

PRIMERJAVA RAZLIČNIH REGRESIJSKIH MODELOV ZA NAPOVEDOVANJE DEBELINSKEGA PRIRAŠČANJA JELKE

A COMPARISON OF ALTERNATIVE TYPES OF REGRESSION MODELS FOR PREDICTING THE DIAMETER INCREMENT OF SILVER FIR

Andrej FICKO¹, Vasilije TRIFKOVIĆ²

(1) Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za gozdarstvo in obnovljive gozdne vire, andrej.ficko@bf.uni-lj.si

(2) Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za gozdarstvo in obnovljive gozdne vire, vasilije.trifkovic@bf.uni-lj.si

IZVLEČEK

V prispevku na primeru jelke predstavljamo sedem regresijskih modelov za modeliranje priraščanja dreves s podatki periodičnih meritev na stalnih vzorčnih ploskvah. Poleg polinomske regresije, modela z dodanim šumom in mešanega linearnega modela, predstavljamo regresijo z naravnimi zlepkami in tri modele z omejenimi odvisnimi spremenljivkami: *truncated regression*, *tobit regression* in *grouped data regression*. Modele lahko uporabimo, kadar se zaradi načina merjenja in zaokroževanja podatkov ter hierarhičnosti podatkov srečamo z rezanimi ali krnjenimi slučajnostnimi spremenljivkami, nezveznostjo odvisne spremenljivke in pristransko oceno prirastka. Pri pojasnitvi debelinskega priraščanja 21.013 jelk na 4.405 ploskvah v obdobju 1990–2014 v raznomernih gozdovih v dinarskih jelovo-bukovjih so vsi modeli pokazali podoben vpliv prsnega premera, sestojne temeljnice, temeljnice debelejših dreves, raznomernosti, nagiba, nadmorske višine in le manjše razlike v regresijskih koeficientih in merah prileganja. Največje povprečne napovedi prirastka daje tobit model, mešani model pa se najbolj prilega podatkom. V primerjavi z drugimi modeli model z zlepkami kaže na počasnejše zmanjševanje prirastka zelo debelih jelk po kulminaciji prirastka.

Ključne besede: prirastek, multipla regresija, statistične metode, tobit model, krnjenje, mešani modeli, jelka, modeli z omejenimi odvisnimi spremenljivkami, stalne vzorčne ploskve

ABSTRACT

We present seven alternative statistical models for modelling tree diameter increment with data from permanent sampling plots. In addition to the polynomial regression model, we present a regression model with added random noise, a mixed linear model, regression with natural splines, and three models with limited dependent variables: truncated regression, tobit regression and grouped data regression. The models may be used when dealing with truncated or censored variables, biased estimation of the increment due to censoring and rounding down, or when having multilevel data. The parametrization of the models was done using 21,013 fir trees on 4,405 plots in the period 1990–2014 in uneven-aged Dinaric fir-beech forests. All models showed a similar effect of tree diameter, stand basal area, basal area of larger trees, diameter structure diversity, altitude and slope. There were only minor differences in the regression coefficients and fit measures. The highest increment predictions were given by the tobit model. The mixed model fit the data best and, compared to the other models, predicted a slower decrease in the growth of large-diameter trees after growth culmination.

Key words: diameter increment, multiple regression, statistical methods, tobit model, censoring, mixed models, silver fir, limited dependent variable models, permanent sampling plots

GDK 564+561.2:174.7Abies(045)=163.6

DOI 10.20315/ASetL.126.6

Prispelo / Received: 5. 10. 2021

Spregeto / Accepted: 28. 10. 2021



1 UVOD

1 INTRODUCTION

Pri preučevanju pojavov v gozdnih ekosistemih nas pogosto zanima, kateri dejavniki vplivajo na pojave. Določen pojav lahko pojasnimo na podlagi funkcijskih zvez z drugimi pojavi, večinoma pa so odnosi zaradi številnih motečih dejavnikov bolj ohlapni in zveze držijo le v povprečju. Eden izmed pojavov, kjer opažamo veliko variabilnost zveze, je rast drevesa. Variabilnost ugotavljamo pri debelinski, temeljnični ali višinski

rasti drevesa, podobno pa lahko trdimo tudi za mnoge druge procese, kot sta mortaliteta ali pomlajevanja sestojev. Na debelinsko rast drevesa vplivajo številni določljivi dejavniki, kot na primer genetska dispozicija osebka, starost osebka, proizvodna sposobnost rastišča, kompeticija, sestojna gostota in struktura. Vpliv teh dejavnikov je morda nelinearen in multiplikativen. Zaradi omejitev v naravi moramo za ugotovitev vpliva dejavnikov uporabljati poskuse. Pri njihovi zasnovi si prizadevamo, da lahko določljive moteče dejavnike že

z zasnovo poskusa v čim večji meri izključimo, vpliv nedoločljivih motečih dejavnikov pa zmanjšamo s slučajnostno izbiro poskusnega materiala, slučajnostno dodelitvijo tretmajev in večjim številom ponovitev poskusa. Poskusna zasnova in lastnosti preučevanih spremenljivk pa vplivajo tudi na izbiro ustreznega statističnega modela za pojasnjevanje vpliva dejavnikov.

Analiza literature s področja modeliranja debelinske rasti dreves za namene upravljanja gozdov kaže, da se za preučevanje debelinskega priraščanja največkrat uporabljajo linearni ali nelinearni regresijski modeli in mešani modeli. Za identifikacijo ključnih dejavnikov debelinske rasti dreves za srednjeevropske drevesne vrste so na primer Rohner in sod. (2018) uporabili podatke iz nacionalne gozdne inventure v Švici in razvili mešane modele. Vključili so 23 pojasnjevalnih spremenljivk in jih združili v štiri glavne kategorije: drevesne, sestojne, rastiščne in klimatske spremenljivke. Ralston in sod. (2003) so z nelinearnimi matričnimi modeli napovedovali rast dreves v raznomernih gozdovih duglazije na podlagi podatkov s stalnih vzorčnih ploskev. Debelinsko rast so napovedovali s pomočjo premera drevesa, gostote sestoja, drevesne vrste, rastiščnega indeksa in sestojne temeljnice. Schelhaas in sod. (2018) so s podatki evropskih gozdnih inventur razvili modele debelinskega priraščanja za nekaj glavnih drevesnih vrst v Evropi, vključujoč drevesne, sestojne, rastiščne in klimatske spremenljivke. Vospernik (2021) je s posplošenimi aditivnimi modeli in podatki nacionalne gozdne inventure pripravila napovedi temeljničnega priraščanja za glavne drevesne vrste v Avstriji. Večina modelov je vključevala sestojne spremenljivke, kot so temeljnica debelejših dreves (BAL), indeks gostote sestoja (SDI), pogosto so bile vključene rastiščne spremenljivke, kot so naklon, pH, depozicija dušika in temperatura, medtem ko so bile podrobnejše klimatske spremenljivke vključene redkeje. Med regresijskimi modeli so se posebej uveljavili mešani linearni modeli, ki združujejo tehnike linearne regresije, analize variance in kovariance (Zar, 2010). S to veliko skupino modelov lahko analiziramo kompleksnejše poskusne zasnove s slučajnostnimi in fiksnimi dejavniki, kjer dejavniki nastopajo kot diskretne ali zvezne spremenljivke, v modelu lahko upoštevamo odvisnost med opazovanji, ki je posledica hierarhične narave poskusa ali pa ponovitev poskusa na istih poskusnih enotah. Medtem ko so mešani modeli in linearna regresija po metodi najmanjših kvadratov v gozdarski literaturi dobro opisani (npr. Mehtätalo and Lappi, 2020), pa opisov in uporabe drugih regresijskih modelov pri modeliranju priraščanja dreves skoraj ne zasledimo. V prispevku zato predstavljamo manjkrat uporabljene

regresijske modele, ki so primerni za modeliranje rasti sestojev v določenih razmerah. Velja sicer opozoriti, da kompleksnejši modeli niso nujno boljši od modelov multiple linearne regresije, upoštevajoč načelo parsimonije (Burnham in Anderson, 2002), namen raziskave in možne posledice kršenja predpostavk modela. Ko ocenjujemo ustreznost modela, ne smemo imeti v mislih le mer prileganja, marveč tudi njegovo praktično uporabo in primernost za napovedi. Tako so Pukkala in sod. (2009) pri preučevanju debelinskega priraščanja dreves v raznodobnih sestojih na Finskem primerjali mešane modele in linearne regresijske modele in zaključili, da je zaradi majhnih razlik v prileganju med obema skupinama modelov, majhnega vpliva slučajnostnih dejavnikov in praktičnosti kasnejše uporabe regresijskih zvez v rastnih simulatorjih smotrnejše uporabiti modele multiple linearne regresije. Tako so v model debelinskega priraščanja posameznih dreves vključili le prsni premer, vsoto temeljnic debelejših dreves, sestojno temeljnico in temperaturno vsoto, ne pa učinka ploskev (Pukkala in sod., 2009). Podobno kot Pukkala in sod. (2009) zaključuje Vospernik (2021) pri modelih temeljničnega priraščanja za vse glavne drevesne vrste v Avstriji. Kljub temu, da se nepojasneni del variabilnosti spremenljivke z vključitvijo slučajnostnih dejavnikov (na primer ploskve) morda nekoliko zmanjša, moramo presoditi, ali je namen raziskave res pojasniti čim večji del variabilnosti v testnih podatkih, ali pa želimo posplošiti vplive neodvisnih spremenljivk na odvisno spremenljivko.

Pred modeliranjem s podatki gozdnih inventur je pomembno preveriti, ali so inventurni podatki združljivi z zahtevami statističnih modelov (Randall in sod., 1988; Vanclay, 1994). Med težavami, s katerimi se srečamo pri modeliranju debelinske rasti s podatki s stalnih vzorčnih ploskev, so najpogostejše naslednje: rezanje ali krnjenje odvisne spremenljivke na določenem merskem pragu (ang. *truncation* in *censoring*), neizpolnjevanje predpostavke o zveznosti odvisne spremenljivke, pristranskost ocen periodičnega prirastka zaradi zaokroževanja meritev navzdol in načina merjenja, vsebinska neutemeljenost zamenjave linearnih modelov z nelinearnimi, nadomeščanje majhnega števila podatkov z združevanjem meritev (ang. *pooling of data*) znaka na istih poskusnih enotah iz različnih časovnih obdobj (ang. *pseudoreplication*) in neupoštevanje hierarhičnosti v podatkih zaradi načina vzorčenja.

V prispevku na praktičnem primeru podatkov stalnih vzorčnih ploskev Zavoda za gozdove Slovenije prikazujemo možnosti statističnega modeliranja z različnimi regresijskimi modeli. Teoretično predstavitev modelov smo zmanjšali na najmanjšo možno mero, saj

so podrobnosti posameznega modela opisane v statističnih učbenikih (npr. Whitlock in Schluter, 2009; Zar, 2010, za modele z omejenimi odvisnimi spremenljivkami Greene, 2003). Namen prispevka je 1) analizirati najpogostejše metodološke izzive pri uporabi podatkov s stalnih vzorčnih ploskev za izdelavo regresijskih modelov priraščanja dreves, 2) izdelati alternativne regresijske modele za debelinsko priraščanje jelke, 3) ugotoviti, kolikšne so razlike v napovedih alternativnih regresijskih modelov na primeru debelinske rasti jelke v dinarskih jelovo-bukovih gozdovih in 4) oceniti uporabno vrednost alternativnih regresijskih modelov za modeliranje razvoja sestojev.

