

TIESINIŲ MIŠRIŲJŲ MODELIŲ TAIKYMAS KELIO DANGOS SAVYBIŲ PAKARTOTINIUOSE MATAVIMUOSE

Jurgita Židanavičiūtė¹, Audrius Vaitkus²

Vilniaus Gedimino technikos universitetas, ¹ Matematinės statistikos katedra,

Vilniaus Gedimino technikos universitetas, ² Kelių tyrimo institutas,

Adresas: ¹ Saulėtekio al. 11, LT-10223 Vilnius, Lietuva,

² Linkmenų g. 28, LT-08217 Vilnius, Lietuva.

El. paštas: ¹ jurgita.zidanaviciute@vgtu.lt, ² audrius.vaitkus@vgtu.lt

Gauta: rugpjūtis 2015

Pataisyta: rugsėjis 2015

Paskelbta: lapkritis 2015

Santrauka. Straipsnyje atlikta Vilniaus Gedimino Technikos universiteto Kelių tyrimo institute surinktų duomenų statistinė analizė, siekiant įvertinti įvairių veiksnių įtaką tiriamos kelio dangos laikomajai gebai. Matavimai atlikti keturiuose skirtinguose kelio ruožuose (klasteriuose), kuriuose laikomoji geba matuojama atskiruose dangos taškuose, ir matavimai tuose pačiuose taškuose kartojami skirtingais metais ir jų sezonais – turimi pakartotiniai matavimai, dėl kurių duomenys yra priklausomi. Priklausomumas gali atsirasti ir dėl klasterių, nes dėl tam tikrų, tik konkrečiam klasteriui būdingų savybių, taškai klasteryje gali būti priklausomi. Dėl šių priežasčių šioje statistinėje analizėje vietoje įprastų tiesinės regresijos modelių yra taikomi tiesiniai mišrieji modeliai (angl. *Linear mixed model*). Pastarieji leidžia įvertinti laikomosios gebos skirtumus ne tik tarp tam tikrų fiksuotų dangos savybių, bet ir atsižvelgti į kiekvieno klasterio ar į tyrimą įtraukto dangos taško klasteryje ypatumus bei pakartotinių matavimų priklausomumą.

Reikšminiai žodžiai: tiesiniai mišrieji modeliai, pakartotiniai matavimai, priklausomi duomenys.

1. Įvadas

Tiesiniai mišrieji modeliai (angl. *Linear Mixed models*, toliau – LMM) yra žingsnis link modernesnių statistinių technikų nei klasikiniai tiesiniai regresijos modeliai, kuriuose laikomasi prielaidos, kad stebėjimai yra gauti iš tos pačios populiacijos ir yra nepriklausomi bei vienodai pasiskirstę. Tiesiniai mišrieji modeliai turi sudėtingesnę, daugialypę, hierarchinę struktūrą – stebėjimai tarp tam tikrų lygių ar klasterių laikomi nepriklausomais, bet stebėjimai tam tikrame klasteryje gali būti susiję, nes jie priklauso tai pačiai subpopuliacijai [1]. Šiais modeliais galima įvertinti ne tik skirtumus tarp klasterių, bet į modelį įtraukti ir žemiausių pagal hierarchiją subjektų, dalyvaujančių tyrime, ypatumus bei atsižvelgti į duomenų priklausomumą, atsiradusį dėl pakartotinių matavimų.

LMM modeliai nėra plačiai taikomi Lietuvoje sutinkamuose tyrimuose, to priežastis gali būti jų ne itin lengvai įsisavinama taikymo metodologija arba trūkumas statistikų bendradarbiavimo su kitos srities, kurioje šiuos modelius būtų galima taikyti, specialistais. Šis tiriamasis darbas atliktas drauge dviejų VGTU padalinių: Kelių tyrimo instituto ir Matematinės statistikos katedros, siekiant įvertinti bandomo kelio dangos konstrukcijos laikomąją gebą, jos skirtumus tarp atskirų dangos dalių, tam tikrų sezoninių veiksnių įtaką jai bei šios savybės dinamiką laike. Šios analizės rezultatai naudingi, vertinant kelio dangos konstrukcijų bei tam tikrų inžinerinių sprendimų, panaudotų tiesiant dangą, efektyvumą, išreiškiamą dangos konstrukcijos laikomąją gebą.

Pirmoje šio straipsnio dalyje pristatoma bendra LMM modelių išraiška, antroje dalyje aprašomi tyrime naudojami duomenys ir taikyta LMM vertinimo metodika. Trečioje dalyje pateikti įvertinti LMM modeliai ir jų interpretacija bei gautos išvados.

2. Tiesiniai mišrieji modeliai

LMM modeliai, kartais dar vadinami hierarchiniais ar modeliais pakartotiniams matavimams, yra svarbi modelių klasė, kurią galima taikyti įvairaus tipo duomenims [6, 7, 8]. Dažnai šių modelių taikymuose pasitaiko klasterizuoti duomenys, pavyzdžiui, mokinių, suskirstytų į klases, ar pacientų iš skirtingų gydymo įstaigų tyrimai, kuriuose klasterius sudaro atitinkamai mokyklos ar gydymo įstaigos. Klasterius taip pat gali sudaryti atsitiktiniai duomenų blokai iš tam

tikrų skirtingų gamybos partijų ir pan. Tuo atveju, kai klasteriai turi tam tikrą hierarchiją (klasės mokyklose ar pan.), tiesiniai mišrieji modeliai vadinami hierarchiniais. LMM modeliai taip pat tinkami ir pakartotinio matavimo tyrimams, kuriuose subjektų savybių matavimai kartojami laike ar prie skirtingų sąlygų. Sudėtingesni tyrimai gali apimti abi paminėtas duomenų savybes, t. y. duomenys suskirstyti į klasterius, kuriuose kintamieji matuojami laike arba prie skirtingų sąlygų (angl. *clustered longitudinal data*). Pastarajam duomenų variantui priklauso ir šiame straipsnyje naudojami duomenys. Tokio tipo duomenys gali pasitaikyti labai įvairiose srityse: nuo inžinerijos iki medicinos, lingvistikos ar žemės ūkio tyrimų [6, 7, 8].

