



Regresja - modele
nielinowe, uogólnione,
addytywne i mieszane

Krótkie przypomnienie

modele liniowe $y=ax+b$ `lm(y~x, data=dataset)`

modele mogą mieć wiele czynników

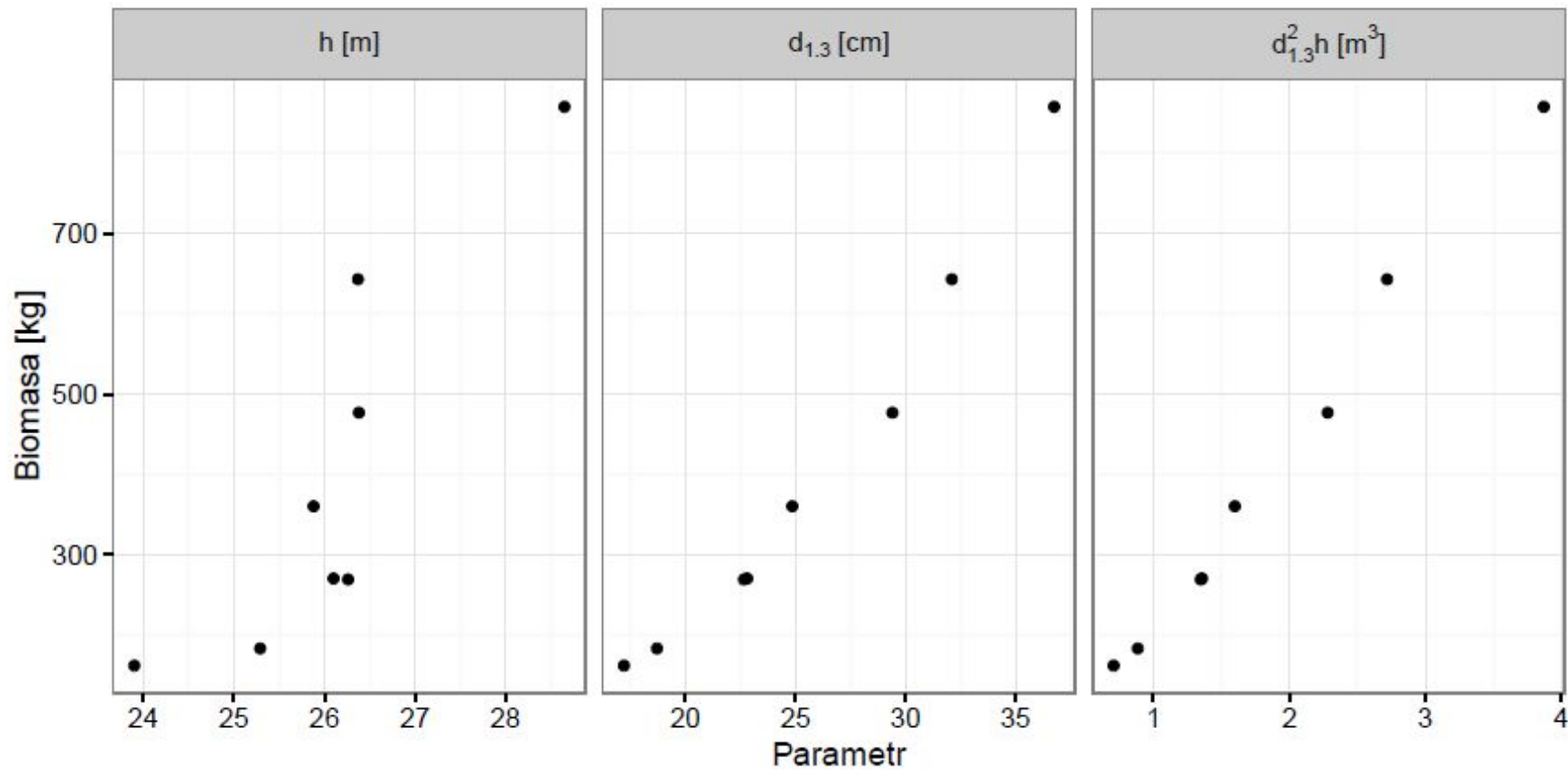
modele mogą być funkcjami liniowymi innymi wielomian 1. stopnia, np. parabole:

$y=ax^2+bx+c$ `lm(y~poly(x,2), dataset)`, $y=a*\log(x)+b$ `lm(y~log(x), dataset)`

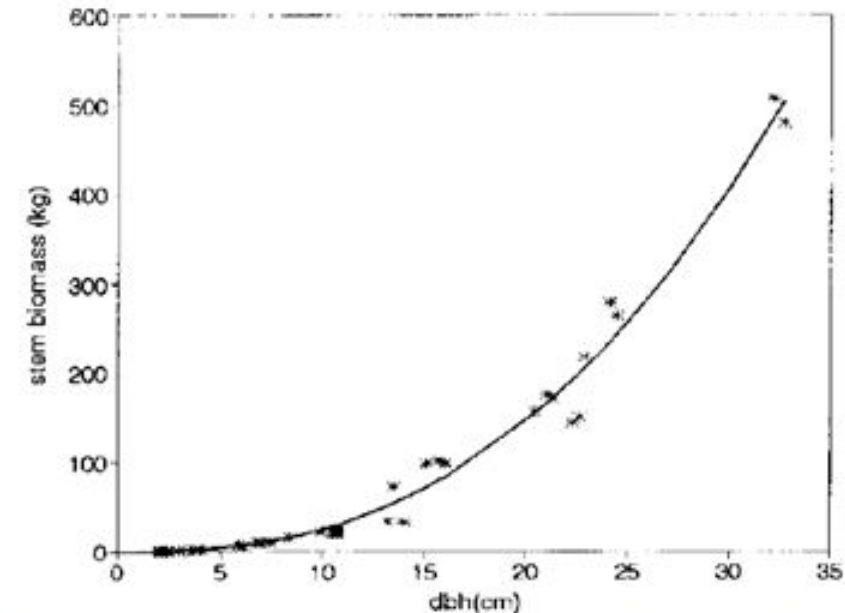
modele mogą mieć wiele zmiennych

$y=ax+by+cz+d$ `lm(y~a+b+c, dataset)`

Czasem lepsza jest krzywa;)



Równania allometryczne



Ann. Sci. For. (1997) 54, 39-50
© Elsevier/INRA

39

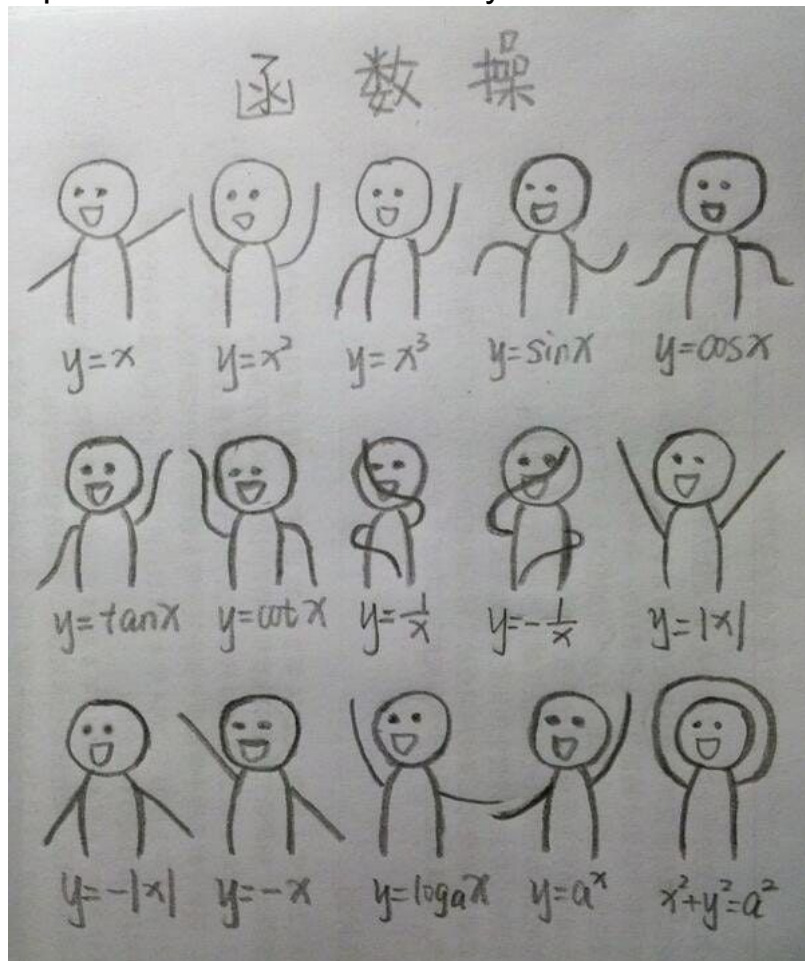
Original article

Allometric relationships for biomass and leaf area of
beech (*Fagus sylvatica* L.)

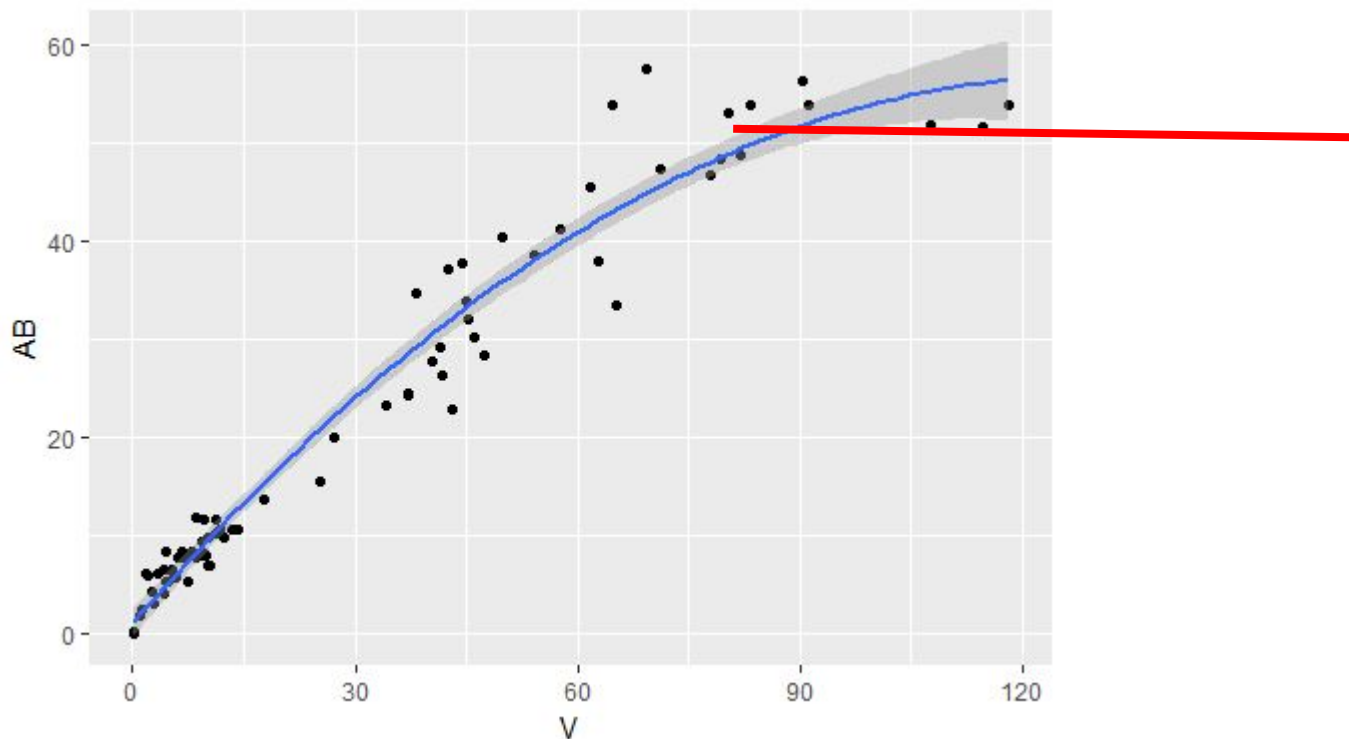
HH Baretlink



https://www.reddit.com/r/funny/comments/21h32s/dancing_math/



Parabole tańczą $y=ax^2+bx+c$



```
ggplot(sosny, aes(x=V,y=AB))+geom_point()+  
geom_smooth(method='lm',formula=y~poly(x,2))
```

```
> AIC(lm(AB~1,data=sosny)) #model zerowy
```

```
[1] 666.9126
```

```
> AIC(lm(AB~V,data=sosny)) #model liniowy
```

```
[1] 467.6902
```

```
> AIC(lm(AB~poly(V,2),data=sosny)) #model kwadratowy
```

```
[1] 422.4431
```


for all biomass components and volume. For each equation we calculated ten regression models:

$$W = a \times D^b$$

$$W = a + b \times D^2$$

$$W = a + b \times \log(D)$$

$$W = a + (b/D)$$

$$W = a \times (D^2 H)^b$$

$$W = a \times D^b \times H^c$$

$$W = a + b \times \log(D^2 H)$$

$$W = a + b \times D^2 + c \times H$$

$$W = a + b \times (D^2 H)$$

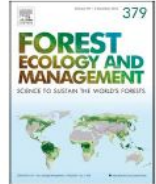
$$W = a + b \times D^2 + c \times H^2$$



Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



How do tree stand parameters affect young Scots pine biomass? – Allometric equations and biomass conversion and expansion factors



Andrzej M. Jagodziński^{a,b,*}, Marcin K. Dyderski^{a,b}, Kamil Gęsikiewicz^a, Paweł Horodecki^a, Agnieszka Cysewska^b, Sylwia Wierczyńska^b, Karol Maciejczyk^b

^a Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland

^b Poznań University of Life Sciences, Faculty of Forestry, Department of Game Management and Forest Protection, Wojska Polskiego 71c, 60-625 Poznań, Poland

(/)

(8)

(9)

(10)

implementacja

liniowy - `lm()` nieliniowy - `nls()`

```
model.liniowy<-lm(masa~D,data=dane)
```

```
model.liniowy<-lm(masa~log(D), data=dane)
```

```
modelnieliniowy<-nls(masa~a*D^b, data=dane, start=list(a=1,b=-1))
```

formuła, dane, (start)

model nieliniowy

```
nls(AB~a*V^b,data=sosny,start=list(a=1,b=1))
```

Nonlinear regression model

model: $AB \sim a * V^b$

data: sosny

a b

2.0119 0.7233

residual sum-of-squares: 1297

Number of iterations to convergence: 6

Achieved **convergence** tolerance: 1.784e-06

czasem model nie osiąga konwergencji -
można kazać R zmienić maksymalną
liczbę iteracji w nls:

```
nls.control(maxiter=500)
```

#domyślne 50 może nie dać rady
#uwaga - wydłuża czas obliczeń!

inna opcja to robust regression

```
library(robustbase)  
nlrob(AB~a*V^b,data=sosny,start=list(a=  
1,b=1))
```

pomaga przy heteroskedastyczności

```
nls(AB~a*V^b,data=sosny,start=list(a=1,b=exp(99999999)))
```

```
Error in numericDeriv(form[[3L]], names(ind), env) :
```

Brakuje wartości lub wyprodukowano wartości nieskończone podczas wyliczania modelu

