pödör, Z., Hirka, A., Führer E., Móricz, N., Rasztovcics, E., Szócs, L., (2013): Időjárásfüggő fluktuáció a tölgy bűcsújáró lepke nyugat-magyarországi populációjánál. Növényvédelmi Tudományos Napok, Budapest, 2013. 02. 19-20.
- Deslauriers, A.; Morin, H. (2005) Intra-annual tracheid production in balsam fir stems and the effect of meteorological variables. *Trees* 19(4), 402-408.
- Deslauriers, A.; Giovannelli, A; Rossi, S.; Castro, G.; Fragnelli, G.; Traversi, L. (2009): Intra-annual cambial activity and carbon availability in stem of poplar. *Tree Physiology* 29(10), 1223–1235.



- Dévényi, D., Gulyás, O., 1988: Matematikai statisztikai módszerek a meteorológiában. Tankönyvkiadó, Budapest. p. 443.
- Dittmar, C., Zech, W., Elling, W. (2003): Growth variations of Common beech (*Fagus sylvatica* L.) under different climatic and environmental conditions in Europe – a dendroecological study. *Forest Ecological Management* 173 (2003), 63-78.
- Druckenbrod, D. L. (2004): Dendroecological reconstructions of forest disturbance history using time-series analysis with intervention detection. *Can. J. For. Res.* (2005) 35, 868–876.
- Edelényi, M., Pödör, Z., Jereb, L. (2011a): Speciális elemzési megközelítés a fák növekedése és az időjárási paraméterek közötti kapcsolatok vizsgálatában. *Agrárinformatika / Agricultural Informatics* (2011) Vol. 2, No. 1, 39-48.
- Edelényi, M., Pödör, Z., Jereb, L., Manninger M. (2011b): Telítődési görbék alkalmazása korlátos növekedési adatsorok vizsgálatában. Országos Gazdaságinformatikai Konferencia, OGIK'2011, Győr
- Edelényi, M., Pödör, Z., Jereb, L., Manninger M. (2011c): Másodlagos idősorokat származtató módszer kifejlesztése és bemutatása erdészeti adatokon. *Acta Agraria Kaposváriensis* (2011) Vol 15 No 3, 39-49.
- Edelényi, M., Pödör, Z., Jereb, L., Manninger, M. (2012): Erdei fák éves növekedésének közelítő leírása függvényekkel. *Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia*. Debrecen, Magyarország, 2012.08.29-2012.08.31.
- Efron, B. (1979): Bootstrap methods: another look at jackknife. *Ann. Stat.*, 7, 1-26.
- Feliksik, E. and Wilczyński S. (2009): The effect of climate on tree-ring chronologies of native and non-native tree species growing under homogenous site conditions. *Geochronometria* 33, 49-57.
- Friedrichs, D. A., Trouet, V., Büntgen, U. Frank, D. C., Esper, J., Neuwirth, B., Löffler, J. (2009): Species-specific climate sensitivity of tree growth in Central-West Germany. *Trees* (2009) 23, 729–739.
- Fritts, H. C. (1976): *Tree rings and climate*. Academic Press, London, pp. 582.
- Fu, T. (2011): A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24 (2011), 164-181.

- Führer E., Horváth L., Jagodics A., Machon A., Szabados I. (2011): Application of a new aridity index in Hungarian forestry practice. *IDŐJÁRÁS* 115(3), 205-216.
- Führer, E., Edelényi, M., Jagodics, A., Jereb, L., Horváth, L., Móring, A., Pödör, Z., Szabados, I., (2012): Az átmérő-növekedés és az időjárás közötti összefüggés egy idős bükkösben. Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.
- Gálos, B., Mátyás, Cs., Jacob, D. (2012): Az erdőtelepítés szerepe a klímaváltozás hatásának mérséklésében. *Erdészettudományi Közlemények*, 2012 2(1), 35–45.
- Gutiérrez, E.; Campelo, F.; Julio Camarero, J.; Ribas, M.; Muntán, E.; Nabais, C. and Freitas, H. (2011): Climate controls act at different scales on the seasonal pattern of *Quercus ilex* L. stem radial increments in NE Spain. *Trees*, Springer-Verlag.