LMM modeliai, kaip ir klasikiniai tiesiniai regresijos modeliai, yra tiesiniai parametrų atžvilgiu, bet nepriklausomi kintamieji juose gali apimti ne tik fiksuotus, bet ir atsitiktinius efektus – dėl šios priežasties šie modeliai vadinami mišriaisiais. Fiksuoti efektai atitinka įprastosios tiesinės regresijos koeficientus modelyje – jais gali būti bet kokio tipo kintamieji, kurių įvertinti parametrai nusako ryšį tarp priklausomo kintamojo ir nepriklausomų kintamųjų. Lyginant su įprastaisiais tiesiniais regresijos modeliais, mišrieji modeliai skiriasi sudėtingesne atsitiktine modelio dalimi, kuri LMM modeliuose aprašo dvejopą atsitiktinumą. Pirmiausia atsitiktinumą LMM modeliuose atsiranda dėl skirtingų subjektų, dalyvaujančių tyrime: kiekvienas žmogus ar kitas subjektas yra labai individualus, todėl tiriant net tas pačias jų savybes, jos gali priklausyti ne tik nuo žinomų ir į tyrimą įtrauktų požymių, bet ir nuo kiekvieno subjekto individualių, tyrėjams nežinomų ypatumų, dėl to LMM modeliuose suteikiama tam tikro parametro išsibarstymo tarp subjektų ar klasterių galimybė, aprašoma atsitiktiniu efektu. Be to, LMM modelių atsitiktinėje dalyje įvairiomis kovariacinės matricos struktūromis [2] galima aprašyti dėl pakartotinių matavimų liekanose atsiradusią priklausomybę. Taigi LMM modelio išraiška i -tajam subjektui (n – iš viso tyrime dalyvaujančių subjektų), kurio tiriamojo kintamojo reikšmių matavimai pakartoti k kartų, įgyja tokią išraišką:

$$Y_i = X_i \beta + Z_i u_i + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

$$u_i \sim N(0, D), \quad \varepsilon_i \sim N(0, R_i)$$

$$Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ \dots \\ Y_{ki} \end{pmatrix}, \quad X_i = \begin{pmatrix} X_{1i}^{(1)} & \dots & X_{1i}^{(p)} \\ X_{2i}^{(1)} & \dots & X_{2i}^{(p)} \\ \dots & \dots & \dots \\ X_{ki}^{(1)} & \dots & X_{ki}^{(p)} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_p \end{pmatrix}, \quad Z_i = \begin{pmatrix} Z_{1i}^{(1)} & \dots & Z_{1i}^{(q)} \\ Z_{2i}^{(1)} & \dots & Z_{2i}^{(q)} \\ \dots & \dots & \dots \\ Z_{ki}^{(1)} & \dots & Z_{ki}^{(q)} \end{pmatrix}, \quad u_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \dots \\ u_{qi} \end{pmatrix}, \quad \varepsilon_i = \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \dots \\ \varepsilon_{ki} \end{pmatrix},$$

čia:

Y_i – tiriamasis (tolydus) kintamasis;

X_i – aiškinamųjų kintamųjų matrica, kurioje kintamųjų skaičius p atitinka į modelį įtrauktų fiksuotų efektų ir jų sąveikų skaičių;

Z_i – atsitiktinių q efektų matrica. Pavyzdžiui, jei modelyje yra tik du atsitiktiniai efektai, iš kurių vienas skirtas laisvo nario atsitiktinumui įvertinti, kitas – krypties koeficiento atsitiktinumui tarp subjektų, tai atsitiktinių efektų matricą Z_i sudarys du stulpeliai – atitinkamai vienetų stulpelis ir matavimo momentų (laiko) reikšmių stulpelis:

$$Z_i = \begin{pmatrix} 1 & \dots & 1 \\ 1 & \dots & 2 \\ \dots & \dots & \dots \\ 1 & \dots & k \end{pmatrix} \text{ su vektoriumi } u_i = \begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix}. \text{ Kadangi daroma prielaida, kad atsitiktinių efektų vidurkis lygus nuliui, o jie}$$

įvertinami tik dispersija ir tarpusavio kovariacija, tai pastarieji dydžiai atvaizduojami matricoje

$$D = \begin{pmatrix} \text{Var}(u_{1i}) & \text{cov}(u_{1i}, u_{2i}) \\ \text{cov}(u_{2i}, u_{1i}) & \text{Var}(u_{2i}) \end{pmatrix}, \text{ kurios įstrižainės elementai atitinka efektų } u_{1i} \text{ ir } u_{2i} \text{ dispersijas, o ne įstrižainės}$$

elementai – kovariaciją tarp šių atsitiktinių efektų. Pavyzdžiui, esanti reikšminga teigiama kovariacija tarp u_{1i} ir u_{2i} reikštų, kad jei tam tikro subjekto efektas u_{1i} didesnis, tai jam ir u_{2i} efektas bus didesnis (ir atvirkščiai);

ε_i – modelio liekanos, kuriose atsiradusi priklausomybė dėl pakartotinių matavimų apibrėžiama kovariacijų matrica R_i :

$$R_i = \begin{pmatrix} \text{Var}(\varepsilon_{1i}) & \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \dots & \text{cov}(\varepsilon_{1i}, \varepsilon_{ki}) \\ \text{cov}(\varepsilon_{1i}, \varepsilon_{2i}) & \text{Var}(\varepsilon_{2i}) & \dots & \text{cov}(\varepsilon_{2i}, \varepsilon_{ki}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(\varepsilon_{1i}, \varepsilon_{ki}) & \text{cov}(\varepsilon_{2i}, \varepsilon_{ki}) & \dots & \text{Var}(\varepsilon_{ki}) \end{pmatrix}$$

LMM modeliams įvertinti yra didelis programinės įrangos pasirinkimas, atitinkamos procedūros realizuotos daugumoje populiariausių statistinių paketų: *SAS*, *SPSS*, *R* ir kt [7]. Šioje statistinėje analizėje naudotas *R* paketas [4] ir jo biblioteka *nlme* [5].

