

TARTU ÜLIKOOL

LOODUS- JA TÄPPISTEADUSTE VALDKOND

MATEMAATIKA JA STATISTIKA INSTITUUT

Joonas - Sander Tamm

**Genotüübi ja keskkonna vahelist koosmõju hindavate meetodite
kirjeldus ja võrdlus NOBALwheat projekti andmete näitel**

Matemaatilise statistika eriala

Magistritöö (30 EAP)

Juhendaja: prof. Tanel Kaart

Juhendaja: Ilmar Tamm (METK)

TARTU 2024

**GENOTÜÜBI JA KESKKONNA VAHELIST KOOSMÕJU HINDAVATE MEETODITE
KIRJELDUS JA VÕRDLUS NOBALWHEAT PROJEKTI ANDMETE NÄITEL**

Magistritöö

Joonas-Sander Tamm

Lühikokkuvõte

Taimede sordiaretuses on oluline, et aretatavatel sortidel oleks kõrge saagikus. Lisaks sellele on tähtis, et saagikus püsiks võimalikult stabiilne ka ebasoodsates kasvutingimustes. Eriti oluline on kliima-resistentsete genotüüpide aretus tänapäeval, kus kliimamuutustega seoses on ilmastikuolud ja seega ka taimede kasvutingimused muutunud ebastabiilsemaks, suurenenud on ekstreemsete ilmastikutingimuste esinemine.

Käesolevas magistritöös seletatakse paljude stabiilsust mõõtvate statistikute tähendusi, tagamaid ning võrreldakse neid omavahel. Lisaks viiakse läbi stabiilsusstatistikute arvutamine ja võrdlus rahvusvahelise NOBALwheat projekti raames Balti riikides ja Norras aastatel 2021-2023 hinnatud 300 suvinisu genotüübi saagikuse andmetel. Töö tulemusena koostati selgitused valitud 19 statistiku interpreteerimiseks ja anti soovitused nende valikuks saagikamate ja stabiilsemate genotüüpide väljaselgitamisel. Andmeanalüüsi tulemusena selgitati välja saagikamate ja muutuvates ilmastikutingimustes stabiilsemate suvinisu sortide aretamiseks sobivam lähtematerjal.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatematika.

Märksõnad: Dispersioonanalüüs, stabiilsus, statistilised meetodid, peakomponentanalüüs

**DESCRIPTION AND COMPARISON OF GENOTYPE AND ENVIRONMENT INTERACTION
MEASURING STATISTICS ON THE BASIS OF DATA FROM NOBALWHEAT PROJECT**

Master thesis

Joonas-Sander Tamm

Abstract

In plant breeding, it is important that the varieties to be bred have a high yield. In addition, it is important that the yield would remain as stable as possible even in unfavorable growing conditions. Nowadays, the breeding of climate-resistant varieties is especially important, when due to climate change, the weather conditions and thus the growing conditions of plants have become more unstable and the occurrence of extreme weather conditions has increased. In this master thesis, different stability measuring statistics were explained and compared with each other. In addition, the measurement and comparison of stability statistics

was carried out on the data from international NOBALwheat project based on yield data of 300 spring wheat genotypes evaluated in the Baltic States and Norway in 2021-2023. Based on data analysis, the highest yielding and stable genotypes were selected as initial material for further spring wheat breeding in changing weather conditions.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics.

Key Words: Analysis of variance, stability, statistical methods, principal component analysis.

Sisukord

Sissejuhatus	5
1 Andmed	6
2 Parameetrilised meetodid	7
2.1 Lineaarne regressioon	8
2.2 AMMI	9
2.3 Zhangi D indeks	10
2.4 AMMI stabiilsusväärtus (ASV)	10
2.5 Genotüübi valikuindeks (GSI)	10
2.6 BLUP	12
2.7 WAASB	13
2.8 WAASBY	14
2.9 Ökovalents	15
2.10 Stabiilsuse dispersioon	16
2.11 Eberharti ja Russeli regressioon	17
2.12 Paremuse määr	20
3 Mitteparameetrilised meetodid	21
3.1 Hühni statistikud	21
3.2 Thennarasu mitteparameetrilised statistikud	22
4 Nobalwheat projekti andmete statistiline analüüs	23
4.1 Andmete esmane analüüs	23
4.2 AMMI mudel	24
4.3 Zhangi D	30
4.4 ASV	30
4.5 GSI	30
4.6 WAASB ja WAASBY	30
4.7 Eberharti ja Russeli regressioon	33
4.8 Lin ja Binns paremuse määr	34
4.9 Ökovalents	38
4.10 Stabiilsuse dispersioon	40
4.11 Hühni mitteparameetrilised statistikud	40
4.12 Thennarasu mitteparameetrilised statistikud	40

5 Erinevad statistikud ja nende omavaheline võrdlus	42
5.1 Peakomponentanalüüs	42
5.2 Millist statistikut valida?	45
5.3 Statistike agregeerimine	45
Kokkuvõte	48
Kasutatud Kirjandus	50
Lisa 1	52
Lisa 2	53
Lisa 3	54
Lisa 4	55
Lisa 5	56

Sissejuhatus

Nisu moodustab umbes 50% kogu teraviljatoodangust Euroopas. Lähiaja kliimamuutusest tingituna on aga paljudes kohtades nisu saagikuse tõus seisma jäänud. Jätkusuutliku nisutootmise jaoks on oluline senisest parema kliimaresistentsusega nisusortide aretamine. Rahvusvahelise suvinisu projekti Nobalwheat – Breeding Toolbox for Sustainable Food System of the NOrdic BALtic Region eesmärgiks oli leida nisu genotüüpe, mis sobiksid lähtematerjaliks Balti riikide ja Norra aretusprogrammidesse senisest saagikamate ja erinevatele kasvutingimustele paremini kohastunud sortide aretamiseks. Projekti lõppeesmärgiks on saada Balti riikides ja Norras saagikate ja muutuvatele kliimatingimustele hästi kohastunud nisusortide valik, mis oluliselt tõstab toidu tootmise jätkusuutlikust nendes regioonides (“NOBALwheat – Breeding Toolbox for Sustainable Food System of the NOrdic BALtic Region | EEA Grants” (n.d.)).

Genotüübi mõju keskmisele saagikusele on lihtsasti uuritav. Genotüübi saab paljudele katselappidele ja erinevates keskkondades maha külvata ning keskmine üle kõigi katsete hindabki genotüübi keskmist saagikust. Keskkonna ja genotüübi koosmõju on aga raskemini hinnatav. Sordiaretuse lihtsustamiseks on kompleksete statistiliste seoste kokkuvõtmiseks ja interpreteerimiseks töötatud välja palju erinevaid statistilise analüüsi meetodeid.

Käesoleva töö eesmärgiks on seletada ja hinnata sordiaretuses kasutatavaid stabiilsust iseloomustavaid statistikuid, anda soovitusel nende rakendamiseks ja selgitada NOBALwheat projektis katsetatud nisu kollektsioonist välja saagikamad ja parema saagi stabiilsusega genotüübid uute sortide aretamiseks.

1 Andmed

Maaelu Teadmuskeskus (METK) osales aastatel 2021-2023 suvinisu projektis NOBALwheat. Projekti raames hinnati põldkatsetes Eestis, Lätis, Leedus ja Norras 300 genotüübi erinevaid omadusi (tabel 2). Katsed külvati kõigil aastatel vahemikus 19.aprill kuni 2.mai ning koristati augustis. Kõigis katsekohtades kasutati lämmastiku normi N120 kg/ha. Katsetes tehti keemilist umbrohu tõrjet ja vajadusel taimehaiguste tõrjet. Katse rajati blokkmeetodiga kahes korduses 5-7 m² lappidele. Genotüübid olid korduste piires randomiseeritud.

Iga riik lisas projekti varasemate katsetulemuste põhjal tema tingimustes kohastunud genotüübid, mida hinnati kõigis neljas riigis. Aretised on pärit doonorriikidest, katsesse võetud sordid on aretatud erinevates riikides st. on pärit ka teistest riikidest peale Baltimaade ja Norra.

Projektis osalesid järgmised organisatsioonid: Maaelu Teadmuskeskus (METK) Eestist, Leedu Põllumajandus- ja Metsandusteaduste Keskus (LAMMC) Leedust, Norra Maaülikool (NMBU) ja aretusfirma Graminor AS Norrast, Agroressursside ja -ökonoomika instituut (AREI). Tabelis 1 on väike väljavõtte töös uuritud genotüüpidest.

Kõigil genotüüpidel mõõdeti lisaks saagikusele veel 13 erinevat tunnust (tabel 2). Kuigi projekti peamine eesmärk on kõrge saagikusega genotüüpide selekteerimine, siis nisul on olulisel kohal ka näiteks gluteeni sisaldus, kuna kõrge gluteeni sisaldusega nisust tehtud sai tuleb kohev ja pehme. Antud uurimistöös kasutati ainult saagikuse kui kõige olulisema aretuseesmärgi andmeid. Puuduvaid andmeid esines vaid ühel juhul: Norra 2023-l aasta 69-nda genotüübi teise korduse saak. Kõik ülejäänud andmed oli täielikud.

Tabel 1: Väljavõtte töös uuritavatest genotüüpidest

Jrk	Töös kasutatud tähistus	Nimi	Doonor	Riik
146	NW146	Sonett	AREI	Läti
82	NW082	KWS Collada	METK	Eesti
246	NW246	DS-13-8-DH	LAMMC	Leedu
72	NW072	Daugana	METK	Eesti
59	NW059	GN11542	Graminor	Norra
33	NW033	Runar	NMBU	Norra
246	NW246	DS-13-8-DH	LAMMC	Leedu
216	NW216	Goplana	LAMMC	Leedu

Tabel 2: NOBALwheat projekti raames mõõdetud tunnuste nimekiri

Tunnus	Mõõtühik
Saak (normaliseeritud 14% niiskuse juurde)	g/m ²
Tärgamise kuupäev (70% taimedest on katselappil tärganud)	päev
Varajane kasvukiirus (taimede maapinna katvus, määratud võrsumise perioodi alguses)	%
Pea loomise kuupäev (70% taimedel on pea tupest väljas)	päev
Põuakindlus (skaala 1-5, kus 1 on põuakindel ja 5 mitte)	subjektiivne vaatlus
Küpsuse kuupäev (taimed on koristuseks valmis)	päev
Taime kõrgus/pikkus	cm
taimede seisukindlus (skaala 1-9, kus 1 on lamandunud ja 9 on täiesti seisukindel)	subjektiivne vaatlus
Vilja mahukaal (tihedus)	kg/l
1000 tera mass	g
proteiinisaldus	%
kleepvalk ehk gluteiini sisaldus	%
tärklise sisaldus	%
sedimentatsioon	ml

2 Parameetrised meetodid

Käesolevas töös on vaadatud ja kasutatud mitmeid erinevaid statistilisi meetodeid. Kuna alandmetest ei ole uurijale teada iga genotüübi geneetilised andmed (geneetiliste alleelide järjestus, genoomid jms.), ehk uurija teab andmeid sordi täpsusega, siis väga täpseid geneetika-põhiseid uurimismeetodeid või statistikuid ei saa siinkohal kasutada. Uurija teeb siinkohal eelduse, et kuna viljapõllul külvati piisavalt palju sama sordi (ka sordi siseselt võivad fenotüüpilised tunnused erineda) taimi, siis kõigi taimede peale kokku võetud keskmine on piisavalt hea näitaja sordi üldiliseks kirjeldamiseks.

Kõigi töös käsitletud koosmõju efekti mõõtvate statistikute nimekiri on ära toodud tabelis 3. Statistikud saab tähenduslikult jagada kaheks: statistikud, mis on mõeldud seletama ainult genotüübi ja keskkonna vahelise koosmõju efekti ehk puhtalt stabiilsust mõõtvad statistikud (näiteks AMMI stabiilsusväärtus, AMMI stabiilsusindeks), ja statistikud, mis mõõdavad nii stabiilsust, kui ka genotüübi keskmist võimekust (paremuse määr, ökovalents). Viimased on just välja kujunenud selle tööga sarnase eesmärgiga teistest uurimustöödest: eesmärgiga leida väga paljude viljasortide seast just kõige stabiilsem ja kõige võimekam sort, mis ühendab endas kõrget keskmist viljakust ja mis suure kindlusega sellest keskmisest madalamat saagikust ei anna.

