Erdészettudományi Közlemények

3. évfolyam 1. szám 2013

137–145. oldal

ERDÉSZETI ADATHALMAZOK ELEMZÉSE ÚJ FÜGGVÉNNYEL

Csanády Viktória

Nyugat-magyarországi Egyetem, Erdőmérnöki Kar

Kivonat

Az irodalmakból eddig ismert telítési, illetve életgörbék helyett egy új, az említett függvényeket jól helyettesítő model illesztésének bemutatása a cél. Alkalmazni az erdészeti kutatásból származó fatermési adatokra alkalmaztuk. Ennek során kitűnik a modell rugalmassága, mivel az új szinuszos telítési függvény akár inflexiós pont nélküli, akár inflexiós ponttal rendelkező esetre is illeszthető. A számítógépes regressziós eljárás által igényelt kezdőértékek megválasztása egyszerű. A modell öt paraméterrel rendelkezik, regressziós meghatározásuk után felhasználásukkal a vizsgált adatsorra fontos és egyben az adatsort jól jellemző értékek számíthatók. Így lehetőség nyílik az adatsorok, jelen esetben akác termőhelyi osztályok alapvető különbségének bemutatására nem csak grafikusan, hanem konkrét számított értékek megadásával.

Kulcsszavak: regresszió számítás, telítési függvény, életgörbe, szinuszos telítési függvény, fatermési adatok

A NEW FUNCTION FOR ANALYSIS OF DATASETS

Abstract

In this paper, we suggest a model of fitting a new type of saturation curve, namely a sinus curve, which can extend the application of saturation curves and life curves for forest yield database. The greatest advantage of the model is its flexibility can be realized in fitting either with or without inflection point. The initial values can easily be given for the regression procedure implemented in computer. The output of the programme consists of five parameters. Beside the usual graphical illustration the algorithm makes possible to show the differences between, for instance, black locust-tree production sites by exact calculation.

Keywords: regression, saturation curve, life curve, sinus curve, forest yield database

BEVEZETÉS

A természetben előforduló különböző folyamatok kísérleti vizsgálata során nyert adatpárhalmazok egyszerű grafikus ábrázolása után az a feladat, hogy megfelelő függvényt találjunk a változás kifejezésére és értelmezésére. Az erdészeti és a faipari kutatásokban is megjelenik ez a probléma, és az adathalmazok vízuális áttekintése során gyakori az a jelenség, hogy a függő változó egy bizonyos határértékhez tart, maximum elérése után vagy anélkül, akár az idő a független változó, akár más paraméter. Ez azt jelenti, hogy telítési vagy több inflexiós ponttal rendelkező életfüggvény alkalmazása látszik célszerűnek. A választott függvény illeszté-

A szerző címe/Correspondence:
9400 Sopron, Bajcsy-Zsilinszky út 4., email: vcsanady@emk.nyme.hu

se a jelen lehetőségeket figyelembe véve a számítógépes statisztikai regressziós programok segítségével lehetséges (Statistica 9) az alább felsorolt feltételek alkalmazásával:

- A függvény fizikailag értelmezhető legyen (polinomos helyettesítés értelmetlen).
- A szereplő paraméterekből a lehető legtöbb információhoz lehessen jutni közvetlenül.
- A nyert korrelációs együttható (R) értéke a lehető legnagyobb legyen.
- A szereplő adathalmaz tartományán pozitív irányban túlmenően is legyen lehetőség értékelésre.

ANYAG ÉS MÓDSZER

Telítési függvények (Kehl és Sípos 2009)

Inflexiós ponttal nem rendelkező telítési görbék:

- *Mitscherlich:* $y(t) = K(1 - e^{-rt})$
- *Bertalanffy:* $y(t) = K(1 - be^{-rt})$
- *Tömquist:* $y(t) = \frac{Kt}{t+a}$, $y(t) = \frac{K(t+a)}{t+a}$

Inflexiós ponttal rendelkező telítési görbék:

- *Verhulst:* $y(t) = \frac{K}{1 + e^{-c(t-m)}}$ $vagy$ $y(t) = \frac{Ke^{ct}}{e^{cm} + e^{ct}}$
- *Pearl – Reed:* $y(t) = \frac{K}{be^{-ct} + 1}$
- *Késleltetett logisztikus trendfüggvény:* $y(t) = \frac{K}{1 + \left(\frac{T}{t}\right)^a}$
- *Négyzetes logisztikus trendfüggvény:* $y(t) = \frac{K^2}{(1 + be^{-ct})^2}$
- *Gompertz:* $y(t) = Ke^{-be^{-ct}}$
- *A 63%-os trendfüggvény:* $y(t) = K - \frac{K}{e^{\left(\frac{T}{t}\right)^a}}$
- *Johnson:* $y(t) = e^{K - \frac{b}{i+t}}$
- *Richards:* $y(t) = \frac{K}{(1 + ve^{-c(t-m)})^{\frac{1}{v}}}$
- *Chapman – Richards:* $y(t) = a(1 - e^{-bt})^c$
- *Colin – Fokasz (módosított Richards függvény):* $y(t) = A + \frac{(K - A)}{(1 + ve^{-c(t-m)})^{\frac{1}{v}}}$

Két inflexiós ponttal rendelkező telítési görbék (életgörbék):

- *Haustein:* $y(t) = \frac{a}{e^{w^{2(r-t)2}}}$
- *Hubbert:* $y(t) = \frac{bue^{b(t-r)}}{(1 + e^{b(t-r)})^2}$

Klasszikus 0 vagy 1 inflexiós ponttal rendelkező telítési görbe:

- *Awrami:* $y(t) = a(1 - e^{-(bx)^c}) + d$

Az új függvény (saját ötlet alapján szerkesztve):

Matematikai alakja (összetett függvény $y = f(g(x))$):

- $y = a \cdot \sin\left(b\left(1 - e^{-(cx)^d}\right)\right) + f$ (SinAwr).

A számítógépi alak (a biztos kezelhetőség érdekében):
$var2 = b4 \cdot \sin\left(b3 \cdot (1 - \exp(-1 \cdot (b2 \cdot var1) \cdot b1))\right) + b0.$

Elemzés: 0, 1 vagy 2 inflexiós ponttal rendelkezik, maximummal vagy anélkül.
$y_{max} = var2_{max} = a + f = b4 + b0$ ha $b3 > \frac{\pi}{2}$; (ha $b3 < \frac{\pi}{2}$ nincs maximum).
$y_{végső} = var2_{végső} = a sinb + f = b4 sinb3 + b0,$

$y_{\max}$ -hoz tartozó $x_{\max}$ számítása:

$$lnx_{max} = lnvar1_{max} = \frac{1}{d} ln\left(\frac{ln2b}{2b - \pi}\right) - lnc = \frac{1}{b1} ln\left(\frac{ln2b3}{2b3 - \pi}\right) - lnb2$$

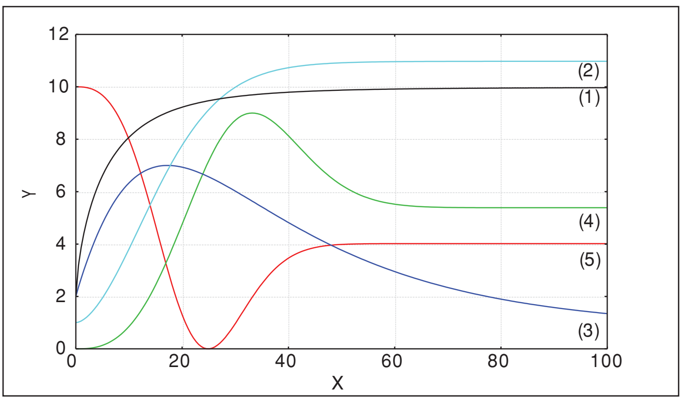
ha $b3 > \frac{\pi}{2}$.

A végső értéktől, azaz a határértéktől, az $y_{végső}$-től közelítőleg 1%-nál kisebb értékkel való eltérés tartományának kezdete: $x_{végső} = var1_{végső}$, ahol

$$lnx_{végső} = lnvar1_{végső} = \frac{ln\,(ln1000)}{d} - lnc = \frac{ln\,(ln1000)}{b1} - lnb2$$. (A technikai gyakorlat szerint.)

A fentiek felsorolásából látható, hogy a szóban forgó függvény minden paramétere (*a, b, c, d, f* illetve *b*4, *b*3, *b*2, *b*1, *b*0) értelmezhető, és a megadott képletekben alkalmazva a fontos és jellemző adatok kiszámolhatók, azaz a vizsgált adathalmaz által megadott folyamat elemezhető és egyértelműen meghatározható.