- Han, J., Camber, M. (2006): *Data Mining, Concepts and Techniques - second edition*. Morgan Kaufmann Publishers, 2006, pp. 772.
- Hirka, A., Szabóky, Cs., Szócs, L., Csóka, Gy. (2011): 50 éves az Erdészeti Fénycsapda Hálózat. *Erdészeti Lapok* CXLVI. évf. 12. Szám.
- Jacob, D. (2001): A note to the simulation of the annual inter-annual variability of the water budget over the Baltic Sea drainage basin. *Meteorology and Atmosphere Physics*, 77, 61-73.
- Jacob, D., Andrae, U., Elgered, G., Fortelius, C., Graham, L.P., Jackson, S.D., Karstens, Koepken, CHR., Lindau, R., Podzun, R., Rockel, B., Rubel, F., Sass, H.B., Smith, R.N.D., Van Den Hurk, B.J.J.M., Yang, X. (2001): A Comprehensive Model Intercomparison Study Investigating the Water Budget during the BALTEX-PIDCAP Period. *Meteorology and Atmospheric Physics* 77 (1–4), 19–43.
- Jansen, M., Zuidema, P. A., Anten, N. P. R., Martínez-Ramos, M. (2012): Strong persistent growth differences govern individual performance and population dynamics in a tropical forest understorey palm. *Journal of Ecology* 100 (5), 1224–1232.
- Jereb, L., Edelényi, M., Pödör, Z. (2013a): EVH I. szintű mintafák levélvesztés, koronaelhalás és rovarkárosítás intenzitás adatainak vizsgálata. *Kutatási jelentés*, pp. 91.
- Jereb, L., Edelényi, M., Pödör, Z. (2013b): Mortalitás visszamenőleges vizsgálata az FNM adatokon, *Kutatási jelentés*, pp. 77.

- Jezik, M.; Blazeneca, M.; Strelcova, K.; Ditmarova, L. (2011): The impact of the 2003–2008 weather variability on intra-annual stem diameter changes of beech trees at a submontane site in central Slovakia. *Dendrochronologia* 29(4), 227-235.
- Jolliffe, I. T. (2002): *Principal Component Analysis*, second edition. Springer, 2002, pp. 518.
- Kahm, M., Hasenbrink G., Lichtenberg-Frate, H., Ludwig, J., Kschischo, M. (2010): grofit: Fitting Biological Growth Curves with R. *Journal of Statistical Software* 33(7), 1-21.
- Kehl, D., Sipos, B. (2009): A telítődési, a logisztikus és az életgörbe alakú trendfüggvények becslése Excel parancsfájl segítségével. *Statisztikai Szemle*, 87 (4), 381-411.
- Keogh, E., Chu, S., Hart, D., Pazzani, M. (1993): Segmenting Time Series: A Survey and Novel Approach. In an Edited Volume, *Data mining in Time Series Databases*, pp. 15.
- Kern, Z., Popa, I. (2007): Climate–growth relationship of tree species from a mixed stand of Apuseni Mts., Romania. *Dendrochronologia* 24 (2007), 109–115.
- Lebourgeois, F.; Cousseau, G. and Ducos, Y. (2004): Climate-tree-growth relationships of *Quercus petraea* Mill. stand in the Forest of Bercé (“Futaie des Clos”, Sarthe, France). *Annals of Forest Science* 61. EDP Sciences, 1–12.
- Lebourgeois, F.; Bréda, N.; Ulrich, E. and Granier, A. (2005): Climate-tree-growth relationships of European beech (*Fagus sylvatica* L.) in the French Permanent Plot Network (RENECOFOR). *Trees* 19. Springer-Verlag, 385–401.
- Liao, C. Y., Podrázsky, V. V., Liu, G. B. (2003): Diameter and height growth analysis for individual White Pine trees in the area of Kostelec nad Černými lesy. *Journal of forest science*, 49, 2003 (12), 544–551.