Statistikuid on palju, ning nende omavaheliste seoste matemaatiline analüüs jääb selle töö mahust välja. On statistikuid, mis võiksid oma teoreetilise tausta poolest sarnased olla (näiteks kõik, mis AMMI meetodika põhinevad). Erinevate statistikute omavaheliste seoste tuvastamiseks kasutatakse Spearmani astakorrelatsiooni. Kõigi eelnevalt äratoodud statistikute järgi saab genotüübid panna järjekorda, kus esimesel kohal on kõige stabiilsem/saagikam/üldiselt parim genotüüp ja teised selle järgnevalt. Erinevate statistikute astakuid võrreldes on võimalik saada ülevaade sellest, kas erinevad statistikud mõõdavad genotüüpide stabiilsust sarnaselt.

Tabel 3: Töös kasutatavate statistiliste meetodite ülevaade

Meetodi nimi	Tähistus	Liik
Saagikuse keskmine	Y	Saagikus
Zhangi D-indeks	DZ	Stabiilsus
AMMI stabiilsusväärtus	ASV	Stabiilsus
Genotüübi valikuindeks	GSI	Üldine
Absoluutsete skooride kaalutud keskmine	WAASB	Stabiilsus
Absoluutsete skooride kaalutud keskmise üldine skoor	WAASBY	Üldine
Ökovalents	w	Stabiilsus
Stabiilsuse dispersioon	sigma2	Stabiilsus
Eberharti ja Russeli regressioon	b0,b1,s2di	Stabiilsus
Paremuse määr	P	Üldine
Hühni-i statistikud	S1,S2	Stabiilsus
Hühni-i statistikud	S3,S6	Üldine
Thennarasu statistikud	NP1,NP2	Stabiilsus
Thennarasu statistikud	NP3,NP4	Stabiilsus

2.1 Lineaarne regressioon

Olgu teoreetilises uuringus kokku p erinevat genotüüpi ja q erinevat keskkonda. Siis kõige tüüpilisem genotüüpi ja keskkonda sisaldav saagikust kujutav lineaarne regressioonimudel avaldub järgnevalt:

$$X_{ij} = \mu + g_i + e_j + (ge)_{ij} + \epsilon_{ij}, \quad (1)$$

kus

X_{ij} - i -nda genotüübi keskmine saagikus j -ndas keskkonnas,

μ - vabaliige, kõigi genotüüpide keskmine saagikus üle kõigi korduste ja kõigi keskkondade,

g_i - i -nda genotüübi mõju keskmisele saagikusele,

e_j - j -nda keskkonna mõju keskmisele saagikusele,

$(ge)_{ij}$ - i -nda genotüübi ja j -nda keskkonna koosmõju (GxE),

ϵ_{ij} - mõõtmata viga, eelduste järgi keskväärtusega 0 ja konstantse hajuvusega.

Juhul, kui andmestik on $n > 1$ kordust ühel genotüübi ja keskkonna kombinatsioonil (rohkem, kui üks kordusvaatlus), siis kasutatakse järgmist mudelit:

$$X_{ijk} = \mu + g_i + e_j + (re)_{ik} + (ge)_{ij} + \epsilon_{ijk}, \quad (2)$$

kus

X_{ijk} - i -nda genotüübi keskmine saagikus j -ndas keskkonnas k -ndal kordusel,

$(re)_{ik}$ - k-nda korduse ja i-nda keskkonna koosmõju,

Kõik teised tähistused on samad eelmise mudeliga.

Koosmõju kordajat $(re)_{ik}$ võib tõlgendada kui süstemaatilist erinevust katselapides, ehk mingis keskkonnas olid esimese korduse ja teise korduse katselapid erinevad mullastiku/päikese/tuule poolest. See koosmõju võiks olla 0 (ehk statistiliselt mitteoluline), sest vastasel juhul võib seada kahtluse alla, kas selles keskkonnas erinevatel kordustel saab ikka keskkonna samaks lugeda.

2.2 AMMI

AMMI (*Additive Main Effects and Multiplicative Interaction*) mudel on algoritm, mis üritab raskesti mõistetavat GxE efekti seletada peakomponentanalüüsiga. Valemis (1) rakendatakse koosmõjudele peakomponentanalüüsi, proovimaks võimalikult palju varieeruvust ära seletada võimalikult vähese arvu peakomponentidega.

Täpsemalt seletades: kõigepealt hinnatakse lineaarne mudel, kuhu on lisatud genotüübi ja keskkonna peamõjud ning genotüübi ja keskkonna koosmõjud. (nagu valemis (1)). Koosmõjude koefitsentidest $(ge)_{ij}$ koostatakse $p \times q$ mõõtmetega tabel. Koosmõjude tabelile rakendatakse peakomponentanalüüsi. Saadud peakomponentanalüüsi tulemustel hinnatakse iga genotüübi jaoks skoor, mis iseloomustab, kui suure osa vastav genotüüp moodustab koosmõjude efektist, ehk kui ebastabiilne vastav genotüüp on. AMMI mudel hinnatakse järgnevalt (Zobel, Wright, and Gauch 1988, lk.389):

$$X_{ij} = \mu + g_i + e_j + \sum_{n=1}^r \lambda_n u_{ni} v_{nj} + \rho_{ij} + \epsilon_{ijk}, \quad (3)$$

kus

r - peakomponentide arv, mis on otsustatud mudelisse jätta $1 \leq r \leq \min(p, q)$, kus p on erinevate genotüüpide arv ja q on erinevate keskkondade arv. Peakomponentide arvu otsustusreeglid on kirjeldatud (Annicchiarico (1997),lk.55),

ρ_{ij} - peakomponentide poolt seletamata jäänud dispersioon GxE-le,

$\sum_{n=1}^r \lambda_n u_{ni} v_{nj}$ - r-i esimese peakomponendi skooride summa vastavale genotüübi ja keskkonna kombinatsioonile. Siin λ_n on n-s omaväärtus, u_{ni} on i-nda genotüübi skoor n-ndal peakomponendil ja v_{nj} on j-nda keskkonna skoor n-ndal peakomponendil (N. N. A. P. Bose Lotan Kumar AND Jambhulkar (2014) lk.4).

Kõik ülejäänud tähistused on samad, nagu eelnevalt mainitud.

i-nda Genotüübi n-nda peakomponendi kujutamiseks graafikul kasutatakse suurust $\sqrt{\lambda_n} u_{ni}$ ja j-nda keskkonna n-nda peakomponendi skoori kujutamiseks graafikutel kasutatakse $\sqrt{\lambda_n} v_{nj}$. Mõlema korrutis kokku annab $\lambda_n u_{ni} v_{nj}$, ehk n-nda peakomponendiskoori (N. N. A. P. Bose Lotan Kumar AND Jambhulkar (2014) lk.5).

Juhul, kui otsustatakse jätta mudelisse kõik peakomponendid, ehk $r = \min(p, q)$, siis $\rho_{ij} = 0$ iga i ja j korral ning AMMI mudel taandub tavalisele lineaarsele regressioonile, kus on kõik koosmõjud lisatud. Sel juhul on AMMI mudeli ennustused ja jäägid võrdsed lineaarse regressioonimudeliga.

2.3 Zhangi D indeks

Zhang'i D-indeks (D_i) baseerub eelmises punktis kirjeldatud peakomponentanalüüsil ning see mõõdab genotüüpide kaugust nullpunktist uues koordinaatsüsteemis. Kaugused arvutatakse tavaliste eukleediliste kaugustena, kusjuures aluseks võetakse kõik peakomponendid, mis on statistiliselt olulised: (L. K. Bose et al. (2014)).

$$D_i = \sqrt{\sum_{n=1}^r u_{ni}^2}, \quad (4)$$

kus kõik tähistused on sarnased, nagu valemis (3).

Mida suurem väärtus on indeksil, ehk mida kaugemal on genotüüp uues koordinaatsüsteemis, seda suurem on tema koosmõju efekt, ehk seda ebastabiilsemaks loetakse genotüüp.

2.4 AMMI stabiilsusväärtus (ASV)

AMMI stabiilsusväärtus (ASV-AMMI stability value) on sarnane Zhang'i D-indeksiga (valem (4)), kus kasutatakse vaid kaht esimest peakomponenti, ning tehakse üks lisaarvestus. AMMI stabiilsusväärtus arvutatakse järgnevalt (Purchase, Hatting, and Deventer (2000) , lk.102):

$$ASV_i = \sqrt{\left[\frac{\lambda_1^2}{\lambda_2^2} u_{1i}\right]^2 + u_{2i}^2}, \quad (5)$$

kus

ASV_i - i -nda genotüübi ASV väärtus,

λ_1 ja λ_2 - vastavalt esimese ja teise peakomponendi omaväärtused.

Muud tähistused saranevad eelmiste valemitega.

Esimene liidetav avaldises on korrutatud kaaluga, mis on mõeldud tasakaalustama fakti, et esimene peakomponent tegelikult seletab rohkem varieeruvust (GxE efekti), kui teine komponent (Purchase, Hatting, and Deventer (2000) lk.102). Statistiku puhul tähendab madal väärtus madalat GxE efekti, ehk genotüüp on stabiilne üldiselt kõigis keskkondades (L. K. Bose et al. 2014).

2.5 Genotüübi valikuindeks (GSI)

Eelnevalt ära toodud ASV ja Zhang'i D statistikute probleemiks on see, et need mõõdavad ainult genotüübi stabiilsust, kusjuures stabiilsus on mõõdetud suhtelisena teiste uuringus osalenud genotüüpide vastu. Genotüüp võib stabiilne olla, kuid kui tema keskmine saagikus on tunduvalt madalam teise ebastabiilsema genotüübi omast, siis isegi halvimas olukorras võib ebastabiilne genotüüp olla tulemuslikum, kui stabiilne. Et võrrelda genotüüpe üksteise vahel üleüldiselt, tuleks ka arvesse võtta nende keskmine saagikus.

Genotüübi valikusindeks on mitteparameetriline statistik, mis mõõdab nii genotüübi stabiilsust, kui ka üldist näitajat. Indeks arvutatakse järgmiselt (Bocianowski, Niemann, and Nowosad 2018 ,lk.7) :

$$GSI_i = RM_i + RA_i,$$

kus

GSI_i - i -nda genotüübi valikuindeks,

RM_i - i -nda genotüübi tunnuse keskmise astak,

RA_i - i -nda genotüübi tunnuse ASV indeksi astak.

Kõikidel uuringus osalenud genotüüpidel mõõdetakse tunnuse keskmine ja arvutatakse ASV eelnevalt toodud meetodi põhjal. Seejärel järjestatakse kõik genotüübid stabiilsuse kahanevas järjekorras ja tunnusel endal kas kasvavas või kahanevas järjekorras (näiteks saagikusel kahanevas aga glükoosisaldusel kasvavas järjekorras, olenevalt sellest, milline on soovitud tunnuse muutumise suund). Madal statistiku väärtus näitab võrdlemisi stabiilset ja/või hea keskmisega genotüüpi, samas kõrge statistiku väärtus näitab madalat keskmist ja/või ebastabiilset genotüüpi. (Bocianowski, Niemann, and Nowosad (2018),lk.7)

Tuues näite selle töö saagikuse uurimisel, siis kõige kõrgema keskmise saagikusega genotüüp 293 saaks oma saagikuse astakuks $RM_{293} = 1$. Samas, kui genotüübi 15 ASV väärtus on kõigist genotüüpidest kõige suurem, siis $RA_{15} = 300$ (kõrge ASV väärtus tähendab ebastabiilset genotüüpi.)

Kui tahetakse mitme erineva tunnuse peale korruga genotüüpe võrrelda ja valida välja genotüüpe, mis kõigis tunnustes sobivaimad on, võib välja arvutada üldise genotüübi valikuindeksi, mis kujutab endast lihtsalt iga tunnuse peal välja arvatud valikuindeksite summeerimist (Bocianowski, Niemann, and Nowosad 2018 ,lk.7) :

$$TGS_i = RM_{i1} + RA_{i1} + \dots + RM_{in} + RA_{in}, \quad (6)$$

kus

TGS_i - i -nda genotüübi üldine valikuindeks,

RM_{in} - i -nda genotüübi n -nda tunnuse keskmise astak,

RA_{in} - i -nda genotüübi n -nda tunnuse ASV indeksi astak.

Sarnaselt eelnevale tähendab madal indeksi väärtus seda, et vastav genotüüp on kõrgel kohal stabiilsuse ja keskmise poolest kõigis uuritavates tunnustes. Iga tunnuse jaoks on eraldi vaja läbi viia AMMI analüüs ja keskmiste astakute järjestus.