3. Duomenys naudojami statistinėje analizėje

Duomenys surinkti VGTU Kelių tyrimo institute. Naudojant krantinčio svorio deflektometrą ir kelio dangos konstrukciją veikiant dinamine apkrova, gaunamas dangos konstrukcijos laikomosios gebos parametras – ekvivalentinis tamprumo modulis E_0 , matuojamas megapaskaliais (MPa) (toliau dangos konstrukcijos laikomoji geba ir ekvivalentinio tamprumo modulis bus naudojami kaip sinonimai ir trumpinami E_0). E_0 matavimai atlikti skirtingais metų laikais keturiuose kelio dangos ruožuose (4-ame, 12-ame, 18-tame ir 24-ame), pasirenkant po 10 atsitiktinių taškų kiekvienoje ruožo apkrautos ir neapkrautos eismo pusės vėžėje (viso trys vėžės apkrautoje ir neapkrautoje eismo pusėse, arba 60 matavimo taškų kiekviename ruože). Pasirinktuose dangos taškuose be E_0 taip pat išmatuojamos ir kitos dangos konstrukcijos savybės: temperatūra ir drėgnis 100 cm, 130 cm ir 150 cm gyliuose, temperatūra dangos paviršiuje ir temperatūra skirtinguose asfaltbetonio sluoksnio gyliuose – 8 cm, 9 cm ir 10 cm. Tuose pačiuose dangos taškuose matavimai yra pakartoti aštuoniais skirtingais laiko momentais su ne vienodu laiko intervalu tarp matavimų: 2013 m. rugpjūčio, 2013 m. spalio, 2013 m. lapkričio, 2014 m. gegužės, 2014 m. liepos, 2014 m. rugsėjo, 2014 m. lapkričio ir 2015 m. balandžio mėn. Šioje pakartotinių matavimų schemoje priklausomas kintamasis yra E_0 arba pagal ankstesnių tyrėjų pasiūlytą metodą – jo perskaičiuotos reikšmės E_{0h9} , pastarosios gautos įvertinant asfalto sluoksnių temperatūros 9 cm gylyje pataisos koeficientą [3]:

$$E_{0h9} = \frac{E_0}{10^{-0,000221 \cdot h_{asf}^{1,0229} (t_{asf} - 20)}}, \quad (2)$$

čia

h_{asf} – asfalto sluoksnio storis (cm);

t_{asf} – matavimo metu 9 cm gylyje užfiksuota asfaltbetonio sluoksnio temperatūra (°C).

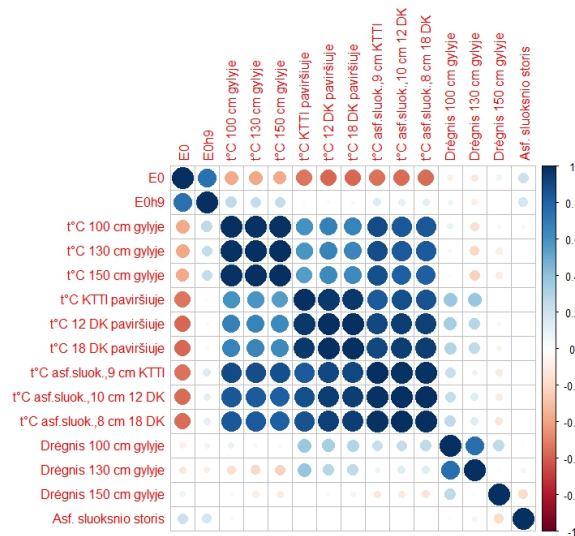
Nepriklausomi kintamieji šiame tyrime yra dviejų tipų: kelio ruožas, kelio pusė, kelio apkrovimo savybė ir asfaltbetonio sluoksnio storis yra požymiai, kurių reikšmės nekinta laike, o įvairių rūšių temperatūros bei drėgnis, kaip ir priklausomo kintamojo reikšmės, keičiasi priklausomai nuo oro sąlygų, buvusių matavimo metu (žr. 1 lentelę). Taigi šioje statistinėje analizėje turime subalansuotus pakartotinius matavimus laike, kur kiekvienas ruožas turi tą patį skaičių matavimo taškų dangoje su vienodu pakartotinių matavimų skaičiumi kiekviename taške.

1 lentelė. Tyrimo kintamųjų tipai

Priklausomas kintamasis	Nepriklausomi / aiškinamieji kintamieji		
	Ruožo požymis	Dangos taškų požymiai	Laike kintantys požymiai
Ekvivalentinis tamprumo modulis E_{0h9} arba E_0	Ruožo identifikacinis numeris	Dangos taško identifikacinis numeris, Kelio pusė (kairė, tarpvėžė ir dešinė), Apkrovimo savybė (Taip / Ne), Asfalto sluoksnio storis (mm)	Laikas (pakartotinių matavimų laiko momentai), Temperatūra (°C) ir drėgnis (%) 100 cm, 130 cm ir 150 cm gyliuose, Temperatūra (°C) dangos paviršiuje, Temperatūra asfaltbetonio sluoksnio 8 cm, 9 cm ir 10 cm gyliuose.

Koreliacijos tarp įvairių veiksnių, įtrauktų į tyrimą, yra vizualizuotos koreliacine matrica (žr. 1 pav.), kurios spalvos intensyvumas atitinka koreliacijos stiprumą: kuo tamsesnė mėlyna, tuo labiau dydžiai priklausomi teigiama koreliacija, kuo tamsesnė raudona, tuo stipresnė neigiama koreliacija tarp dydžių. Iš šio grafiko matome, kad pradinės E_0 reikšmės beveik vienodu stiprumu, tiesa ne itin dideliu (apie -0,5), neigiamai susijęs su visomis temperatūromis dangoje ir temperatūromis asfalto sluoksniuose, šiek tiek mažesnė E_0 koreliacija su temperatūra tam tikrame gylyje (apie -0,3). Visos minėtos E_0 koreliacijos su temperatūromis yra neigiamos, tai reiškia, kad temperatūrai didėjant, dangos laikomoji geba mažėja. Reiktų pastebėti, kad įvairios temperatūros tarpusavyje yra labai susijusios (koreliacija > 0,9), dėl to nėra labai didelio skirtumo, kuri iš jų bus įtraukta į modelį. Antroje ir trečioje matricos eilutėje matome, kad

perskaičiuotose reikšmėse E_{0h9} priklausomybė nuo temperatūrų pašalinta, tačiau pastebėta labai nežymi teigiama E_0 priklausomybė nuo temperatūros tam tikrame dangos konstrukcijos gylyje – koreliacija apie 0,25. Su asfalto sluoksnio storiumi E_0 koreliacija teigiama, bet nedidelė, siekia 0,2. Nuo drėgnio E_0 priklausomybė nepastebėta – koreliacija beveik nulinė.