Parametry startowe

literatura

brute force - dopasowanie metodą prób i błędów

dobra rada - zacząć od 1, -1, potem małe i duże cyfry

*przy funkcjach potęgowych - wykorzystać postać linearyzowaną

$$y=a*x^b$$

$$\log(y)=\log(a)+b*\log(x) \quad \text{czyli} \quad \text{lm}(\log(y) \sim \log(x), \text{data=dane})$$

nie ma r^2 :(i co ja teraz zrobię?

```
liczania modelu
> summary(nls(AB~a*v^b,data=sosny,start=list(a=1,b=-2)))

Formula: AB ~ a * v^b

Parameters:
      Estimate Std. Error t value Pr(>|t|)
a    2.01191    0.24523    8.204 4.88e-12 ***
b    0.72327    0.02917   24.793  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.159 on 75 degrees of freedom

Number of iterations to convergence: 11
Achieved convergence tolerance: 1.965e-06

>
```



policzę AIC i porównam z modelem zerowym

```
> AIC(lm(AB~1,data=sosny)) #model zerowy
```

```
[1] 666.9126
```

```
> AIC(lm(AB~V,data=sosny)) #model liniowy
```

```
[1] 467.6902
```

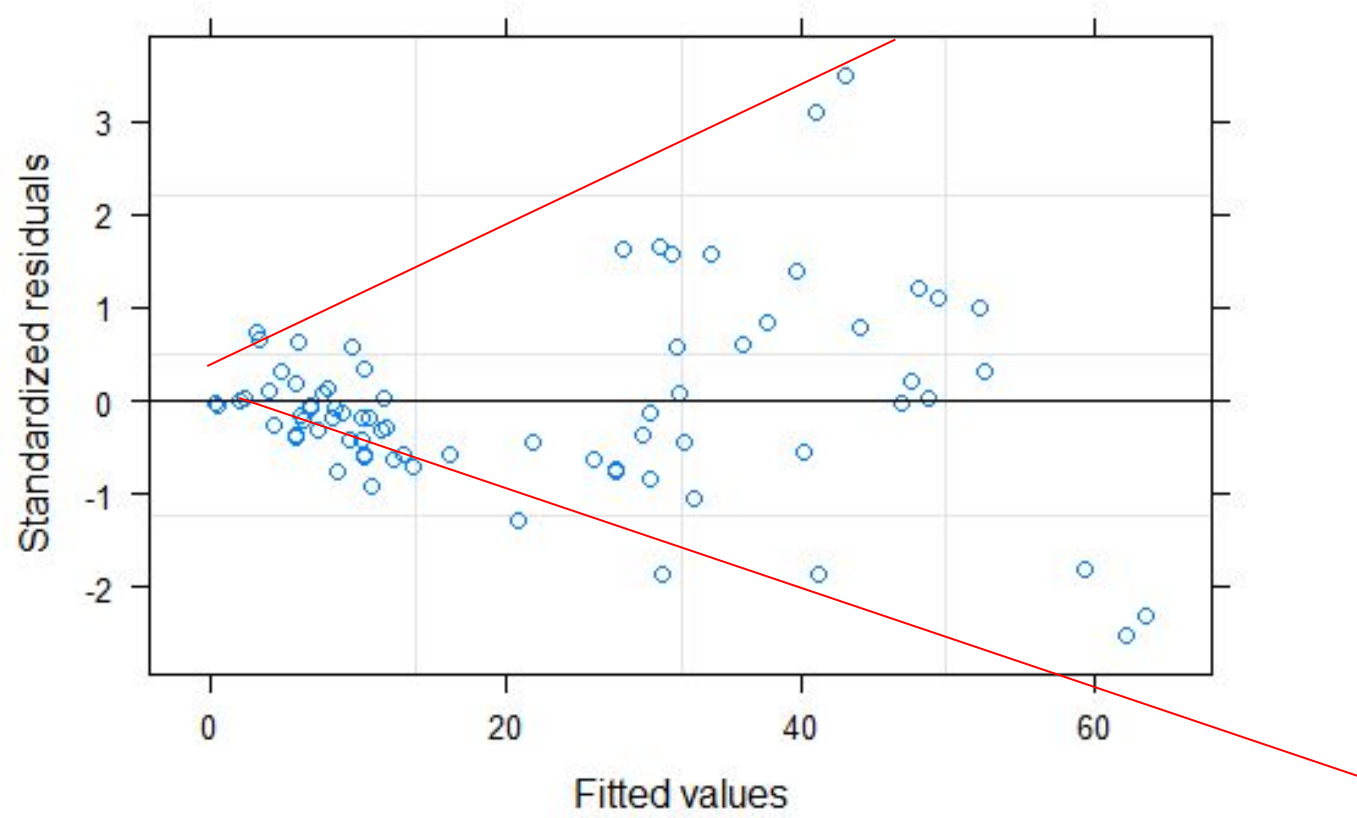
```
> AIC(lm(AB~poly(V,2),data=sosny)) #model kwadratowy
```

```
[1] 422.4431
```

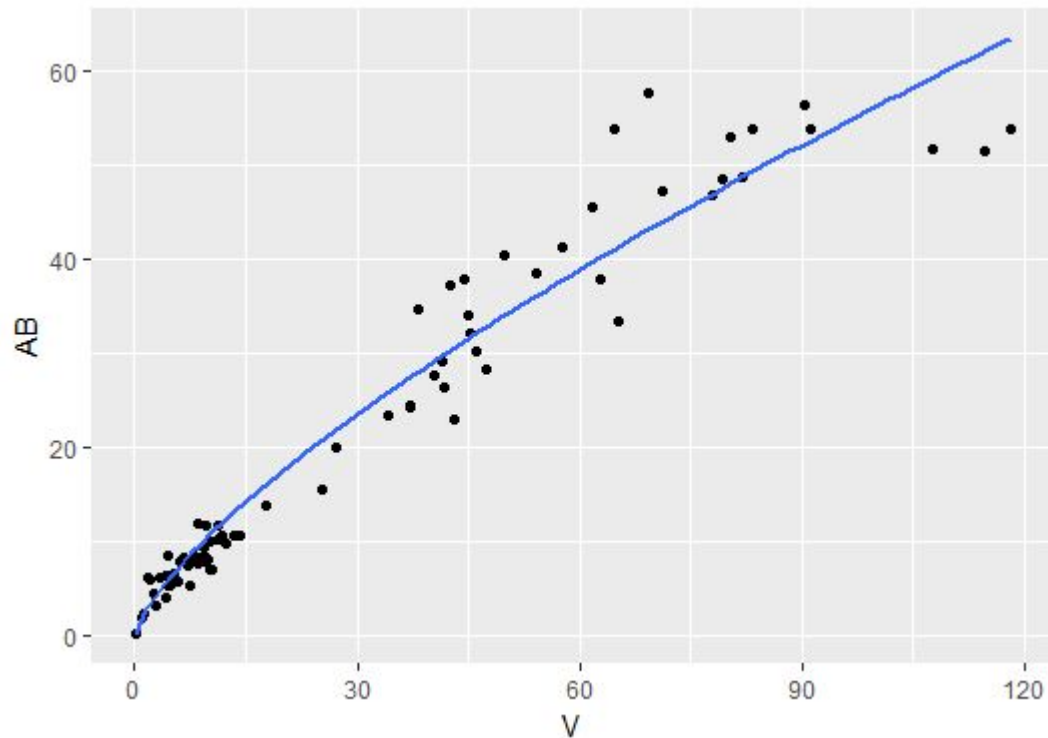
```
> AIC(nls(AB~a*V^b,data=sosny,start=list(a=1,b=-2)))
```

```
[1] 441.9817
```

diagnostyka modelu:




```
ggplot(sosny, aes(y=AB,x=V))+geom_point()  
+geom_smooth(method='nls',formula=y~a*x^b,  
method.args=list(start=list(a=1,b=1)),se=F)
```



se=F to coś bez czego nie ruszy
zawsze zwróci error!
parametry startowe - jak
wcześniej;)

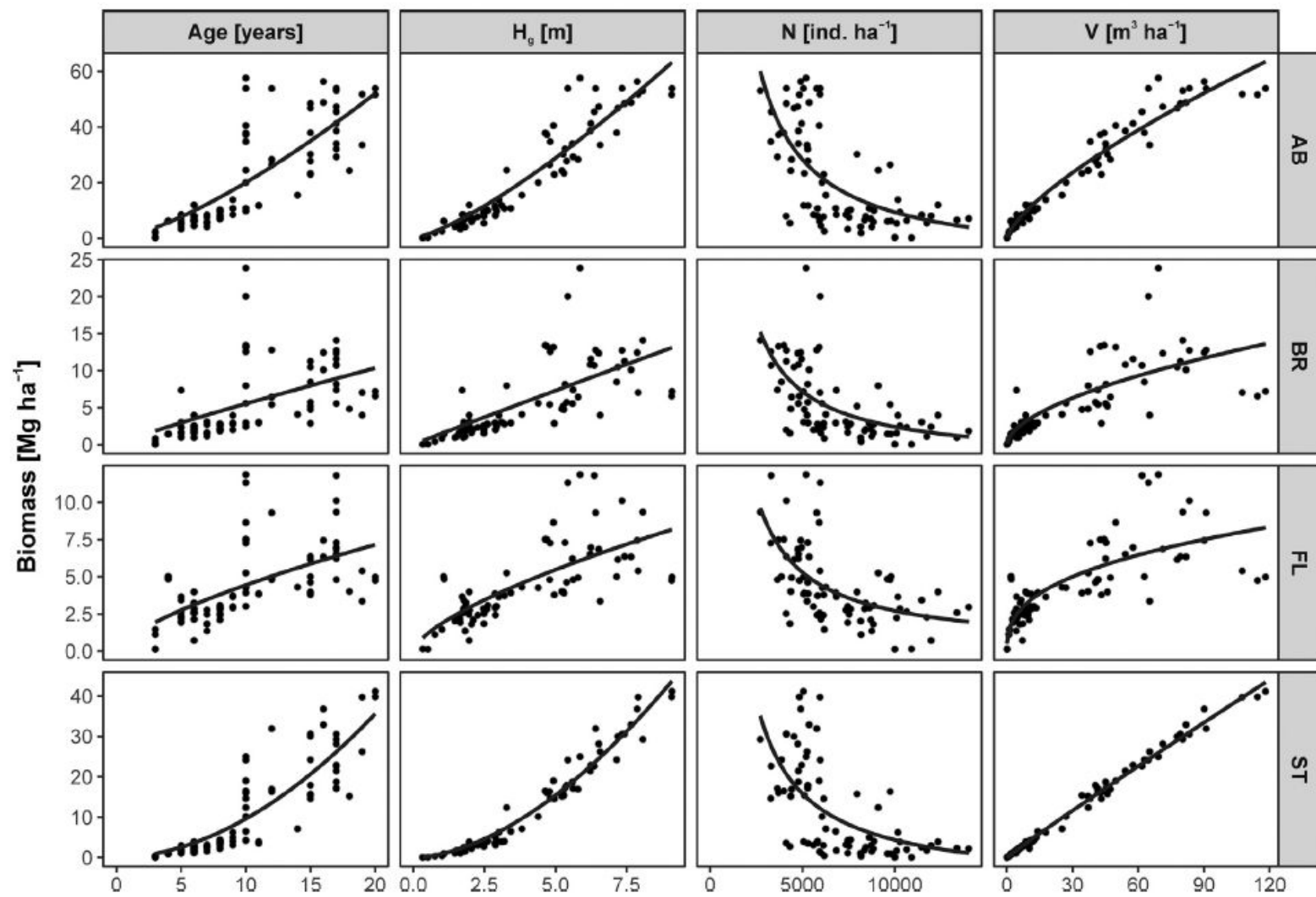
dlaczego tak?
za dużo punktów małych!
więcej nie znaczy lepiej;)

$$BCEF = a \times z^b \quad (\text{e. g. Peichl and Arain, 2007}) \quad (11)$$

$$BCEF = a + b \times e^{-z \times c} \quad (\text{e. g. Lehtonen et al. , 2004; Peichl and Arain, 2007; Jagodziski et al. , 2017}) \quad (12)$$

$$BCEF = a + b/z \quad (\text{e. g. Teobaldelli et al. ,2009; Wojtan et al. , 2011}) \quad (13)$$

$$BCEF = a + b/z^c \quad (\text{e. g. Teobaldelli et al. , 2009}) \quad (14)$$



jak prezentować wyniki?

Table 7
Relationships between tree stand characteristics (predictors) and BCEFs for particular biomass components [Mg m^{-3}].

Biomass component	Predictor	Model type (Eq. no.)	a	SE	b	SE	c	SE	RMSE	R ²	AIC	AIC ₀
AB	A	(12)	0.61691	0.06542	5.33448	1.03961	0.35147	0.05317	< 0.0001	0.709	44.473	134.220
	H _g	(12)	0.62412	0.05746	3.26428	0.35316	0.87235	0.10597	< 0.0001	0.729	39.084	–
	N	(11)	0.00376	0.00537	0.63451	0.15950	–	–	0.0119	0.186	120.599	–
	V	(12)	0.67209	0.04559	2.09454	0.15847	0.22621	0.02754	< 0.0001	0.764	28.410	–
BR	A	(12)	0.07917	0.07343	0.99264	0.22166	0.18209	0.06920	< 0.0001	0.438	–42.693	–2.843
	H _g	(12)	0.10830	0.06310	0.69410	0.13060	0.45800	0.18110	< 0.0001	0.378	–34.930	–
	N	(13)	0.46840	0.07462	–1080.0	422.1	–	–	< 0.0001	0.081	–7.283	–
	V	(12)	0.16099	0.03062	0.55175	0.08779	0.14350	0.03963	< 0.0001	0.428	–41.339	–
FL	A	(12)	0.09562	0.05773	5.67080	1.45618	0.42193	0.07180	< 0.0001	0.637	42.380	115.452
	H _g	(12)	0.08925	0.05133	3.09855	0.38049	1.01347	0.13006	< 0.0001	0.682	31.089	–
	N	(13)	0.91070	0.15630	–3192.17130	884.35630	–	–	< 0.0001	0.150	32.470	–
	V	(12)	0.12295	0.04423	1.86903	0.16629	0.26766	0.03813	< 0.0001	0.700	27.958	–
ST	A	(12)	0.37645	0.01009	6.30763	9.27127	1.11956	0.48019	< 0.0001	0.316	–174.461	–149.642
	H _g	(12)	0.38054	0.00548	3.65437	1.24671	5.30087	0.92841	< 0.0001	0.736	–246.734	–
	N	(11)	0.14574	0.08838	0.11308	0.06883	–	–	< 0.0001	0.035	–150.317	–
	V	(13)	0.37748	0.00526	0.03547	0.00240	–	–	< 0.0001	0.747	–252.212	–

skubany, skąd masz R^2 ?

R^2 dotyczy tylko modeli liniowych

w pozostałych przypadkach można podać (zaznaczając to wyraźnie) - **pseudo- R^2**

czasem nazywane współczynnikiem efektywności modelu (model efficiency; MEf)
lub Nash-Sutcliffe Efficiency (NSE)

bo czym jest tak naprawdę R^2 ?