Lisaks on võimalik lisada kordajad iga tunnuse astakule, et rõhutada ühtesid tunnuseid või vähendada teiste tunnuste mõju. Üldist valimisindeksit saaks siis arvutada järgnevalt:

$$TGS I_i = a_1 RM_{i1} + a_1 RA_{i1} + \dots + a_n RM_{in} + a_n RA_{in},$$

kus

a_n - vabalt valitud kordaja, mis näitab n-nda tunnuse tähtsust uuringus.

Kõik teised tähistused on samad, nagu eelnevalt.

Individaalsete tunnuste astakute korrutamisel mingisuguse kordajaga, saab esile tuua spetsiifiliste tunnuste olulisust või vähendada teiste vähemoluliste, kuid siiski huvi pakkuvate tunnuste mõju üleüldisele valimisindeksile. Kui veel lisaks nõuda, et $\sum_{i=1}^n a_i = 1$, siis saab ka võrrelda genotüüpide valimisindeksi käitumist olenevalt kaalude muutmisest, kusjuures $a_i = 0$ korral saab i-nda tunnuse uurimisest ka välja jätta (sarnaselt WAASBY indeksile, mida tutvustatakse järgnevalt). Eesmärgiks on kokkuvõttes ühese statistiku pakkumine, millega saab võrrelda erinevaid genotüüpe sõltuvalt uuringute eesmärkidest.

2.6 BLUP

Kogu järgnev statistiline meetod on kirjutatud T.Olivoto artikli põhjal (Olivoto et al. 2019 , lk.2950).

Teistsugune lähenemine on genotüübi efekte kujutada mitte fikseerituna vaid juhuslikuna. Et seda illustreerida, võime võtta tavalise lineaarse regressiooni mudeli (1), ja oletada, et genotüübi efekt on juhuslik. Siis saaks mudeli kirjutada tavalise lineaarse segamudelina:

$$X = Yb + Zu + \epsilon,$$

kus

X - $[n = (\sum_{j=1}^q (pk))] \times 1$ - Vaadeldava tunnuse vektor (saak) $X = [x_{111}, x_{112}, \dots, x_{pqb}]^t$,

b - $(qk) \times 1$ vektor tundmatuid fikseeritud efekte (keskkonna peamõju ja keskkond:kordus koosmõjud) $b = [\mu + e_{11}, e_{12}, \dots, e_{qb}]^t$,

u - $(m = p + pq) \times 1$ vektor tundmatuid juhuslikke efekte (genotüübi peamõju ja genotüüp:keskkond koosmõjud) $u = [g_1, g_2, \dots, g_p, (ge)_{11}, \dots, (ge)_{pq}]$,

Y - $n \times (qk)$ nullidest ja ühtedest koosnev disainimaatriks, mis seob vektori X vektoriga b ,

Z - $n \times m$ nullidest ja ühtedest koosnev disainimaatriks, mis seob vektori X vektoriga u ,

ϵ - $n \times 1$ vektor juhuslikke vigasid $\epsilon = [\epsilon_{111}, \dots, \epsilon_{pqb}]^t$.

Vektorid u ja ϵ on eelduste kohaselt normaalfaotusega keskväärusega 0 ja kovariatsiooni maatriksitega G ja R nii, et

$$\begin{bmatrix} u \\ \epsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} \right),$$

kus

$$G = \begin{bmatrix} \hat{\sigma}_g^2 I_g & 0 \\ 0 & \hat{\sigma}_{ge}^2 I_{ge} \end{bmatrix}$$

ja $R = \hat{\sigma}_e^2 I_n$, kus $\hat{\sigma}_g^2, \hat{\sigma}_{ge}^2, \hat{\sigma}_e^2$ on vastavalt genotüübi, genotüübi-keskkonna koosmõju ja juhuslikku vea dispersioon. I_g, I_{ge}, I_n on ühikmaatriksid vastavate dimensioonidega. Vektorid b ja u hinnatakse segamudeli võrrandist.

$$\begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} Y^t R^{-1} Y & Y^t R^{-1} Z \\ Z^t R^{-1} Y & Z^t R^{-1} Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} Y^t R^{-1} x \\ Z^t R^{-1} x \end{bmatrix},$$

kus tähistused $^{-1}$ ja $^{-}$ näitavad pöördmaatriksit ja üldistatud pöördmaatriksit. Dispersiooni komponendid on tavaliselt tundmatud ja nende asemel kasutatakse hinnanguid (\hat{G} ja \hat{R}), mis saadakse EM-algoritmi (*Expectation-Maximization*) kasutades. Nende algoritmide täpsem kirjeldamine jääb sellest tööst välja, kuid eeldades tasakaalustatud andmeid (kõiki korduseid on võrdselt), siis saab i -nda genotüübi efekti (\hat{g}_i) vektoris u välja kirjutada ANOVA tähistuste järgi järgmiselt:

$$\hat{g}_i = h_g^2(\bar{x}_i - \bar{x}_{..}),$$

kus $h_g^2 = (\hat{\sigma}_{ge}^2 + q\hat{\sigma}_g^2)/(\hat{\sigma}_{ge}^2 + \hat{\sigma}_e^2 + q\hat{\sigma}_g^2)$. i -nda genotüübi BLUP hinnatakse järgmiselt: $BLUP_i = \mu + \hat{g}_i$. i -nda genotüübi ja j -nda keskkonna koosmõju (\hat{g}_{ij}) vektoris u hinnatakse järgmiselt:

$$\hat{g}_{ij} = h_g^2(\bar{x}_i - \bar{x}_{..}) + h_{ge}^2(x_{ij} - \bar{x}_i - \bar{x}_{.j} + \bar{x}_{..}),$$

kus h_g^2 on defineeritud eelnevalt ja $h_{ge}^2 = \hat{\sigma}_{ge}^2/(\hat{\sigma}_{ge}^2 + \hat{\sigma}_e^2)$. BLUP i -ndale genotüübile j -ndas keskkonnas on siis $BLUP_{ij} = \bar{x}_{.j} + \hat{g}_{ij}$.

2.7 WAASB

Kogu järgnev statistiline meetod on kirjeldatud T.Olivoto poolt (Olivoto et al. 2019, lk.2952).

Juhul, kui AMMI mudelis on esimese peakomponendi poolt seletatud varieeruvus väike, siis võib AMMI hinnata stabiilsust ebatäpselt, kuna vähese arvu peakomponentide kasutamisel jääb suur osa varieeruvusest välja jäetud peakomponentidesse. Selle parandamiseks pakutakse välja WAASB (*Weighted Average of Absolute Scores from singular value decomposition of BLUPs for the GEI effects generated by an LMM*) indeks. WAASB-i arvutamisel kasutatakse kõiki peakomponente ehk see kasutab rohkem informatsiooni, kui AMMI mudel, seega seletab WAASB rohkem genotüübi ja keskkonna koosmõju, kui AMMI mudel. i -nda genotüübi WAASB indeks arvutatakse välja järgmiselt:

$$WAASB_i = \sum_{n=1}^r |u_{in} * \lambda_n^2| / \sum_{n=1}^r \lambda_n^2,$$

kus

$WAASB_i$ - i-nda genotüübi absoluutsete skooride kaalutud keskmine,

λ_n^2 - n-nda peakomponendi poolt seletatud koosmõju dispersioon (omaväärtuse ruut),

u_{in} - i-nda genotüübi skoor n-ndas peakomponendis.

Üldiselt näeb valem sarnane välja AMMI stabiilsusväärtusega (valem (5)), kus GxE efekti proovitakse seletada vastava genotüübi skooridega ning skoorid on korrutatud läbi suurusega, mis iseloomustab, kui palju vastav peakomponent kogu varieeruvusest seletab. Erinevus seisneb siin selles, et WAASB kasutab alati kõiki peakomponente, mitte ainult kahte või statistiliselt olulisi komponente. “Kauguse” arvutamine on ka veidi modifitseeritud. Ruutjuure võtmise ruutude summast on asendatud absoluutväärtuste summaga. Lisaks on tavalise lineaarse mudeli asemel kasutatud segamudeleid, mis on üldistus lineaarsetest mudelitest ning mis lubab fikseeritud efekte asendada juhuslike efektidega. Genotüübid järjestatakse WAASB-i väärtuste järgi kasvavas järjekorras, kus madal väärtus näitab stabiilset genotüüpi. Graafiliselt kujutatakse ühel teljel uuritava tunnuse väärtusi ning teisel teljel WAASB väärtuseid. Sellise kujutamise eelis on see, et ühel graafikul saab uurida nii stabiilsust kõikidelt peakomponentidelt ja ka keskmist tulemust.

2.8 WAASBY

Et valida genotüüpe, mis on nii stabiilsed, kui ka kõrge keskmise tulemusega, tuleks kasutada ära informatsiooni nii keskmise kohta kui stabiilsuse kohta. Selle jaoks saame luua WAASBY indeksi. Indeksi põhimõte on väga sarnane eelnevalt toodud GSI indeksile. Kõigepealt skaleerime tunnuse ja WAASB-i indeksi samale skaalale, näiteks skaalale 0-st 100-ni, et nad oleks võrreldavad. Järgmised teisendused tehakse WAASB skooridele ja tunnuse väärtustele (Olivoto et al. 2019 , lk.2952):

$$rG_i = \frac{100 - 0}{G_{max} - G_{min}} \times (G_i - G_{max}) + 100$$

ja

$$rW_i = \frac{0 - 100}{W_{max} - W_{min}} \times (W_i - W_{max}) + 0,$$

kus rG_i ja rW_i on teisendatud väärtused tunnusele ja WAASB-i skoorile i-ndal genotüübil, ja G_i ja W_i on tunnus ja WAASB väärtus i-ndale genotüübile.

G_{max} , G_{min} on kogu andmestiku peale suurim ja vähim mõõdetud tunnuse väärtus ning W_{min} , W_{max} on samad näitajad WAASB indeksile. Juhul, kui võtta $G_i = G_{max}$, siis näeme, et $rG_i = 100$, ehk kõrge tunnuse väärtus (näiteks kõrge saagikus) tähendab kõrget rG_i väärtust. $G_i = G_{min}$ korral on avaldise väärtus 0, ehk madal rG_i väärtus

tähendab väikest tunnust (saagikust). Kui $W_i = W_{min}$, ehk genotüüp on stabiilne, siis $rW_i = 100$ ning vastupidi $rW_i = 0$, kui on ebastabiilne genotüüp (W_i on kõrge). Need kombineeritakse kokku üheks WAASBY-i indeksiks järgnevalt :

$$WAASBY_i = \frac{(rG_i \times \theta_Y) + (rW_i \times \theta_S)}{\theta_Y + \theta_S},$$

kus

$WAASBY_i$ - i-nda genotüübi WAASBY indeks ,

θ_y, θ_S - vastavalt tunnuse ja stabiilsuse kaalud.

Sarnaselt, nagu TGSi (valem (6)) puhul, antakse siin sõltuvalt uuringu eesmärgist ja vajadustest stabiilsusele ja tunnusele erinevad kaalud. Tavaliselt valitakse kaalud nii, et nad summeeruks kokku 100-ks, mis lihtsustab kaalude interpreteerimist. Kaalud määravad, kui oluline on genotüübi stabiilsus või genotüübi üldine saagikus. Üleüldiselt tähendab kõrge WAASBY väärtus stabiilset ja/või kõrge tunnuse väärtusega genotüüpi ning madala WAASBY väärtusega vastupidist (Olivoto et al. (2019), lk.2952).