- Lindholm, M., Lehtonen, H., Kolström, T., Meriläinen, J., Eronen, M. & Timonen, M. (2000): Climatic signals extracted from ring-width chronologies of Scots pines from the northern, middle and southern parts of the boreal forest belt in Finland. *Silva Fennica* 34 (4), 317–330.
- Macias, M., Andreu, L., Bosch, O., J. JULIO Camaero, J. J., Gutierrez, E. (2006): Increasing aridity is enhancing silver fir (*abies alba mill.*) Water stress in its south-western distribution limit. *Climatic Change* (2006) 79, 289–313.

- MacQueen, J. B. (1967): Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Symposium on Math, Statistics, and Probability. Berkeley, CA: University of California Press, 281-297.
- Mäkinen, H.; Nöjd, P. and Mielikäinen, K. (2000): Climatic signal annual growth variation of Norway spruce (*Picea abies*) along a transect from central Finland to the Arctic timberline. Canadian Journal of Forest Research 30, 769–777.
- Mäkinen, H.; Nöjd, P.; Kahle, H. P.; Neumann, U.; Tveite, B.; Mielikäinen, K.; Röhle, H. and Spiecker, H. (2002): Radial growth variation of Norway spruce (*Picea abies* (L.) Karst.) across latitudinal and altitudinal gradients in central and northern Europe. Forest Ecology and Management 171, 243–259.
- Mäkinen, H.; Nöjd, P.; Kahle, H. P.; Neumann, U.; Tveite, B.; Mielikäinen, K.; Röhle, H. and Spiecker, H. 2003: Large-scale climatic variability and radial increment variation of *Picea abies* (L.) Karst. in central and northern Europe. Trees 17, 173–184.
- Manninger, M. (2004): Erdei fák éves és korszaki növekedésmenete és kapcsolódása egyes ökológiai tényezőkhöz. In: Mátyás Cs., Vig P. (ed.): Erdő és Klíma IV., Nyugat-magyarországi Egyetem, Sopron, 151-162.
- Manninger, M. (2008): A növekedés és a csapadék összefüggései az alföldi mérések alapján, In: Szulcsán G. (ed.) Kutatói nap, tudományos eredmények a gyakorlatban, Alföldi Erdőkért Egyesület kiadványa, Szeged, 50-53.
- Manninger M., Edelényi, M., Pödör, Z., Jereb, L.(2011a): Alkalmazott elemzési módszerek a környezeti tényezők fák növekedésére gyakorolt hatásának vizsgálatában. Erdészettudományi Közlemények, 1(1), 59-70.
- Manninger, M., Edelényi, M., Pödör, Z., Jereb, L. (2011b): The effect of temperature and precipitation on growth of beech (*Fagus sylvatica* L.) in Mátra Mountains, Hungary. Applied Forestry Research in the 21st Century conference, Prága-Pruhonice, 2011.09.13-15.
- Manninger, M., Edelényi, M., Pödör, Z., Jereb, L. (2012): A hőmérséklet és a csapadék hatása a bükk növekedésére. Meteorológiai Társaság XXXIV. Vándorgyűlés és VII. Erdő és Klíma Konferencia. Debrecen, Magyarország, 2012.08.29-2012.08.31.

- Mares, I., Mares, C. (1994): Statistical methods for estimating the signal to noise ratio. In: Contemporary Climatology (Proceedings of the meeting of the Commission on Climatology of the International Geographical Union). Brno. 373–379.
- Martín-Benito, D.; Cherubini, P.; del Río, M. and Cañellas, M. (2008): Growth response to climate and drought in *Pinus nigra* Arn. trees of different crown classes. *Trees* 22, 363–373.
- Mátyás, Cs. (2010): Forecasts needed for retreating forests (Opinion). *Nature: International Weekly Journal of Science* 464:(7293) p. 1271.
- Mátyás, Cs., Berki, I., Czúcz, B., Gálos, B., Móricz, N., Rasztoivits, E. (2011): Assessment and projection on climate change impacts in SE European forests: a case study of common beech (*Fagus sylvatica* L.). *Les: Revija Za Lesno Gospodarstvo* 63: pp. 142-153.