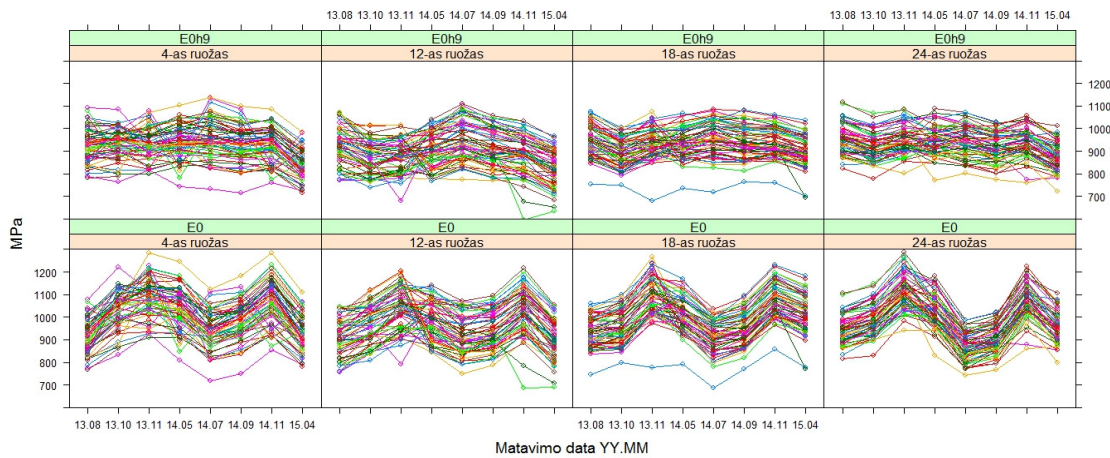
1 pav. Skirtingų veiksnių ir dangos konstrukcijos ekvivalentinio tamprumo modulio E_0 bei perskaičiuotų E_{0h9} reikšmių koreliacinė matrica

4. Mišriojo tiesinio modelio specifikacija ir įvertinimas

LMM modeliu galima įvertinti daug sudėtingesnes duomenų struktūras nei įprastiniais tiesinės regresijos modeliais, bet dėl to LMM modelių specifikacija yra sudėtingesnė ir apima daugiau žingsnių nei įprastoje tiesinėje regresijoje. Šiame tyrime yra naudojama iš viršaus į apačią (angl. „*top-down*“) modelio parinkimo strategija [5]. Pirmiausia vertinamas LMM modelis su visais į tyrimą įtrauktais fiksuotais ir galimais atsitiktiniais efektais. Antrame LMM modelio specifikavimo žingsnyje įvertinamas į modelį įtrauktų atsitiktinių efektų statistinis reikšmingumas – tikrinama hipotezė, ar atsitiktinis efektas, įtrauktas į modelį, gali būti pašalintas iš jo (pavyzdžiui, $H_0: \sigma_{u_i}^2 = 0, H_1: \sigma_{u_i}^2 > 0$). Šiai hipotezei patikrinti, iš pradinio modelio pašalinus tiriamą atsitiktinį efektą, modelis pervertinamas iš naujo bei, naudojant tikėtinumo santykio kriterijų (angl. *Likelihood ratio test*), palyginamas su pradiniu [7]. Trečiame modelio specifikavimo žingsnyje pasirenkama tinkama kovariacijų matrica modelio liekanoms. Pradiniame modelyje pagal nutylėjimą liekanoms yra numatyta pastovi dispersija kiekvienu matavimo momentu, bet perrinkimo būdu galima išmėginti ir kitas kovariacinės matricos struktūras [2, 7]. Tam, kad įvertintume, kuri iš jų yra tinkamesnė, kaip ir antrame žingsnyje, naudojame tikėtinumo santykio kriterijų dviem modeliams palyginti. Paskutiniame LMM modelio specifikavimo žingsnyje peržiūrime, ar modelyje yra statistiškai nereikšmingų parametru fiksuotiems efektams ir, jei yra, pašaliname juos iš modelio bei modelio parametrus pervertiname iš naujo.

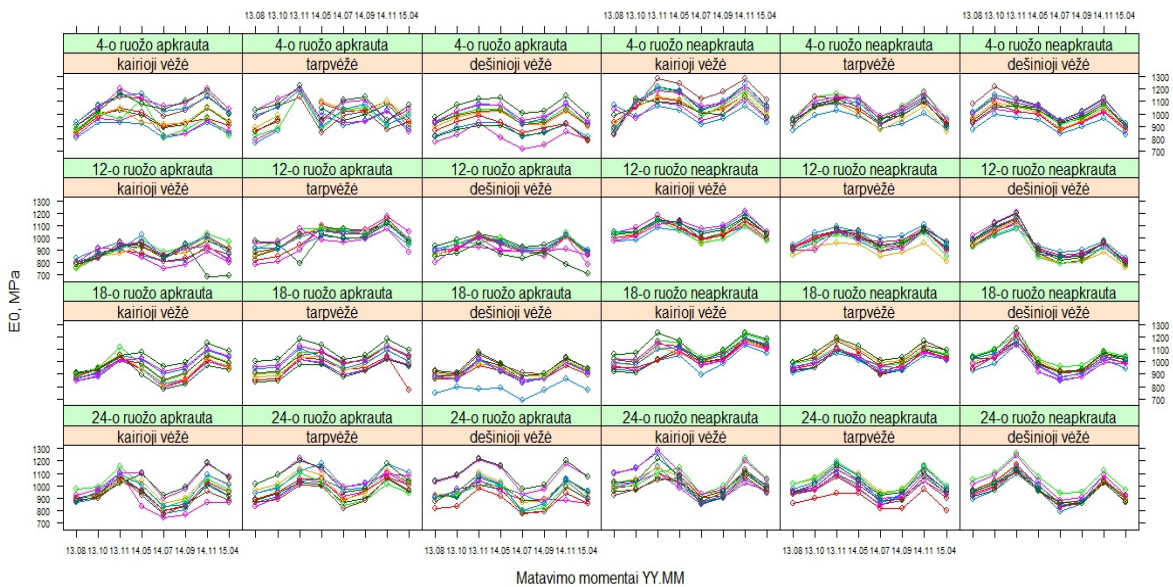
Fiksuotų ir atsitiktinių efektų parinkimas. Paprastai LMM modelio fiksuotų efektų įtraukimą į modelį apsprendžia keliama tyrimo tikslai (pavyzdžiui, identifikuoti priklausomojo kintamojo skirtumus tarp skirtingų dangos pusių ir pan.), o atsitiktinių efektų parinkimas nulemiamas tam tikros tyrimo specifikos (pavyzdžiui, žinoma, kad duomenys skirstomi į tam tikrus klasterius ar blokus). Papildomai atsitiktinių efektų nustatymui gali būti naudinga ir duomenų aprašomoji – grafinė analizė, kurioje galima pastebėti priklausomo kintamojo skirtumus tarp tiriamų subjektų ar klasterių. Tam, kad šiame tyrime išsiaiškintume, kokia kitimo tendencija būdinga E_0 ir kaip jo reikšmės pasiskirsto tarp skirtingų dangos taškų, toliau nagrinėsime E_{0h9} ir E_0 reikšmių dinamiką bendrai ir kiekviename matavimo taške atskirai. 2 pav. matomas E_0 būdingas sezoniskumas laike, o pagal 2-ą formulę perskaičiuotoms reikšmėms E_{0h9} – šiek tiek sugludinta kreivė, kurioje išryškėja reikšmių mažėjimo tendencija, labiau pastebima paskutiniais matavimo momentais. E_0 ir E_{0h9} vidurkiai per visą stebėjimo laikotarpį atitinkamai lygūs $984,4 \pm 100,91$ MPa ir $926,8 \pm 78,34$ MPa. Tačiau 2 pav. taip pat galima pastebėti, kad tiek E_{0h9} , tiek E_0 reikšmės net ir tame pačiame ruože turi ryškų išsibarstymą tarp taškų. Šios statistinės analizės tikslas – išsiaiškinti, kas turi didžiausią įtaką tokiam E_{0h9} ir E_0 reikšmių