2.9 Ökovalents

Ökovalents on G.Wricke (Becker and Léon 1988, lk.3) poolt välja pakutud stabiilsust kirjeldav statistik. See mõõdab, kui palju iga spetsiifilise genotüübi GxE efekt ruutu võetud ja summeeritud üle kõigi keskkondade moodustab kogu GxE efektist. Statistikut arvutatakse järgnevalt:

$$w_i = \sum_{j=1}^q (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2, \quad (7)$$

kus

w_i - i-nda genotüübi ekovalents väärtus,

q - keskkondade arv,

X_{ij} - i-nda genotüübi keskmine j-ndas keskkonnas,

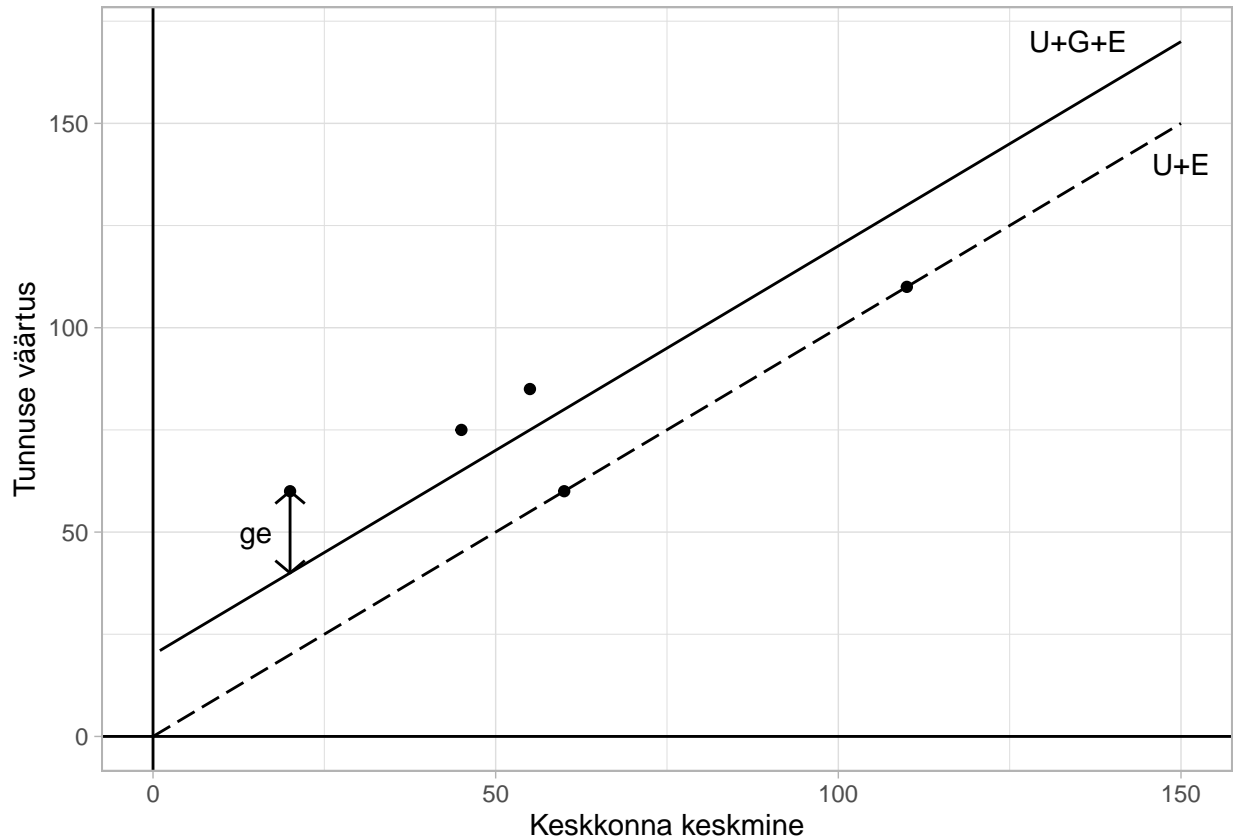
$\bar{X}_{i.}$ - i-nda genotüübi üldine keskmine,

$\bar{X}_{.j}$ - j-nda keskkonna üldine keskmine,

$\bar{X}_{..}$ - üldine keskmine tunnusel.

Mida väiksem väärtus ökovalentsil on, seda stabiilsem on genotüüp. Ökovalents mõõdab ainult stabiilsust, ning on mõeldud testima genotüüpide üldist adapteerimisvõimet. Genotüüpide vahelisi tunnuste keskmiste erinevusi ei saa selle statistikuga võrrelda.

Joonis 1 seletab ökovalentsi tähendust. Y-telg kujutab tunnuse väärtuse prognoosi ja X-telg kujutab j-nda keskkonna keskmist. Kriipsjoon tähistab tunnuse väärtuse prognoosi keskkonnas üle kõigi genotüüpide. Pidev joon kujutab ühe genotüübi prognoosi sõltuvalt keskkonna keskmisest. Punktid tähistavad tegelikke andmeid. Kui genotüüp oleks stabiilne (kui genotüübi ja keskkonna vahelist koosmõju ei eksisteeri), siis oleksid kõik punktid pideva joone peal. Kõik erinevused keskmisest loetakse koosmõju efektiks ning mida suuremad erinevused on ühel genotüübil, seda ebastabiilsemaks seda loetakse. Koosmõju efektiks loetakse kõigi erinevuste summat üle kõigi genotüüpide. Ökovalentsiks on kauguste ruutude summa.



Joonis 1: Keskkonna ja genotüübi koosmõju kujutamine (Graafik tehtud Becker (1988) näitel)

2.10 Stabiilsuse dispersioon

Stabiilsuse dispersioon on Shukla (1972) poolt välja pakutud stabiilsust kirjeldav statistik (Shukla (1972) lk 240).

$$\hat{\delta}_i^2 = \frac{1}{(q-1)(p-1)(p-2)} \left[p(p-1) \sum_j^q (\bar{X}_{ij} - \bar{X}_i - \bar{X}_{.j} + \bar{X}_{..})^2 - \sum_i^p \sum_j^q (X_{ij} - \bar{X}_i - \bar{X}_{.j} + \bar{X}_{..})^2 \right], \quad (8)$$

kus

$\hat{\delta}_i^2$ - i-nda genotüübi stabiilsuse dispersioon,

p - genotüüpide arv,

q - keskkondade arv.

Ülejäänud tähistused on samad valemiga (7).

Stabiilsuse dispersioon, ning ka selle valem, on väga sarnane ja tihedalt seotud ökovalentsi valemiga. Kui vaadata valemit (8), siis esimene summa on ökovalentsi valem i-ndale genotüübile, ehk i-nda genotüübi panus GxE efekti. Teine summa on kõigi genotüüpide GxE panuste summa, ehk kokku GxE ruutude summa. Kõik ülejäänud liikmed peale individuaalse GxE efektide summa genotüübile i on võrdsed kõigil genotüüpidel. Selle tulemusel on stabiilsuse dispersiooni statistiku järgi genotüüpide järjestus alati sama, mis ökovalentsi järgi järjestus. Seega erinevus, mis stabiilsuse dispersiooni ja ökovalentsi vahel eksisteerib, on statistiku interpretatsioon (Becker and Léon (1988)).

2.11 Eberharti ja Russeli regressioon

Eberhart pakkus välja järgmise lähenemise regressioonimudelite poolt kujutamaks nii genotüüpide efekti, kui ka genotüüpide stabiilsust keskkondades (Eberhart and Russell (1966) lk.37):

$$X_{ij} = b_{i0} + b_{i1}I_j + \delta_{ij},$$

kus

X_{ij} - i-nda genotüübi keskmine j-ndas keskkonnas,

b_{i0} - i-nda genotüübi keskmine üle kõigi keskkondade,

I_j - j-nda keskkonna indeks (ehk j-nda keskkonna keskmine mõju),

b_{i1} - regressioonikordaja, mis mõõdab i-nda genotüübi vastuvõtlikkust keskkonnast tulenevatele muutustele,

δ_{ij} - i-nda genotüübi ja j-nda keskkonna hinnangu kaugus regressioonisirgest (ehk mudeldamatta dispersioon).

Genotüüpi i kirjeldavad sellise regressiooni puhul regressioonikordaja b_{i1} ja hälvete ruutude keskmine s_{di}^2 . Need arvutatakse välja järgnevalt (Becker and Léon 1988 , lk.5):

$$b_{i1} = 1 + \frac{\sum_j (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})(\bar{X}_{.j} - \bar{X}_{..})}{\sum_j (\bar{X}_{.j} - \bar{X}_{..})^2},$$

ja

$$s_{di}^2 = \frac{1}{q-2} \left[\sum_j (X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X}_{..})^2 - (b_{i1} - 1)^2 \sum_j (\bar{X}_{.j} - \bar{X}_{..})^2 \right],$$

kus

X_{ij} - i-nda genotüübi keskmine j-ndas keskkonnas,

\bar{X}_i - i-nda genotüübi keskmine,

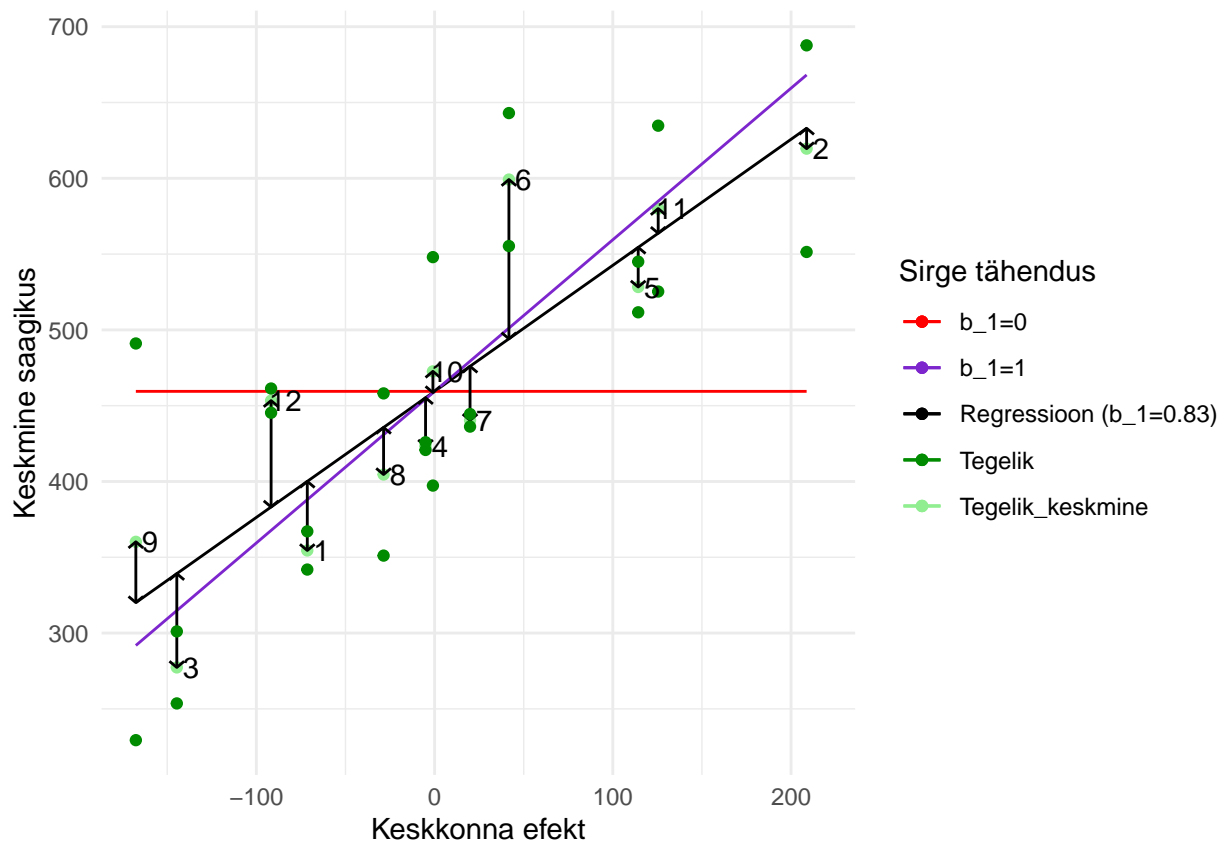
$\bar{X}_{.j}$ - j-nda keskkonna keskmine,

$\bar{X}_{..}$ - üldine keskmine,

q - keskkondade arv.

Mõlemat statistikut kasutatakse genotüüpide stabiilsuse hindamiseks, olenevalt sellest, mida stabiilsuse all mõistetakse. Kuigi s_{di}^2 on väga tugevalt seotud regressiooni jääkliikmega i-ndale genotüübile (ehk dispersioon, mida regressioon ei seleta), kujutab b_{i1} genotüübi tundlikust erinevate keskkonna tingimuste suhtes. Mida lähemal on b_{i1} väärtus 0-le, seda vähem mõjutab keskkondade erinevus genotüübi tunnust. Parameetri väärtuse $b_{i1} = 1$ korral on oodata genotüübi saagikuselt täpselt samasugust efekti, nagu seda on keskkonna efekt. Kui $b_{i1} > 1$, on genotüüp adapteerunud spetsiifilistesse keskkondadesse. Positiivsete kordajatega keskkondades on oodata väga suurt tõusu, aga samas ka negatiivsete kordajatega keskkondades on oodata väga suurt langust saagikusel. s_{di}^2 näitab, kui palju erinevad keskmised väärtused regressiooni sirgel olevatest väärtustest, ehk see näitab, kui palju varieeruvust jääb mudeli poolt seletamata.