1-suma kwadratów odchyłeń modelu/suma kwadratów odchyłeń rzeczywistych

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}.$$

* Jak obliczyć R² ręcznie?

```
model<-nls(AB~a*V^b,data=sosny,start=list(a=1,b=-2))
```

```
1-(sum(residuals(model)^2)/sum((sosny$AB-mean(sosny$AB))^2))
```

czyli 1 -

suma kwadratów reszt (reziduów): $\text{sum}(\text{residuals}(\text{model})^2)$

przez

suma kwadratów odchyłeń $(\text{sosny\$AB} - \text{mean}(\text{sosny\$AB}))^2$

GLM

uogólniony model liniowy - generalized linear model

dlaczego uogólniony?

model liniowy zakłada rozkład normalny, GLM uogólnia metodę na inne rozkłady

po co? do innych rozkładów

(przypominamy sobie przykłady)

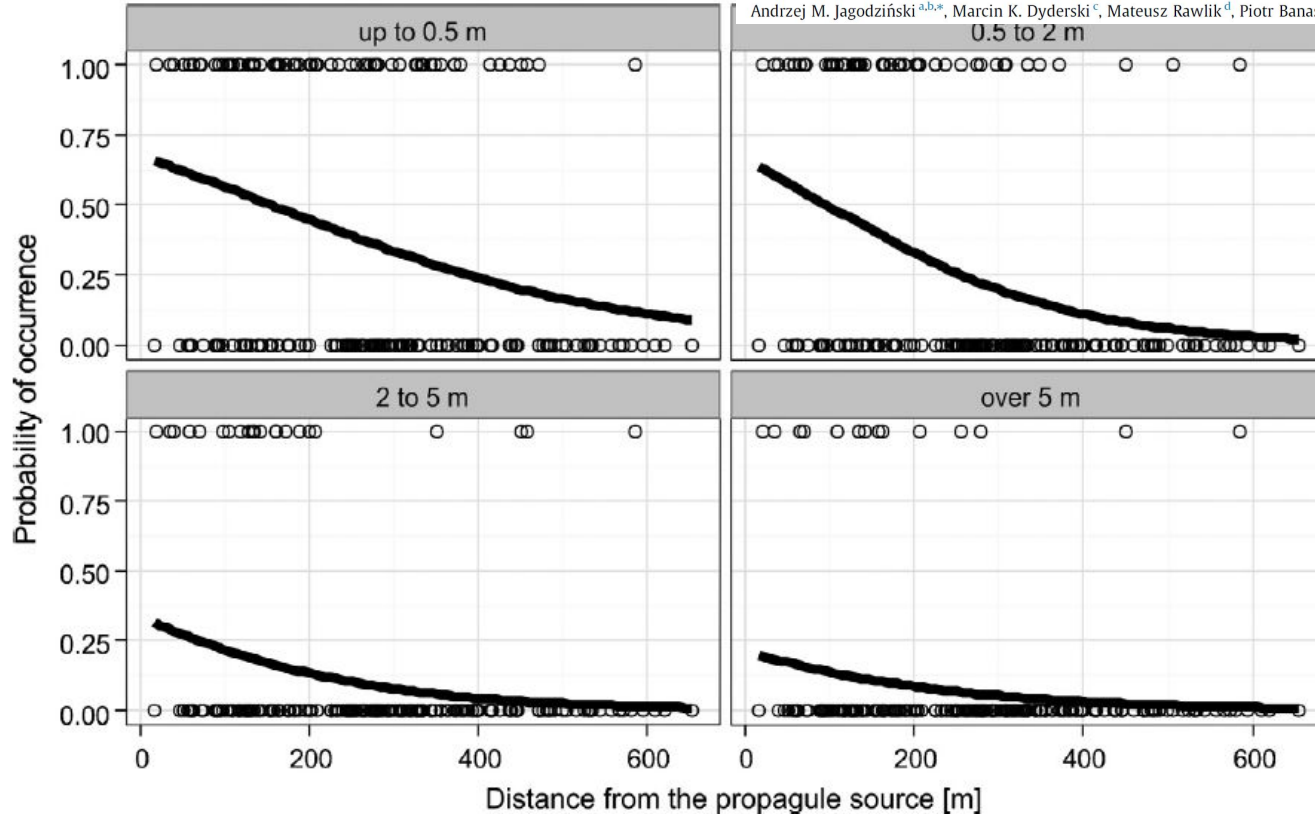
GLM z innymi rozkładami

`glm(formula, data, family='rozkład')`

`family='poisson'` - rozkład Poissona (dyskretny - liczby naturalne - liczba osobników, liczba gatunków)

`family=binomial(link='logit')` - regresja logistyczna (0/1 - np. przeżycie, występowanie lub brak - klasyfikacja binarna)

regresja logistyczna



Plantation of coniferous trees modifies risk and size of *Padus serotina* (Ehrh.) Borkh. invasion – Evidence from a Rogów Arboretum case study

Andrzej M. Jagodziński^{a,b,*}, Marcin K. Dyderski^c, Mateusz Rawlik^d, Piotr Banaszcak^e



```

> model<-glm(szens~odl,family=binomial(link="logit"))
> summary(model)

Call:
glm(formula = szans ~ odl, family = binomial(link = "logit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.4374  -0.9748  -0.7137   1.1419   1.9901

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.661478   0.318846   2.075 0.038024 *
odl          -0.004263   0.001120  -3.806 0.000141 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

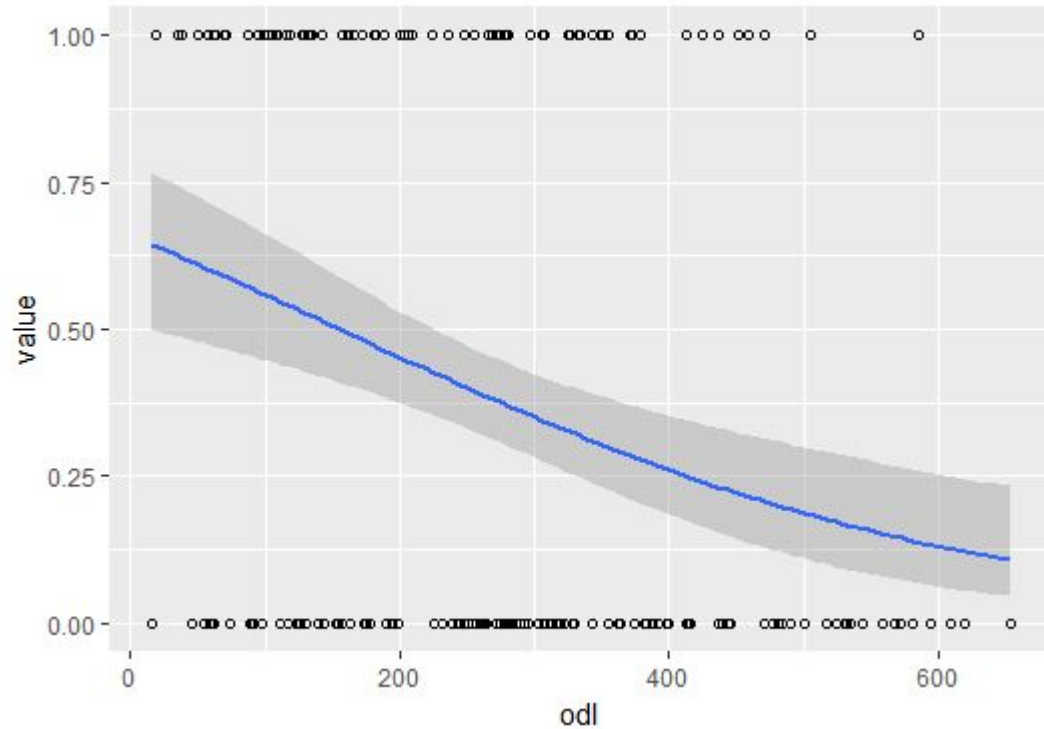
    Null deviance: 268.48  on 200  degrees of freedom
Residual deviance: 252.23  on 199  degrees of freedom
AIC: 256.23

Number of Fisher Scoring iterations: 4

> predict(model)

```

```
ggplot(data=fig4df, aes(x=odl, y=value))+geom_point(shape=1)  
+geom_smooth(method="glm", method.args=c(family="binomial"))
```



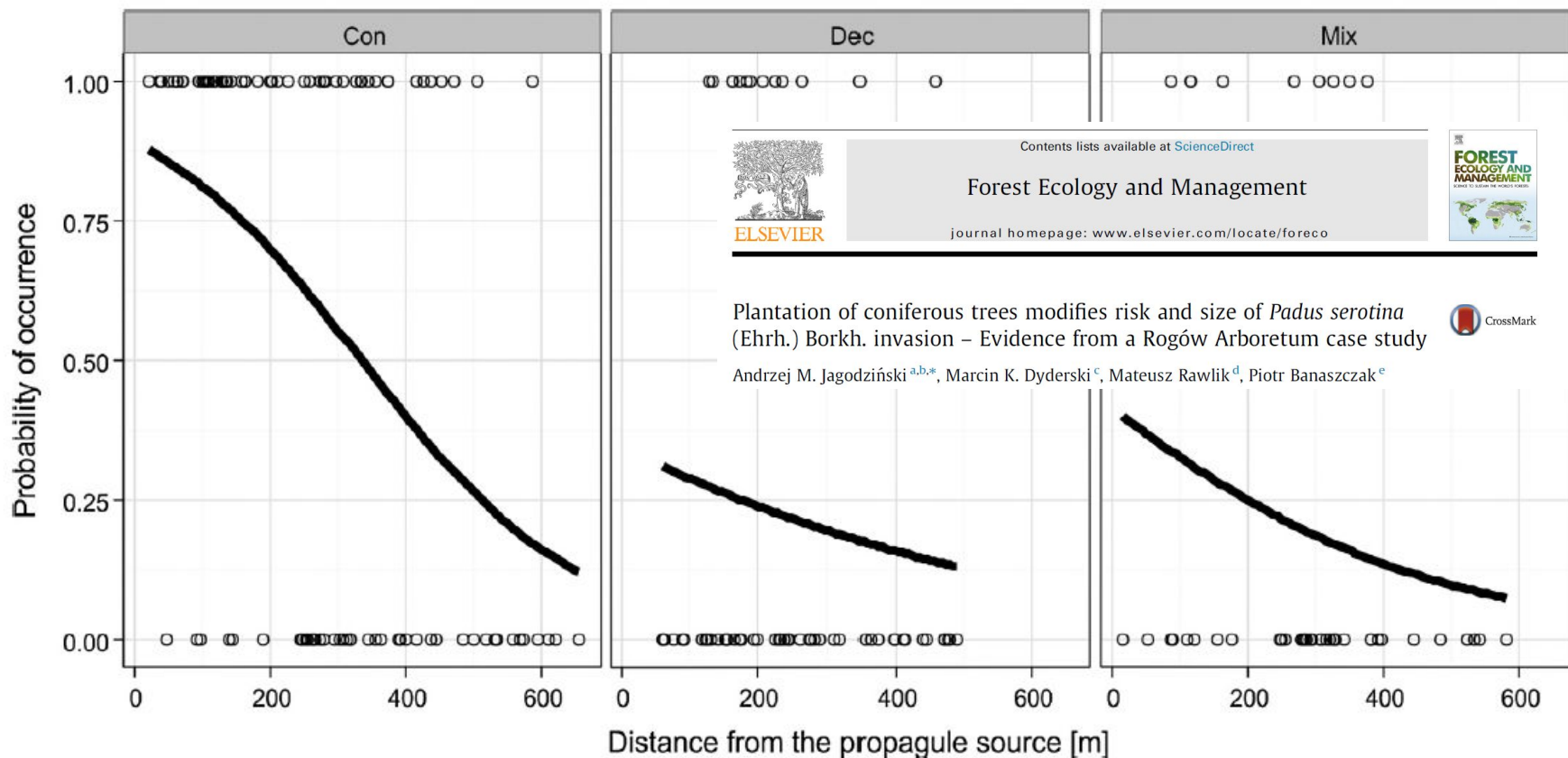


Fig. 4. Effect of coniferous (Con), deciduous (Dec) and mixed (Mix) tree stands on the probability of colonisation by black cherry ($p < 0.001$).

GLM z rozkładem Poissona

```
model<-glm(prunusc~richness,data=prunus)
```

```
summary(model)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.5654	-1.7546	-0.7967	0.2663	15.6279

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.61911	1.70456	-0.950	0.3482
richness	0.19327	0.08731	2.214	0.0329 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 17.97445)

Null deviance: 771.10 on 39 degrees of freedom
Residual deviance: 683.03 on 38 degrees of freedom
AIC: 233.02

```
model<-glm(prunusc~richness,data=prunus, family=poisson)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.3062	-1.5175	-1.1999	-0.7244	8.0114

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.46686	0.38989	-3.762	0.000168 ***
richness	0.09892	0.01544	6.408	1.48e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 248.48 on 39 degrees of freedom
Residual deviance: 203.69 on 38 degrees of freedom
AIC: 244.38

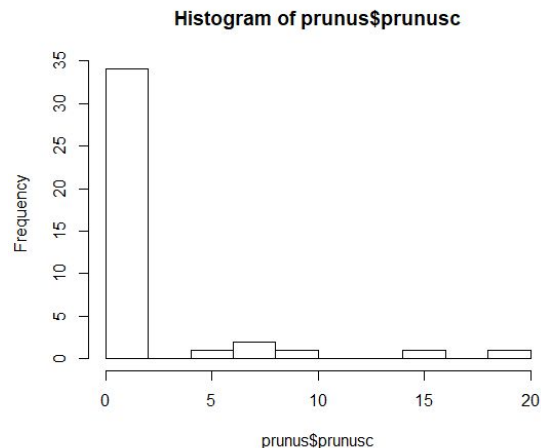
Number of Fisher Scoring iterations: 7

zwiększyło się AIC

teoretycznie jest to gorszy model, ale!

do wykonania GLM jesteśmy uprawnieni
mając dane o rozkładzie normalnym

logika i założenia modelu > cyferki



Problem z zerami

Rozkład Poissona - rozkład teoretyczny

Rzeczywistość - puste próby, brak występowania

Usunąć? informacja biologiczna - jak wnioskować o korniku nie badając miejsc gdzie go nie ma?

Zostawić - inflacja zer - model mając wiele zer "idzie na łatwiznę" - częściej przewiduje zero niż by wynikało z rzeczywistości

*Problem z zerami - zero-inflated Poisson models

Tzw. modele złożone (hurdle models) - składają się z dwóch części:

1. modeluje prawdopodobieństwo wystąpienia (binomial) - czy w ogóle coś będzie
2. modeluje liczebność - czyli właściwy Poisson

Jak to zrobić?

```
library(pscl)
```

```
model3<-zeroinfl(prunusc~richness|richness, data=prunus)
```

#kreska jest ważna

#zakładamy że prunusc zależy od richness, a obecność również richness

#można tu dać różne parametry, np. prunusc~richness|1 lub

```
prunusc~a+richness|richness+N
```


Wynik

Pearson residuals:

Min	1Q	Median	3Q	Max
-0.9613	-0.4931	-0.4152	-0.2490	7.7762

Count model coefficients (poisson with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.92540	0.39316	2.354	0.01858 *
richness	0.04274	0.01545	2.766	0.00567 **

Zero-inflation model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.52589	1.01872	2.479	0.0132 *
richness	-0.08254	0.04757	-1.735	0.0828 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 11

Log-likelihood: -64.92 on 4 Df

AIC(model, model2, model3)

	df	AIC
model	3	233.0214
model2	2	244.3800
model3	4	137.8487

Choć GLM Poisson (2) był gorszy niż GLM zakładający rozkład normalny to biorąc pod uwagę inflację zer mamy jeszcze lepsze dopasowanie

zgodność wszystkich założeń modelu daje dobry wynik!