A jobb áttekinthetőség érdekében az alábbiakban egy összetett grafikus ábracsoport látható az egyes görbéket megadó függvények feltüntetésével:



1. ábra: *A SinAwr függvény grafikonjai*
Figure 1: *Graph of function SinAwr*

Az egyes görbék áttanulmányozása alapján megállapítható:

Az $y = 8\sin(1{,}5(1 - \exp(-(0{,}08x) \cdot 0{,}7))) + 2$ (1)

függvény olyan egyszerű telítési függvény, melynek nincs inflexiós pontja.

Az $y = 10\sin(1{,}5(1 - \exp(-(0{,}04x) \cdot 2{,}7))) + 1$ (2)

függvény olyan telítési függvény, melynek a határérték előtt van inflexiós pontja.

Az $y = 5\sin(3{,}4(1 - \exp(-(0{,}035x) \cdot 0{,}95))) + 2$ (3)

függvény olyan speciális függvény, melynek a maximuma után van inflexiós pontja és határértéke.

Az $y = 9\sin(2{,}5(1 - \exp(-(0{,}03x) \cdot 2{,}7)))$ (4)

függvény olyan különleges eset, melynek a maximuma előtt és után is van inflexiós pontja és azt követően határértéke.

Az $y = -10\sin(2{,}5(1 - \exp(-(0{,}04x) \cdot 2{,}7))) + 10$ (5)

függvény olyan speciális eset, melynek minimuma van, és ez előtt és után is rendelkezik inflexiós ponttal és ezt követően határértékkel.

A felsoroltakból egyértelműen látható, hogy a szóban forgó új függvény regressziós alkalmazási területe igen széles körű és kedvező a későbbiekben megadott gyakorlati felhasználások során nyert kezdőértékek egyszerű adatai miatt.

Gyakorlati alkalmazások

Tekintettel arra, hogy nem a korábbiakban felsorolt 17 féle függvény különböző fokú értékelése a dolgozat témája, hanem az új rugalmas függvény (SinAwr) széles körű alkalmazási lehetőségének bemutatása, célszerűen erdészeti vonatkozású adatsorokat kerestünk azzal a feltétellel, hogy a szakirodalmi adatok eredete és előállításának módja (símítás, korrekció, javítás) nem tartozik a témához, azok valós értékeknek tekinthetők.

Az irodalomban (Rédei és mtsai 2011) fellelhető és véletlenszerűen kiválasztott adathalmazok akácos erdőterületre vonatkoznak, a következőkben felsoroltak szerint:

1. táblázat: *Akác 1*

Table 1: *Black locust 1*

	I. fatermési osztály					
	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6
1	5,000	7,200	5,200	41,000	8,300	0,000
2	10,000	13,100	10,200	121,000	12,900	17,500
3	15,000	17,600	15,300	169,000	14,300	17,000
4	20,000	20,800	19,400	217,000	14,500	15,400
5	25,000	23,100	22,800	259,000	14,200	13,000
6	30,000	24,700	25,600	294,000	13,600	10,600
7	35,000	25,800	28,000	323,000	12,900	8,700
8	40,000	26,600	30,100	350,000	12,300	7,600
9	45,000	27,300	32,100	378,000	11,700	7,500

2. táblázat: *Akác 4*
Table 2: *Black locust 4*

	IV. fatermési osztály					
	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6
1	5,000	4,900	3,400	22,000	4,300	0,000
2	10,000	8,900	6,700	62,000	6,400	8,500
3	15,000	11,900	10,300	89,000	7,100	8,400
4	20,000	14,200	13,200	114,000	7,200	7,600
5	25,000	15,700	15,600	136,000	7,100	6,500
6	30,000	16,800	17,500	154,000	6,800	5,300
7	35,000	17,600	19,100	169,000	6,400	4,300
8	40,000	18,100	20,600	183,000	6,100	3,800
9	45,000	18,600	22,000	198,000	5,800	3,800

Az 1. és 2. táblázatokban szereplő adatok dimenziója: var1=kor (év), var2=magasság (m), var3=átmérő (cm), var4=fatérfogat (m^3), var5=átlagnövedék (m^3/év), var6=folyónövedék (m^3/év), mely adatok a vizsgált állományban szereplő átlag- vagy összes értékek. A függvénykiválasztásnál mindkét termőhelyi osztállyal kapcsolatban független változónak a kort (var1) választottuk, függő változóként pedig a többi öt paraméter egyikét, azaz a var2=f(var1), var3=f(var1), var4=f(var1), var5=f(var1) és var6=f(var1) függvények regressziójára került soraz említett új függvény (SinAwr) alkalmazásával.