2 OMEJITVE PRI UPORABI PODATKOV S STALNIH VZORČNIH PLOSKEV ZA MODELIRANJE DEBELINSKEGA PRIRASTKA Z REGRESIJSKIMI MODELI

2 LIMITATIONS IN USING PERMANENT SAMPLING PLOT DATA FOR MODELLING DIAMETER INCREMENT WITH REGRESSION MODELS

2.1 Rezanje in krnjenje

2.1 Truncation and censoring

Do rezanja slučajnostne spremenljivke pride, kadar vrednosti nad in/ali pod določeno mejno vrednostjo c niso vključene v vzorec. Pri rezanju od spodaj pri vrednosti c (ang. *truncation from below*) ima latentna slučajnostna spremenljivka y^* v vzorcu le vrednosti $y > c$, vrednosti $y \leq c$ niso vključene. Pri rezanju od zgoraj (ang. *truncation from above*) je slučajnostna spremenljivka omejena na zgornji meji c ; vrednosti $y \leq c$ ne izmerimo. Možno je rezanje na obeh straneh. Vzorec, ki prihaja iz rezane porazdelitve (ang. *truncated distribution*), je rezan vzorec. Posledice rezanja se kažejo v sistematično večji ali manjši aritmetični sredini (odvisno od mesta rezanja) in manjši variabilnosti slučajnostne spremenljivke v primerjavi s spremenljivko, ki ni rezana. Ker vrednosti slučajnostne spremenljivke y^* nad ali pod mejno vrednostjo c ne izmerimo, ne poznamo tudi vrednosti neodvisnih spremenljivk v tem območju. To pomeni, da kadar spremenljivka z rezano porazdelitvijo nastopa kot odvisna spremenljivka v linearni regresiji, lahko z metodo najmanjših kvadratov dobimo pristranske regresijske napovedi.

S problemom rezanja od spodaj se pri modeliranju debelinskega priraščanja srečamo pri debelinskem prirastku, ki je v splošnem lahko le nenegativen (a glej na primer Baker in sod., 2002 in Fichtner in sod., 2020). Z debelinskim prirastkom mislimo na tekoči debelinski prirastek zdravih dreves v gozdovih zmernega pasu v časovnem obdobju vsaj enega leta, izračunan kot raz-

lika med začetnim in končnim prsnim premerom drevesa. Porazdelitev prirastkov zaradi rezanja pri 0 ni simetrična, ampak rezana. Primerna porazdelitev za tako spremenljivko je rezana porazdelitev, na primer rezana normalna porazdelitev (ang. *truncated normal distribution*), formalno ustrezen regresijski model za spremenljivke z rezano porazdelitvijo pa je rezana regresija ali *truncated regression*. Rezanje pri $c = 0$ je z vidika modeliranja prirastka manj problematično, kadar je povprečni prirastek, izračunan po zgornjem načinu, mnogo večji od 0 in je njegova porazdelitev simetrična in je tako porazdelitev napak bližje normalni porazdelitvi. Tak primer bi bila na primer hitreje rastoča drevesa v enomernem sestoju. Domneva normalnosti pa je manj ustrezna, kadar gre za počasi rastoča drevesa s porazdelitvijo, ki je asimetrična v desno. To je, ko je povprečje blizu 0 in imamo posamezna drevesa v velikim prirastkom, kar se kaže kot dolgi desni rep porazdelitve.

Na videz podoben, a vsebinsko precej drugačen pojav kot rezanje je krnjenje spremenljivke (ang. *censoring*). Pri krnjenju domnevamo, da obstajajo vrednosti latentne spremenljivke y^* , ki so pod ali nad mejno vrednostjo c , a so v vzorcu cenzurirane na mejno vrednost c . To pomeni, da imajo vrednosti $y > c$ v primeru krnjenja od zgoraj ali $y < c$ v primeru krnjenja od spodaj v vzorcu vedno vrednost c in ne svoje prave vrednosti. Regresijski model, ki upošteva krnjenje, je model tobit I krnjene regresije (ang. *censored regression*). Kadar domnevamo, da obstajajo vrednosti neodvisnih spremenljivk, ki močno zavirajo ali pospešujejo rast dreves, a zaradi načina merjenja teh učinkov na prirastek nikoli ne zaznamo, je pravilneje uporabiti model tobit I. Zaviranje ali pospeševanja rasti je v teoriji lahko tolikšno, da se izraža v zelo negativni (majhni) ali zelo pozitivni (veliki) vrednosti latentnega prirastka y^* , a naš izmerjeni prirastek ima pri vrednostih nad ali pod mejno vrednostjo vedno le vrednost c .

Pri modeliranju debelinskega priraščanja s pomočjo stalnih vzorčnih ploskev bi do krnjenja debelinskega prirastka lahko prišlo pri meritvah debelih dreves zaradi premajhnih premerk, kjer bi zelo debela drevesa dobila prsni premer, ki je enak spodnji meji navzgor odprtega razreda oziroma maksimalni vrednosti na premerki. Do krnjenja od spodaj pa bi lahko prišlo, kadar domnevamo, da je izmerjeni ničelni prirastek (končni premer $d_2 =$ začetni premer d_1) izraz delovanja različnih jakosti vplivnih dejavnikov, ki segajo od šibko pozitivnega učinka (debelinski prirastek $id < 1$ cm / 10 let) do mortalitete drevesa ali celo skrčenja debela, a se v merskem smislu kažejo v enakem dogodku, to je $d_2 = d_1$. Če za negativni periodični prirastek dreves

v gozdovih zmernih širin težje najdemo fiziološko razlago in gre negativne vrednosti prirastka obravnavati kot mersko napako, pa prirastek pod 1 cm lažje razumemo kot rezultanto različnih, nesorazmernih učinkov vplivnih dejavnikov na prirastek. Regresija namreč domneva sorazmernost učinkov na celotnem intervalu vrednosti vplivnih dejavnikov, a samo iz meritev prirastkov ni mogoče ugotoviti sorazmernosti učinkov dejavnikov na rast, kadar so ti dejavniki dejavniki minimuma. Poseben primer krnjenja je tudi merjenje prsnih premerov s centimetrsko natančnostjo in zaokroževanjem navzdol, kjer prsne premere in iz njih izračunane prirastke razumemo kot v celoti na celoštevilске vrednosti krnjene podatke. Primer tovrstnega popolnega krnjenja zaradi zaokroževanja na centimetrске debelinske stopnje in za to primeren model obravnavamo posebej v naslednjem poglavju 2.2.

2.2 Neizpolnjevanje predpostavke o zveznosti odvisne spremenljivke

2.2 Violating the assumption of a continuous dependent variable

Ena izmed predpostavk linearnega regresijskega modela je zveznost odvisne spremenljivke. Pri izračunu debelinskega prirastka iz dveh zaporednih meritev prsnih premerov s centimetrsko premerko na stalnih vzorčnih ploskvah Zavoda za gozdove Slovenije dobimo le celoštevilске vrednosti. Kadar je zaloga vrednosti tako izračunanega prirastka velika, lahko predpostavimo, da celoštevilске vrednosti aproksimirajo vrednosti pravega prirastka y^* dovolj dobro. Težava nastopi takrat, kadar je prirastek majhen in ima zato spremenljivka majhno zalogo vrednosti, na primer 0, 1 in 2 cm. Z večjo verjetnostjo za kršitev predpostavke o zveznosti se zato utegnemo srečati pri modeliranju prirastka v enomernih čistih sestojih, kjer je variabilnost prirastka majhna. Pri rešitvi tega problema si lahko pomagamo s splošno prakso pri obravnavi urejenostnih spremenljivk kot razmernostnih: urejenostno spremenljivko obravnavamo kot razmernostno, kadar ima ta vsaj pet kategorij in imamo velik vzorec. Na primer, če smo iz zaporednih meritev prsnih premerov s centimetrsko premerko izračunali desetletni prirastek dreves 0, 1, 2, 3 in 4 cm, lahko privzamemo, da je spremenljivka zvezna.

Druga rešitev zgornjega problema je, da odvisno spremenljivko obravnavamo kot večkrat krnjeno spremenljivko in uporabimo regresijo z grupiranimi podatki. Regresija z grupiranimi podatki (ang. *grouped data regression*) je različica modela tobit, v katerem so podatki v celoti krnjeni. Pri tem modelu domnevamo, da imamo namesto pravih vrednosti prirastka y^* na voljo

le grobe podatke o prirastku po velikostnih razredih y , na primer $y = 1, 2, \dots, J$ (Greene, 2012):

$$\begin{aligned} y = 1, & \text{ če } -\infty < y^* < a_1 \\ y = 2, & \text{ če } a_1 \leq y^* < a_2 \\ y = 3, & \text{ če } a_2 \leq y^* < a_3 \\ & \dots \\ y = J, & \text{ če } a_{J-1} \leq y^* < +\infty \end{aligned} \quad (1)$$

pri čemer so a_1, a_2, \dots, a_{J-1} meje intervalov in domnevamo, da so največje in najmanjše vrednosti y znane. Notranje meje v takšnem modelu določimo tako, da intervali ustrezajo načinu zaokroževanja prsnih premerov pri meritvah s centimetrskimi premerkami.

2.3 Sistematična napaka ocen debelinskega prirastka

2.3 Bias in the diameter increment estimate

Prsne premere na stalnih vzorčnih ploskvah zaokrožujemo navzdol na centimeter natančno (ZGS, 2010). Pri izračunu debelinskega prirastka na podlagi dveh zaporednih meritev prsnih premerov s centimetrsko premerko prihaja do sistematičnega podcenjevanja prirastka. Lahko predpostavimo, da je ta razlika v povprečju 0,5 cm oziroma 3 do 6 % premera. Procentualno je napaka večja pri tanjših drevesih. Ker se napaka premera s kvadratom prenaša na oceno volumna, je napaka v oceni volumna zaradi zaokroževanja navzdol od 2 % do 8 % (Hočevar, 1995).