Table 1 Zero-inflated Poisson generalized linear models of natural regeneration of *Picea abies* density for three age-height classes: seedlings (germinated in the current year), short saplings (≥ 1 year old, < 50 cm height), and tall saplings (50–130 cm height). AIC of final models: seedlings 1552.0 ($AIC_0 = 1674.0$), short saplings 3785.7 ($AIC_0 = 4246.6$), and tall saplings 1113.9 ($AIC_0 = 1271.5$)

Seedlings				
Term	Estimate	SE	z	$P(> z)$
Count model coefficients (Poisson with log link)				
(Intercept)	2.176	0.046	47.428	< 0.001
Distance to the riverbank	0.295	0.035	8.519	< 0.001
Soil pH	−0.221	0.069	−3.199	0.001
Zero-inflation model coefficients (binomial with logit link)				
(Intercept)	0.323	0.161	2.012	0.044
Short saplings				
Term	Estimate	SE	z	$P(> z)$
Count model coefficients (Poisson with log link)				
(Intercept)	2.969	0.028	107.166	< 0.001
Distance to the riverbank	0.106	0.017	6.214	< 0.001
DIFN	0.042	0.017	2.403	0.016
Soil pH	−0.578	0.036	−16.110	< 0.001
Zero-inflation model coefficients (binomial with logit link)				
(Intercept)	−0.640	0.168	−3.817	< 0.001
Tall saplings				
Term	Estimate	SE	z	$P(> z)$
Count model coefficients (Poisson with log link)				
(Intercept)	1.680	0.076	22.214	< 0.001
Distance to the riverbank	0.202	0.036	5.563	< 0.001
DIFN	−0.103	0.052	−1.982	0.048
Soil pH	−0.663	0.087	−7.595	0.000
Zero-inflation model coefficients (binomial with logit link)				
(Intercept)	0.519	0.170	3.058	0.002



Inne rozkłady i warianty modeli złożonych

Rozkład Rayleigha - używany do modelowania zagęszczeń zwierzyny w lesie (choć hurdle models są lepsze)

Rozkład beta - do modelowania proporcji (w zakresie 0-1) - np. udział form życiowych, proporcje grup wiekowych, itp.

Warianty zero-inflated, one-inflated, zero-and-one-inflated

pakiet gamlss - dowolne rozkłady

Impacts of soil conditions and light availability on natural regeneration of Norway spruce *Picea abies* (L.) H. Karst. in low-elevation mountain forests


Marcin K. Dyderski¹  • Anna Gazda² • Mariusz Hachułka³ • Paweł Horodecki¹ • Izabela L. Kałucka^{3,4} • Jacek Kamczyc⁵ • Marek Malicki⁶ • Remigiusz Pielech² • Michał Smoczyk⁷ • Maciej Skorupski⁵ • Sylwia Wiercholska⁸ • Andrzej M. Jagodziński¹

Table 2 Zero- and one-inflated beta regression models of natural regeneration of *Picea abies* age-height class proportions in the regeneration layer: seedlings (germinated in the current year), short saplings (≥ 1 year old, < 50 cm height), and tall saplings (50–130 cm height). AIC of final models: seedlings 245.0 (AIC₀ = 247.5), short saplings 293.3 (AIC₀ = 299.4), and tall saplings 207.4 (AIC₀ = 210.0)

Seedlings				
Term	Estimate	SE	<i>t</i>	<i>P</i> (> <i>t</i>)
μ estimation, link function: logit				
(Intercept)	− 0.438	0.186	− 2.354	0.020
Distance to the riverbank	− 0.026	0.013	− 2.075	0.040
σ estimation, link function: logit				
(Intercept)	0.205	0.126	1.625	0.106
ν estimation, link function: log				
(Intercept)	0.438	0.166	2.647	0.009
τ estimation, link function: log				
(Intercept)	− 2.148	0.399	− 5.379	< 0.001
Short saplings				
Term	Estimate	SE	<i>t</i>	<i>P</i> (> <i>t</i>)
μ estimation, link function: logit				
(Intercept)	4.467	1.375	3.248	0.001
pH	− 1.002	0.351	− 2.855	0.005
σ estimation, link function: logit				
(Intercept)	0.051	0.105	0.490	0.625
ν estimation, link function: log				
(Intercept)	− 0.393	0.173	− 2.274	0.024
τ estimation, link function: log				
(Intercept)	− 1.374	0.244	− 5.626	< 0.001
Tall saplings				
Term	Estimate	SE	<i>t</i>	<i>P</i> (> <i>t</i>)
μ estimation, link function: logit				
(Intercept)	− 4.763	1.705	− 2.794	0.006
pH	0.956	0.434	2.202	0.029
σ estimation, link function: logit				
(Intercept)	− 0.035	0.138	− 0.251	0.802
ν estimation, link function: log				
(Intercept)	0.684	0.170	4.018	< 0.001
τ estimation, link function: log				
(Intercept)	− 2.342	0.468	− 5.001	< 0.001



Jak to pokazać na obrazku?

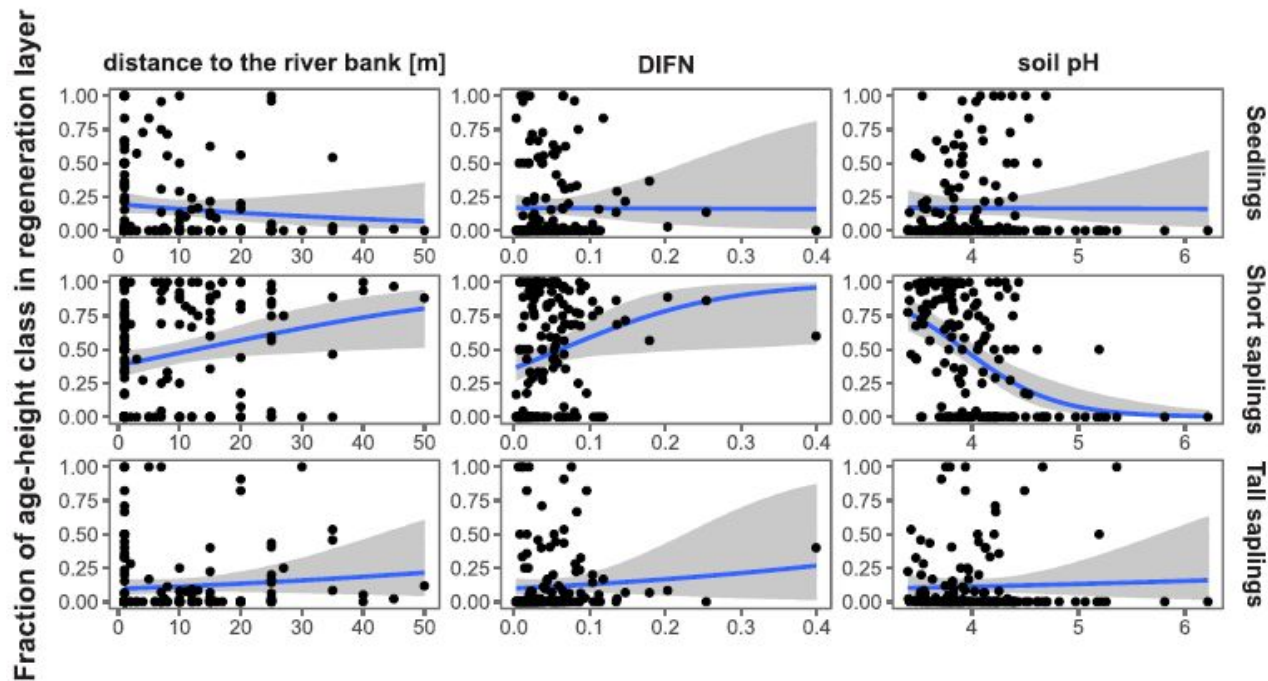


Fig. 2 Relationships between proportion of age-height classes of natural regeneration of *Picea abies* and three environmental variables. Classes: seedlings (germinated in the current year), short saplings (≥ 1 year old, < 50 cm height), and tall saplings (50–130 cm height). Lines represent fitted

generalized linear models with binomial distribution for visualization of general trends. See Table 2 for zero- and one-inflated regression models and their parameters

W prosty sposób
prostym GLMem

piszemy że linia jest
przybliżeniem, a model
jest w tabelce

Testy statystyczne a GLM

Test chi-kwadrat?

Binomial GLM?

Rozbudowa o inne parametry

Porównanie z modelem zerowym (intercept-only, $y \sim 1$)

data.frame: 0/1 występowanie, zmienna grupująca (wrzucamy i 0 i 1)

```
mm<-glm(Ficavern~OLDFR,data=afis, family=binomial(link='logit'))
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.1390	-0.7832	-0.5448	-0.5448	1.9905

data=afis

dataset do ćwiczeń;)

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.8326	0.2408	-7.609	2.75e-14 ***
OLDFRoldforest	1.7416	0.3861	4.510	6.47e-06 ***
OLDFRrecentforest	0.8081	0.3938	2.052	0.0402 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 259.40 on 241 degrees of freedom

Residual deviance: 238.45 on 239 degrees of freedom

AIC: 244.45

Number of Fisher Scoring iterations: 4

Uogólnione modele addytywne - GAM

Generalized Additive Models

Po co? nieparametryczne - nie zawsze rozkłady odpowiadają rzeczywistym danym i założeniom modeli

Co zamiast funkcji liniowych? Funkcje sklepane (splines)

Zamiast składników liniowych - składniki nieliniowe

Funkcje sklejane (splines)

dla $x \in (-\infty, -10)$ $y = 12x + 2$

dla $x \in [-10, 2)$ $y = 17^x$

dla $x \in [2, 7)$ $y = 2 - x^3$

dla $x \in [7, \infty)$ $y = 2x^2 - 3.122$

mamy cztery funkcje, sklejone w jedną (wygładzoną)

Mamy trzy punkty przegięcia - w których funkcja zmienia bieg

Przykład zastosowania

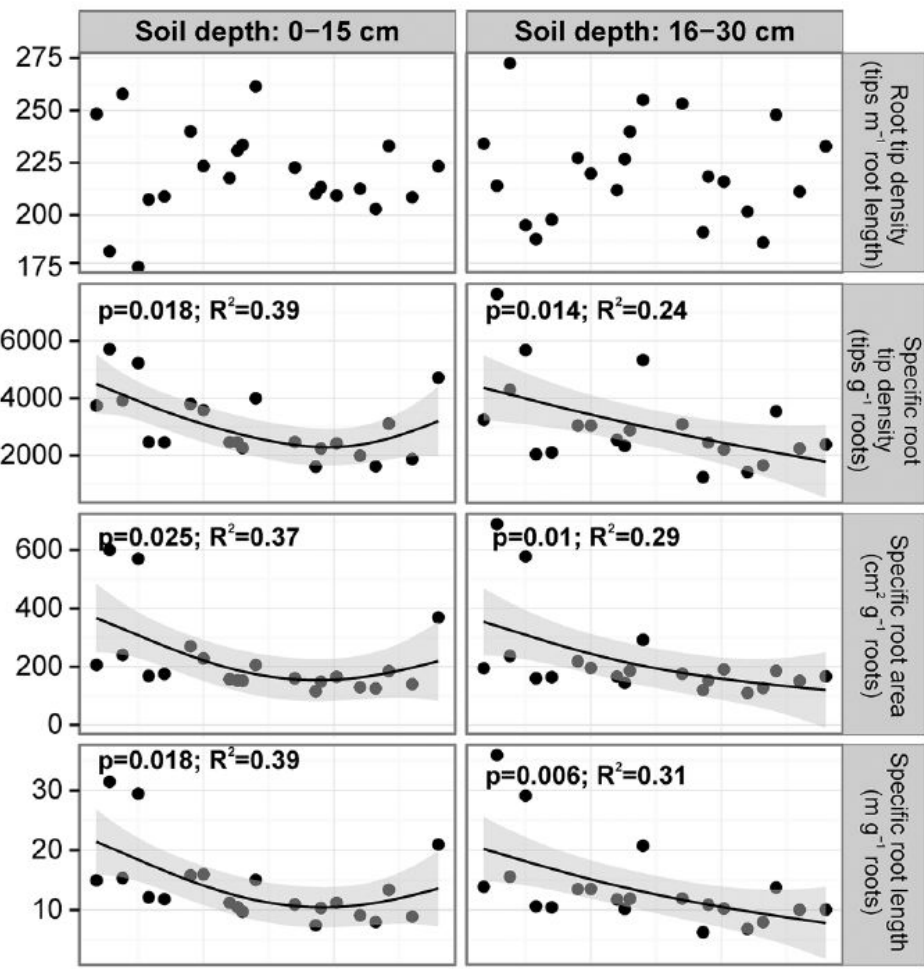
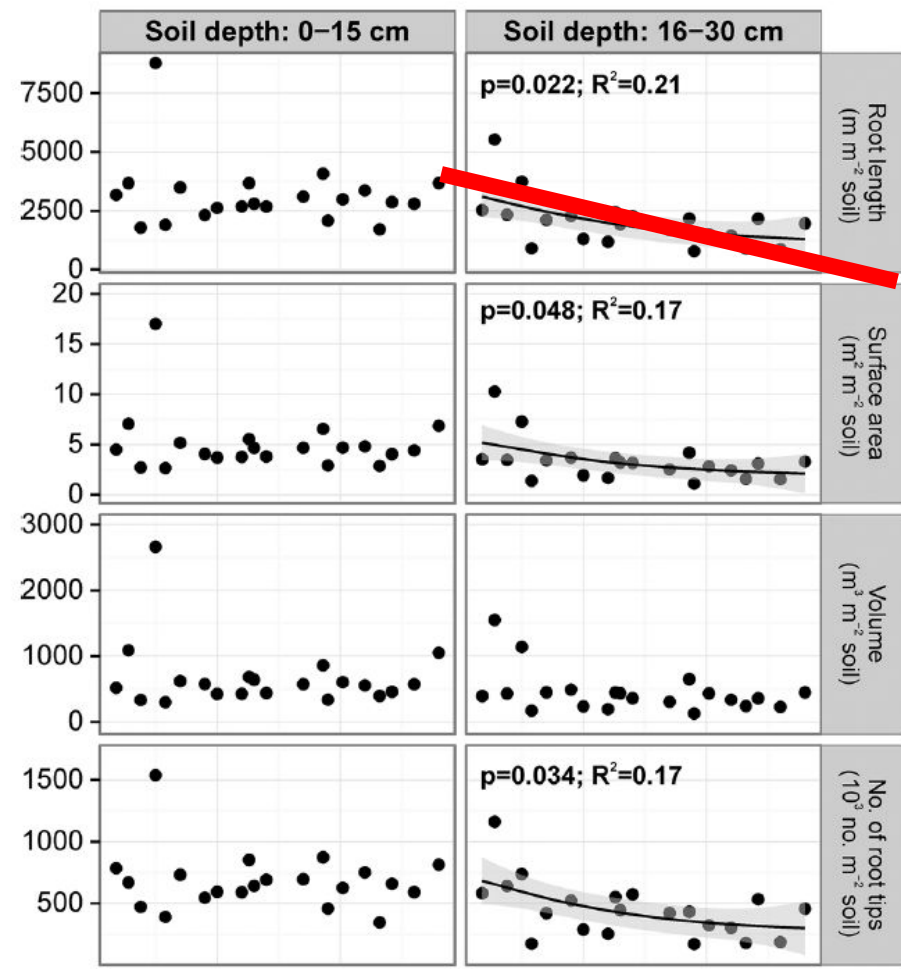
wzorce rozwoju korzeni drobnych

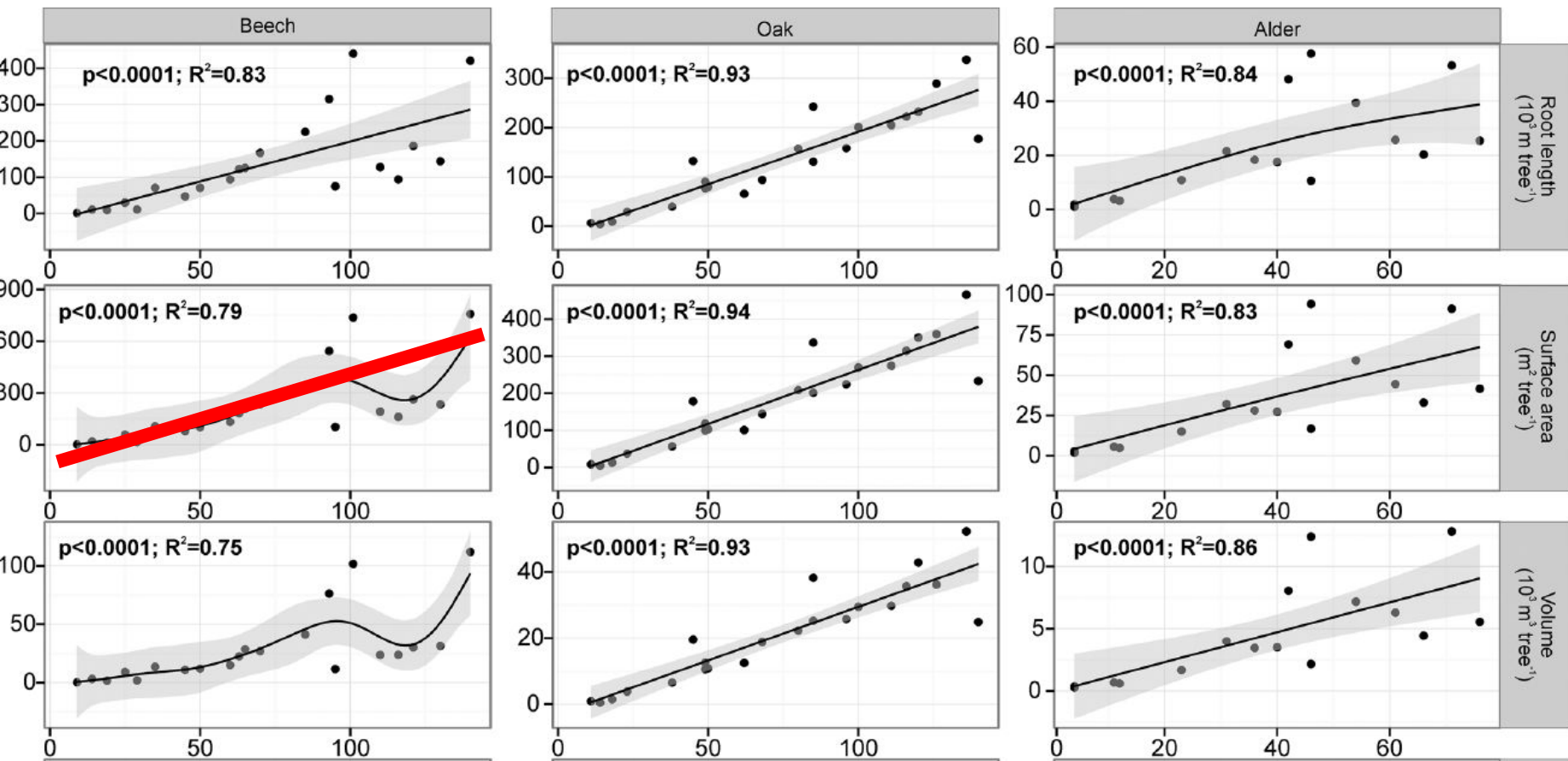


RESEARCH ARTICLE

Tree Age Effects on Fine Root Biomass and Morphology over Chronosequences of *Fagus sylvatica*, *Quercus robur* and *Alnus glutinosa* Stands

Andrzej M. Jagodzinski^{1,2*}, Jędrzej Ziółkowski², Aleksandra Warnkowska², Hubert Prais²