A számítógépi alak (k=2; 3; 4; 5; 6):

$$vark = b4 \cdot \sin\left(b3 \cdot (1 - \exp(-1 \cdot (b2 \cdot var1) \cdot b1))\right) + b0$$

A regressziós eljárással nyert paraméterértékeket a következő, sorrendbe állított táblázatok mutatják, majd ezeket követik a görbék grafikonjai.

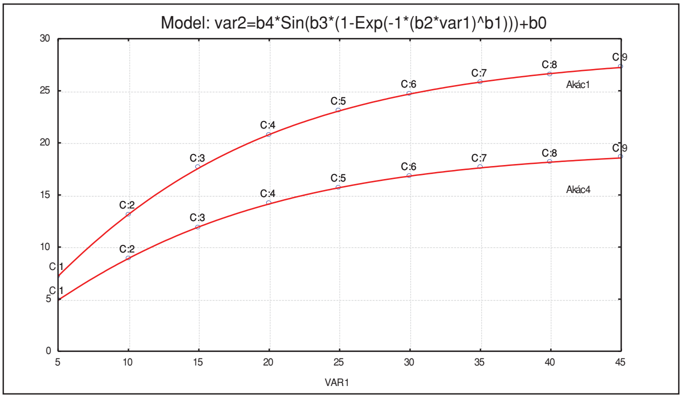
3.táblázat: *Az Akác 1 illesztési adatai*
Table 3: *Black locust 1, fitted parameters*

Akác1	Meghatározott paraméterek / Kezdőértékek					Korrelációs együttható
	b4	b3	b2	b1	b0	R
var2=f(var1)	31,38590	1,155323	0,048784	1,076554	0,122526	0,999986890
	1	1	1	1	1	
var3=f(var1)	55,66011	0,840869	0,029335	1,071664	-0,494438	0,999873374
	1	1	1	1	1	
var4=f(var1)	751,7262	2,160109	0,004994	0,648291	-98,1050	0,999754527
	50	1	1	1	1	
var5=f(var1)	25,39753	2,405057	0,058462	0,668584	-10,8658	0,999919056
	4	1	1	1	1	
var6=f(var1)	359,2486	1,854706	0,245289	0,573719	-341,150	0,998031548
	1	1	1	1	1	

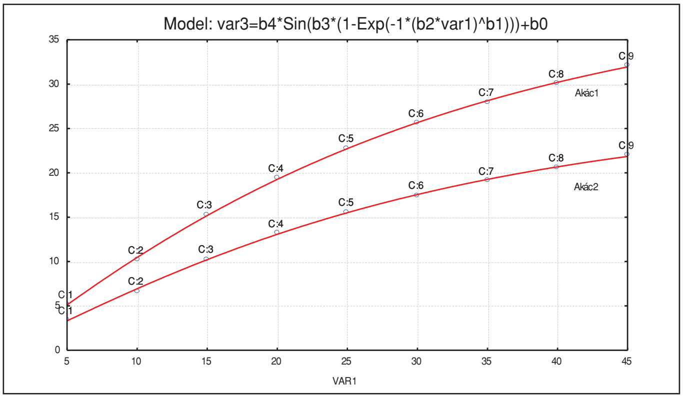
4.táblázat: Az Akác 4 illesztési adatai
Table 4: Black locust 4, fitted parameters

Akác4	Meghatározott paraméterek					Korrelációs együttható
	Kezdőértékek					
	b4	b3	b2	b1	b0	R
var2=f(var1)	21,04890	1,168576	0,048090	1,091107	0,253086	0,999969874
	1	1	1	1	1	
var3=f(var1)	36,07375	0,864967	0,030458	1,128914	–0,192557	0,999764834
	1	1	1	1	1	
var4=f(var1)	368,8762	2,198491	0,005692	0,673153	–47,5826	0,999870091
	50	1	1	1	1	
var5=f(var1)	9,945070	2,534462	0,050328	0,721599	–2,71321	0,999613173
	4	1	1	1	4	
var6=f(var1)	121,3702	1,906059	0,191742	0,633440	–112,495	0,997712571
	4	1	1	1	1	