Joonis 2 seletab Eberharti ja Russeli regressiooni. Tumerohelised punktid tähistavad esimese genotüübi mõõdetud andmepunkte, helerohelised punktid tähistavad keskkondade keskmiseid, ning jooned erinevaid sirgeid olenevalt regressiooni kordajatest. Must sirge näitab empiirilist (ehk vähima dispersiooniga) regressioonisirget, punane sirge tähistab regressiooni sirget, kui $b_1 = 1$ ja sinine tähistab regressiooni, kui $b_1 = 0$. Eberhart ja Russeli regressioon proovib sobitada sirge olemasolevatele punktidele nii, et punktide kauguste ruutude summa sirgest oleks võimalikult väike. Sirge on defineeritud kahe parameetriga: vabaliige ja tõus. Vabaliikmeks kutsutakse b_0 ehk sirge väärtus punktis $x = 0$. Tõusuks nimetatakse b_1 väärtust ehk kui palju mõjutab keskkonna erisus selle genotüübi saagikust, ehk kui vastuvõtlik genotüüp keskkonnale on. Igale genotüübile mõõdetakse vabaliige ja tõus, mis tähendab, et kokku määratakse 2 korda niipalju tundmatuid, kui palju on genotüüpe. Keskkonna indeksid, mis tähistavad x-telge, on mõõdetud puhtalt andmete pealt, seega neid ei loeta tundmatute alla.



Joonis 2: Eberharti ja Russeli regressioon genotüübi 1 näitel.

Sarnaselt hälbe keskmisele kasutatakse ka määramise koefitsenti (r_i^2), mille pakkus välja Pinthus (1973) (Becker and Léon (1988) lk.5):

$$r_i^2 = 1 - \frac{s_{di}^2}{s_{xi}^2},$$

kus

$$s_{xi}^2 = \frac{\sum_j (X_{ij} - \bar{X}_i)^2}{q-1}$$

ja s_{di}^2 on eelnevalt defineeritud.

Määramise koefitsenti kasutatakse, kuna hälvete ruutude keskmine sõltub mõõtühikust. Määramise koefitsenti kasutamisel saadakse lahti sellest, et erinevaid mõõtühikuid kasutades saab erinevas suuruses statistiku väärtuseid, ehk standardiseeritakse s_{di}^2 väärtust (Becker and Léon (1988) lk.5).

2.12 Paremuse määr

Paremuse määr (*Superiority measure*) on Lin & Binns (1988) poolt loodud statistik, mis proovib ühe parameetriga hinnata nii genotüübi stabiilsust, kui ka saagikust. Olgu X_{ij} tunnuse väärtus i -nda genotüübi ja j -nda keskkonna korral (või selle keskmine üle korduste) ning olgu M_j maksimaalne tunnuse väärtus keskkonnas j . Siis P_i on i -nda genotüübi keskmine kaugus maksimaalsest tunnuse väärtusest igas keskkonnas (Lin and Binns (1988) lk.194).

$$P_i = \sum_{j=1}^q (X_{ij} - M_j)^2 / (2q),$$

kus q on keskkondade arv. Eelneva valemi saab lahti kirjutada järgmiselt:

$$P_i = [n(\bar{X}_i - \bar{M})^2 + \sum_{j=1}^n (X_{ij} - \bar{X}_i - M_j + \bar{M})^2] / (2n),$$

kus $\bar{X}_i = \sum_{j=1}^n X_{ij} / n$ ja $\bar{M} = \sum_{j=1}^n M_j / n$

Kui me eeldame, et M_j kujutab endast hüpoteetilist genotüüpi, siis esimene osa valemist on genotüübi efekti erinevuste ruutude summa ja teine osa valemist on GxE efekti erinevuste ruutude summa. Seega P_i võiks kujutada endas genotüübi ja GxE efekti summat. M_j -na kujutatakse hüpoteetilist kultuuri: see on kultuur, mis saavutab maksimaalse väärtuse igas keskkonnas ning P_i väärtus on kaugus sellisest hüpoteetilisest kultuurist. Kui P_i väärtus on väike, siis see kultuur on lähedane hüpoteetilisele kultuurile, mis on parim kõigis keskkondades (Lin and Binns (1988) lk.194).

Üks eelis paremuse määra juures on, et selle välja arvutamiseks ei pea igas keskkonnas iga genotüübiga katsed tegema. Kauguse maksimaalsest saab võtta nendes keskkondades, kus selle genotüübiga mõõtmiseid tehtud on ja keskmist kaugust saab arvutada ka vähemate keskkondade pealt. Rohkemates keskkondades katsete tegemisega saab siiski täpsemaid väärtuseid ja rohkem infot genotüüpide kohta.

Paremuse määr on ühte pidi võrdlus kahe genotüübi vahel. Üks nendest on hüpoteetiline genotüüp, mida ei eksisteeri, kuid siiski on võrdlus kahe genotüübi vahel. Kui kaugus ei ole "liiga suur", siis loetakse genotüübid statistiliselt mitte-erinevateks, ehk genotüüp on tõesti piisavalt sarnane sellise hüpoteetilise genotüübiga, mis on parim kõigis keskkondades (Lin and Binns (1988) lk.195).

3 Mitteparameetrilised meetodid

Parameetriliste stabiilsusstatistikute kasutamiseks tuleb tavaliselt teha eeldusi vigade jaotuste ja sõltumatuse kohta. Mitte-parameetriliste statistikutega ei ole seda probleemi, ning mitte-parameetrilisi statistikke ei mõjuta ka erindid niipalju, kui need mõjutavad parameetrilisi statistikke (Becker and Léon (1988), lk.7). Lisaks saab mitte-parameetrilisi statistikke kasutada väga lihtsalt ka selliste suuruste võrdlemisel, mis ei ole ilmtingimata samades ühikutes. Näiteks on võimalik saagikust võrrelda taime pikkusega, ja uurida, kas saagikad taimed on samal ajal ka pikad, kuigi üht mõõdetakse kilogrammides ja teist meetrites.

3.1 Hühni-i statistikumid

Olgu meil q erinevat keskkonda ja p erinevat genotüüpi. Keskkonniti on meil võimalik järjestada genotüüpide tulemused (saagikused) kasvavas järjekorras nii, et kõige väiksema saagikusega genotüüp on kohal 1, järgmine genotüüp saab koha 2 ja nii edasi. Tähistagu r_{ij} i -nda genotüübi astakut j -ndas keskkonnas. Kasutades eelnevalt määratletud astakuid pakkus Hühni välja järgnevad statistikumid stabiilsuse mõõtmiseks (Hühni and Nassar (1989)):

$$S_i^{(1)} = \frac{2 \sum_{j=1}^{q-1} \sum_{k=j+1}^q |r_{ij} - r_{ik}|}{q(q-1)},$$

$$S_i^{(2)} = \sum_{j=1}^q \frac{(r_{ij} - \bar{r}_{i.})^2}{q-1},$$

kus $\bar{r}_{i.} = \sum_{j=1}^q \frac{r_{ij}}{q}$

Eelnevate statistikute arvutusvalemid põhinevad eeldusel, et stabiilse genotüübi astak ei tohiks keskkonniti suuresti varieeruda. Mõlema statistiku puhul tähendab suur statistiku väärtus ebastabiilset genotüüpi.

Hühni pakkus välja ka järgnevad statistikumid, mis kasutavad astakuid, kuid mis on lisaks stabiilsusele mõeldud mõõtma ka keskmist saagikust (Sabaghnia, Dehghani, and Sabaghpour (2006), lk.1102):

$$S_i^{(3)} = \frac{\sum_j (r_{ij} - \bar{r}_{i.})^2}{\bar{r}_{i.}},$$

$$S_i^{(6)} = \frac{\sum_j |r_{ij} - \bar{r}_{i.}|}{\bar{r}_{i.}}.$$

Põhimõtte sarnaneb puhtalt stabiilsust mõõtvate statistikutega, kuid lugeja on asendatud genotüübi keskmisega üle kõigi keskkondade. Kui genotüüp on kõrge keskmise saagikusega, siis on ka keskmine astak üle keskkondade kõrge, ehk läbi jagatakse suure arvuga, mis vähendab statistiku väärtust.

3.2 Thennarasu mitteparameetrilised statistikud

Tennarasu (1995) stabiilsus-statistikud kasutavad keskmistatud genotüüpide astakuid igas keskkonnas. Olgu x_{ij} i -nda genotüübi (keskmine) tulemus j -ndas keskkonnas ja \bar{x}_i i -nda genotüübi keskmine tulemus üle kõigi keskkondade. Siis $x_{ij}^* = x_{ij} - \bar{x}_i$ on i -nda genotüübi keskmistatud tulemus j -ndas keskkonnas. Pärast keskmistamist jääb vaatlusest järgi vaid GxE efekt ja juhuslik viga, seega astakud keskmistatud väärtuste järgi mõõdavad ainult stabiilsust ja mitte saagikust. Kasutades keskmistatud astakuid, pakkus Tennarasu välja järgmised statistikud (Sabaghnia, Dehghani, and Sabaghpour (2006) lk.1102):

$$NP_i^{(1)} = \frac{1}{q} \sum_{j=1}^q |r_{ij}^* - M_i^*|$$

$$NP_i^{(2)} = \frac{1}{q} \left(\sum_{j=1}^q |r_{ij}^* - M_i^*| / M_i \right)$$

$$NP_i^{(3)} = \frac{\sqrt{\sum (r_{ij}^* - \bar{r}_i)^2 / q}}{\bar{r}_i}$$

$$NP_i^{(4)} = \frac{2}{q(q-1)} \left[\sum_{j=1}^{q-1} \sum_{j^*=j+1}^q |r_{ij}^* - r_{ij^*}^*| / \bar{r}_i \right]$$

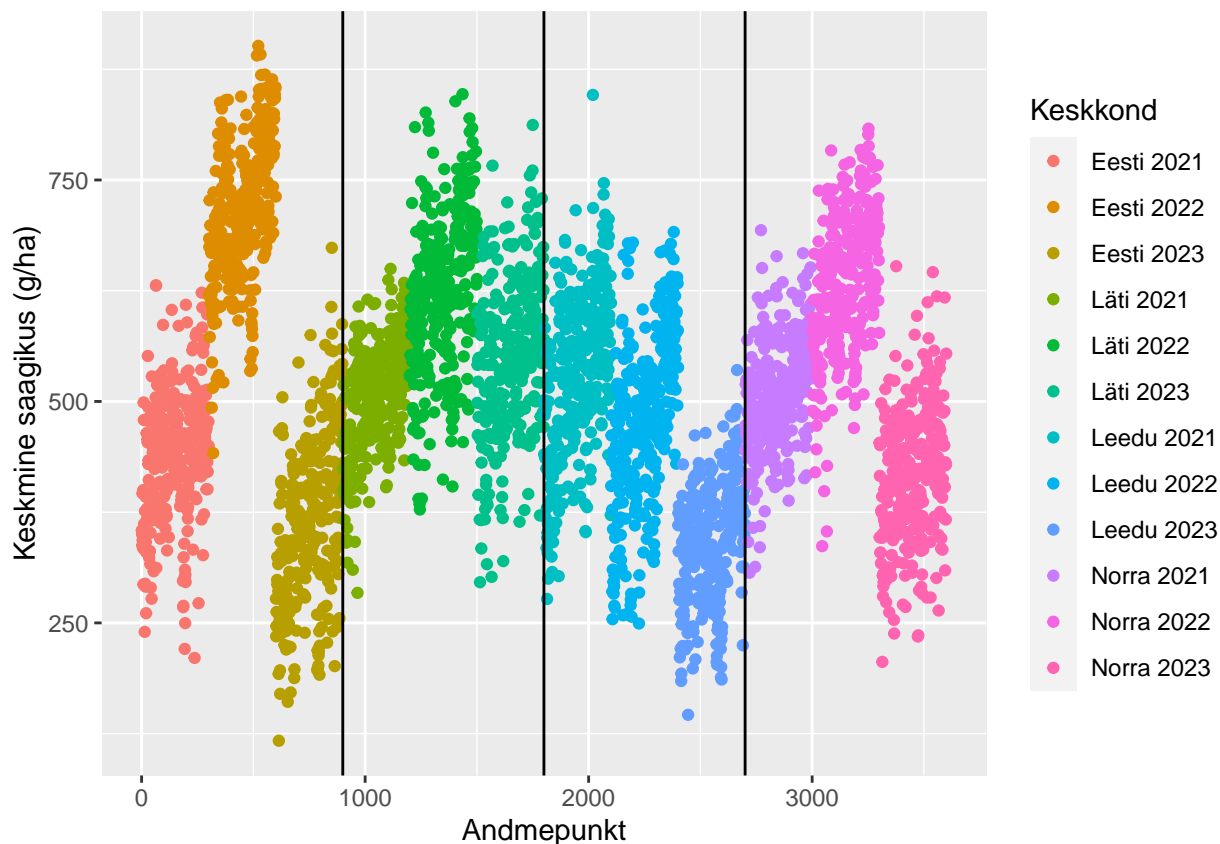
Eelnevalt välja toodud valemites tähistab r_{ij}^* keskmistatud väärtuse x_{ij}^* astakut j -nda keskkonna siseselt, \bar{r}_i ja M_i^* on i -nda genotüübi astakute keskmine ja mediaan üle kõigi keskkondade, ning \bar{r}_i ja M_i on astakute keskmine ja mediaan originaalsete väärtuste järgi. Alamindeks i näitab i -ndat genotüüpi ja alamindeks j näitab j -ndat keskkonda. Kõigi statistikute korral näitab madal statistiku väärtus stabiilset genotüüpi.