```

> model1<-gam(AB~s(V),data=sosny)
> library(mgcv)
> model1<-gam(AB~s(V),data=sosny)
> summary(model1)

```

```

Family: gaussian
Link function: identity

```

```

Formula:
AB ~ s(V)

```

```

Parametric coefficients:

```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	21.5752	0.3845	56.11	<2e-16 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Approximate significance of smooth terms:

```

	edf	Ref.df	F	p-value
s(V)	3.701	4.546	460.4	<2e-16 ***

```

---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

R-sq.(adj) =  0.965   Deviance explained = 96.7%
GCV = 12.125   Scale est. = 11.385       n = 77

```

```

>

```

w modelu zamiast V
mamy s(V), czyli
funkcję sklejaną

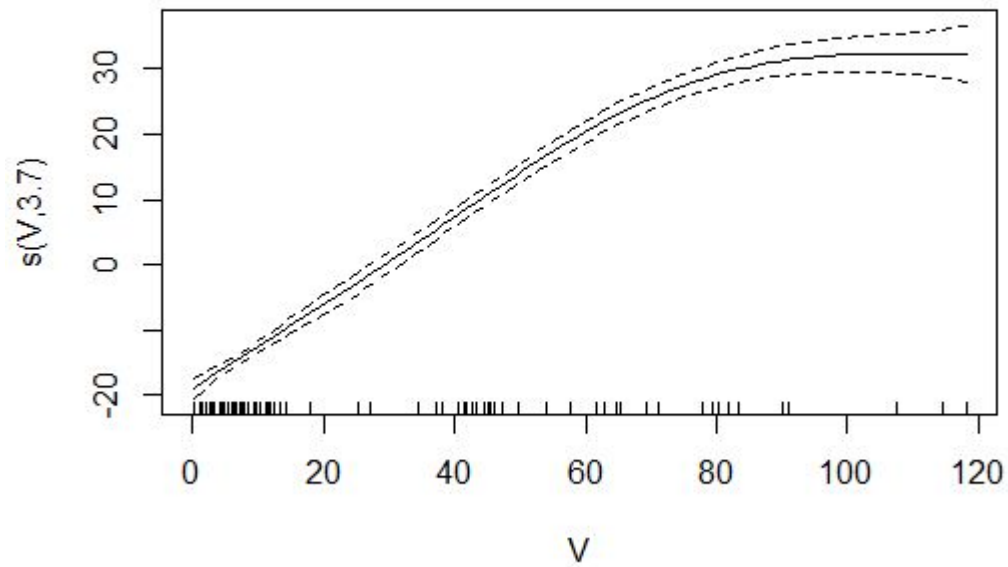
jest to wielomian
stopnia 3,7
(patrzemy na edf)

edf - estimated degrees
of freedom

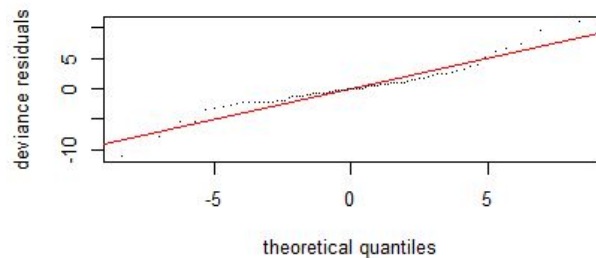
Ref.df - reference
degrees of freedom

F - statystyka testowa

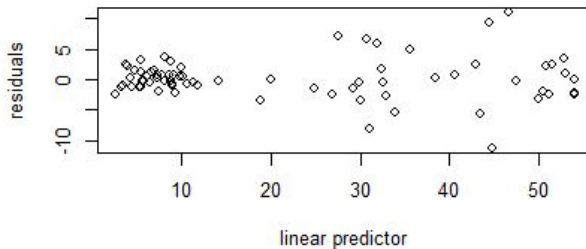
plot(model1)



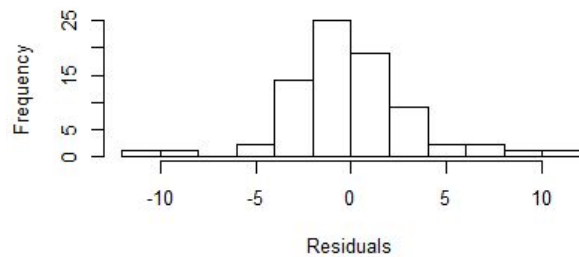
```
par(mfrow=c(2,2))  
gam.check(model1)
```



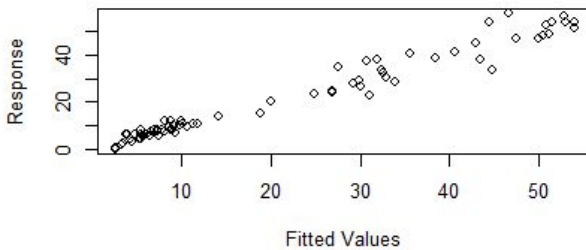
Resids vs. linear pred.



Histogram of residuals



Response vs. Fitted Values



```
> AIC(model1)
```

```
[1] 412.3526
```

```
> AIC(model0)
```

```
[1] 666.9126
```


poprzednie modele

```
> AIC(lm(AB~1,data=sosny)) #model zerowy
```

```
[1] 666.9126
```

```
> AIC(lm(AB~V,data=sosny)) #model liniowy
```

```
[1] 467.6902
```

```
> AIC(lm(AB~poly(V,2),data=sosny)) #model kwadratowy
```

```
[1] 422.4431
```

```
> AIC(nls(AB~a*V^b,data=sosny,start=list(a=1,b=-2))) #model potęgowy
```

```
[1] 441.9817
```

```
>AIC(gam(AB~s(V),data=sosny)) #GAM
```

```
[1] 412.3526
```

```
> plot(modelbr)
> modelbr<-gam(BR~s(Age),data=sosny)
> summary(modelbr)
```

Family: gaussian
Link function: identity

Formula:
BR ~ s(Age)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.4899	0.3988	13.77	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

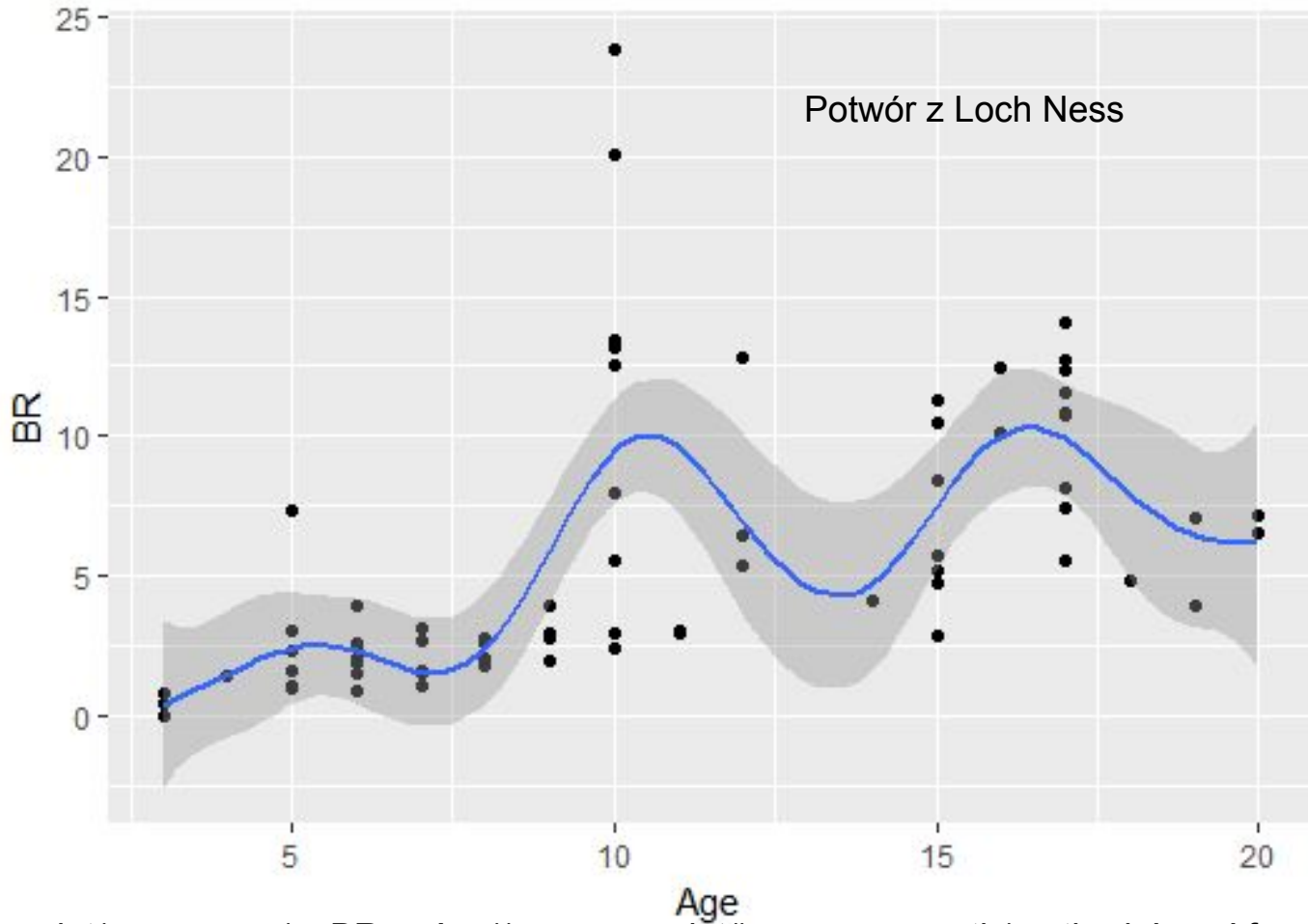
Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(Age)	7.751	8.566	9.005	3.45e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.486 Deviance explained = 53.9%
GCV = 13.814 Scale est. = 12.244 n = 77

```
> |
```



```
ggplot(sosny, aes(y=BR,x=Age))+geom_point()+geom_smooth(method='gam',formula=y~s(x))
```

With **four** parameters I can fit an elephant, and with **five** I can make him wiggle his trunk.

Drawing an elephant with four complex parameters

Jürgen Mayer
Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden,
Germany

Khaled Khairy
European Molecular Biology Laboratory, Meyerhofstraße, 1, 69117 Heidelberg, Germany

Jonathan Howard
Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden,
Germany

(Received 20 August 2008; accepted 5 October 2009)

We define four complex numbers representing the parameters needed to specify an elephantine shape. The real and imaginary parts of these complex numbers are the coefficients of a Fourier coordinate expansion, a powerful tool for reducing the data required to define shapes. © 2010 American Association of Physics Teachers
[DOI: 10.1119/1.3254017]

A turning point in Freeman Dyson's life occurred during a meeting in the Spring of 1953 when Enrico Fermi criticized the complexity of Dyson's model by quoting Johnny von Neumann: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Since then it has become a well-known saying among physicists, but nobody has successfully implemented it.

To parametrize an elephant, we note that its perimeter can be described as a set of points $(x(t), y(t))$, where t is a parameter that can be interpreted as the elapsed time while going along the path of the contour. If the speed is uniform, t becomes the arc length. We expand x and y separately² as a Fourier series

$$x(t) = \sum_{k=0}^{\infty} (A_k^x \cos(kt) + B_k^x \sin(kt)), \quad (1)$$

$$y(t) = \sum_{k=0}^{\infty} (A_k^y \cos(kt) + B_k^y \sin(kt)), \quad (2)$$

where A_k^x , B_k^x , A_k^y , and B_k^y are the expansion coefficients. The lower indices k apply to the k th term in the expansion, and the upper indices denote the x or y expansion, respectively.