Pri modeliranju lahko neizpolnjevanje predpostavke o zveznosti odvisne spremenljivke odpravimo tako, da povsod tam, kjer je prirastek nenegativen, prsnim premerom ob prvi (d_1) in drugi meritvi (d_2) dodamo slučajnostno komponento ρ z enakomerno porazdelitvijo v intervalu $[0,1]$ in tako dobimo prsna premera z realno zalogo vrednosti (d_{ri}):

$$d_{ri} = d_i + \rho, \text{ pri čemer } \rho \sim U[0,1], i = 1, 2 \quad (2)$$

Razlika med tako povečanima prsnima premeroma je novi debelinski prirastek, ki ustreza predpostavki o zveznosti znaka in vsebuje popravek zaradi zaokroževanja navzdol. Takšen popravek sicer ne upošteva odvisnosti v slučajnostnih komponentah prve in druge meritve na istem drevesu, vendar lahko pri velikem številu podatkov domnevamo, da je povprečni učinek zaokroževanja navzdol v populaciji tako pri prvi kot pri drugi meritvi enak 0,5 cm, kar ustreza aritmetični sredini slučajnostne spremenljivke ρ z enakomerno porazdelitvijo v intervalu od 0 do 1: $\rho \sim U[0,1]$. Kadar je prirastek negativen ($d_2 < d_1$), je novi debelinski prirastek enak slučajnostni komponenti ρ ($d_{ri} = \rho$). Slednji popravek sicer

pomeni krnjenje od spodaj, zato bi v primeru večjega števila negativnih prirastkov veljalo premisliti, ali krniti ali raje odstraniti osamelce. Če se odločimo za odstranitev osamelcev, moramo odstraniti ne le spodnje, marveč tudi zgornje osamelce, sicer s popravkom vnesemo dodatno sistematično napako (Vanclay, 1994).

2.4 Osamelci in linearnost zvez

2.4 Outliers and linear relationships

Regresija je občutljiva na osamelce (Zar, 2010). Ena izmed predpostavk v modelu linearne regresije je, da so ostanki približno normalno porazdeljeni. Zato moramo pred začetkom regresijske analize najprej napraviti analizo kakovosti podatkov ter po potrebi izločiti vrednosti, ki se močno ločijo od drugih, na primer tiste, ki so ekstremni osamelci in posledica merskih napak. Po regresiji analiziramo homoskedastičnost modela in porazdelitev ostankov.

Včasih mere prileganja kažejo boljše prileganje nelinearnih funkcij. Ustrezno obliko zveze med odvisno in neodvisnimi spremenljivkami v modelu z več napovednimi spremenljivkami lahko izberemo s pomočjo grafa parcialnih ostankov (na primer Component Plus Residual Plots v programskem okolju RStudio, ki jih nariše funkcija `crPlots()` iz paketa `car` (Fox in Weisberg, 2019)). Parcialne ostanke $e_{i;x_j}$ v odvisnosti od x_{ij} grafično prikažemo in na njih prikažemo še gladilnik LOWESS (ang. *locally weighted scatterplot smoothing*).

Pri izbiri oblike zveze imamo več alternativ. Pogosto se zgodi, da nelinearnosti ne moremo opisati s polinomsko regresijo dovolj nizke stopnje. V takem primeru lahko vrednosti napovedne spremenljivke razdelimo na odseke in na posameznem odseku uporabimo polinomsko regresijo nižje stopnje. Polinome zlepimo in dobimo regresijo zlepkov (ang. *regression splines*). Zlepek je funkcija, ki opiše krivuljo na izbranih odsekih napovedne spremenljivke x . Odseki so določeni z vozlišči. Na posameznem odseku odvisnost spremenljivke y od x opiše polinom p -te stopnje. Vendar pa je pri uporabi nelinearnih oblik zvez med y in x_i potrebna previdnost. Vpliv prsnega premera na debelinski prirastek bi namesto s polinomom druge stopnje lahko ponazorili tudi s polinomom višje stopnje in verjetno dobili nižjo celotno napako (RMSE), vendar je večkratno spremembo hitrosti priraščanja vsebinsko težje utemeljiti, posebej na robovih podatkovne množice, kjer polinomi kažejo večjo fleksibilnost. Pri modeliranju se moramo zavedati, da se bo vsaka funkcija z večanjem števila parametrov vse bolj prilagajala podatkom. Wel (1975, cit. po Burnham in Anderson, 2002) je na primer samo s 36 parametri v sinusni Fourierovi vrsti narisal slona. Velja namreč, da v regresijskih modelih ni omejitev gle-

de števila neodvisnih spremenljivk m , dokler velja, da je število podatkov $n \geq m + 2$ (Zar, 2010). Pri iskanju najboljših funkcij, ki ponazori zvezo med okoljem in priraščanjem drevesa, ne moremo brez biološke utemeljenosti. Superiornost nelinearnega modela mora biti vsebinsko utemeljena, nikakor ne zgolj z boljšimi merami prileganja (Zar, 2010: 448–449). Uporaba nelinearnih modelov (na primer Gompertzove regresije) je pri modeliranju debelinske rasti smiselna, a zaradi lažje interpretacije vpliva dejavnikov velja pred izbiro nelinearne regresije preveriti polinomsko regresijo ali regresijo zlepkov (Oddi in sod., 2019). Velja, da mnoge nelinearne večparametrne funkcije lahko zamenjamo s polinomsko obliko ali zlepkami, ki zadovoljivo odsevajo dejansko biološko odvisnost.

2.5 Združevanje podatkov iz različnih obdobj in hierarhičnost podatkov

2.5 Pooling data from different time periods and hierarchical data

Včasih imamo na voljo le majhno število opazovanj pojava. Problema majhnih vzorcev ne moremo rešiti tako, da pojav opazujemo večkrat na istih osebkih, zatem pa meritve obravnavamo kot neodvisna opazovanja. Pojav je izredno pogost v ekoloških raziskavah, saj naj bi po nekaterih podatkih bila vsaka osma študija opravljena na podatkih, ki so rezultat psevdoponovitev poskusa (Hurlbert, 1984; Whitlock in Schluter, 2009: 97–98). Psevdoponovitve pomenijo kršitev predpostavke o neodvisnosti opazovanj, saj bi s takšnimi podatki pričakovali odvisnost regresijskih ostankov. Takšno »povečevanje« vzorca (n) umetno zmanjšuje standardno napako ocen regresijskih parametrov, saj velja, da je $\lim_{n \rightarrow \infty} \frac{s}{\sqrt{n}} = 0$. Posledica tega je večja verjetnost za napako prve vrste, to je, da odkrijemo statistično značilen vpliv spremenljivke, čeprav vpliva ni.

Do podobne napake lahko pride, kadar ne upoštevamo hierarhičnosti v podatkih. Določene neodvisne spremenljivke v regresijskem modelu so na ravni drevesa in jih moramo obravnavati kot dejavnike na prvem nivoju (prsni premer, socialni položaj, oblika krošenj), določene spremenljivke pa so na ravni ploskve ali kakšne druge večje prostorske ravni in jih moramo obravnavati kot dejavnike na višjem nivoju (sestojna temeljnica, gostota sestojja). Vrednosti spremenljivk na drugem nivoju niso rezultat neodvisnih opazovanj, saj bodo vsa drevesa na določeni ploskvi dobila enake vrednosti sestojne temeljnice. Pri hierarhičnosti podatkov in/ali avtokorelaciji moramo zato ustrezno prilagoditi variančno-kovariančno matriko oziroma uporabiti mešane modele, ki skupno variabilnost odvisne spremenljivke pravilno razporedijo po virih.

3 METODE

3 METHODS

3.1 Študijsko območje in priprava podatkov

3.1 Study area and data preparation

Regresijske modele debelinskega priraščanja smo izdelali na podlagi dveh zaporednih meritev prsnih premerov 21.027 jelk na 4.907 stalnih vzorčnih ploskvah v raznomernih dinarskih jelovo-bukovih gozdovih (gozdni tip 641, slika 1). Prve meritve so bile opravljene od 1990 do 2003, druge meritve pa v obdobju od 2000 do 2014. Najprej smo na podlagi različnih mer sestojne zgradbe (Ginijev indeks, koeficient variacije prsnih premerov) (Pommerening, 2002; Lexerod in Eid., 2006; Danescu in sod., 2016) oblikovali podatkovno zbirko za raznomerne gozdove. Kot najustreznejši kriterij za razmejitev raznomernih sestojev od enomernih se je izkazal Ginijev indeks diverzitete prsnih premerov, ki se je v največji meri ujema z ocenami razvojne faze pri opisih sestojev in na stalnih vzorčnih ploskvah. Za primerjavo skladnosti Ginijevega indeksa in opisnih mer smo v prvem koraku z metodo voditeljev (*k-means*) stalne vzorčne ploskve na podlagi podobnosti Ginijevega indeksa združili v dve skupini. Maksimalna vrednost prve skupine 0,3293 je bila mejna vrednost za razmejitev enomernih in raznomernih sestojev. Mejna vrednost se ujema z ugotovitvami iz literature (Lexerod in Eid., 2006; Danescu in sod., 2016), ki kažejo, da imajo raznomerni sestoji vrednosti Ginijevega indeksa nad 0,30. Zatem smo s Sokal-Michenerjevo mero ujemanja oziroma Randovim indeksom (Pfitzner in sod., 2009) preverili ujemanje med izračunanimi in-

deksi na ploskvi z opisnimi ocenami sestojne zgradbe na stalnih vzorčnih ploskvah in v sestojni karti. Posamično do šopasto raznomerne sestoje, skupinsko do gnezdasto raznomerne sestoje in tipične prebiralne sestoje smo uvrstili v raznomerne sestoje, druge pa v enomerne sestoje. Najvišji indeks podobnosti 0,62 je bil med Ginijevim indeksom in oceno sestojne zgradbe ob opisih sestojev.

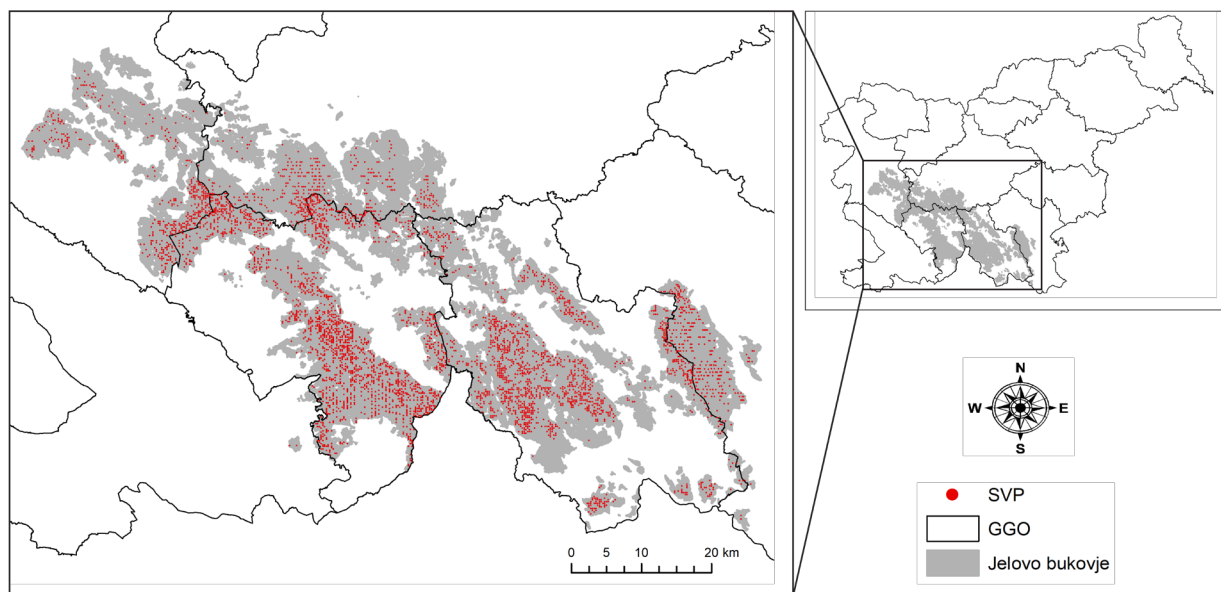
Zaradi obsežnosti alfanumeričnih datotek smo podatke pripravili s pomočjo programa R (version 4.0.2). Regresijske modele smo izdelali s pomočjo ekonometrične programske opreme Nlogit 5.0, analize, za katere so bili na voljo paketi ali možnosti obdelave v drugih paketih, pa smo ponovili v R (version 4.0.2).