4 Nobalwheat projekti andmete statistiline analüüs

Kõikide analüüside tegemisel on kasutatud programmeerimiskeelt R R Core Team (2023). Graafikute tegemisel on kasutatud R-i paketti ggplot2 (Wickham (2016)). Kõik funktsioonid statistikute arvutamiseks on pakettidest “metan” ja “agricolae” (de Mendiburu (2021)), (Olivoto and Lúcio (2020)).

4.1 Andmete esmane analüüs

Jooniselt 3 on selgelt näha, et keskmine saagikus erineb keskkonniti. Ka ühe riigi raames võivad erinevatel aastatel genotüüpide saagikused olulisel määral erineda. See tähendab, et genotüüpide valikul on oluline valida sellised genotüüpe, mis on stabiilsed paljudes eri keskkondades. On näha, et viimase aastaga (2023) on kõigis riikides üldine saagikus võrreldes aastaga 2022 langustrendis. Samas on ka näha, et keskkonna siseselt võib saagikus genotüüpide vahel erineda rohkem kui kahekordselt. Sellest võiks järeldada, et nii keskkond kui genotüübid mõjutavad olulisel määral saagikust.



Joonis 3: Genotüüpide keskmine saagikus keskkondade lõikes

Graafikute vaatamise asemel võime ka vaadata veidi täpsamat ANOVA väljundit ja uurida, mil määral seletavad

erinevad tegurid saagikuse dispersiooni andmetes.

Vaatame tabelit 4, mis on tehtud valemi (2) põhjal. Keskkond tähistab erinevaid koha ja aasta kombinatsioone, Genotüüp tähistab erinevaid sorte. Plokk tähistab ühe ja sama genotüübi ja keskkonna kombinatsiooni korduseid (kokku tehti igas keskkonna ja genotüübi kombinatsioonis kaks katset). Kooloniga märgitud kujutavad kahe erineva tunnuse koosmõju. Keskkonna ja genotüübi koosmõju on see GxE efekt, mida kõigi käesolevas töös olevate statistikutega proovitakse mõõta. Eelnevast tabelist näeme, et keskkonna sisse toomine tõesti seletab väga suure osa saagikuse varieeruvusest. Rohkem, kui pool (55,0%) kogu saagikuse varieeruvusest on ära seletatud keskkonna mõju poolt. Väga suure osa varieeruvusest seletab ära ka genotüübi peamõju (16,5%). Teoreetiliselt ei tohiks kordustel statistilist mõju olla, kuna ploki mõju eksisteerimine tähendaks mingit süstemaatilist erinevust saagikuse erinevuses, mis on katse läbiviimisel toimunud. Antud analüüsis tuli see mõju statistiliselt oluline, kuid ta seletab väga väikese osa varieeruvusest (2,3%) ning tema olulisust võib seletada suure valimimahuga. Mida suurem valimimaht, seda tõenäolisemalt muutuvad ka väiksemad muutused keskmisest statistiliselt oluliseks, isegi kui need muutused suured ei ole.

Keskkonna ja genotüübi koosmõju (GxE) on statistiliselt oluline, kusjuures see seletab märgatava osa (15,8%) kogu dispersioonist. Genotüübi ja keskkonna koosmõju seletab peaaegu sama palju, kui genotüüpide mõju, mis märgib just selle olulist mõju saagikuse hindamisel. Sellele efektile ei ole lihtsasti interpreteeritavat seletust. Kõikidest teistest mõjudest erinevalt on tal väga suur vabadusastmete arv, mis teeb selle interpreteerimise eriti raskeks. Järgnevalt kasutame erinevaid statistilisi meetodeid leidmaks GxE efektile paremaid seletusi ja kasutatavaid järeldusi.

4.2 AMMI mudel

AMMI analüüsi läbiviimiseks kasutatakse R-i tarkvara ning “agricolae” pakettist AMMI käsku “AMMI”. (de Mendiburu 2021)

Tavaliselt kasutatakse AMMI analüüsis kahte esimest peakomponenti, kuna kaht komponenti on lihtne kujutada ja tõlgendada kahemõtmelisel joonisel ning sellest on hea näha genotüüpide ja keskkondade vahelisi erinevusi koosmõju efektile. Tabelist 5 on aga näha, et statistiliselt olulisi peakomponente (peakomponentide statistilise olulisuse teste on kirjeldatud (Annicchiarico 1997, lk.55) on kaheksa ja et vähemalt 60% koosmõju dispersioonist oleks seletatud, tuleks võtta vähemalt viis peakomponenti. Nii palju peakomponente on aga raske korraga graafiliselt kujutada. Kui kaasata kõik statistiliselt olulised komponendid (8 peakomponenti), oleks peakomponentide poolt kirjeldatud 87,7%

Tabel 4: ANOVA analüüsi tulemused koosmõjuga mudelile

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Proportion of variance
Keskkond	11	82712378	7519307.087	1714.9709	0	0.5498
Genotüüp	299	24803758	82955.711	18.9202	0	0.1649
Keskkond:Plokk	12	3422768	285230.671	65.0542	0	0.0228
Keskkond:genotüüp	3289	23769329	7226.917	1.6483	0	0.1580
Jääk	3587	15727238	4384.510	NA	NA	0.1045

koosmõjude dispersioonist.

On selge, et ainult esimese kahe peakomponendi peale graafiku tegemisel jääb suurem osa koosmõju efekti seletamata. Kuna rohkem, kui kahe telje korraga graafikul kujutamine on keeruline, siis järgnevalt on tehtud neli erinevat graafikut, kus ühel teljel on esimene peakomponent (sest see seletab siiski kõige suurema osa dispersiooni) ning teisel teljel on sellele oluliselt järgnevad komponendid.

Meeldetuletuseks toome välja keskkonna tähistused :

1- Eesti 2021 2- Eesti 2022 3- Eesti 2023 4- Läti 2021 5- Läti 2022 6- Läti 2023 7- Leedu 2021 8- Leedu 2022 9- Leedu 2023 10- Norra 2021 11- Norra 2022 12 - Norra 2023

Mõned seletused järgnevate graafikute interpreteerimiseks. (Wodebo et al. 2023, ptk.2.5.2)

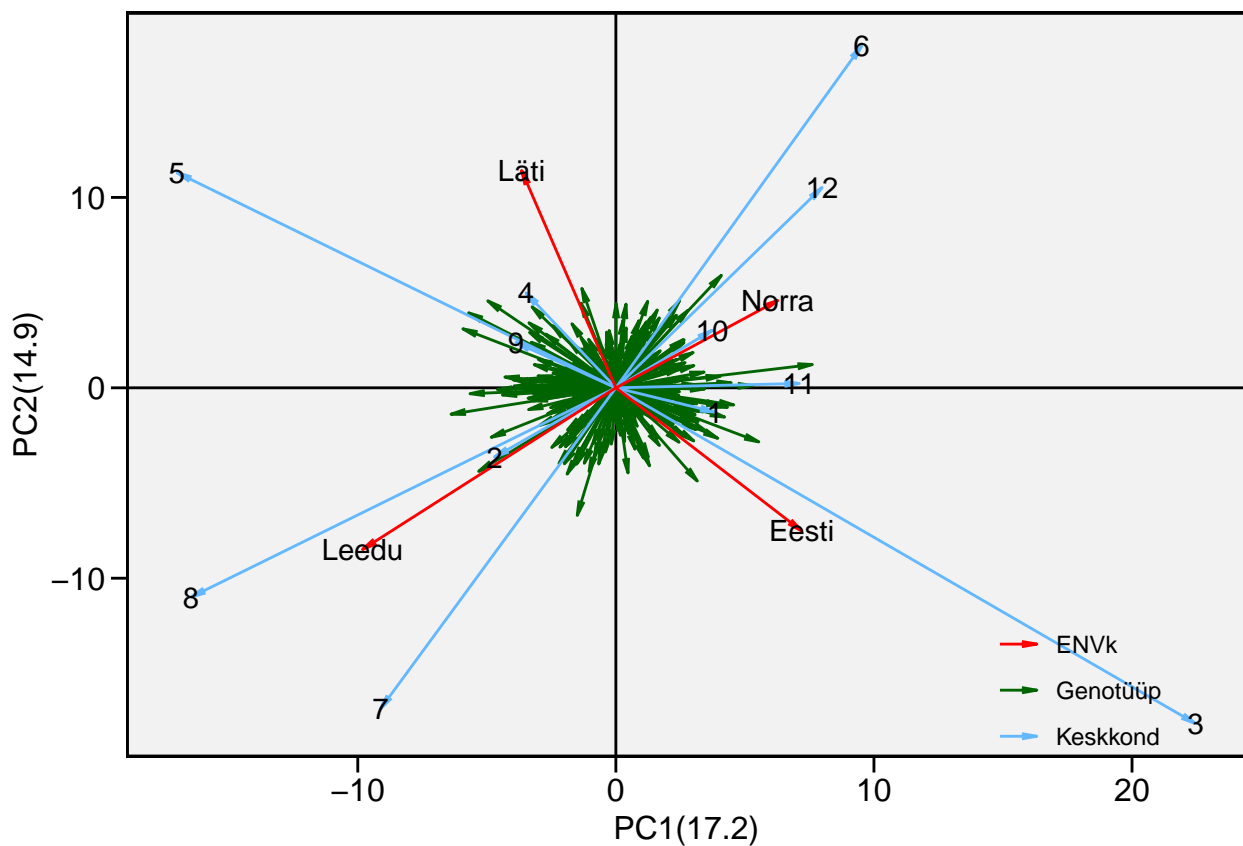
- Iga roheline vektor tähistab üht genotüüpi, ning iga sinine vektor tähistab üht keskkonda. Genotüübid ei ole lahterdatud, kuna selle tegemisel muutuks graafik loetamatuks. Punasega on proovitud anda keskmist hinnangut sama riigi kolmele aastale kokku. Kõigi samale riigi kolmele erinevale aastale vastavate vektorite koordinaadid on kokku liidetud ja jagatud kolmega.
- Mida pikem on keskkonna vektor, seda ebastabiilsem üleüldiselt on see keskkond (selles keskkonnas on igalt genotüübilt oodata suurt erinevust (nii positiivset, kui negatiivset) selle genotüübi üldkeskmisest). Mida lähemal on vektor nullpunktile, seda väiksem on selles keskkonnas üldine varieeruvus. Kõigilt genotüüpidelt on oodata selles keskkonnas nende üldkeskmisele lähedast tulemust.
- Mida pikem on genotüübi vektor, seda ebastabiilsem on see genotüüp keskmiselt (igas keskkonnas on sellelt genotüübilt oodata suurt erinevust selle genotüübi keskmisest). Mida lähemal on genotüüp null punktile, seda stabiilsem on see genotüüp. Kõigis keskkondades on oodata sellelt genotüübilt tema keskmise lähedast tulemust.
- Kaks genotüüpi, mis on üksteisele lähedal, käituvad kõigis keskkondades samasuguselt. Kahes keskkonnas, mis on üksteisele lähedal, käituvad kõik genotüübid sarnaselt (nii ebastabiilsuse mõistes kui ka stabiilsuse mõistes).