Using this expansion of the x and y coordinates, we can analyze shapes by tracing the boundary and calculating the coefficients in the expansions (using standard methods from Fourier analysis). By truncating the expansion, the shape is smoothed. Truncation leads to a huge reduction in the information necessary to express a certain shape compared to a pixelated image, for example. Székely *et al.*³ used this approach to segment magnetic resonance imaging data. A similar approach was used to analyze the shapes of red blood cells,⁴ with a spherical harmonics expansion serving as a 3D generalization of the Fourier coordinate expansion.

The coefficients represent the best fit to the given shape in the following sense. The $k=0$ component corresponds to the center of mass of the perimeter. The $k=1$ component corresponds to the best fit ellipse. The higher order components

trace out elliptical corrections analogous to Ptolemy's epicycles.⁵ Visualization of the corresponding ellipses can be found at Ref. 6.

We now use this tool to fit an elephant with four parameters. Wei⁷ tried this task in 1975 using a least-squares Fourier sine series but required about 30 terms. By analyzing the picture in Fig. 1(a) and eliminating components with amplitudes less than 10% of the maximum amplitude, we obtained an approximate spectrum. The remaining amplitudes were

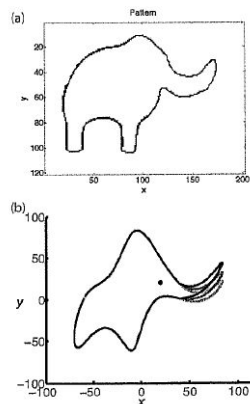


Fig. 1. (a) Outline of an elephant. (b) Three snapshots of the wagging trunk.

~John vonNeumann

Czy jest to biologicznie uzasadnione?

Co się dzieje w wieku 10 lat?

Czy coś takiego przejdzie?

na obronie doktoratu - zależy od audytorium;)

w dobrym czasopiśmie - nie bardzo

im większy stopień wielomianu/głębokość splinu tym lepsze dopasowanie

na czym nam zależy?

overfitting

Overfitting i ekstrapolacja

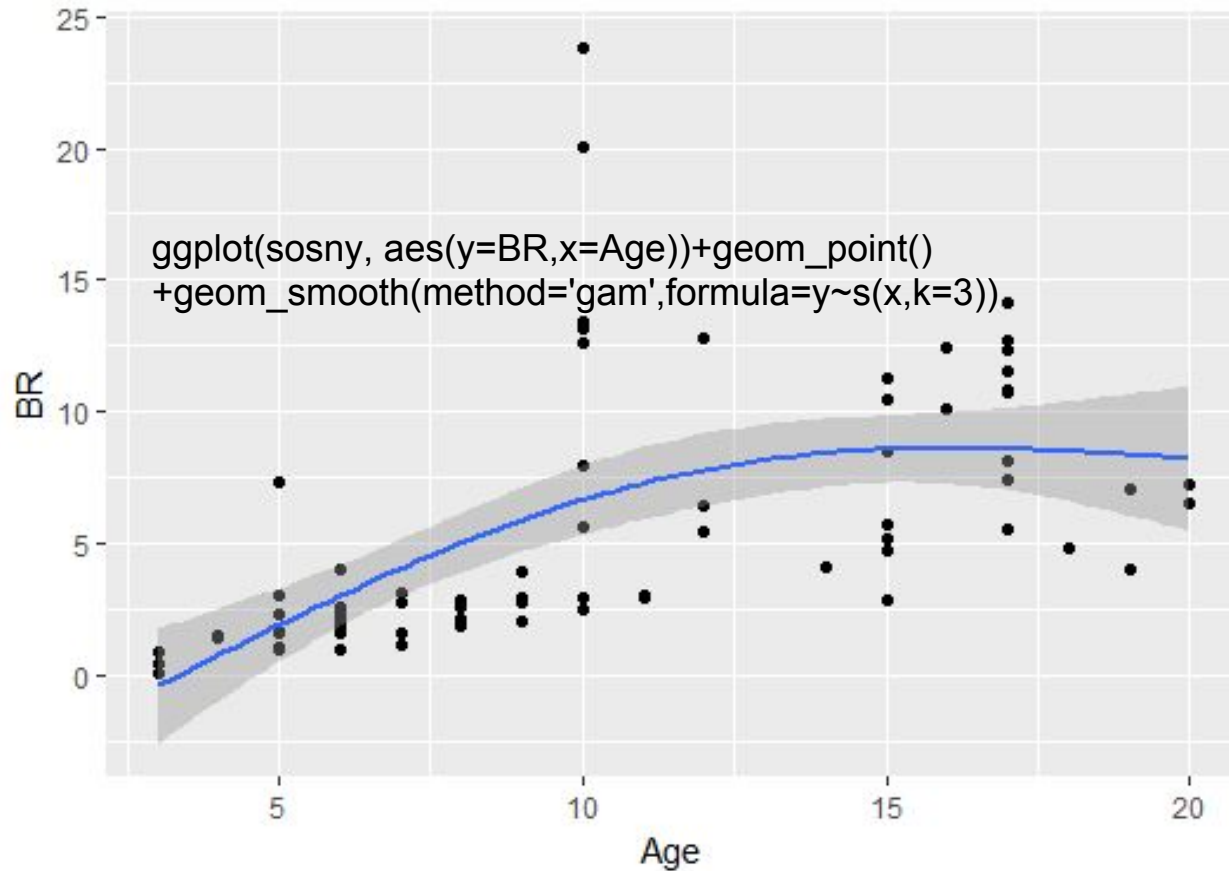
overfitting - model dobrze działa na zbiorze treningowym, ale źle na testowym
brak możliwości ekstrapolacji

Po co robimy modele? Aby coś uogólnić, wyciągnąć trend

Oddzielić ziarno od plew - wiedzę biologiczną od elementów lokalnych i szumu

Stąd założenia o niezależności prób, wielkość próby, powtórzenia ...

rozwiązanie - ograniczyć spline



```
', formula=y~s(x,k=3))  
> modelbr<-gam(BR~s(Age,k=3), data=sosny)  
> summary(modelbr)
```

Family: gaussian
Link function: identity

Formula:
BR ~ s(Age, k = 3)

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.4899	0.4488	12.23	<2e-16 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

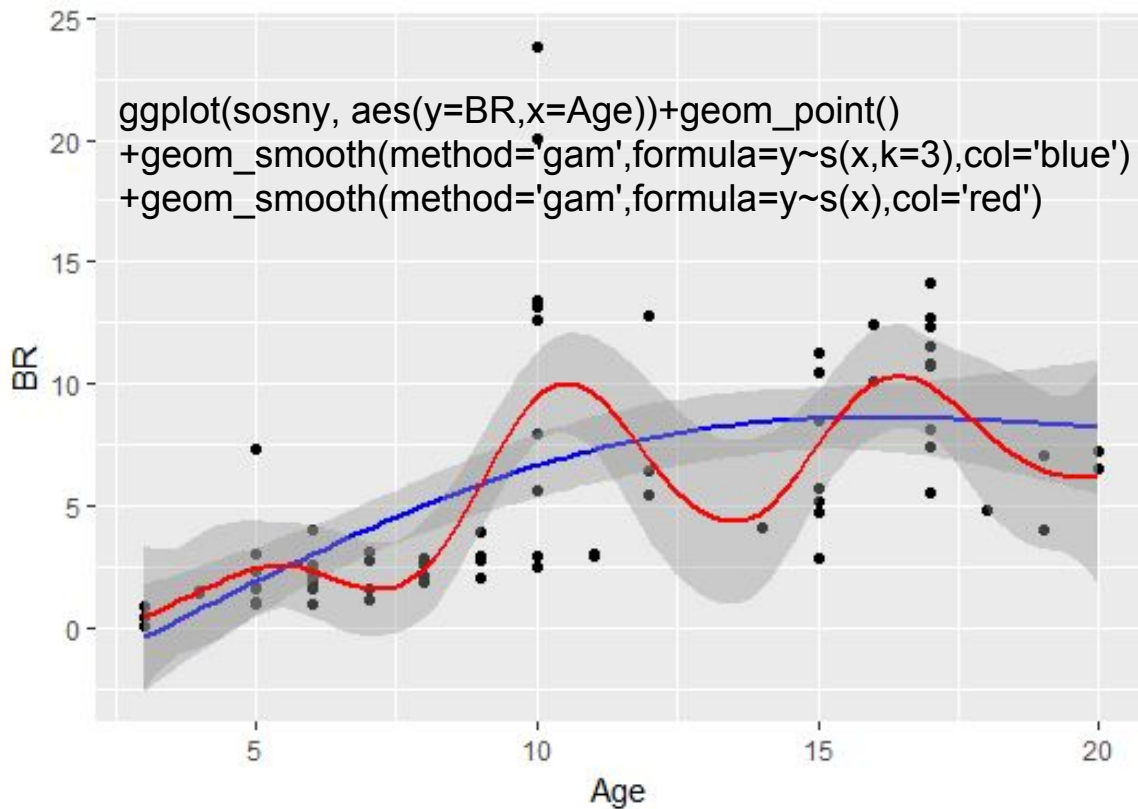
	edf	Ref.df	F	p-value
s(Age)	1.87	1.983	22.44	6.6e-08 ***

signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.349 Deviance explained = 36.5%
GCV = 16.112 Scale est. = 15.512 n = 77

```
> |
```


Tracimy trochę R2 ale model jest bardziej sensowny



Modele liniowe z efektami losowymi i stałymi (LMM)

a.k.a. modele mieszane, modele z efektem losowym, powtarzalne pomiary,

linear mixed-effect models/ generalized linear mixed-effects model,

wymysł recenzenta, #\$\$!@!@*6, s*****, WTF, <płacz>

...czy na pewno takie złe jak nam się wydaje?

efekt stały a efekt losowy

stały - związany z działaniem czynnika

losowy - związany z elementami które powinny być niezależne, a mogą mieć wpływ, np. wariant doświadczenia, powtórzenie, termin, itp.

Przykład

3 terminy badań na 6 blokach po 10 poletek 4 gatunków drzew - światło i odczyn

światło, odczyn, gatunek drzewa - efekty stałe

termin badań, poletko, blok - efekty losowe

Po co efekty losowe

niezależność obserwacji - możliwość wnioskowania i ekstrapolacji

uwypuklenie trendu z uwzględnieniem kontekstu

ocena zmienności na różnych poziomach hierarchii

“a co jeśli jakieś poletko jest odwiedzane przez psy i koty - to też może mieć wpływ na zbiorowiska mszaków epifitycznych przez dopływ biogenów i wydrapywanie?” - plot-specific effects

Prosty przykład na początek

```
hotspots<-read.csv('datasety/hotspots.csv',sep=';')  
mod<-lm(plants~mammals,data=hotspots)  
summary(mod)
```

Call:

lm(formula = plants ~ mammals, data = hotspots)

Residuals:

Min	1Q	Median	3Q	Max
-631.69	-192.69	-82.69	133.72	1294.67

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	48.746	28.716	1.698	0.0903 .
mammals	31.824	3.114	10.219	<2e-16 ***

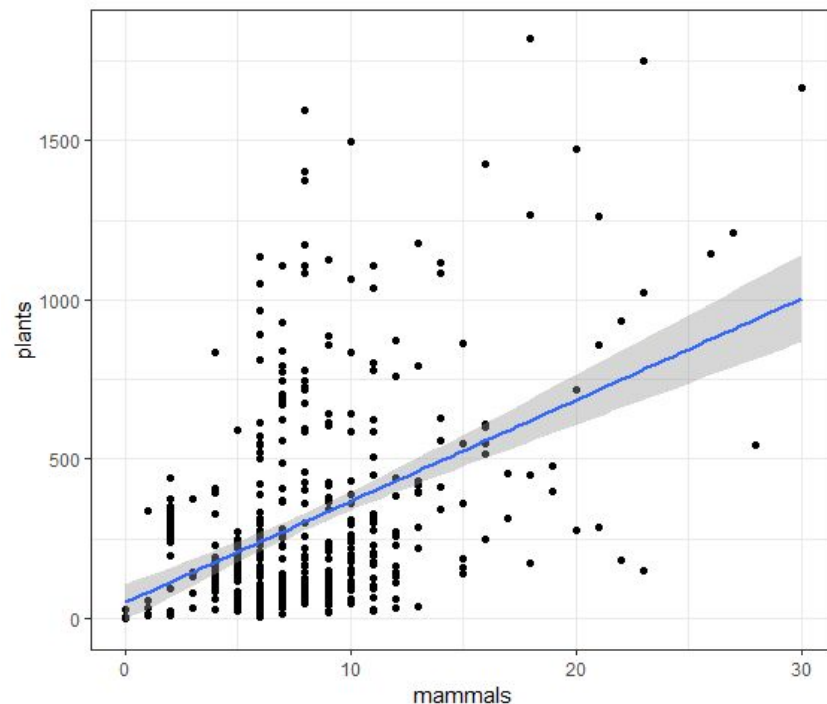
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 294.8 on 447 degrees of freedom

(160 observations deleted due to missingness)

Multiple R-squared: 0.1894, Adjusted R-squared: 0.1876

F-statistic: 104.4 on 1 and 447 DF, p-value: < 2.2e-16



```
ggplot(hotspots, aes(x=mammals,y=plants))+geom_point()  
+theme_bw()+geom_smooth(method='lm')
```

```
mod<-lm(plants~mammals+mainl,data=hotspots)
```

Call:
lm(formula = plants ~ mammals + mainl, data = hotspots)

Residuals:

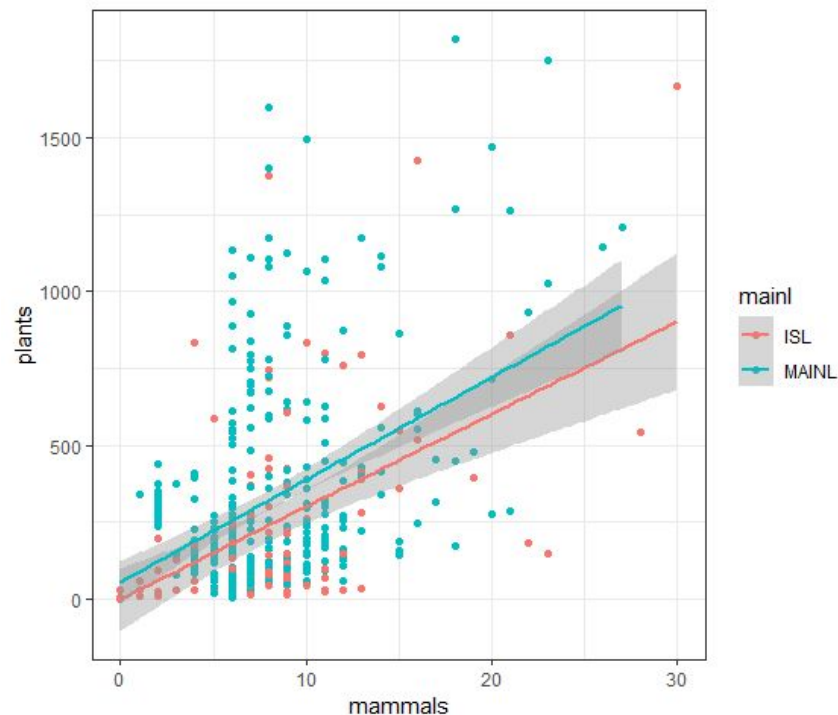
Min	1Q	Median	3Q	Max
-574.32	-203.29	-80.29	130.46	1275.97

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-18.392	39.241	-0.469	0.640
mammals	32.249	3.101	10.400	<2e-16 ***
mainlMAINL	82.437	33.056	2.494	0.013 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 293 on 446 degrees of freedom
(160 observations deleted due to missingness)
Multiple R-squared: 0.2005, Adjusted R-squared: 0.1969
F-statistic: 55.93 on 2 and 446 DF, p-value: < 2.2e-16



```
ggplot(hotspots, aes(x=mammals,y=plants,col=mainl))  
+geom_point()+theme_bw()+geom_smooth(method='lm')
```