- Maxime, C. and Hendrik, D. (2011): Effects of climate on diameter growth of co-occurring *Fagus sylvatica* and *Abies alba* along an altitudinal gradient. *Trees* 25, 265–276.
- Mazerolle, M. J.: AICcmodavg (2012) Model selection and multimodel inference based on (Q)AIC(c). <http://CRAN.R-project.org/package=AICcmodavg>
- McCarthy, J. W., Weetman, G. (2006): Age and Size Structure of Gap-Dynamic, Old-Growth Boreal Forest Stands in Newfoundland *Silva Fennica* 40(2), 209–230.
- Michelot, A., Bréda, N., Damesin, C., Dufrene, E. (2012): Differing growth responses to climatic variations and soil water deficits of *Fagus sylvatica*, *Quercus petraea* and *Pinus sylvestris* in a temperate forest. *Forest Ecology and Management* 265 (2012), 161–171.
- Milei, O. (2013): Gazdaságelemzés, különös tekintettel a fa- és Bútoripar helyzetére. Faipari bútorigipari erdészeti ágazati párbeszéd bizottság részére, Budapest, pp. 101.
- Molnár, J. (2003): A légnomási mező szerkezete és módosulása a Kárpát-medence térségében. Doktori értekezés, Debrecen, pp. 170.
- Molnár, J., Izsák, T. (2011) Trendek és töréspontok a léghőmérséklet Kárpátalja idősorában. *Légekör* 56(2), 49-54.
- Monserud, R. A. (1986): Time-Series Analyses of Tree-Ring Chronologies. *Forest Science*, 32 (2), 349-372.
- Montgomery, D. C., Peck, A. E., Vining, G. G. (2012): Introduction to linear regression analysis (fifth edition). Published by John Wiley & Sons, pp. 672.

- Norton, D.A. 1983. Modern New Zealand chronologies: *Nothofagus solandri*. *Tree Ring Bulletin*, 43, 1-17.
- Novák, J.; Slodičák, M.; Kacálek, D. and Dušek, D. (2010): The effect of different stand density on diameter growth response in Scots pine stands in relation to climate situations. *Journal Of Forest Science* 56 (10). Czech Academy of Agricultural Sciences, 461–473.
- Oberhuber, W.; Kofler, W.; Pfeifer, K.; Seeber, A.; Gruber, A. and Wieser, G. (2008): Long-term changes in tree-ring–climate relationships at Mt. Patscherkofel (Tyrol, Austria) since the mid-1980s. *Trees* 22, 31–40.
- Packalen, T., Marques, A. F., Rasinmäki, J., Rosset, C., Mounir, F., Rodriguez, L. C. E., Nobre, S. R. (2013): Review. A brief overview of forest management decision support systems (FMDSS) listed in the FORSYS wiki. *Forest Systems* 2013 22(2), 263-269.
- Pálfai I., Herceg Á. (2011): Droughtness of Hungary and Balkan Peninsula. *Riscuri Si Catastrofe* 10(2), 145-154.
- Pärn, H. (2003): Radial growth response of scots pine to climate under dust pollution in northeast Estonia. *Water, Air, and Soil Pollution* 144. Kluwer Academic Publishers. Netherlands, 343–361.
- Péczely, Gy., 1979: Éghajlattan. Szeged. Utánnomás: Nemzeti Tankönyvkiadó, Budapest 1996. pp. 336.
- Pełech-Pilichowski, T., Duda, J. T. (2010): A two-level algorithm of time series change detection based on a unique changes similarity method. *Proceedings of the International Multiconference on Computer Science and Information Technology*, 259–263.
- Petráš, R., Mecko, J. (2011): Effect of climatic factors on the dynamics of radial increments of Norway spruce, European beech and sessile oak. *Journal of Forest Science*, 57, 2011 (7), 293–302.