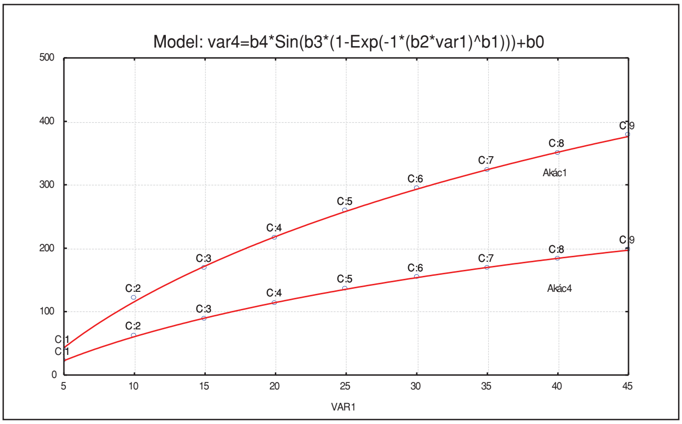
A kezdőértékeket a „Statistica 9” programtól függetlenül, az adatpárok előzetes áttekintése alapján választottuk meg, az összes esetben egyszerű becsléssel, a matematikai függvénytranszformációs szabályokra támaszkodva.



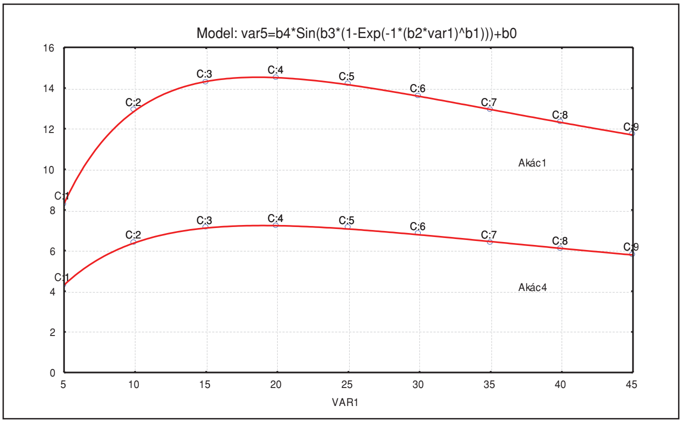
2. ábra: *A famagasság a kor függvényében*
Figure 2: *Function height-age*



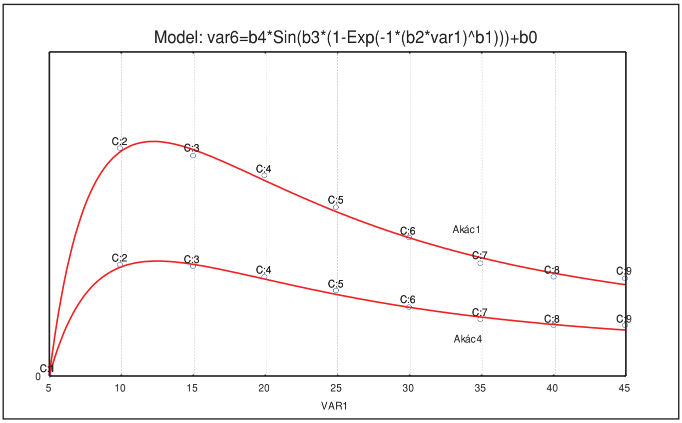
3. ábra: *Az átmérő a kor függvényében*
Figure 3: *Function diameter-age*



4. ábra: *A fatérfogat a kor függvényében*
Figure 4: *Function volume-age*



5. ábra: *Az átlagnövedék a kor függvényében*
Figure 5: *Function average increment-age*



6. ábra: *A folyónövedék a kor függvényében*
Figure 6: *Function increment-age*

A 2–6. ábrák egyértelműen mutatják, hogy a tíz különböző görbét ugyanaz a függvény adta, azonos kezdőértékek mellett, kifogástalan korrelációs együtthatóval, a görbék meglehetősen különböző alakja ellenére, aminek oka a függvény összetett alakja és az öt szereplő paraméter, valamint az adatpárokat jelző pontsorozat feltehető előzetes korrigálása, simítása. A felhasznált irodalomban (Rédei és mtsai 2011) erre nézve nincs közlés. Ha a pontsorozatot nem érinti előzetes korrigálás, a kezdőértékek meghatározása az említett szabályok szerint akkor is egyszerű. Ezzel kapcsolatban vannak az új függvényre vonatkozó alkalamzási adatok más vizsgálatokra, ami nem képezi jelen dolgozat tárgyát.