pasiskirstymui tarp skirtingų matavimo taškų ruože. Kaip pagrindiniai veiksniai išskiriami šie fiksuoti parametrai: transporto apkrovos intensyvumas, eismo juosta ir kt. Tačiau, net ir priimant visiškai vienodas sąlygas, skirtumai tarp taškų vis tiek pastebimi (žr. 3 ir 4 pav.), vadinasi, be žinomų sąlygų, yra ir tam tikri nestebimi ir į statistinę analizę neįtraukiami atsitiktiniai veiksniai, turintys įtakos E_{0h9} ir E_0 reikšmėms (žr. modelius M1 ir M2).

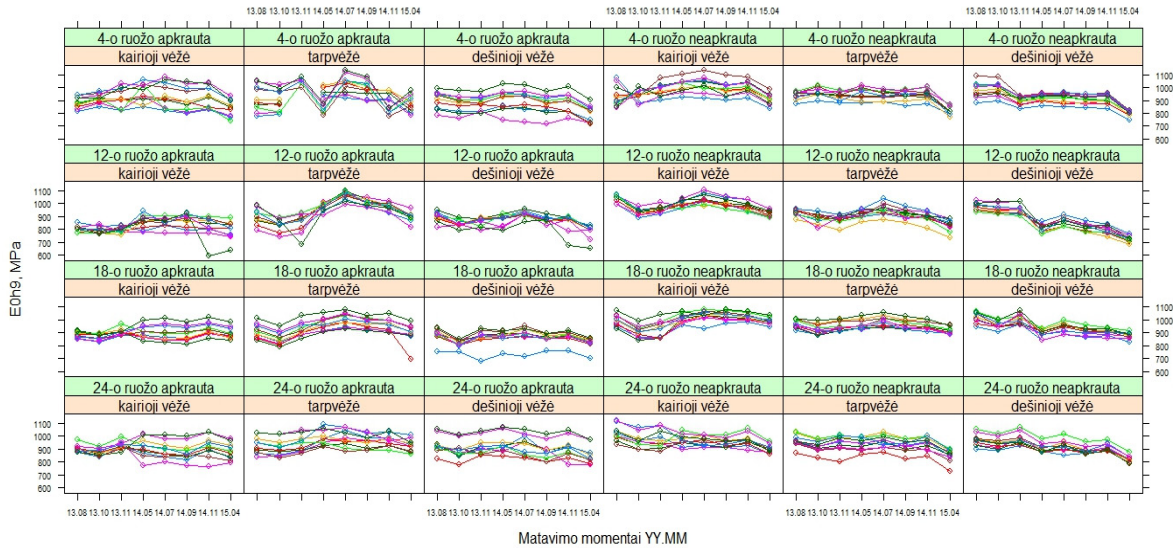


2 pav. Ekvivalentinio tamprumo modulio E_0 ir E_{0h9} dinamika dangos pavieniuose matavimo taškuose

Modelis M1. Pradėsime nuo paprastesnio modelio M1, kuriame priklausomas kintamasis – E_{0h9} . Pastarojo reikšmės yra gautos, perskaičiuojant pradinio E_0 reikšmės pagal kelių tyrėjų naudojamą 2-ą formulę [3], kuri pašalina asfaltbetonio 9 cm gylyje buvusios temperatūros poveikį ir atsižvelgia į asfalto sluoksnio storį. Šis E_0 reikšmių koregavimas turėtų sugludinti pradinį E_0 reikšmių dinamiką laike, panaikinant matomą sezoniškumą duomenyse. Bet kaip matome 2 pav. ne visais atvejais pavyko išvengti sezoniškumo, dėl to daroma prielaida, kad E_0 reikšmės priklauso ne tik nuo temperatūros asfaltbetonio sluoksnyje, bet ir nuo kitų sezoninių faktorių, kurie 2-ąja formule nepašalinami. Kadangi E_{0h9} reikšmės gautos, naudojant ne tiesinę išraišką (2-a formulė), tai į M1 modelį neįtraukiame papildomų sezoninių kintamųjų, nes tokio modelio rezultatai pradinėms ekvivalentinio tamprumo modulio reikšmėms būtų sunkiai interpretuojami. M1 modeliu tik įvertinsime, ar egzistuoja E_{0h9} reikšmių statistiškai reikšmingi skirtumai tarp atskirų kelio dalių, ar reikšminga tiesinė E_{0h} dinamika laike ir kaip šios reikšmės priklauso nuo dangos apkrovimo. Visi šie efektai modelyje traktuojami kaip fiksuotieji, nes jie nekinta tarp atskirų taškų, o yra susiję tik su tam tikromis fiksuotomis kelio dangos savybėmis.