- Petritán, A. M., von Lüpke, B., Petritán, I. C. (2009): Influence of light availability on growth, leaf morphology and plant architecture of beech (*Fagus sylvatica* L.), maple (*Acer pseudoplatanus* L.) and ash (*Fraxinus excelsior* L.) saplings. *Eur J Forest Res* (2009) 128, 61–74.
- Pichler, P. and Oberhuber, W. (2007): Radial growth response of coniferous forest trees in an inner Alpine environment to heat-wave in 2003. *Forest Ecology and Management* 242, 688–699.

- Pieczka, I., Pongrácz, R., Bartholy, J. (2011): Comparison of Simulated Trends of Regional Climate Change in the Carpathian Basin for the 21st Century Using Three Different Emission Scenarios. *Acta Silv. Lign. Hung.*, Vol. 7 (2011), 9–22.
- Piovesan, G.; Biondi, F.; Bernabei, M.; Filippo, A. and Schirone, B. (2005): Spatial and altitudinal bioclimatic zones of the Italian peninsula identified from a beech (*Fagus sylvatica* L.) tree-ring network. *Acta Oecologica* 27, 197–210.
- Power DJ, 2002. Decision support systems: concepts and resources for managers. Westport, Conn, Quorum Books, pp. 251.
- Pödör, Z., Kiss, B., Csóka, Gy., Jereb, L. (2013a): Egyes nagylepkefajok Magyarországi fogási adatainak lehetséges klímafüggése – vizsgálati módszertan és előzetes eredmények. Kari Tudományos Konferencia, Sopron, 2013.12.10.
- Pödör, Z., Csóka, Gy., Kiss, B. (2013b): Simple- and Multivariate data analysis of light trap catching data by a systematic window procedure. Decision Support System Workshop and ForestDSS Community of Practice, Lisbon, 4.-6. 12. 2013.
- Pödör, Z., Manninger, M., Jereb, L. (2014a): Application of Sigmoid Models for Growth Investigations of Forest Trees. Applied Mathematics and Applications Conference, Conference Proceedings, Advanced Computational Methods for Knowledge Engineering, Advances in Intelligent Systems and Computing 282, 353-364.
- Pödör, Z., Edelényi, M., Jereb, L. (2014b): Systematic Analysis of Time Series – CReMIT. Infocommunication Journal, VI(1), 16-22.
- Ramachandran, K. M., Tsokos, C. P. (2009): Mathematical Statistics with Applications. Elsevier Academic Press, pp. 803.
- Rennols K. (1995): Forest height growth modelling. *Forest Ecology and Management* 71 (1995), 217-225.
- Reynolds, K. M., Twery, M., Lexer, M. J., Vacik, H., Ray, D., Shao, G., Borges, J. G. (2008): Handbook on Decision Support Systems 2, 499-533.
- Rybniček M.; Čermák P.; Kolář T.; Přemyslovská E., Žid T. 2009: Influence of temperatures and precipitation on radial increment of Orlické hory Mts. spruce stands at altitudes over 800 m a.s.l. *Journal Of Forest Science* 55 (6), 257–263.

- Sánchez-González, M., Tomé, M., Montero, G. (2005): Modelling height and diameter growth of dominant cork oak trees in Spain. *Ann. For. Sci.* 62 (2005), 633–643.
- Savva, Y.; Oleksyn, J.; Reich B. P.; Tjoelker G. M.; Vaganov A. E. and Modrzynski, J. (2006): Interannual growth response of Norway spruce to climate along an altitudinal gradient in the Tatra Mountains, Poland. *Trees* 20, 735–746.
- Scharnweber, T., Manthey, M., Criegee, C., Bauwe, A., Schroder, C., Wilmking, M. (2011): Drought matters – Declining precipitation influences growth of *Fagus Sylvatica* L. and *Quercus robur* L. in north-eastern Germany, *Forest Ecology and Management* 262(6), 947-961.
- Seo, J.-W.; Eckstein, D.; Jalkanen, R.; Schmitt, U. (2011): Climatic control of intra- and inter-annual wood-formation dynamics of Scots pine in northern Finland. *Environmental and Experimental Botany* 72(3), 422–431.