EREDMÉNYEK ÉS MEGVITATÁSUK

A táblázatokban megadott változók kapcsolatának szorosságát jelző korrelációs együtthatók – $0{,}99771<R<0{,}99999$ – a látható grafikonok görbéinek pontos illeszkedése egyértelműen megadja a lehetőséget az akác termőhelyi osztályok értékelésére. Különös tekintettel arra is, hogy a regresszióhoz alkalmazott függvény paramétereinek felhasználásával fontos jellemző számértékeket kaphatunk a megkülönböztetésre és osztályozásra a korábbiakban megadott képletek alkalmazásával:

Minden esetben k=2; 3; 4; 5; 6

$$vark_{max} = b4 + b0 \text{ ha } b3 > \frac{\pi}{2}$$

$$vark_{végső} = b4 \cdot \sin b3 + b0$$

$$In\ var1_{max} = \frac{1}{b1} ln\left(\frac{ln2b3}{2b3-\pi}\right) - lnb2 \text{ ha } b3 > \frac{\pi}{2}$$

$$In\ var1_{végső} = \frac{\ln(ln1000)}{b1} - lnb2$$

formulákból számítva a következő értékeket kaphatjuk:

5. táblázat: *A var2=f(var1) esetén*
Table 5: *var2=f(var1)*

	var2max(m)	var2végső(m)	var1max(év)	var1végső(év)
Akác1	–	28,8	–	123
Akác4	–	19,6	–	122

6. táblázat: *A var3=f(var1) esetén*
Table 6: *var3=f(var1)*

	var3max(m)	var3végső(cm)	var1max(év)	var1végső(év)
Akác1	–	41,3	–	232
Akác4	–	27,3	–	182

7. táblázat: *A var4=f(var1) esetén*
Table 7: *var4=f(var1)*

	var4max(m3)	var4végső(m3)	var1max(év)	var1végső(év)
Akác1	569	465	203	>200
Akác4	321	249	245	>200

8. táblázat: *A var5=f(var1) esetén*
Table 8: *var5=f(var1)*

	var5max(m3/év)	var5végső(m3/év)	var1max(év)	var1végső(év)
Akác1	14,5	6,2	18,6	305
Akác4	7,2	3,0	19,0	137

9. táblázat: *A var6=f(var1) esetén*
Table 9: *var6=f(var1)*

	var6max(m3/év)	var6végső(m3/év)	var1max(év)	var1végső(év)
Akác1	18,1	3,7	12,2	118
Akác4	8,9	2,1	9,1	110

A fenti táblázatokban szereplő kulcsadatok eleve megadják a lehetőséget arra, hogy grafikus ábrázolás nélkül is egyértelműen meghatározható legyen a két termőhelyi osztály, Akác1 és Akác4 alapvető különbsége a regressziós eljárással kiszámolt konkrét értékek bemutatásával. Ezek az eredmények természetesen az eredeti kiinduló adatsorokat megjelentető szerzők (Rédei és mtsai 2011) számára adhatnak lehetőséget a továbbfejlesztésre az erdészeti szakmai kiértékelés során.

ÖSSZEFOGLALÁS

E közlemény fő témája egy új, könnyen és biztonságosan alkalmazható regressziós függvény bemutatása és valós adatsorokon való kipróbálása volt, minden részletprobléma feltárásával. Ennek alapján megállapítható:

- Az egy függő és egy független változót tartalmazó új függvény (SinAwr) regresziós statisztikai használata egyszerű a könnyen megválasztható ún. kezdőértékek miatt.
- Telítési és nem telítési jellegű adatsorokra egyaránt alkalmazható.
- A függvény a paraméterek értékeitől függően tartalmazhat 2, 1 vagy 0 inflexiós pontot.
- A regressziós illesztésből nyert paraméterekből fontos jellemző adatok számíthatók ki egyszerű módszerrel a megadottak szerint.
- A ma már széles körben elterjedt számítógépes alkalmazások miatt az eljárás gyors.
- Javasolható széles körű kísérleti alkalmazásra.

FELHASZNÁLT IRODALOM

Kehl F. és Sipos B. 2009: A telítődési, a logisztikus és az életgörbe alakú trendfüggvények becslése Excel parancsfájl segítségével. Statisztikai Szemle, 87 (4): 381–411.

Rédei K.; Csiha I.; Keserű Zs.; Kamandiné Végh Á. és Rásó J. 2011: Nyírségi akácosok táji fatermféstáblája. Erdészettudományi Közlemények, 1 (1):115–124.

Statistica 9.: StatSoft: *Statistica* statisztikai adatelemző, analitikai szoftvercsalád

Érkezett: 2013. február 25.
Közlésre elfogadva: 2013. június 28.