3 pav. Ekvivalentinio tamprumo modulio E_0 dinamika atskiruose dangos taškuose pagal kelio puses ir apkrovimo lygį



4 pav. Ekvivalentinio tamprumo modulio E_{0h9} dinamika atskiruose dangos taškuose pagal kelio puses ir apkrovimo lygį

Lyginant E_{0h9} reikšmes tarp ruožų (žr. 2 pav.), matome, kad jo vidutinės reikšmės ir dinamika vizualiai šiek tiek skiriasi: visuose keturiuose ruožuose pastebėtas E_{0h9} reikšmių mažėjimas laike, bet ne visuose ruožuose vienodai greitai tai vyksta – krypties koeficientas skirtingas. Atsižvelgiant į tai, į M1 modelį įtraukiami šie atsitiktiniai efektai: laisvas narys ir krypties koeficientas tarp ruožų. Šie atsitiktiniai efektai padeda įvertinti E_{0h9} skirtumus tarp ruožų, kurie gali atsirasti dėl tam tikrų skirtingų inžinerinių sprendimų, taikytų tiesiant vieno ar kito ruožo dangą. Pastaroji informacija nėra įtraukta į statistinius duomenis, bet yra žinoma, kad tuo aspektu ruožų dangos konstrukcijos skiriasi. Taip pat iš 2 pav. galime pastebėti, kad E_{0h9} reikšmės tarp taškų kiekviename ruože irgi gana stipriai išsibarstę, dėl ko galime įtarti tam tikrą fiksuotais parametrais neaprašomą atsitiktinumą tarp taškų ruože. Pastarajam atsitiktinumui aprašyti į modelį įtrauksime atsitiktinį laisvą narį atskiriems taškams ruože. Atsitiktinis krypties koeficientas atskiriems taškams ruože nėra įtraukiamas į modelį, nes E_{0h9} dinamika visuose pasirinkto ruožo taškuose gana panaši. Galutinai vertinamas M1 modelis, kuriame be fiksuotų kintamųjų kelio pusės, apkrovimo ir laiko įtraukiami atsitiktiniai efektai, susiję tiek su kelio ruožais, tiek su dangos taškais ruožuose. Dėl pakartotinių matavimų modelio liekanose atsiradusiai priklausomybei užrašyti tinkamiausia pasirodė įstrižaininė kovariacijų matrica, su skirtingomis liekanų dispersijomis kiekvienu laiko momentu. Bendra M1 modelio išraiška, kuria aprašomas j -tojo ruožo ($j=1,\dots,4$) i -tajame taške ($i=1,\dots,60$) laiko momentu t , ($t=1,\dots,8$) ekvivalentinio tamprumo modulius E_{0h9} , pateikta 3-oje formulėje:

$$E_{0h9}_{ij} = \beta_0 + \beta_1 L_t + \beta_2 K_{ij} + \beta_3 T_{ij} + \beta_4 A_{ij} + \beta_5 L_t \times A_{ij} + \beta_6 L_t \times K_{ij} + \beta_7 L_t \times T_{ij} + u_{0j} + u_{1j} \times L + u_{0ij} + \varepsilon_{ij} \quad (3)$$

čia:

L – laikas (matavimo momentai);

K – indikatorinis kintamasis, žymintis kairę vėžę;

T – indikatorinis kintamasis, žymintis tarpvėžę;

A – indikatorinis kintamasis, žymintis apkrautą eismo juostą;

$L \times A$ – laiko ir apkrovos sąveika, t.y. šis požymis vertina skirtumus tarp dangos laikomosios gebos dinamikos apkrautoje ir neapkrautoje kelio pusėse;

$L \times K$ – laiko ir kairės vėžės sąveika, t. y. šis požymis vertina skirtumus tarp dangos laikomosios gebos dinamikos kairėje ir dešinėje kelio pusėse;

$L \times T$ – laiko ir tarpvėžės sąveika, t. y. šis požymis vertina skirtumus tarp dangos laikomosios gebos dinamikos tarpvėžėje ir dešinėje kelio pusėse;

Parametrai $\beta_0 - \beta_7$ atitinka fiksuotus efektus modelyje: laisvą narį, laiką, kelio pusę, apkrovą ir šių efektų sąveikas.

Likę, fiksuotais parametrais neįvertinti E_{0h9} , skirtumai duomenyse aprašomi šiais atsitiktiniais parametrais modelyje:

u_{0j} ir u_{1j} – atsitiktiniai efektai, aprašantys laisvo nario ir krypties koeficiento skirtumus tarp ruožų. Šių atsitiktinių efektų vidurkis lygus nuliui, o jų dispersija ir kovariacija pateikiama D matricoje, kurios įstrižainės elementai atitinka atsitiktinių efektų u_{0j} ir u_{1j} dispersiją; o ne įstrižainės elementai – kovariaciją tarp šių atsitiktinių efektų:

$u_{0ij} \sim N(0, \sigma_{laisvas\ narys\ tarp\ taškų\ (ruože)})$ – atsitiktinis efektas, susijęs su laisvo nario skirtumais tarp atskirų ruožo taškų;

ε_{ij} – modelio liekanos. Šiame modelyje daroma prielaida, kad tam tikro subjekto liekanų vektorius ε_{ij} yra atsitiktinis kintamasis, pasiskirstęs pagal daugiamačių normalųjį skirstinį su nuliniu vidurkiu ir įstrižainine dispersijų matrica R_{ij} :

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{1ij} \\ \dots \\ \varepsilon_{8ij} \end{pmatrix} \sim N(0, R_{ij}), R_{ij} = \begin{pmatrix} r_{11} & 0 & 0 & 0 \\ 0 & r_{22} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & r_{88} \end{pmatrix}.$$

Visi fiksuoti parametrai M1 modelyje yra statistiškai reikšmingi (žr. 2 lentelę). Prie laiko ir apkrovos kintamųjų esantys neigiami parametrai rodo atitinkamai dangos laikomosios gebos reikšmingą mažėjimą laike bei apkrovos daromą neigiamą įtaką laikomajai gebai. Apkrautuose taškuose E_{0h9} yra vidutiniškai 88,45 MPa mažesnis nei neapkrautuose. Taip pat skiriasi apkrautų ir neapkrautų taškų laikomosios gebos dinamika laike – apkrautuose taškuose ji mažėja lėčiau nei neapkrautuose (žr. 2 lentelėje β_5 parametras). Pašalinus laiko poveikį, dar egzistuoja E_{0h9} reikšmingi skirtumai tarp skirtingų kelio pusių: dangos laikomoji geba kairėje ir tarpvėžėje yra mažesnė nei dešinėje eismo pusėje.