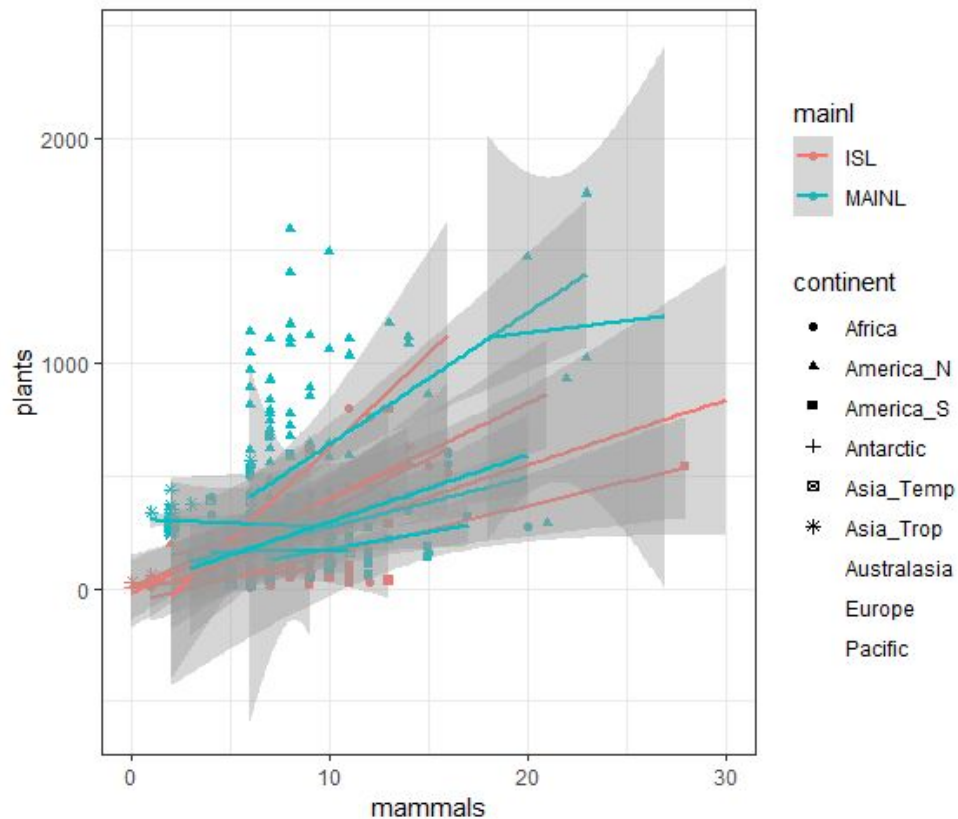
czy dołożyć kontynent?

Czy robi nam różnice?

O co nam chodzi?

różne nachylenia i różne położenia

slope i intercept zależny od continent



```
ggplot(hotspots, aes(x=mammals,y=plants,col=mainl,shape=continent))  
+geom_point()+theme_bw()+geom_smooth(method='lm')
```

Dodajemy losowy intercept

normalnie $\text{plants} = a * \text{mammals} + b$
teraz
 $\text{plants} = a * \text{mammals} + b + u_2$
 u_2 - losowy intercept

```
mod2<-lmer(plants~mammals+(1|continent),hotspots)
```

```
summary(mod2)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method ['lmerModLmerTest']

Formula: $\text{plants} \sim \text{mammals} + (1 \mid \text{continent})$

Data: hotspots

REML criterion at convergence: 6242.7

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.8175	-0.5247	-0.0337	0.3143	4.3142

Random effects:

Groups	Name	Variance	Std.Dev.
continent	(Intercept)	21998	148.3
Residual		62794	250.6

Number of obs: 449, groups: continent, 9

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	30.684	57.399	12.012	0.535	0.603
mammals	36.700	3.055	441.698	12.011	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)
mammals -0.421

O co chodzi z zapisem z kreską?

mammals|continent oznacza że mammals ma losowy slope i intercept dla każdego continent dla mammals

normalnie $\text{plants} = a * \text{mammals} + b$

teraz

$\text{plants} = (a + u_1) * \text{mammals} + b + u_2$

u_1 - losowy slope, u_2 - losowy intercept

```
mod3<-lmer(plants~mammals+(mammals|continent),hotspots)
```

```
summary(mod3)
```

Linear mixed model fit by REML. t-tests use Satterthwaite's method

['lmerModLmerTest']

Formula: plants ~ mammals + (mammals | continent)

Data: hotspots

REML criterion at convergence: 6220.9

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9728	-0.5086	-0.0979	0.3215	4.4747

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
continent	(Intercept)	7088.0	84.19	
	mammals	357.5	18.91	-0.41
	Residual	59102.2	243.11	
Number of obs: 449, groups: continent, 9				

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	42.905	40.798	8.532	1.052	0.3218
mammals	33.215	7.429	5.960	4.471	0.0043 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)
mammals -0.549

Jak wyciągnąć wartości

```
coef(mod2)
```

```
$continent
```

```
(Intercept) mammals
```

```
Africa      -52.99065 36.69969  
America_N    223.33035 36.69969  
America_S   -179.55621 36.69969  
Antarctic   -58.33926 36.69969  
Asia_Temp   -34.16470 36.69969  
Asia_Trop   161.22696 36.69969  
Australasia 159.63777 36.69969  
Europe      -80.81611 36.69969  
Pacific     137.82662 36.69969
```

```
attr("class")
```

```
[1] "coef.mer"
```

```
coef(mod3)
```

```
$continent
```

```
(Intercept) mammals
```

```
Africa      25.878318 26.99799  
America_N    52.495373 57.20669  
America_S     4.901014 16.86633  
Antarctic    19.770299 22.02693  
Asia_Temp    58.166601 19.92525  
Asia_Trop   186.721396 22.31839  
Australasia  17.782051 48.76824  
Europe       24.924300 27.16040  
Pacific     -4.492198 57.66827
```

```
attr("class")
```

```
[1] "coef.mer"
```

Co otrzymujemy?

Model bez efektów losowych - część uogólnioną

Efekty losowe - czynniki warunkujące przebieg krzywych dla kontynentów

Informacje o zmienności tych efektów w ramach modelu:

SD i variance

Który lepszy?

```
> AIC(mod, mod2, mod3)
      df    AIC
mod    4 6380.127
mod2   4 6250.657
mod3   6 6232.945
```

Ile procent wyjaśnia?

```
> library(MuMIn)
> r.squaredGLMM(mod2)
      R2m      R2c
[1,] 0.2410538 0.4379521
> r.squaredGLMM(mod3)
      R2m      R2c
[1,] 0.2039087 0.4537038
```

R2m - marginal R2; R2c - conditional R2

R2m - % zmienności wyjaśnionej przez fixed effects

R2c - % zmienności wyjaśnionej przez fixed + random effects

random effects - R2c-R2m

LMM w układzie hierarchicznym

4 obiekty, po 50 poletek w każdym

`lmer(biomasa ~ land.use.history + invasion + (1 | variant:plot), dane)`

plot zagnieżdżony w variant - czyli efekty losowe na poziomie plot i variant

zakładamy że dla każdego mamy random intercept

SD tych modyfikacji interceptów wynosi 0,0837

R2c i R2m - z `r.squaredGLMM()`



Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



Effects of land use change and *Quercus rubra* introduction on *Vaccinium myrtillus* performance in *Pinus sylvestris* forests

Beata Wozniwoda^{a,*}, Marcin K. Dyderski^b, Andrzej M. Jagodziński^b

^a Department of Geobotany and Plant Ecology, Faculty of Biology and Environmental Protection, University of Lodz, Banacha 12/16, PL-90-237 Łódź, Poland

^b Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, PL-62-035 Kórnik, Poland



Table 2

Differences in dry biomass and biomass allocation of *V. myrtillus* shoots (ramets) between types of pine forest assessed using linear mixed effect models.

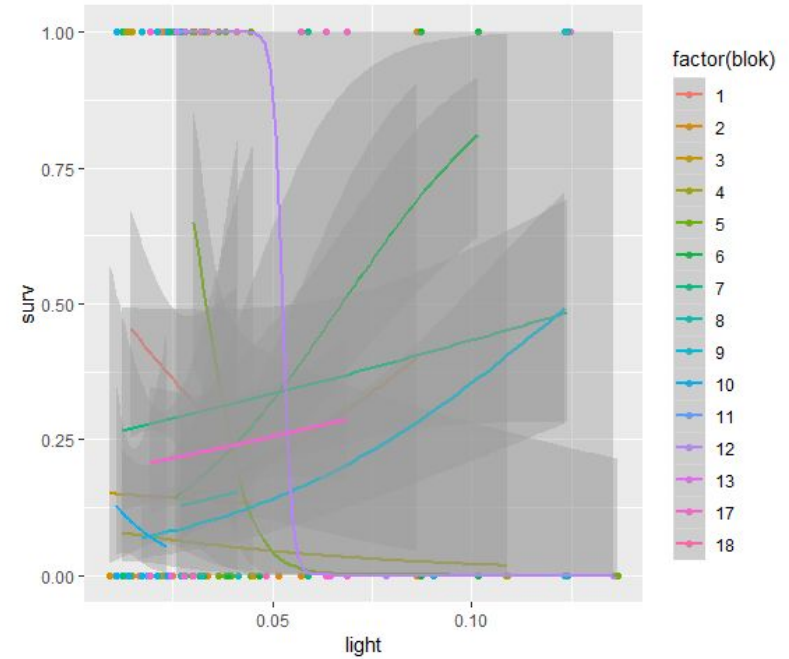
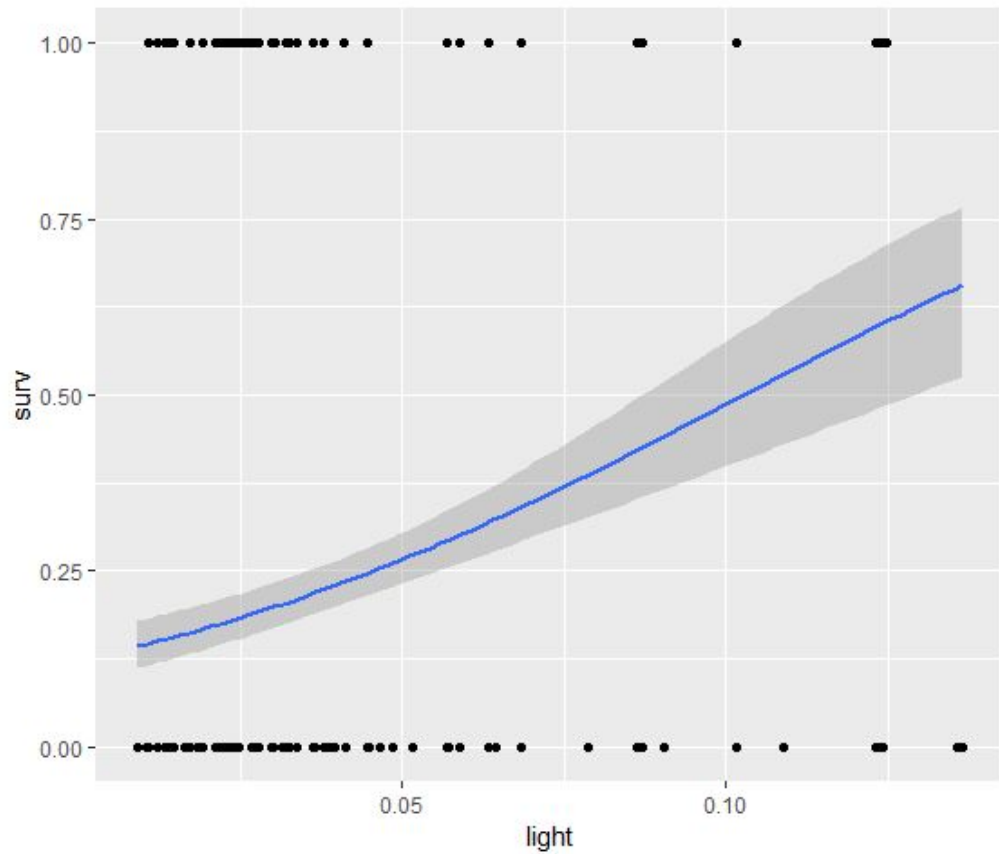
Dry shoot mass (n = 400)	Random effects	Variance	SD	Mixed model parameters	–
–	Plots nested in variant	0.0070	0.0837	R_m^2	0.1159
–	Residuals	0.1557	0.3946	R_c^2	0.1540
–	Fixed effects	Estimate	SE	t	Pr(> t)
–	(Intercept)	0.745	0.041	18.110	< 0.001
–	Land use history – recent forest	–0.243	0.048	–5.110	< 0.001
–	Invasion – <i>Q. rubra</i>	–0.162	0.048	–3.401	0.002

Uogólnione LMM - Generalized Linear Mixed-effects Models (GLMM)

Ta sama zasada co w przypadku GLM - możemy uogólnić LMM na inne rozkłady

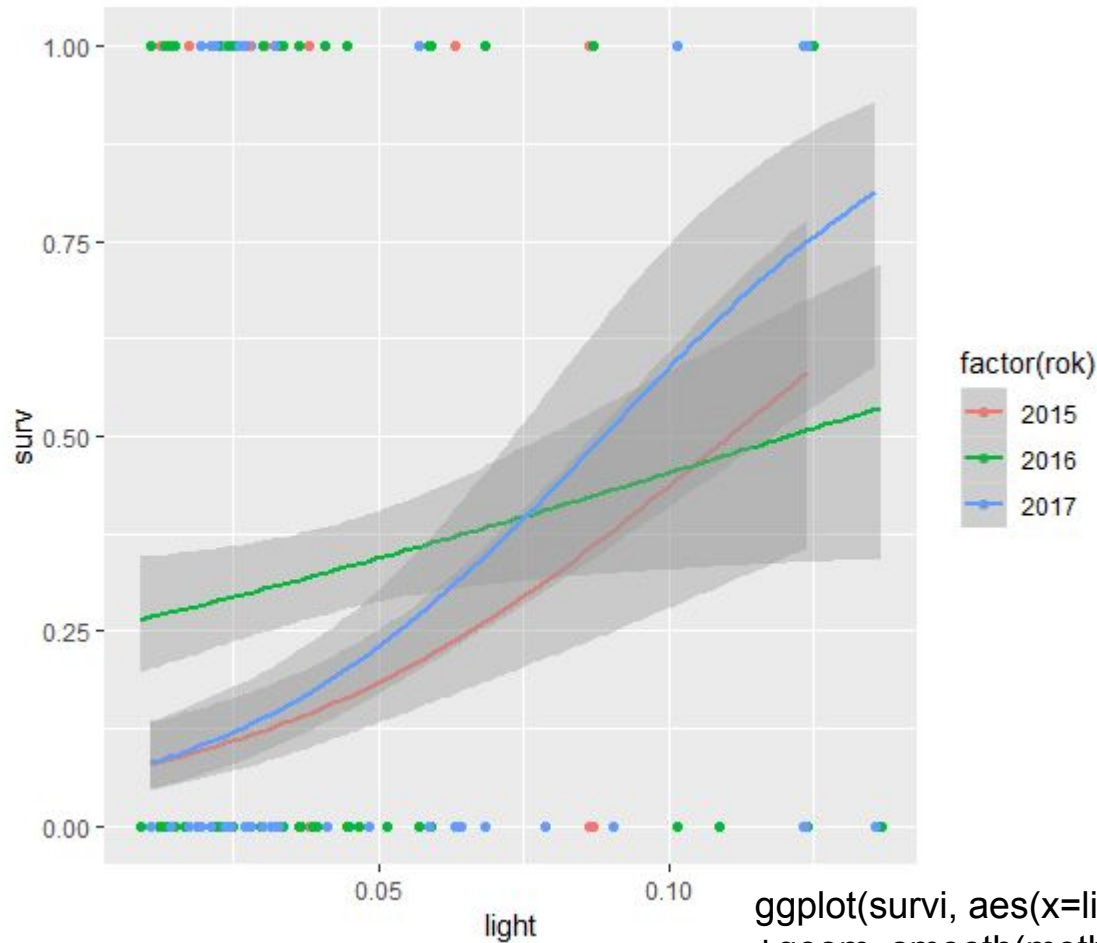
Przykład - przeżywalność siewek dębu czerwonego

Trzy lata, poletka w ramach bloków - efekty losowe