3.2 Odvisna in neodvisne spremenljivke

3.2 Dependent and independent variables

Odvisna spremenljivka je bil desetletni debelinski prirastek v cm (*id*). Razpon vrednosti je bil med 0 in 15 cm, povprečje je znašalo 3,8 cm (preglednica 1). Zaradi neenake variabilnosti prirastka glede na prsni premer smo odvisno spremenljivko transformirali s korensko transformacijo.

Kot neodvisne spremenljivke smo vključili spremenljivke, ki so dostopne ob gozdni inventuri in jih bo mogoče uporabiti za zagon sestojnega simulatorja. Tako nismo vključili talnih lastnosti, klimatskih spremenljivk in drugih težje dostopnih spremenljivk. Po predhodnih testiranjih večjega nabora spremenljivk smo v modele vključili naslednje spremenljivke oziroma njihove transformacije, kadar je bilo to potreb-



Slika 1: Študijsko območje: jelovo-bukovi gozdovi v Sloveniji (SVP: stalne vzorčne ploskve; GGO: gozdnogospodarsko območje)

Fig. 1: Study area: Dinaric silver fir-beech forests (SVP: permanent sample plots; GGO: forest management regions)

Preglednica 1: Pregled in opis spremenljivk

Spremenljivka / Variable ¹	Transformacija / Transformation ²	Arit. sredina / Mean	St. odklon / S. D.	Min. / Min.	Max. / Max.
id (cm 10a ⁻¹)	sqrt (id)	3,8	2,6	0,0	14,8
d ₁ (cm)	d ₁ ²	36,6	17,2	10,0	88,5
G ₁ (m ² ha ⁻¹)	sqrt (G ₁)	33,1	10,3	2,3	84,5
BAL (m ² ha ⁻¹)		18,6	12,2	0,0	77,6
Dg ₁ (cm)		29,8	6,8	14,0	64,0
GINI		0,4	0,1	0,3	0,7
NMV (m) ³		834,8	169,7	189,0	1330,0
NAGIB (°)		14,6	7,2	0,0	47,0
k		2,1	0,2	1,1	2,5

Table 1: Overview and description of variables included in the model

¹ V modelih grouped regression (GROUP) je bila odvisna spremenljivka kategorična krnjena spremenljivka, ki ponazarja velikostne razrede prirastka / In grouped regression models (GROUP), the dependent variable was categorical and censored.

² Mere centralne tendence in razpršenosti se nanašajo na netransformirane spremenljivke / Measures of central tendency and dispersion refer to untransformed variables.

³ NMV: altitude; NAGIB: slope; k: the volume of a tree with a reference diameter of 45 cm as an indicator of site productivity; k is the parameter in the adapted Schäffer's and Algan's one-parameter tree volume functions used by the Slovenia Forest Service in timber volume estimations.

no, da bi zadostili pogojem normalnosti in homoskedastičnosti (glej poglavje 2.3): prsni premer ob prvi meritvi (d_1), kvadrat prsnega premera (d_1^2), sestojna temeljnica ob prvi meritvi (G_1), vsota temeljnic debelejših dreves (BAL), srednjemeljnični premer ob prvi meritvi (Dg_1), Ginijev indeks diverzitete prsnih premerov (GINI), nagib (NAGIB), nadmorska višina (NMV) in volumen drevesa premera 45 cm (k) na rastišču skladno z veljavnimi tarifnimi razredi v enotnih tarifah kot kazalec bonitete rastišča.

3.3 Regresijski modeli

3.3 Regression models

Izdelali smo sedem alternativnih regresijskih modelov, najprej model polinomske regresije po metodi najmanjših kvadratov s korensko transformacijo odvisne spremenljivke, linearnimi členi za neodvisne spremenljivke in polinomom druge stopnje za prsni premer (ang. *ordinary least squares regression*, OLS, v stats R package, R Core Team, 2020). Sledili so modeli polinomske linearne regresije po metodi najmanjših kvadratov z dodanim šumom v odvisni spremenljivki (OLS_NOISE), regresija naravnih zlepkov (ang. *natural spline regression*, SPLINE) v splines R package (R Core Team, 2020, mešani linearni model s slučajnostno konstanto (ang. *random intercept mixed effect linear model*, MIXED) v lme4 R package (Bates in sod., 2015), in trije modeli z omejenimi odvisnimi spremenljivkami, ki smo jih izdelali v programu NLOGIT 5 (Greene, 2012): *truncated regression* (TRUNC), *censored regression* (TOBIT) in *grouped data regression* (GROUP). Modele

z omejenimi odvisnimi spremenljivkami smo ponovno testirali v R z uporabo *truncreg* paketa (Croissant in Zeileis, 2018) in *censReg* paketa (Henningsen, 2020).

Pri vseh modelih smo pred zagonom modela odstranili osamelce, ki so imeli standardizirani ostanek po absolutni vrednosti večji od 3 in opravili logične kontrole podatkov. V modelu OLS nismo pred zagonom modela naredili nobenih drugih korakov razen opisanih transformacij spremenljivk (preglednica 1). V modelu OLS_NOISE smo za drevesa z nenegativnim prirastkom prsnim premerom v prvi in drugi meritvi z Excel funkcijo *rand()*, ki s pomočjo Mersenne Twisterjevega generatorja psevdoslučajnostnih števil vrne enakomerno porazdeljeno naključno realno število, večje ali enako 0 in manjše od 1, dodali slučajnosti šum. Novi prirastek smo izračunali kot razliko med tako popravljenima premeroma. Negativnih prirastkov ni bilo. Sledila je regresija po metodi najmanjših kvadratov z istimi spremenljivkami kot pri modelu OLS.

V modelu SPLINE smo napravili regresijo z naravnimi zlepkami. Zlepek je funkcija, s katero opišemo zvezo med odvisno spremenljivko in neodvisno spremenljivko na izbranem odseku neodvisne spremenljivke. Vpliv spremenljivke x na y na posameznem odseku opišemo s polinomom p-te stopnje, na prvem in zadnjem odseku pa predpostavimo linearni odnos med y in x. Zatem smo pripravili mešani linearni model z vključeno slučajnostno komponento (MIXED), kjer smo poleg hierarhičnosti upoštevali tudi učinek ploskve. Učinek ploskve smo upoštevali tako, da smo med slučajnostne dejavnike v mešanem modelu dodali regresijsko

konstanto. S tem smo dopustili variabilnost konstant v regresijskih enačbah po ploskvah, ne pa tudi variabilnosti smernih koeficientov za spremenljivke v modelih. Za takšen model smo se odločili, ker predmet zanimanja v našem primeru niso bila rastišča ali drug del nepojasnjene variabilnosti, ki ga nosijo ploskve.

V modelu TRUNC smo upoštevali, da prirastek y ne more doseči negativnih vrednosti ($c = 0$) in v LIMDEP uporabili funkcijo TRUNCATION. V tobit modelu smo v LIMDEP uporabili ukaz TOBIT in enako mejno vrednost c kot pri rezani regresiji. V modelu GROUP smo uporabili ukaz GROUPEDDATA in opredelili velikostne razrede prirastka z ukazom LIMITS = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14. Velikostni razredi 1, 2, 3,...,14 ustrezajo cenzuriranemu debelinskemu prirastku y^* v intervalih $-\infty < y^* < 1$ cm, $1 \text{ cm} \leq y^* < 2$ cm, $2 \text{ cm} \leq y^* < 3$ cm, ..., $14 \text{ cm} \leq y^* < +\infty$. S tem smo izmerjeni prirastek 0 cm (velikostni razred 1) obravnavali kot cenzurirano vrednost dejanskih prirastkov y^* , ki segajo od zelo majhnih prirastkov do malo manj kot centimetrskih, izmerjeni prirastek 1 cm (velikostni razred 2) smo obravnavali kot cenzurirano vrednost dejanskih prirastkov y^* , ki segajo od centimetrskih do malo manj kot dvocentimetrskih prirastkov itn. Privzeta porazdelitev latentne spremenljivke y^* v vseh modelih z omejenimi odvisnimi spremenljivkami je bila normalna.

Medsebojna primerjava modelov, ki prihajajo iz različnih družin, imajo različne odvisne spremenljivke, ki so posledica transformacij ali zahtev modela, ter zaradi različnega izračuna mer prilaganja, je otežena. Kljub temu smo za vse modele ocenili prilaganje z naslednjimi deloma primerljivimi kazalniki: celotna napaka (RMSE), ki meri povprečno razliko med dejanskimi vrednostmi in regresijskimi napovedmi, determinacijski koeficient R^2 in različni psevdo R^2 (opisi

izračunov so v opombah pod preglednico 8), Akaikejev informacijski kriterij AIC, ki je mera parsimonije, za polni model z vključitvijo vseh spremenljivk in model samo s konstanto, in logaritem funkcije verjetja, ki meri prilaganje statističnega modela, katerega parametri so dobljeni z metodo največjega verjetja.

Primerjavo dveh gnezdenih modelov, modela m_1 s spremenljivkami in modela m_2 s samo konstanto, smo opravili s testom razmerja verjetij (LR test) $LR = -2\ln(L(m_1) / L(m_2)) = 2(LL(m_2) - LL(m_1))$, ki se porazdeljuje v hi-kvadrat porazdelitvi s stopinjami prostosti, ki so razlika med stopinjami prostosti v m_2 in m_1 . Prav tako smo grafično analizirali porazdelitev ostankov in izdelali napovedi prirastka v odvisnosti od prsnega premera. Validacije modelov na neodvisnih podatkih nismo opravili, ker je namen prispevka zgolj primerjati modele in njihove napovedi. Modeli so zato veljavni le za populacijo, v kateri smo vzorčili.