Modelis M2. Šiame modelyje vertinamas toks pat tiesinis mišrusis modelis kaip ir M1 (su tais pačiais fiksuotais ir atsitiktiniais efektais), bet jame priklausomas kintamasis yra E_0 reikšmės, iš kurių nepašalintas temperatūrų ir asfalto storio faktorius (žr. 2 pav.), dėl to papildomai į modelį įtraukiama temperatūra, asfalto sluoksnio storis ir temperatūros ir drėgnio sąveika kaip atskiri fiksuoti parametrai modelyje – atitinkamai β_8, β_9 ir β_{10} . Kaip jau buvo minėta anksčiau, neperskaičiuotos E_0 reikšmės neigiama koreliacija susijusios su įvairiomis temperatūromis, bet kadangi įvairių rūšių temperatūros stipriai koreliuoja tarpusavyje (koreliacija $> 0,9$), tai į modelį įtrauksime tik vieną iš jų – temperatūrą asfaltbetonio sluoksnyje 9 cm gylyje (KTTI), kuri pagal 2-ą formulę yra naudojama prieš tai modelyje priklausomojo kintamojo E_{0h9} reikšmėms glodinti. Modelio M2 vertinama išraiška pateikta (4) formulėje, kurioje visi pažymėjimai atitinka 3-iosios formulės žymėjimus:

$$\begin{aligned} E_{0ij} = & \beta_0 + \beta_1 L_t + \beta_2 K_{ij} + \beta_3 T_{ij} + \beta_4 A_{ij} + \beta_5 L_t \times A_{ij} + \beta_6 L_t \times K_{ij} + \beta_7 L_t \times T_{ij} + \\ & \beta_8 TEMP + \beta_9 ASF + \beta_{10} TEMP \times DREGNIS + \\ & + u_{0j} + u_{1j} \times L + u_{0ij} + \varepsilon_{ij} \end{aligned} \quad (4)$$

Visi fiksuoti efektai modelyje M2 lieka statistiškai reikšmingais su panašiais parametų įverčiais kaip ir modelyje M1 (žr. 2 lentelę). Papildomas kintamasis temperatūra turi neigiamą efektą kelio dangos laikomajai gebai ($\beta_8 = -8,41$, $p < 0,001$), o asfalto sluoksnio storis – teigiamą ($\beta_9 = 4,37$, $p = 0,0001$), tai reiškia, kad asfaltbetonio sluoksnio 9 cm gylyje temperatūrai padidėjus 1 laipsniu, E_0 vidutiniškai sumažėja 8,41 MPa, o asfalto sluoksniui padidėjus 10 mm, kelio dangos laikomoji geba vidutiniškai padidėja beveik 44 MPa, kai visų kitų parametų reikšmės modelyje nekinta. Temperatūros ir drėgnio sąveika modelyje yra nereikšminga ($p = 0,391$).

2-oje lentelėje taip pat pateikti M1 ir M2 modelių atsitiktinių parametų įverčiai (dispersijos ir koreliacijos). Matome, kad ruožo atsitiktiniai parametrai yra priklausomi tarpusavyje, jų koreliacija M1 modelyje lygi $-0,594$, o M2 modelyje $-0,34$, tai reiškia, kad jei tam tikrame ruože didesnis laisvas narys, tai mažesnis krypties koeficientas. Kitaip sakant, jei kelio dangos laikomoji geba tam tikrame ruože yra didesnė, tai ji lėčiau mažėja laike (arba atvirkščiai – šiuo atveju koreliacija neparodo priešingumo). Reikšmingiems atsitiktiniams skirtumams tarp taškų tam tikrame ruože gali turėti įtakos skirtingas apkrovimo lygis, kuris į tyrimo duomenis neįtrauktas – žinoma tik, kad taškas kelio dangoje yra apkrautas arba ne, bet nežinoma, kaip stipriai jis apkrautas.

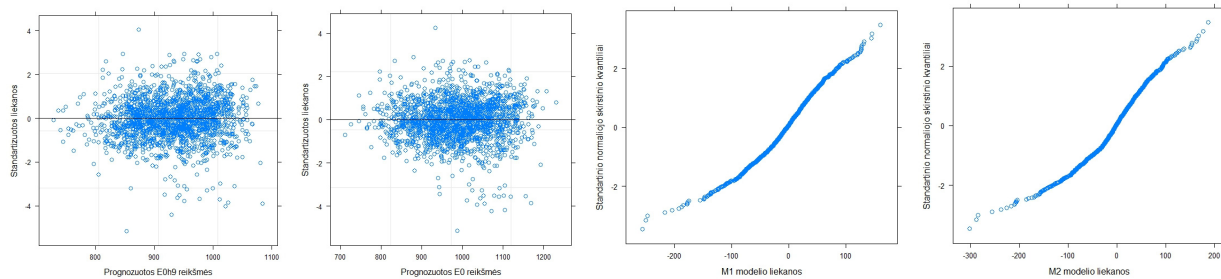
Modelių M1 ir M2 adekvatumas. Apie M1 ir M2 modelių adekvatumą galime spręsti iš modelių liekanų, kurios pateiktos 5 pav. Matome, kad tiek M1, tiek M2 modelio liekanos artimos normaliajam skirstiniui, bet taip pat neigiamoje liekanų dalyje pastebimos tam tikros išskirtys, iš kurių galime spręsti, kad į modelį įtraukti efektai nevisiškai

paaiškina priklausomojo kintamojo kitimą. Iš kitos pusės, tiek M1, tiek M2 modeliai yra skirti išsiaiškinti egzistuojančius priklausomojo kintamojo skirtumus tarp įvairių dangos dalių ir jo priklausomybę nuo tam tikrų sezoninių veiksnių, bet šie modeliai nėra skirti E_0 reikšmių prognozei pagal į modelį įtrauktus požymius.