```
ggplot(survi, aes(x=light, y=surv, col=factor(blok)))+geom_point()+
+geom_smooth(method='glm',method.args=list(family='binomial'))
```

```
ggplot(survi, aes(x=light, y=surv))+geom_point()+
+geom_smooth(method='glm',method.args=list(family='binomial'))
```



Co się działo w tych latach?
rok oznaczenia tasiemką,
przeżywalność oceniona po roku

2015 - pierwsze tasiemkowanie
2016 - dobra pomoc w terenie
2017 - susza w 2018

```
ggplot(survi, aes(x=light, y=surv, col=factor(rok)))+geom_point()+  
+geom_smooth(method='glm',method.args=list(family='binomial'))
```


Call:
glm(formula = surv ~ light, family = binomial(link = "logit"),
data = survi)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.4627	-0.6787	-0.6265	-0.5605	1.9640

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.978	0.159	-12.444	< 2e-16 ***
light	19.233	2.859	6.727	1.73e-11 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 750.10 on 681 degrees of freedom
Residual deviance: 704.07 on 680 degrees of freedom
AIC: **708.07**

Number of Fisher Scoring iterations: 4

Generalized linear mixed model fit by maximum likelihood (Laplace
Approximation) [glmerMod]
Family: binomial (logit)
Formula: surv ~ light + (1 | plot:blok) + (1 | rok)
Data: survi
AIC BIC logLik deviance df.resid
678.8 696.9 -335.4 670.8 678

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.0560	-0.4657	-0.3609	-0.2568	2.9111

Random effects:

Groups	Name	Variance	Std.Dev.
plot:blok	(Intercept)	0.89750	0.9474
rok	(Intercept)	0.09808	0.3132

Number of obs: 682, groups: plot:blok, 127; rok, 3
Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.243788	0.001140	-1968	<2e-16 ***
light	19.003078	0.001139	16677	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)	r.squaredGLMM(mod)	R2m	R2c
light 0.000	theoretical	0.06727710	0.2839641
convergence code: 0	delta	0.04502403	0.1900380

```
mod2<-glmer(surv~light+(light|plot:blok)+(1|rok),family=binomial(link='logit'),survi)
mod3<-glmer(surv~light+(light|plot:blok)+(light|rok),family=binomial(link='logit'),survi)
mod4<-glmer(surv~light+(1|plot:blok)+(light|rok),family=binomial(link='logit'),survi)
```

```
AIC(mod.lm, mod,mod2,mod3,mod4)
```

	df	AIC
mod.lm	2	708.0652
mod	4	678.8019
mod2	6	682.7923
mod3	8	681.2299
mod4	6	677.2307

Wniosek - random slop zależny od bloku i plotu - nie za bardzo, intercept - tak,
random slop zależny od roku - tak

Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
Family: binomial (logit)
Formula: surv ~ light + (1 | plot:blok) + (light | rok)
Data: survi

AIC	BIC	logLik	deviance	df.resid
677.2	704.4	-332.6	665.2	676

Scaled residuals:

Min	1Q	Median	3Q	Max
-1.6123	-0.4963	-0.3524	-0.2451	3.0508

	R2m	R2c
r.squaredGLMM(mod4)		
theoretical	0.07901921	0.2946918
delta	0.06523120	0.2432712

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
plot:blok	(Intercept)	0.8019	0.8955	
rok	(Intercept)	0.4788	0.6919	
	light	58.6229	7.6566	-1.00

Number of obs: 682, groups: plot:blok, 127; rok, 3

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.3249	0.4699	-4.948	7.51e-07 ***
light	20.7508	6.2316	3.330	0.000869 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)
light -0.891

convergence code: 0
boundary (singular) fit: see ?isSingular

*Testy statystyczne a GLMM

test chi-kwadrat albo sparowany test Wilcoxona może zastąpić

Binomial GLMM

uwzględnienie efektów losowych

informacja o ilości wyjaśnionej zmienności

więcej o modelach mieszanych

<https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf>

<https://www.r-bloggers.com/linear-mixed-models-in-r/>

<https://www.r-bloggers.com/getting-started-with-mixed-effect-models-in-r/>

<http://www.biecek.pl/WZUR/PrzemekBiecek2009.pdf>

<https://libra.ibuk.pl/book/39524> - podręcznik P. Biecka

Wysublimowanie czy prostota?

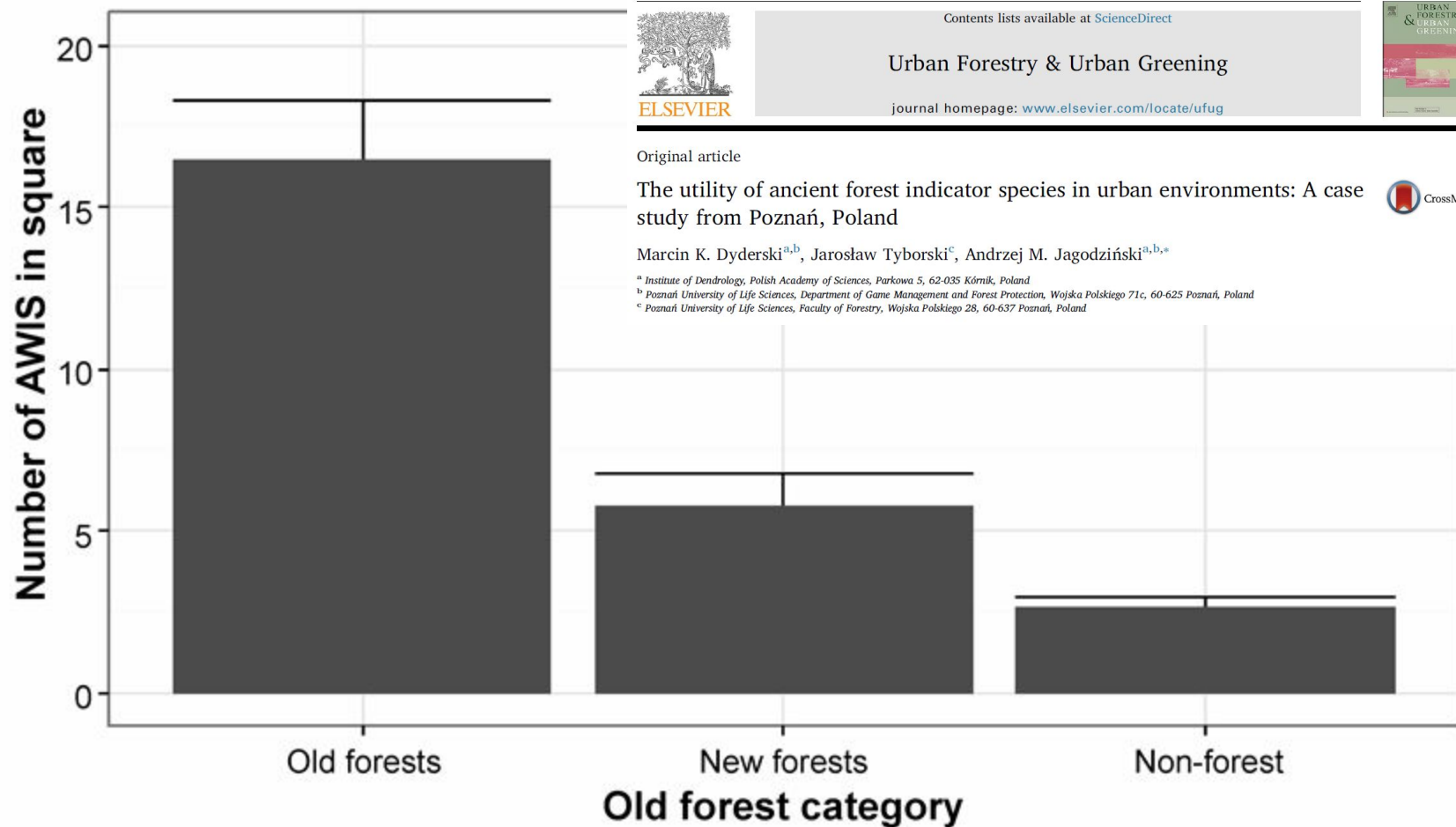
Cel

Użytkownik

Jakość danych

Widzimi się recenzentów

LM/GLM vs. Spatially-explicit Poisson GLM (model mieszany z efektami losowymi dla macierzy sąsiedztwa)



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Urban Forestry & Urban Greening

journal homepage: www.elsevier.com/locate/ufug



Original article

The utility of ancient forest indicator species in urban environments: A case study from Poznań, Poland

Marcin K. Dyderski^{a,b}, Jarosław Tyborski^c, Andrzej M. Jagodziński^{a,b,*}

^a Institute of Dendrology, Polish Academy of Sciences, Parkowa 5, 62-035 Kórnik, Poland

^b Poznań University of Life Sciences, Department of Game Management and Forest Protection, Wojska Polskiego 71c, 60-625 Poznań, Poland

^c Poznań University of Life Sciences, Faculty of Forestry, Wojska Polskiego 28, 60-637 Poznań, Poland



Table 1

Effect of distance from seed source and road and microhabitats on occurrence of *Q. rubra* juveniles, assessed by a spatially-explicit generalized linear model, with binomial distribution of the dependent variable.

Parameter	Estimate	SE	t-Value	P (> t)
<i>Fixed effects</i>				
(Intercept)	-1.3548	0.1262	-10.735	< 0.001
Distance to the public road	-0.0045	0.0009	-5.169	< 0.001
Distance to the seed source	-0.0146	0.0019	-7.531	< 0.001
Microhabitat – under shrubs	3.2067	0.1598	20.068	< 0.001
Microhabitat – in clumps of <i>V. myrtillus</i>	0.1436	0.1633	0.879	0.379
<i>Dispersion parameters</i>				
For fixed effects	1	–	–	–
For random effects	86.63	–	–	–
<i>Dispersion model</i>				
1/CAR.tau	4.4617	0.7941	–	–
– CAR.rho/CAR.tau	0.0192	0.0123	–	–
CAR.tau (estimated spatial variance component)	0.2241	–	–	–
CAR.rho (estimated spatial correlation)	-0.0043	–	–	–

Table 2

Effect of distance from seed source and road and microhabitats on number of *Q. rubra* juveniles, assessed by a spatially-explicit generalized linear model, with Poisson distribution of the dependent variable.

Parameter	Estimate	SE	t-Value	P (> t)
<i>Fixed effects</i>				
(Intercept)	-1.4532	0.1051	-13.833	< 0.001
Distance to the public road	-0.0044	0.0007	-6.128	< 0.001
Distance to the seed source	-0.0115	0.0015	-7.698	< 0.001
Microhabitat – under shrubs	2.2410	0.0973	23.024	< 0.001
Microhabitat – in clumps of <i>V. myrtillus</i>	0.2477	0.1312	1.888	0.059
<i>Dispersion parameters</i>				
For fixed effects	1	–	–	–
For random effects	6.889	–	–	–
<i>Dispersion model</i>				
1/CAR.tau	1.9299	0.1723	–	–
– CAR.rho/CAR.tau	0.0092	0.0011	–	–
CAR.tau (estimated spatial variance component)	0.5182	–	–	–
CAR.rho (estimated spatial correlation)	-0.0048	–	–	–



Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco



Tu z kolei model pomógł ominąć mankamenty i wyjawić czyste zjawisko

Choć recenzenci mówili że jest za skomplikowany;)

Propagule pressure, presence of roads, and microsite variability influence dispersal of introduced *Quercus rubra* in temperate *Pinus sylvestris* forest

Beata Woźniowska^{a,*}, Anastazja Krzyżanowska^a, Marcin K. Dyderski^{b,c}, Andrzej M. Jagodziński^{b,c}, Ewa Stefańska-Krzaczek^d



Effect size

Zamiast zabawy w istotność można pokazać estimate i 95% CI (przedział ufności)

```
mod<-lm(plants~mammals+mainl,data=hotspots)
```

```
mod2<-lmer(plants~mammals+(1|continent),hotspots)
```

```
#model niemieszany
```

```
confint(mod)
```

```
2.5 % 97.5 %
```

```
(Intercept) -95.51193 58.72750
```

```
mammals 26.15442 38.34272
```

```
mainlMAINL 17.47285 147.40202
```

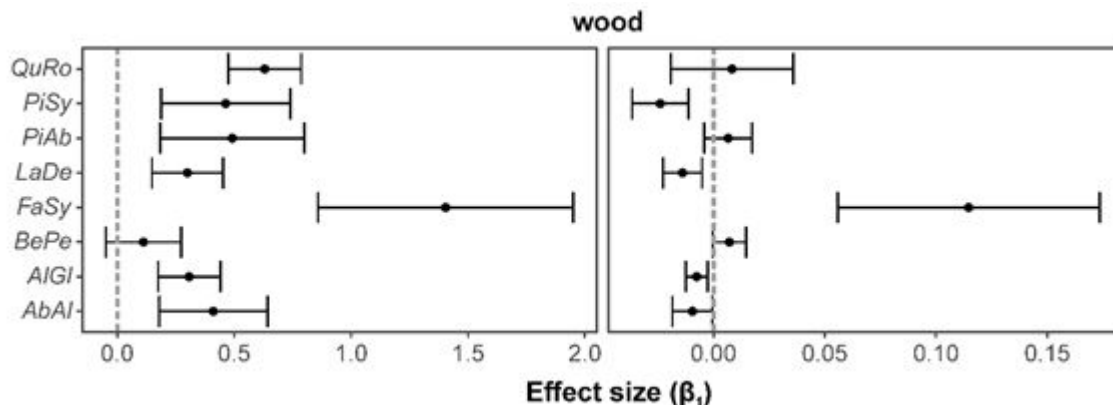
```
confint(mod2) #model mieszany
```

```
.sig01 88.14137 247.59839
```

```
.sigma 234.63689 267.78245
```

```
(Intercept) -84.75104 146.50362
```

```
mammals 30.71091 42.67729
```



Co dalej? Asymetryczne CI, rozkłady *posteriori* - statystyka Bayesowska;)

Podsumowanie

typ relacji i wiedza o naturze badanego zjawiska

znajomość narzędzi vs. znajomość biologii

założenia matematyczne i założenia logiczne (np. nieujemna biomasa)

dokładność vs. aplikowalność (overfitting, zależności)

paradygmat ciągłego rozwoju

co dalej: macierze filogenetyczne, statystyka bayesowska, machine learning, deep learning, IML/XAI...

... all models are approximations. Essentially, **all models are wrong, but some are useful**. However, the approximate nature of the model must always be borne in mind....

~George E. P. Box