4 REZULTATI

4 RESULTS

4.1 Primerjava regresijskih modelov

4.1 Comparison of regression models

Vsi modeli kažejo na podoben in istosmeren vpliv dejavnikov. Ugotovili smo, da je vpliv prsnega premera na prirastek v vseh modelih nelinearen in nemonoton; prirastek narašča s prsnim premerom drevesa do debeline 40 do 50 cm (odvisno od uporabljenega modela; slika 2), zatem se prirastek z naraščanjem prsnega premera zmanjšuje. V regresiji zlepkov kažejo nelinearen vpliv tudi sestojna temeljnica (G), vsota temeljnic debelejših dreves (BAL), Ginijev indeks diverzitete prsnih premerov (GINI) in nadmorska višina (NMV). Drugi modeli kažejo, da se z naraščajočo temeljnico in temeljnico debelejših dreves prirastek zmanjšuje. Uči-

Preglednica 2: Primerjava modela polinomske regresije po metodi najmanjših kvadratov (OLS) in polinomske regresije z dodanim šumom (OLS_NOISE)

Table 2: A comparison of ordinary least squares regression (OLS) to OLS with random noise added to the dependent variable (OLS_NOISE)

	OLS					OLS_NOISE				
	β	Sig.	SE	95% CI lower	95% CI upper	β	Sig.	SE	95% CI lower	95% CI upper
Const.	1,31733	***	0,08181	1,15697	1,47769	1,61034	***	0,05893	1,49483	1,72585
d_1	0,05969	***	0,00117	0,05740	0,06199	0,05155	***	0,00108	0,04944	0,05366
d_1^2	-0,00059	***	0,00001	-0,00062	-0,00056	-0,00050	***	0,00001	-0,00052	-0,00047
Sqrt (G_1)	-0,19111	***	0,00851	-0,20778	-0,17443	-0,17812	***	0,00774	-0,19329	-0,16295
BAL	-0,00511	***	0,00091	-0,00689	-0,00332	-0,00373	***	0,00083	-0,00535	-0,00211
Dg_1	0,00439	***	0,00095	0,00253	0,00626	0,00345	***	0,00086	0,00176	0,00514
GINI	1,09871	***	0,09074	0,92086	1,27657	1,05144	***	0,08205	0,89061	1,21226
NMV	-0,00023	***	0,00003	-0,00028	-0,00017	-0,00022	***	0,00003	-0,00027	-0,00017
NAGIB	-0,00841	***	0,00064	-0,00967	-0,00715	-0,00788	***	0,00059	-0,00903	-0,00674
k	0,08549	**	0,02605	0,03442	0,13656	1,61034	***	0,05893	1,49483	1,72585

Preglednica 3: Regresija z naravnimi zlepkami (SPLINE)

	SPLINE				
	β	Sig.	SE	95% CI lower	95% CI upper
Const.	2,34577	***	0,10015	2,14947	2,54206
ns(d ₁ , 3)1	0,54205	***	0,02672	0,48966	0,59443
ns(d ₁ , 3)2	1,62185	***	0,05453	1,51497	1,72873
ns(d ₁ , 3)3	0,39060	***	0,05893	0,27508	0,50611
ns(sqrt(G ₁), 4)1	-0,91321	***	0,05148	-1,01412	-0,81230
ns(sqrt(G ₁), 4)2	-0,84512	***	0,04730	-0,93784	-0,75240
ns(sqrt(G ₁), 4)3	-1,71581	***	0,12403	-1,95893	-1,47270
ns(sqrt(G ₁), 4)4	-0,66609	***	0,11440	-0,89033	-0,44185
ns(BAL, 3)1	-0,12945	***	0,03929	-0,20646	-0,05244
ns(BAL, 3)2	-0,31957	***	0,07926	-0,47492	-0,16422
ns(BAL, 3)3	-0,55672	***	0,11004	-0,77242	-0,34103
Dg ₁	0,21604	***	0,05455	0,10912	0,32296
ns(GINI, 3)1	0,09636	**	0,03108	0,03545	0,15728
ns(GINI, 3)2	0,42782	***	0,05387	0,32223	0,53341
ns(GINI, 3)3	0,56839	***	0,09197	0,38812	0,74866
ns(NMV, 4)1	-0,45689	***	0,06828	-0,59072	-0,32306
ns(NMV, 4)2	-0,28526	***	0,04664	-0,37668	-0,19383
ns(NMV, 4)3	-0,93471	***	0,15435	-1,23726	-0,63217
ns(NMV, 4)4	-0,33518	***	0,06235	-0,45740	-0,21296
NAGIB	-0,44946	***	0,03423	-0,51656	-0,38236
k	0,05689	*	0,02395	0,00995	0,10382

Table 3: Natural spline regression (SPLINE)

nek gostote in posledično konkurenčnih odnosov, ki jih nakazuje srednjetelemjnični premer (Dg_1), kaže, da se z naraščanjem srednjega premera prirastek povečuje. Izjema je rezana regresija, kjer je vpliv te spremenljivke negativen, a mejno značilen. Raznomernost sestojev (GINI) pospešuje prirastek jelke.

Modela OLS in OLS_NOISE se v napovedih vpliva dejavnikov pričakovano malo razlikujeta (preglednica 2). Razlike v koeficientih so posledica dodajanja slučajnostne komponente v odvisno spremenljivko, s čimer smo zadostili predpostavki o zveznosti odvisne spremenljivke. Ostanke v modelu OLS_NOISE so zato ustrezneje porazdeljeni (slika 2), celotna napaka napovedi (RMSE) je manjša, Akaikejev informacijski kriterij in logaritem funkcije verjetja sta v modelu OLS_NOISE manjša kot v modelu OLS, kar kaže na boljše prilaganje modela s šumom (preglednica 8).

Interpretacija regresije z zlepkami je na podlagi regresijskih koeficientov težja, zato pri tovrstni regresiji navadno uporabljamo grafične prikaze, s katerimi lahko zlepljene polinome najboljše predstavimo. Pozitiven linearni učinek na prirastek imata srednjetelemjnični premer (Dg_1) in rastišče (k), negativen linearni učinek ima samo nagib (NAGIB). Od osem vključenih spremenljivk jih ima pet nelinearen učinek, od tega NMV, BAL

in G negativen, druge spremenljivke pa pozitiven. Zaradi kompleksnosti interpretacije zlepkov smo rezultate modela SPLINE grafično predstavili (slika 2, slika 3).

V mešanem linearnem modelu s slučajnostno konstanto (MIXED, preglednica 4) nismo z vključitvijo ploskve kot slučajnostnega dejavnika bistveno izboljšali prilaganje modela (preglednica 8). Ugotovili smo, da obstajajo statistično značilne, a majhne razlike v regresijskih konstantah po ploskvah ($ui = 0,08072$, $p < 0,01$) in da obstaja zmerna korelacija prirastkov jelk znotraj ploskve ($ICC = 0,21557$). S tem smo pokazali, da je uporaba mešanega linearnega modela utemeljena, a da je vpliv ploskve na debelinski prirastek jelke majhen.

Model rezane regresije (TRUNC) in tobit model (TOBIT) kažeta podobne regresijske koeficiente, vpliv bonitete rastišča pri obeh modelih ni značilen (preglednica 5). Ker nas v modelih z omejenimi odvisnimi spremenljivkami navadno ne zanimajo učinki spremenljivk na latentno spremenljivko y^* , ampak učinki neodvisnih spremenljivk na omejeno odvisno spremenljivko y , v našem primeru na debelinski prirastek, ki je rezan (ali krnjen) pri 0, prikazujemo prilagojene učinke (ang. *marginal effects*). Zaradi rezanja in krnjenja dajejo modeli z omejenimi spremenljivkami v povprečju višje napovedi prirastkov kot modeli, v ka-

Preglednica 4: Mešani linearni model s slučajnostno konstanto (MIXED)

	MIXED				
	β	Sig.	SE	95% CI lower	95% CI upper
Const.	1,34010	***	0,09667	1,15034	1,52963
d_1	0,05665	***	0,00104	0,05461	0,05869
d_1^2	-0,00055	***	0,00001	-0,00057	-0,00052
Sqrt(G_1)	-0,18208	***	0,00874	-0,19922	-0,16494
BAL	-0,00376	***	0,00078	-0,00530	-0,00223
Dg_1	0,00383	***	0,00104	0,00178	0,00587
GINI	1,03994	***	0,10577	0,83258	1,24728
NMV	-0,00019	***	0,00004	-0,00026	-0,00012
NAGIB	-0,00806	***	0,00081	-0,00965	-0,00647
k	0,07718	**	0,03337	0,01176	0,14265
Ostanek / Residual	0,29375				
Sluč. konst. / Random intercept (ui)	0,08072				
Korelacija med meritvami znotraj ploskev / Intraclass correlation (ICC)	0,21557				
McFadden Pseudo R-squared	0,27451				

Table 4: Mixed linear model with random intercept (MIXED)

terih ne upoštevamo omejenosti (slika 4, preglednica 7). Aritmetična sredina odvisne spremenljivke je zaradi rezanja ali krnjenja od spodaj pri modelih z omejenimi spremenljivkami večja kot pri modelih OLS ali MIXED. Rezultati kažejo, da z modeli, ki ne upoštevajo omejenosti odvisne spremenljivke (OLS, OLS_NOISE in MIXED), nekoliko podcenjujemo prirastek.

Pri regresiji z grupiranimi podatki (GROUP) je vpliv spremenljivk večji (preglednica 6). Kadar je cilj zgolj opisno pojasniti odvisno spremenljivko, je interpretacija regresije z grupiranimi podatki podobna interpretaciji drugih modelov, težje pa bi ta model uporabili v

rastnih simulatorjih. Model GROUP ocenjuje centimetrške prirastke z nekoliko večjo napako $\pm 0,78$ cm.

Velikih razlik v prileganju modelov podatkom nismo zaznali (preglednica 8). Vsi modeli so značilno boljši kot modeli brez spremenljivk (LR test značilen). Pojasnenost variabilnosti prirastka je največja pri mešanem modelu, kjer nam je uspelo pojasniti 27,5 % variabilnosti. Povprečna natančnost napovedi v analiziranih modelih je od 0,51 cm do 0,78 cm. Če sodimo po celotni napaki, najboljše prileganje kaže mešani model s slučajnostnim faktorjem ploskev (RMSE = 0,51 cm). Med modeli OLS je model z vključenim šumom boljši

Preglednica 5: Dva regresijska modela iz skupine modelov z omejenimi odvisnimi spremenljivkami (truncated in tobit regresija)