2 lentelė. M1 ir M2 modelių įvertinti parametrai

Modelio parametrai	Parametrų įverčiai (Standartinė paklaida), parametro reikšmingumas p	
Fiksuoti parametrai:	M1 modelis	M2* modelis
β_0 (Laisvas narys)	1 002,65 (11,19), $p < 0,001$	454,96 (79,89), $p < 0,001$
β_1 (Laikas)	-17,27 (1,76), $p < 0,001$	-18,72 (1,99), $p < 0,001$
β_2 (Kairė vėžė)	-21,97 (10,04), $p = 0,029$	-21,31 (10,08), $p = 0,0356$
β_3 (Tarpvėžė)	-21,64 (10,06), $p = 0,0324$	-32,13 (10,14), $p = 0,0017$
β_4 (Apkrauta pusė)	-88,45 (8,21), $p < 0,001$	-115,11 (8,71), $p < 0,001$
β_5 (Laikas \times apkrauta pusė)	9,33 (1,01), $p < 0,001$	10,03 (1,08), $p < 0,001$
β_6 (Laikas \times kairė vėžė)	14,16 (1,24), $p < 0,001$	14,99 (1,32), $p < 0,001$
β_7 (Laikas \times tarpvėžė)	15,58 (1,24), $p < 0,001$	17,09 (1,32), $p < 0,001$
β_8 (Temperatūra)	-	-8,41 (0,17), $p < 0,001$
β_9 (Asfalto sluoksnio storis)	-	4,37 (0,48), $p < 0,001$
β_{10} (Temperatūra \times drėgnis)	-	-0,027 (0,03), $p = 0,391$
Atsitiktinių efektų parametrai:	$\sigma_{\text{laisvo nario tarp ruožų}} = 15,21$	$\sigma_{\text{laisvas narys tarp ruožų}} = 21,88$
standartinis nuokrypis $\sqrt{\sigma^2}$ ir	$\sigma_{\text{krypties koef. tarp ruožų}} = 2,87$	$\sigma_{\text{krypties koef. tarp ruožų}} = 3,37$
koreliacija (matricos	$cor_{\text{laisvo nario ir krypties koef. tarp ruožų}} = -0,59$	$cor_{\text{laisvo nario ir krypties koef. tarp ruožų}} = -0,34$
D elementai)	$\sigma_{\text{laisvo nario tarp taškų (ruože)}} = 49,15$	$\sigma_{\text{laisvo nario tarp taškų (ruože)}} = 45,98$

* $\beta_0 - \beta_9$ parametrų įverčiai lentelėje pateikti pašalinus nereikšmingą temperatūros ir drėgnio sąveiką modelyje.



5 pav. M1 ir M2 modelių standartizuotos liekanos

5. Išvados

Atlikta statistinė analizė padėjo išryškinti tam tikrus kelio dangos laikomosios gebos, kuri nusakoma ekvivalentiniu tamprumo moduliui, skirtumus atskirose dangos dalyse, įvertinti jos dinamiką laike bei priklausymą nuo tam tikrų sezoninių veiksnių. Gauti rezultatai naudingi kelių inžinieriams, priimant tam tikrus sprendimus, gerinančius dangos konstrukcines savybes.

Rezultatai parodė, kad kelio dangos laikomoji geba laikui bėgant silpnėja. Taip pat ją silpnina kelio apkrova bei tam tikri inžineriniai sprendimai lemia, kad dangos geba skiriasi tarp kelio pusių. LMM modelio galimybės leido pastebėti ir egzistuojančius atsitiktinius skirtumus tarp dangos taškų, kurie nėra paaiškinami į modelį įtrauktais fiksuotais efektais. Pastarasis aspektas būtų likęs nepastebėtas, jei būtume naudoję įprastinius tiesinės regresijos

modelius, kuriuose tam tikros savybės apibendrinamos visiems stebėjimams. Viena iš priežasčių, kad net ir įvertinus fiksuotus parametrus, dangos laikomoji geba taškuose skiriasi, gali būti tam tikri, į tyrimą neįtraukti veiksniai, t. y. ruožo konstrukcinės ypatybės ar apkrovos dydis, tenkantis kiekvienam taškui.

Literatūra

1. Demidenko, E. 2013: Mixed models theory and applications with R. John Wiley & Sons.
2. Littell R.C., and Pendergast J. and Natarajan, R. 2000: Modelling covariance structure in the analysis of repeated measures data, *Statistics in medicine* *Statist. Med.* 19:1793–1819, [interaktyvus], [žiūrėta 2015-08-28]. Prieiga per internetą: <http://www.stat.unm.edu/~bedrick/glm/Littell_StatMed.pdf>
3. Motiejūnas A., and Paliukaitė M., and Vaitkus A., and Čygas D., and Laurinavičius, A. 2010: Research on the dependence of asphalt pavement stiffness upon the temperature of payment layers. *The Baltic Journal of Road and Bridge Engineering*, vol. 5, no. 1.
4. Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team 2015. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-122, [interaktyvus], [žiūrėta 2015-08-28]. Prieiga per internetą: <<http://CRAN.R-project.org/package=nlme>>.
5. R Core Team. R: A Language and Environment for Statistical Computing.
6. Wang Z., and Goonewardene, L. A. 2004: The use of MIXED models in the analysis of animal experiments with repeated measures data, *Canadian Journal of animal science*, 2004, 84: 1–11.
7. West B. T., and Welch K.B, and Galecki, A.T 2015: *Linear Mixed Models: A Practical Guide Using Statistical Software* (Second Edition);
8. Winter, B. 2013: Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499 [interaktyvus], [žiūrėta 2015-08-28]. Prieiga per internetą: <<http://arxiv.org/pdf/1308.5499.pdf>>

APPLICATION OF MIXED LINEAR MODELS IN THE ANALYSIS OF ROAD SURFACE FEATURES

Jugita Židanavičiūtė, Audrius Vaitkus

Abstract. The data were collected by researchers at the Road Research Institute, in a study investigating the impact of different factors on road surface strength. In this statistical analysis, we apply linear mixed models (LMMs) to clustered longitudinal data, in which the units of analysis (points in the road) are nested within clusters (sample of four different road segments), and repeated measures of road strength in these different points are collected over time with unequally spaced time intervals. The data are balanced – each cluster has the same number of units, which are measured at the same number of time points. Because of correlated data and different clusters in which data could be correlated, linear regression models are not appropriate here, and therefore linear mixed models are applied.

Keywords: mixed linear models, repeated measurements, clustered longitudinal data.