- Sit, V. and Poulin-Costello, M. (1994): Catalog of curves for curve fitting. BC Ministry of Forests Biometrics Information Handbook No. 4. pp. 110.
- Spiegel, M. R. (1995): *Statisztika, Elmélet és gyakorlat*, PANEM-McGraw-Hill, Budapest, 1995, pp. 546.
- Szabados I. (2002): Az évgyűrűszélesség és egyes termőhelyi tényezők kapcsolata NYME-disszertáció, pp. 114.
- Szabados I. (2008): A csapadék hatása a cser évgyűrűméretére *Erdészeti Kutatások Budapest*, Vol. 92, 121-128.
- Szeőke, K. és Csóka, Gy. (2012): Jövevény kártevő ízeltlábúak áttekintése Magyarországon lepkék (lepidoptera). *Növényvédelem* 48 (3), 105-115.
- Thorp, T. G., Anderson, P., Camilleri, M. (1995): Avocado tree growth cycles - a quantitative model. *Proceedings of The World Avocado Congress III*, 76-79.
- Tuovinen, M. (2005): Response of tree-ring width and density of *Pinus sylvestris* to climate beyond the continuous northern forest line in Finland. *Dendrochronologia* 22, 83–91.
- Tusnádi G. és Ziermann M. (szerk): *Idősorok analízise*, Műszaki Könyvkiadó, Budapest, 1986, pp 339.



- van der Werf, G. W.; Sass-Klaassen, G. W. U. and Mohren, G. M. J. (2007): The impact of the 2003 summer drought on the intra-annual growth pattern of beech (*Fagus sylvatica* L.) and oak (*Quercus robur* L.) on a dry site in the Netherlands. *Dendrochronologia* 25, 103–112.
- Veress K. (2007): A Newton és Gauss-Newton módszerek alkalmazása egyenletrendszerek megoldására és nemlineáris optimalizálásra. Egyetemi jegyzet, Szeged, pp. 46.
- Wilczyński, S. and Podlaski, R. (2007): The effect of climate on radial growth of horse chestnut (*Aesculus hippocastanum* L.) in the Swietokrzyski National Park in central Poland. *Journal of Forest Research* 12, 24–33.
- Wikström, P., Edenius, L., Elfving, B., Eriksson, L. O., Lamas, T., Sonesson, J., Ohman, K., Wallerman, J., Waller, K., Klinteback, F. (2011): The HEUREKA forestry decision support system: An overview. *Mathematical and Computational Forestry&Natural-Resource Sciences*, Vol. 3, Issue 2, 87–94.
- Wipfler, P., Seifert, T., Biber, P., Pretzsch, H. (2009): Intra-annual growth response of adult Norway spruce (*Picea abies* [L.] KARST.) and European beech (*Fagus sylvatica* L.) to an experimentally enhanced, free-air ozone regime. *Eur J Forest Res* (2009) 128, 135–144.
- WMO, 2000: A Meteorológiai Világszervezet állásfoglalása az éghajlat 2000. évi állapotáról. WMO-No. 913. Geneva. pp. 12.
- WMO, 2011: A Meteorológiai Világszervezet állásfoglalása az éghajlat 2011. évi állapotáról. WMO-No. 1085. Geneva. pp. 24.
- Zhang, Y., Wilmking, M., Gou, X. (2008): Changing relationships between tree growth and climate in Northwest China. *Plant Ecol* (2009) 201, 39–50.
- Zeide, B. (1993) Analysis of Growth Equations. *Forest Science* 39(3), 594-616.
- Zhang, Y., Wilmking, M., Gou, X. (2011): Dynamic relationships between *Picea crassifolia* growth and climate at upper treeline in the Qilian Mts., Northeast Tibetan Plateau, China. *Dendrochronologia* 29 (2011), 185–199.
- Zwietering, M. H.; Jongenburger I.; Rombouts F. M.; Van't Riet K. (1990) Modeling of the Bacterial Growth Curve. *Applied and Environmental Microbiology* 56(6), 1875-1881.