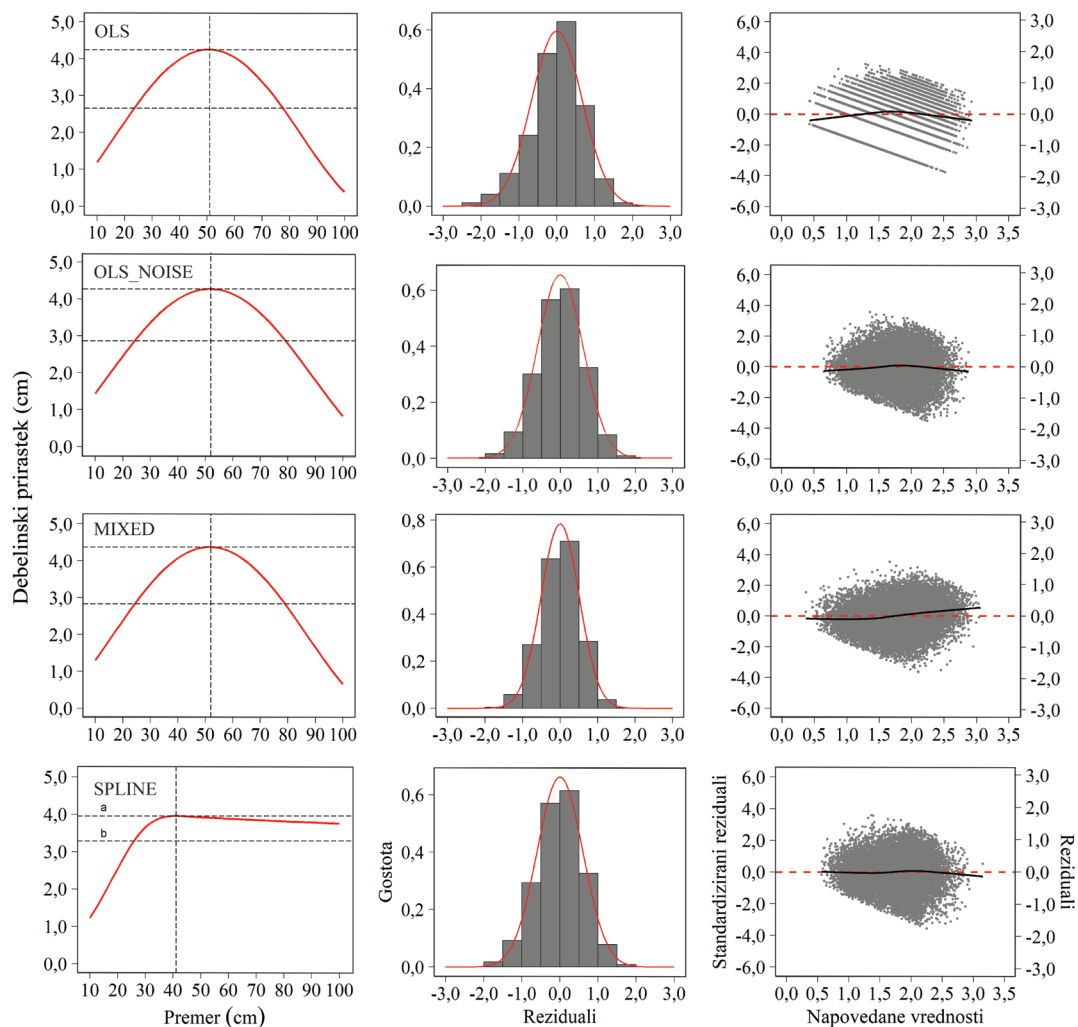
	TRUNC					TOBIT				
	β	Sig.	SE	95% CI lower	95% CI upper	β	Sig.	SE	95% CI lower	95% CI upper
Const.										
d_1	0,05458	***	0,00110	0,05242	0,05675	0,05303	***	0,00107	0,05093	0,05513
d_1^2	-0,00053	***	0,00000	-0,00056	-0,00051	-0,00051	***	0,00000	-0,00054	-0,00049
Sqrt(G_1)	-0,17949	***	0,00781	-0,19480	-0,16419	-0,18047	***	0,00770	-0,19557	-0,16538
BAL	-0,00415	***	0,00084	-0,00581	-0,00250	-0,00340	***	0,00082	-0,00502	-0,00178
Dg_1	-0,00356	*	0,00088	-0,00183	0,00529	0,0086	***	0,00086	0,00140	0,00478
GINI	1,07604	***	0,08368	0,91202	1,24005	0,08213	***	0,08213	0,87863	1,20059
NMV	-0,00022	***	0,000003	-0,00027	-0,00017	-0,00022	***	0,00003	-0,00027	-0,00017
NAGIB	-0,00807	***	0,00055	-0,00924	-0,00882	-0,00788	***	0,00058	-0,00902	-0,00674
k	0,03749	N.S.	0,02412	-0,00979	0,08476	0,03476	N.S.	0,02359	-0,01147	0,08099

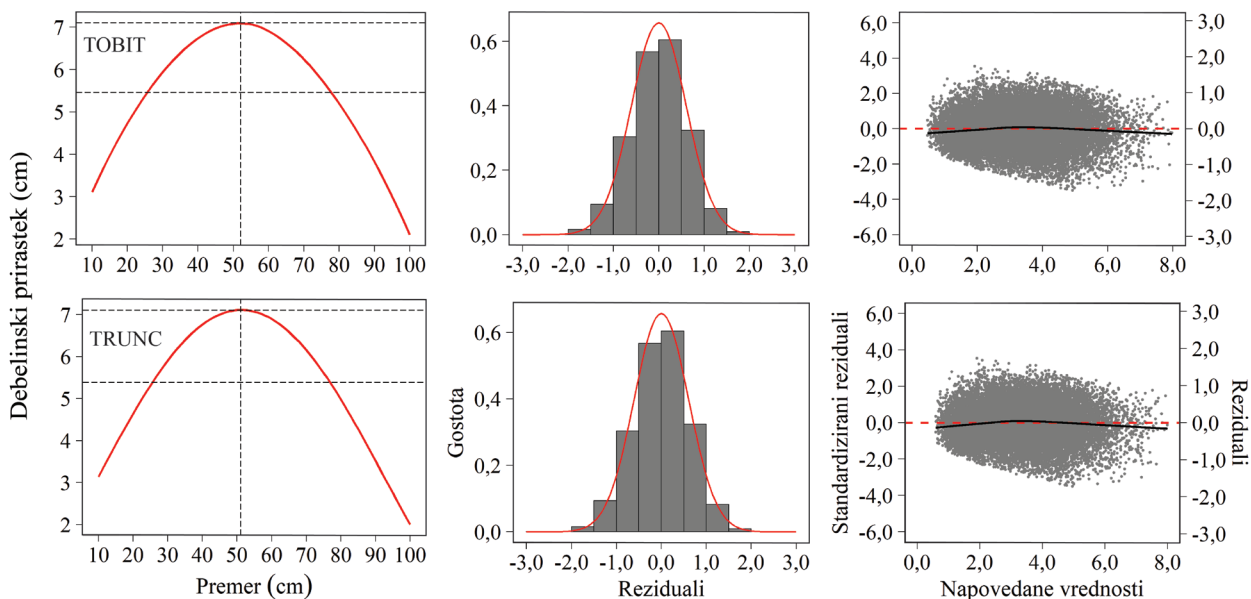
Table 5: Two limited dependent variable models (truncated and tobit regression)

Prilagojene vrednosti učinkov. Standardni odklon razlik med napovedmi in latentno spremenljivko (Scale factor) za TRUNC = 0,9838, za TOBIT = 0,9986. Marginal effects. Scale factor for marginal effects = 0.9838 and 0.9986 for the TRUNC model and the TOBIT model, respectively.

Preglednica 6: Tobit model krnjene regresije, v katerem so podatki v celoti krnjeni (GROUP)**Table 6:** Grouped regression models with completely censored data (GROUP)

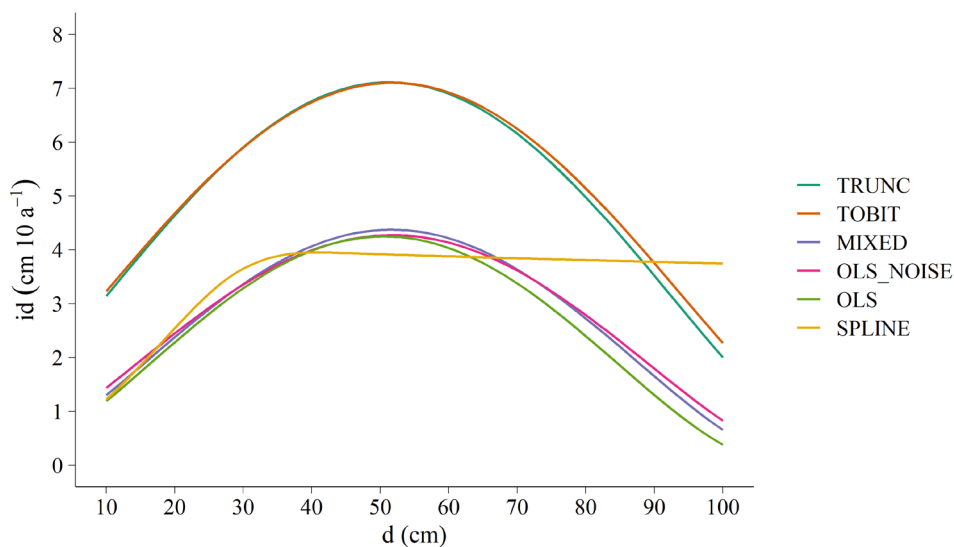
	GROUP				
	β	Sig.	SE	95% CI lower	95% CI upper
Const.	3,63608	***	0,29426	3,05935	4,21282
d_1	0,19688	***	0,00427	0,18851	0,20525
d_1^2	-0,00186	***	0,00005	-0,00197	-0,00175
Sqrt(G_1)	-0,77119	***	0,03043	-0,83083	-0,71155
BAL	-0,00919	***	0,00328	-0,01554	-0,00272
Dg_1	0,01101	***	0,00343	0,00429	-0,01772
GINI	4,12548	***	0,32582	3,48689	4,76407
NMV	-0,00093	***	0,00010	-0,00113	-0,000073
NAGIB	-0,03239	***	0,00231	-0,03693	-0,02786
k	0,16152	N.S.	0,09367	-0,02208	0,34512
Standardni odklon napake / Disturbance standard deviation					
Sigma	2,36365	***	0,01235	2,33944	2,38786

**Slika 2:** Napoved prirastka v odvisnosti od prsnega premera ter porazdelitev ostankov v modelu polinomske regresije (OLS), modelu polinomske regresije z dodanim šumom (OLS_NOISE), mešanem modelu s slučajnostno konstanto (MIXED) in modelu z naravnimi zlepci (SPLINE) (a - maksimum, b - povprečje)**Fig. 2:** Diameter increment prediction as a function of dbh and distribution of residuals in the polynomial regression model (OLS), the polynomial regression model with added noise (OLS_NOISE), the mixed model with a random intercept (MIXED) and the model with natural splines (SPLINE) (a - maximum, b - mean)



Slika 3: Napoved prirastka v odvisnosti od prsnega premera ter porazdelitev ostankov v tobit modelu (TOBIT) in modelu rezane regresije (TRUNC)

Fig. 3: Diameter increment prediction dependent on dbh, distribution of residuals in the tobit model (TOBIT) and truncated regression model (TRUNC)



Slika 4: Primerjava napovedanih vrednosti debelinskega prirastka v odvisnosti od prsnega premera pri modelih z zvezno odvisno spremenljivko

Fig. 4: Comparison of diameter increment predictions by regression model with a continuous dependent variable

Preglednica 7: Povprečne in ekstremne vrednosti napovedanega debelinskega prirastka (\bar{d}) v cm/10 let

Table 7: Mean, maximum and minimum of predicted tree diameter increment (\bar{d}) in cm/10 years

	Modeli linearne regresije		Naravni zleпки	Mešani modeli	Modeli z omejenimi odvisnimi spremenljivkami		
	Linear regression models		Natural spline regression	Mixed models	Limited dependent variable models		
	OLS	OLS_NOISE	SPLINE	MIXED	TRUNC	TOBIT	GROUP
Povp. / Mean	2,66	2,86	3,29	2,84	5,39	5,47	4,15
Min. / Min	0,38	0,61	0,41	0,47	2,00	2,27	-1,47
Maks. / Max	4,25	4,27	3,95	4,37	7,11	7,10	12,87