Tabel 5: Koosmõju peakomponentide tulemused AMMI mudelist

	percent	acum	Df	Sum.Sq	Mean.Sq	F.value	Pr.F	Proportion of full variance
PC1	17.2	17.2	309	4091830.2	13242.169	3.02	0.000	2.720
PC2	14.9	32.1	307	3545619.6	11549.249	2.63	0.000	2.357
PC3	12.0	44.1	305	2843449.2	9322.784	2.13	0.000	1.890
PC4	11.5	55.6	303	2727558.3	9001.842	2.05	0.000	1.813
PC5	10.3	65.9	301	2453391.7	8150.803	1.86	0.000	1.631
PC6	8.0	73.9	299	1906986.4	6377.881	1.45	0.000	1.268
PC7	7.1	81.0	297	1687197.0	5680.798	1.30	0.001	1.122
PC8	6.7	87.7	295	1599020.3	5420.408	1.24	0.004	1.063
PC9	5.5	93.3	293	1311617.4	4476.510	1.02	0.399	0.872
PC10	3.9	97.2	291	933024.2	3206.269	0.73	1.000	0.620
PC11	2.8	100.0	289	668731.9	2313.951	0.53	1.000	0.445

- Nurk vektorite (nii genotüüpide kui keskkondade) vahel näitab seoseid nende kahe vahel. Kui keskkond ja genotüüp on samasuunalised (väike nurk nende vahel), siis annab genotüüp keskmisest kõrgemaid tulemusi selles keskkonnas. 90 kraadine nurk tähendab seotust nende kahe vahel ehk keskmist tulemust. Rohkem kui 90 kraadine nurk (vastassuunalised vektorid) tähendab keskmisest halvemat tulemust.
- Genotüübi ja keskkonna vahelist seost saab mõõta, kui joonestada projektsioon genotüübilt keskkonna vektorile. Mida kaugemal positiivses suunas on projektsioon keskkonna vektoril, seda parem on genotüüp selles keskkonnas. Negatiivne suund vektoril tähendab halvemat tulemust.

Jooniselt 4 on näha, et leiduvad üksikud genotüübi vektorid, mis on rohkem adapteerunud (pikemate nooltega), kuid enamus genotüübi vektoreid on võrreldes keskkonna vektoritega keskme lähedal. Kui võrrelda riikide kaupa keskmistatud vektoreid, siis peaaegu iga vektor on risti teise vekoriga. Eesti ja Norra koondvektorid on veidi samasuunalised, mis tähendab, et need keskkonnad kutsuvad esile sarnaseid koosmõju efekte genotüüpides. Norra ja Leedu vektorid on omavahel peaaegu täielikult vastassuunalised.

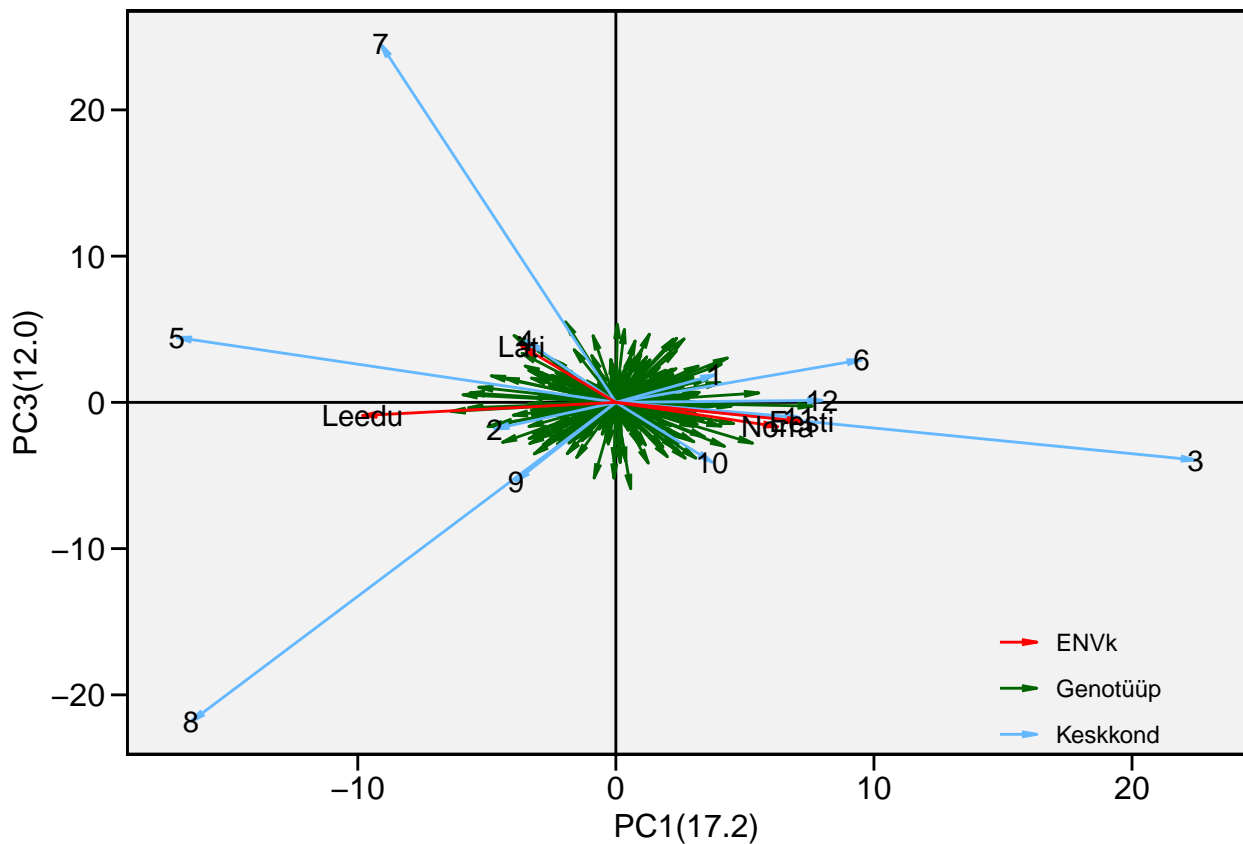


Joonis 4: AMMI mudeli esimese ja teise komponendi graafiline võrdlus (kokku 32,1 protsenti)

Vaadates iga keskkonna individuaalseid vektoreid, siis ainult Norra omad on võrdlemisi koondunud lähestikku. Eesti,

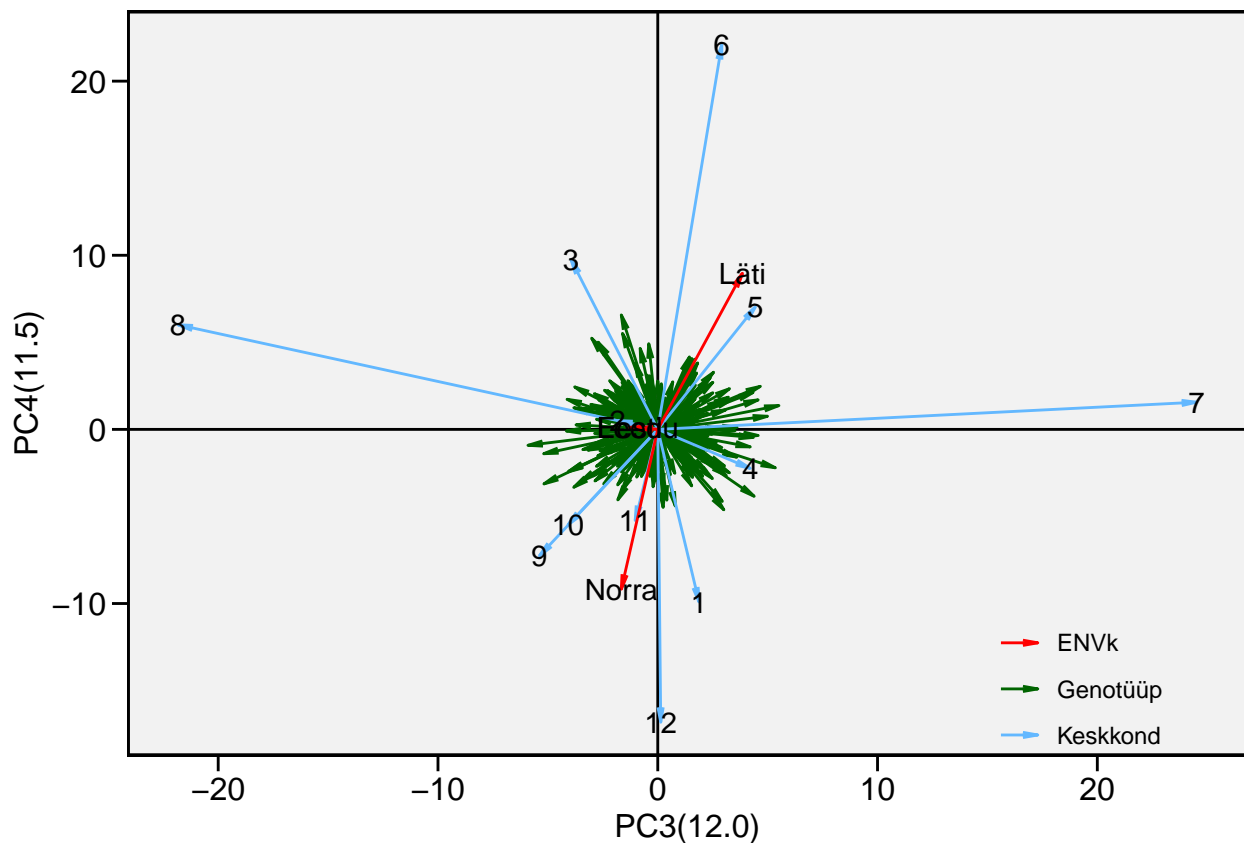
Läti ja Leedu individuaalsete aastate vektorid on võrdlemisi laiali, mis tähendab, et nendes piirkondades keskkond erines aastate võrdluses oluliselt, mis teeb koondkeskkonna mõtte kaheldavaks.

Jooniselt 5 näeme, et genotüüpide vektorid on samamoodi keskpunkti ümber koondunud. Erinevus esimese joonisega on see, et Norra ja Eesti koondvektorid on nüüd väga lähedastikku üksteisele. Kuigi Norra vektorite puhul on selline koondvektor mõistlik hinnang, sest Norra erinevate aastate vektorid on üksteise lähedastikku, siis Eesti puhul see nii ei ole. Aastal 2021 ja 2023 on vektorid Norra omaga risti, kuid aastal 2022 on see väga erinev teistest vektoritest. Leedu ja Läti vektorid aastate lõikes on väga erinevad üksteisest, seega nende järgi koondvektori mõte on kaheldav.