Preglednica 8: Mere prileganja regresijskih modelov**Table 8:** Model fit criteria

	Modeli linearne regresije			Mešani modeli	Modeli z omejenimi spremenljivkami		
	Linear regression models			Mixed models	Limited dependent variable models		
	OLS	OLS NOISE	OLS SPLINE	MIXED	TRUNC	TOBIT	GROUP
R ²	0,2668	0,2518	0,2653	0,2745	0,2558 ¹	0,2535 ²	0,9636 ¹
RMSE	0,6690	0,6088	0,6032	0,5110	0,6070	0,6065	0,7781
-LL	21360,61	19378,02	19174,97	18682,59	19174,00	19305,00	45830,00
AIC	42743,23	38778,03	38393,94	37389,18	38370,00	38634,00	91682,00
-LL _c	24628,34	22431,49	22431,49	21790,88	22317,00	22431,00	48824,00
AIC _c	49260,69	44866,98	44866,98	43587,77	44638,00	44867,00	97653,00
LR test	6536***	6107***	6513***	6217***	3143***	3126***	2994***

- LL: -Logaritem funkcije verjetja / -Log-likelihood. LLc: -Logaritem funkcije verjetja za model s konstanto / -Log-likelihood of the model with the intercept. AIC: Akaikejev informacijski kriterij / Akaike Information Criterion. AIC ob enakem logaritmu funkcije verjetja slabše vrednoti model z več parametri / AIC penalizes for complexity (number of parameters). AIC_c: AIC modela samo s konstanto / AIC of the model with the intercept.

¹ Ocena pojasnjene variabilnosti s kvadriranjem korelacijskega koeficienta med napovedanimi in dejanskimi vrednostmi prirastka / Rough estimation by correlating the dependent variable with the predicted value and squaring the result.

² Ocena pojasnjene variabilnosti s primerjavo variance napovedanih vrednosti in variance odvisne spremenljivke na celotni podatkovni bazi ($R_{ANOVA}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$, cf. Greene, 2012: E-1083). / Anova based fit measure which divides the variance of the estimated conditional mean by the variance of the observed variable using a full sample of observations ($R_{ANOVA}^2 = \frac{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$, cf. Greene, 2012: E-1083).

kot model brez šuma samo s fiksnimi faktorji (OLS). Če ocenjujemo kakovost modelov po več kriterijih, pa je najboljši regresijski model z grupiranimi podatki, kjer smo 14 centimetrskih razredov prirastka napovedali s povprečno točnostjo $\pm 0,78$ cm, s čimer smo dosegli 96,4% povezanost med dejanskimi vrednostmi in napovedmi (preglednica 8).

5 DISKUSIJA

5 DISCUSSION

5.1 Presoja regresijskih modelov

5.1 Assessment of regression models

Vpliv drevesnih, sestojnih in rastiščnih spremenljivk na debelinsko priraščanje jelke, kot ga kažejo regresijski modeli (preglednice 2 do 6), je skladen z literaturo (Trasobares in sod., 2004; Pukkala in sod., 2009; Schelhaas in sod., 2018). Pri mešanem modelu smo ugotovili statistično značilen učinek ploskev, kar je glede na velik vzorec ($n = 4.405$ ploskev) pričakovano. Zaradi nepraktičnosti uporabe mešanih modelov v rastnih simulatorjih, kjer ne moremo vključiti slučajnostnih dejavnikov oziroma je njihova vključitev pri nelinearnih modelih problematična, menimo podobno kot Pukkala in sod. (2009) in Vospernik (2021), da je za namen napovedovanja primerneje oblikovati modele brez slučajnostnih dejavnikov.

Nekoliko presenečajo razlike med modeli pri napovedih največjega prirastka v odvisnosti od prsnega pre-

mera (slika 4, preglednica 7). Vendar pa je pri primerjavi modelov pravilneje opazovati povprečne vrednosti napovedi, saj so regresijski parametri v modelu izračunani tako, da se modelske napovedi v povprečju najbolj ujemajo s podatki. Prav tako se moramo zavedati, da pri multipli regresiji poteka ocena regresijskih parametrov za vse spremenljivke hkrati in da je vrednost regresijskih parametrov izračunana tako, da kriterialna funkcija (vsota najmanjših kvadratov ali logaritem funkcije verjetja) doseže svoj ekstrem, zato opazovanje napovedi pri spreminjanju ene same spremenljivke v modelu ne daje prave slike o napovedni moči modela. Velike razlike med modeli v napovedih največjega prirastka glede na prsni premer pa so v največji meri posledica togosti polinomov druge stopnje. Kot kaže slika 4, zleпки dopuščajo mnogo bolj prilagodljive oblike za modeliranje nemonotonih zvez. V konkretnem primeru je prav nezmožnost parabole, da ponazori počasi pojemajoči prirastek po kulminaciji, razlog za izrazi-tejši odziv prirastka na prsni premer in višje napovedi največjega prirastka. To velja za vse modele, kjer je zveza med prsnim premerom in prirastkom v obliki parabole, razlike pa so večje pri modelih z omejenimi odvisnimi spremenljivkami, saj je aritmetična sredina odvisne spremenljivke v teh modelih zaradi rezanja ali krnjenja od spodaj večja kot pri modelih brez krnjenja.

Prileganje modelov je glede na to, da smo vključili le osnovne, pretežno sestojne spremenljivke in da smo

analizirali le pretežno raznomerne sestoje, zadovoljivo (prim. $R^2 = 0,26$ v Klopčič in sod., 2010 za celotno Slovenijo). Nekoliko boljše prileganje so ob vključitvi večjega števila klimatskih in talnih spremenljivk dosegli v primerljivih modelih v tujini ($R^2 = 0,29$ in $0,40$ v Schelhaas in sod., 2018 in Vospernik, 2021), vendar je v primeru vseevropskega modela za jelko (Schelhaas in sod., 2018) bil uporabljen nelinearni Gompertzov model, v primeru temeljnične rasti jelke v Avstriji (Vospernik, 2021) pa je bil uporabljen posplošeni aditivni model. Visok odstotek »pojasnjene variabilnosti« v modelu GROUP ni presenetljiv, saj v modelu GROUP R^2 predstavlja kvadrat Pearsonovega korelacijskega koeficienta med napovedanimi in dejanskimi vrednostmi, ki pa so celoštevilске. Model GROUP z izjemo najnižjega razreda razume izmerjeni prirastek kot cenzurirano vrednost dejanskih prirastkov nad izmerjenim prirastkom, zato napoveduje prirastke enako dobro za vse velikostne kategorije. Tega ne moremo trditi za model OLS, kjer se z oddaljevanjem od srednje vrednosti levo in desno povečuje širina intervala zaupanja za posamično in povprečno napoved, in sicer v obliki hiperbole (Košmelj, 2007; Whitlock in Schluter, 2009).

Glede na potencialne napake pri delu s podatki stalnih vzorčnih ploskev in možne posledice pri modeliranju se zastavlja vprašanje, kateri model je najboljši in kako v čim večji meri zmanjšati možne posledice napak. Enotnega odgovora na to ni, saj je izbira modela odvisna od teoretičnega izhodišča, kaj želimo preveriti, obsega napak in stopnje krnjenja, zahtevane točnosti napovedi in mnogih drugih dejavnikov. Zdi pa se, da bi se mnogim potencialno ogrožajočim napakam lahko izognili s sodobnejšo terensko inventuro, kjer bi meritve prsnih premerov opravljali z večjo natančnostjo, merske napake pa bi identificirali že na terenu. S tem bi podatki gozdnih inventur postali še bolj uporabni za raziskovalne namene in ne le za oceno gozdnih fondov.

5.2 Alternative parametričnih regresijskih modelov

5.2 Alternatives to parametric regression models

Nameni raziskovanja priraščanja dreves so različni. Če je predmet zanimanja predvsem napovedovanje odvisne spremenljivke na območju, ki ga določajo vrednosti neodvisnih spremenljivk in želimo posplošiti učinke neodvisnih spremenljivk na odvisno spremenljivko, je vsekakor smiselno uporabiti katero izmed parametričnih metod. Prednost regresijske analize je namreč v tem, da omogoča posploševanje iz vzorca na populacijo. V določenih primerih pa ne moremo zado- stiti predpostavkam (ne)linearnih regresijskih mode-

lov ali pa je naše raziskovalno vprašanje predvsem po- jasnjevalne narave in lahko uporabimo bolj posplošene oblike zvez ali neparametrične metode.

Posplošeni aditivni modeli (*Generalized Additive Models (GAM)*) omogočajo neparametrično prileganje podatkom z manj strogimi predpostavkami o dejanskem razmerju med odvisno in neodvisnimi spremenljivkami, s čimer lahko izboljšamo prileganje modela podatkom, vendar izgubimo jasnost interpretacije. V GAM zadostuje, da je zveza med Y in X gladka. Multivariatni adaptivni regresijski zleпки (*Multivariate adaptive regression splines (MARS)*) je neparametrična regresija, kjer zlepmo linearne funkcije in polinome višjega reda v zlepike tako, da je prileganje zleпка najboljše. Metoda tako avtomatsko poišče nelinearnosti v podatkih in jo lahko uporabljamo kot preliminarno analizo pred parametrično regresijo. Slabost metode je v velikem številu modelov, ki jih na ta način lahko izdelamo, kar otežuje odločitev za končni model. Podobna metoda je odsekoma linearna regresija (*Piecewise linear regression*), s katero lahko poiščemo območja neodvisnih spremenljivk, kjer je linearni odziv odvisne spremenljivke podoben. Z metodo lahko poiščemo točke preloma v prirastku in značilne razlike v nagibu premic. V preliminarni analizi utegnejo biti koristne tudi različne verzije lokalne regresije in gladilniki, kot so LOWESS (*Locally weighted scatterplot smoothing*), LOESS (*locally estimated scatterplot smoothing*), *Kernel average smoother* in druge, s katerimi si pomagamo predvsem pri vizualizaciji odnosov med spremenljivkami ali pri analizi homoskedastičnosti modela. Med polparametričnimi metodami so uporabna regresijska drevesa, ki so po delovanju najbolj podobna odsekoma linearni regresiji, pri čemer algoritem CART (Breiman in sod., 1984) rekurzivno razdeli podatke na več binarnih razcepov (vozlišč), dokler niso doseženi omejitveni kriteriji. Model ima veliko razlagalno, a omejeno napovedno vrednost. Podobno lahko rečemo tudi za metode iz skupine strojnega učenja, kot so umetne nevronske mreže (*Artificial neural networks (ANN)*), *Support vector machine (SVM)* in druge sorodne metode, kot so modelna drevesa (*Model trees (MT)*), ansambl modelnih dreves (*Model trees ensembles (BMT)*) in naključni gozdovi regresijskih dreves (*Random regression forests (RF)*) (npr. Jevšenak in Skudnik, 2021).

6 SKLEP

6 CONCLUSIONS

Pri uporabi inventurnih podatkov s stalnih vzorčnih ploskev se moramo zavedati omejitvev, zato moramo pred izbiro ustreznega statističnega modela preveriti izpolnjenost predpostavk in uporabiti najustre-

znejši model. V raziskavi smo obravnavali več kršitev modelov, s katerimi se srečujemo pri modeliranju z inventurnimi podatki, in predstavili ustrezne modele za določeno kršitev. Na podlagi podatkov o debelin-skem priraščanju jelke v raznomernih jelovo bukovih gozdovih nismo ugotovili zelo velikih razlik med preučevanimi modeli v vplivu posameznih spremenljivk. Nekoliko večje so bile razlike med modeli z omejenimi odvisnimi spremenljivkami (*tobit*, *truncated* in *grouped regression*) in modeli s privzeto normalno porazdelitvijo spremenljivk (linearni, polinomski, mešani modeli in model z zleпки). Potrdili smo znano lastnost modelov z omejenimi odvisnimi spremenljivkami, da pri omejenosti od spodaj ti modeli dajejo v povprečju večje napovedi kot modeli brez omejitev, zato bi jih kazalo v večji meri uporabiti, kadar je stopnja rezanja ali krnjenja spremenljivk velika. Model z naravnimi zleпки se je najbolj prilagal podatkom, vpliv spremenljivk v regresiji z zleпки je bil tudi biološko utemeljen, zato je regresija z zleпки poleg mešanega linearnega modela dobra izbira za statistično modeliranje. Predstavljeni modeli ne ponazarjajo vseh možnih izbir in sestav, ampak le glavne tipe modelov. Pri izboru modela je poleg izpolnjenosti predpostavk treba največ pozornosti nameniti vsebinski utemeljenosti modela in uporabnosti rezultatov.

7 POVZETEK

7 SUMMARY

In regression-based growth models, growth predictions are based on underlying assumptions of growth distribution. The problem of truncation and censoring of forest inventory data is mentioned in many growth and yield textbooks. However, the significance of bias in tree diameter predictions caused by truncation and censoring has not been examined. Ignoring truncation or censoring at zero in the distribution of the diameter increment may lead to an underestimation of growing potential or to an erroneous prediction of the impact of stand, site and competition variables. In this paper we examined the possibilities of statistical modelling of diameter increment with alternative regression models using the practical example of the permanent sample plot data provided by the Slovenia Forest Service. The aim of this paper is to 1) analyze the most common methodological challenges in using data from permanent sample plots for regression models, 2) test several limited dependent variable models, 3) determine the differences in diameter increment predictions by different regression models using data on the growth of fir on 4.405 plots in the period 1990–2014 in

uneven-aged Dinaric fir-beech forests and 4) evaluate the appropriateness of different models for modelling diameter increment. We discuss in detail the following challenges and their possible solutions: censoring and truncation of the dependent variable at a certain measurement threshold, failure to meet the assumption of continuity of the dependent variable, bias in periodic increment estimates due to rounding diameter measurements downwards, replacing linear models with nonlinear ones, replacing a small amount of data by pooling data from different time periods, and disregarding hierarchy in data.

We compared a polynomial regression model (OLS) to a regression model with added random noise (OLS_NOISE), a mixed linear model (MIXED), regression with natural splines (SPLINE), and three models with limited dependent variables: truncated regression (TRUNC), tobit regression (TOBIT) and grouped data regression (GROUP). All models showed a similar effect of tree diameter, stand basal area, basal area of larger trees, diameter structure diversity, altitude and slope, and there were only minor differences in the regression coefficients and fit measures. Limited dependent variable models generally showed higher predictions of increment than the other models. The mean predicted decadal diameter increment was 2.66 cm, 2.86 cm, 3.29 cm, 2.84 cm, 5.39 cm, 5.47 cm and 4.15 cm for the OLS, OLS_NOISE, SPLINE, MIXED, TRUNC, TOBIT and GROUP models, respectively. The root mean square error (RMSE) was 0.6690 cm, 0.6088 cm, 0.6032 cm, 0.5110 cm, 0.6070 cm, 0.6065 cm and 0.7781 cm for the respective models. The tobit model predicted a slower decrease in the growth of large-diameter trees after growth culmination, but no other greater discrepancies were found between the models with respect to the size of the effects.

8 ZAHVALA

8 ACKNOWLEDGEMENTS

Prispevek je nastal v okviru projektov Razvoj modelov za gospodarjenje z gozdovi v Sloveniji (CRP V4-2014) in Digitalizacija kmetijskih gospodarstev za načrtovanje gospodarjenja z gozdovi (DIGIGOZD). Projekt CRP V4-2014 sofinancirata Ministrstvo za kmetijstvo, gozdarstvo in prehrano in Javna agencija za raziskovalno dejavnost Republike Slovenije. Projekt DIGIGOZD je financiran v okviru PRP 2014-2020, podukrepa 16.2 Razvoj novih proizvodov, praks, procesov in tehnologij Evropskega partnerstva za inovacije (EIP), odločba št. 1119/2020/11. Zahvaljujemo se vsem financerjem za finančno podporo.

9 VIRI

9 REFERENCES

- Baker T.R., Affum-Baffoe K., Burslem D.F.R.P., Swaine M.D. 2002. Phenological differences in tree water use and the timing of tropical forest inventories: conclusions from patterns of dry season diameter change. *Forest Ecology and Management*, 171: 261–274.
- Bates D., Mächler, M., Bolker B., Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1: 1–48.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 2017. Classification and regression trees. *Classification and Regression Trees*: 1–358.
- Burnham K.P., Anderson D.R. 2002. Model selection and multimodel inference: a practical information theoretic approach. New York, Springer: 514 str.
- Croissant Y., Zeileis A. 2018. truncreg: Truncated Gaussian Regression Models. R package version 0.2-5. <https://CRAN.R-project.org/package=truncreg> (20. 8. 2021).
- Dănescu A., Albrecht A.T., Bauhus J. 2016. Structural diversity promotes productivity of mixed, uneven-aged forests in southwestern Germany. *Oecologia*, 182, 2: 319–333.
- Fichtner A., Schnabel F., Bruelheide H., Kunz M., Mausolf K., Schuldt A., Härdtle W., von Oheimb G. 2020. Neighbourhood diversity mitigates drought impacts on tree growth. *Journal of Ecology*, 108: 865–875.
- Fox J., Weisberg S. 2019. An {R} companion to applied regression. 3rd ed. Thousand Oaks CA, Sage.
- Greene W.H. 2003. *Econometric analysis*. 5th ed.. Upper Saddle River, NJ, Prentice Hall.
- Greene W.H. 2012. LIMDEP version 10.0: econometric modeling guide. Bellport, NY, Econometric Software, Inc.
- Henningsen A. 2020. censReg: censored regression (Tobit) models: R package version 0.5-32. <https://CRAN.R-project.org/package=censReg> (20. 7. 2021).
- Hočevar M. 1995. Dendrometrija-gozdna inventura: nelektorirano študijsko gradivo. Ljubljana, Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za gozdarstvo in obnovljive gozdne vire: 274 str.
- Hurlbert S.H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54: 187–211.
- Jevšenak J., Skudnik M. 2021. A random forest model for basal area increment predictions from national forest inventory data. *Forest Ecology and Management*, 479: 118601.
- Klopčič M., Matijašič D., Bončina A. 2010. Značilnosti debelinskega priraščanja jelke v Sloveniji. *Gozdarski vestnik*, 68: 204–213.
- Košmelj K. 2007. *Uporabna statistika*. 2. dopolnjena izd.. Ljubljana, Biotehniška fakulteta. http://www.bf.uni-lj.si/fileadmin/groups/2721/Uporabna_statistika_okt_2007/Uporabna_statistika_01.pdf (21.5.2021)
- Lexerod N.L., Eid T. 2006. An evaluation of different diameter diversity indices based on criteria related to forest management planning. *Forest Ecology and Management*, 222: 17–28.
- Mehtätalo L., Lappi J. 2020. *Biometry for forestry and environmental data with examples* in R. Chapman and Hall/CRC: 426 str.
- Oddi F.J., Miguez F.E., Ghermandi L., Bianchi L.O., Garibaldi L.A. 2019. A nonlinear mixed-effects modeling approach for ecological data: using temporal dynamics of vegetation moisture as an example. *Ecology and Evolution*, 9: 10225–10240.
- Pfützner D., Leibbrandt R., Powers D. 2009. Characterization and evaluation of similarity measures for pairs of clusterings. *Knowledge and Information Systems*, 19: 361–394.
- Pommerening A. 2002. Approaches to quantifying forest structures. *Forestry*, 75: 305–324.
- Pukkala T., Lahde E., Laiho O. 2009. Growth and yield models for uneven-sized forest stands in Finland. *Forest Ecology and Management*, 258, 3: 207–216.
- R Core Team. 2020. R: a language and environment for statistical computing. Vienna, R Foundation for Statistical Computing <https://www.R-project.org/> (25. 7. 2021).
- Ralston R., Buongiorno J., Schulte B., Fried J. 2003. Non-linear matrix modeling of forest growth with permanent plot data: the case of uneven-aged Douglas-fir stands. *International Transactions in Operational Research*, 10, 5: 461–482.
- Randall B.L., Ek A.R., Hahn J.T., Buchman R.G. 1988. STEMS model projection capability with incomplete tree list input data. *Northern Journal of Applied Forestry*, 5: 190–194.
- Rohner B., Waldner P., Lischke H., Ferretti M., Thürig E. 2018. Predicting individual-tree growth of central European tree species as a function of site, stand, management, nutrient, and climate effects. *European Journal of Forest Research*, 137, 1: 29–44.
- Schelhaas M.J., Hengeveld G.M., Heidema N., Thürig E., Rohner B., Vacchiano G., Vayreda J., Redmond J., Socha J., Fridman J., Tomter S., Polley H., Barreiro S., Nabuurs G.J. 2018. Species-specific, pan-European diameter increment models based on data of 2.3 million trees. *Forest Ecosystems*, 5: 21 str.
- Trasobares A., Pukkala T., Miina J. 2004. Growth and yield model for uneven-aged mixtures of *Pinus sylvestris* L. and *Pinus nigra* Arn. in Catalonia, north-east Spain. *Annals of Forest Science*, 61: 9–24.
- Vanclay J.K. 1994. *Modelling forest growth and yield: applications to mixed tropical forests*. Model. Forest growth yield Applications to Mixed Tropical Forest: 312 str.
- Vospersnik S. 2021. Basal area increment models accounting for climate and mixture for Austrian tree species. *Forest Ecology and Management*, 480, 118725.
- Whitlock M., Schluter D. 2009. *The analysis of biological data*. Roberts and Company Publishers, Greenwood Village, Colorado, USA: 704 str.
- Zar J.H. 2010. *Biostatistical analysis*. Prentice-Hall, Upper Saddle River, New Jersey, USA: 690 str.
- ZGS. 2010. Navodila za snemanje na stalnih vzorčnih ploskvah. Zavod za gozdove Slovenije, Ljubljana.