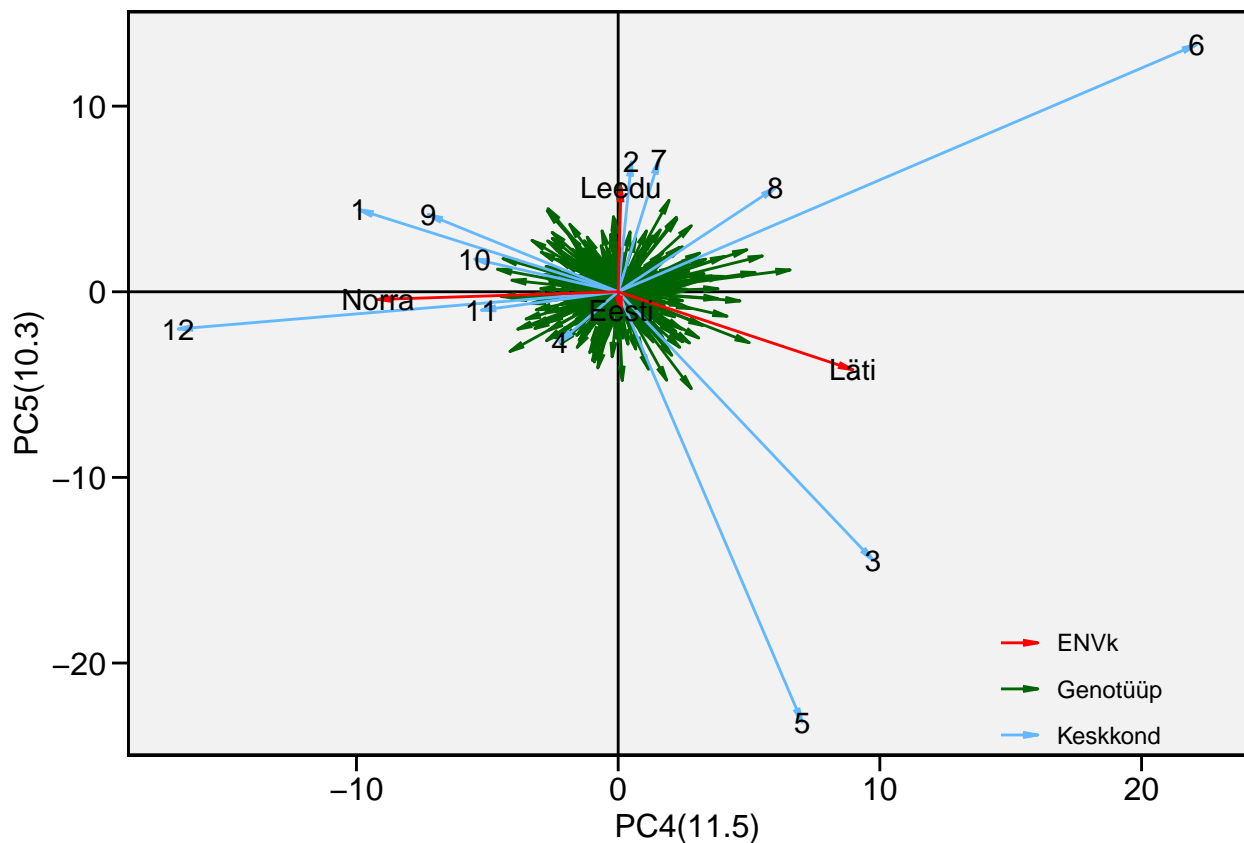
Joonis 5: AMMI mudeli esimese ja kolmanda peakomponendi graafiline võrdlus (kokku 29,2 protsenti)

Jooniselt 6 on näha erinevusi teiste graafikutega. Eesti ja Leedu koondvektorid on keskpunkti lähedal. Eesti vektorid on ka kõigil aastatel suhteliselt lähedal keskpunktile, mis ütleb, et need kaks peakomponenti ei seleta väga hästi koosmõju efekti või seda ei eksisteeri Eestis. Leedu koondvektor on samuti keskpunkti lähedal, kuid selle individuaalsed vektorid on laiali, mis teeb koondvektori väheusutavaks. Läti ja Norra vektorid on võrdlemisi lähedastikku, seega on nende koondvektorid usutavad. Koondvektorid on neil riikidel vastassuunalised, seega Norra ja Läti on koosmõju poolt üksteisele vastanduvad keskkonnad.



Joonis 6: AMMI mudeli kolmanda ja neljanda peakomponendi graafiline võrdlus (kokku 23,5 protsenti)

Sarnaselt joonisega 6 saab joonise 7 pealt näha, et Eesti koosmõju ei ole hästi modelleeritud nende peakomponentide poolt. Norra koondvektor on kõige usutavam, sest selle aastate vektorid (10+11+12) on samasuunalised. Leedu vektorid (7+8+9) on veidi rohkem killustunud, seega koordineksisse tuleks skeptiliselt suhtuda. Läti vektorid (4+5+6) on kõige enam killustunud, seega koordineks ei ole usaldusväärne. Siiski on Norra vektor üldiselt vastassuunaline või risti Leedu ja Läti vektoritega ja Eesti vektoriga pigem risti.



Joonis 7: AMMI mudeli neljanda ja viienda peakomponendi graafiline võrdlus (kokku 21,8 protsenti)

Kõikide eelnevate graafikute peale kokku (mis viie peakomponendi peale kokku seletab 65,9 % GxE efektist) saab öelda, et Eestis ja Norras on kas sarnased GxE efektid või ollakse üksteisega risti. Nende riikide keskmised ja ka individuaalsed vektorid kipuvad samas suunas olema. Läti ja Leedu vektorid on võrreldes Eesti ja Norra omadega poolenisti vastassuunalised ja poolenisti risti, ehk nendes riikides on oodata väiksemat saaki, kuid ka mingit omapärast teist efekti. Läti vektor kipub olema rohkem risti ja Leedu vektor kipub olema rohkem vastassuunaline Eesti ja Norra vektoriga.

Kahemõõtmeliste graafikute vaatamine võib anda meile aimu, millised genotüübid on millistes keskkondades head, kuid see on subjektiivne graafikutelt järelduste tegemine. Kuna individuaalseid peakomponente oli palju, ning üksik peakomponent seletas vähe dispersiooni, siis joonised võisid üksteisest tunduvalt erineda ning järelduste tegemine on raske. Veidi täpsemaid ja analüütilisemaid stabiilsusstatistikuid kasutades saab leida üldiselt stabiilsemaid ja volatiilsemaid genotüüpe.

4.3 Zhangi D

Stabiilsusindeksis kasutatakse kuni statistiliselt olulisi peakomponente indeksi arvutamisel. Eelnevalt on näha, et kuni kaheksas peakomponent on statistiliselt oluline. Kaheksa peakomponenti seletavad kokku 87,7 % kogu GxE efektist.

D_i välja arvutamise eeskiri on välja toodud eelnevalt valemis (4). D(8) tähistab siinkohal, et Zhangi D statistiku arvutamisel on kasutatud kaheksat peakomponenti. Tabelist 6 on näha, et Zhangi D parameetri järgi on kõige stabiilsemad genotüübid 299,255 ja 23.

4.4 ASV

ASV väärtuse väljaarvutamise eeskiri on eelnevalt ära toodud valemis (5). ASV järgi (tabel 7) on stabiilsemad genotüübid 243, 195 ja 214.

4.5 GSI

Mida madalam on GSI indeks, seda parem nii stabiilsuse, kui ka keskmise saagikuse poolest genotüüp on. Vaadates GSI indekseid (tabel 8), siis genotüüp 243 on tunduvalt parem kui teised (sellel on väike GSI väärtus isegi võrreldes teise kohaga). Genotüübid 291,142, 240, 228 on kõik enamvähem võrdsete GSI väärtustega ning genotüüp 287 juures toimub juba suurem hüpe.

Suured GSI väärtused tähistavad madalat saagikust ja väga ebastabiilseid genotüüpe. GSI indeksi järgi (tabel 9) on kõige halvemad genotüübid 271,298,263 ja 274.

4.6 WAASB ja WAASBY

Kuna peakomponentanalüüsis selgus (tabel 5), et esimesed kaks peakomponenti seletavad võrdlemisi vähe dispersiooni, siis on võimalik, et WAASB poolt antud hinnangud, mis kasutavad kõiki peakomponente, on täpsemad ja seletavad paremini GxE efekti. Käsus `waasb()` jäetakse vaikselt vaikselt fikseeritud efektiks keskkond ja juhuslikuks efektiks genotüüp.

WAASBY mudelis saab erinevaid kaale anda genotüübi stabiilsusele ja genotüübi võimekusele, olenevalt uuringu eesmärkidest. Vaikselt antakse mõlemale võrdsed kaalud, kuid kaalusid muutes ühest ekstreemsusest teise ekstreem-

Tabel 6: Kuus kõige madalama D(8) väärtusega genotüüpi

GEN	DZ
299	0.2477038
255	0.2525498
23	0.2590561
264	0.2787218
268	0.3138955
237	0.3393384

Tabel 7: Kuus kõige madalama ASV väärtusega genotüüpi

GEN	ASV
243	0.0890029
195	0.2017990
214	0.2444504
56	0.2483647
121	0.2730699
172	0.3844823

Tabel 8: Genotüüpide ASV ja saagikuse astakud ja GSI väärtused viimaste kasvavas järjekorras.

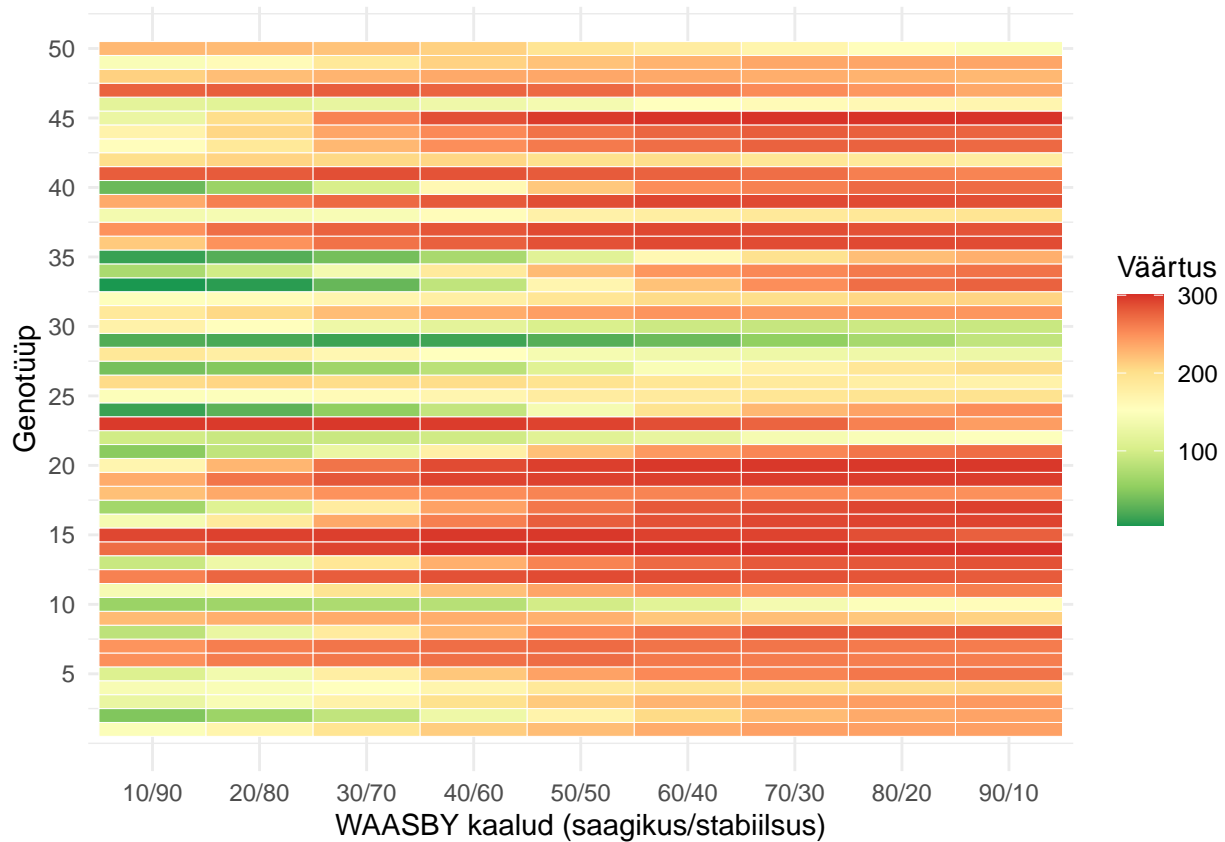
GEN	Y_R	ASV_R	GSI
243	4	1	5
291	13	24	37
142	21	21	42
240	16	27	43
228	36	9	45
287	20	42	62

Tabel 9: Genotüüpide ASV ja saagikuse astakud ja GSI väärtused viimaste kahanevas järjekorras.

GEN	Y_R	ASV_R	GSI
6	257	263	520
89	248	277	525
12	274	281	555
118	263	292	555
14	298	258	556
15	271	300	571

susse saab uurida, kas kõrge WAASBY indeksiga genotüübid jäävad samale kohale, või muutuvad nende asukohad märgatavalt. Suur kohtade muutus tähendab seda, et genotüüp on ühest aspektist (saagikus või stabiilsus) hea, kuid teisest aspektist võrdlemisi kehv.

Graafikul 8 on ära toodud esimese 50 genotüübi WAASBY indeksite astakud sõltuvalt sellest, kuidas on jagatud WAASBY indeksi kaalud. Mida “rohelisem” on genotüüp kogu x-telje vältel, seda parem on genotüüp nii stabiilsuse, kui ka saagikuse poole pealt. On näha, et leidub genotüüpe (Näiteks 14,15 ja 41), mis on kogu x-telje peal “punased”, mis on halva genotüübi tunnus. Samas leidub ka genotüüpe (Näiteks 29), mis on enamasti “roheline”, ehk need tähistavad väga heasid genotüüpe. Lisaks on ka genotüüpe, mille astak tugevalt sõltub kaaludest. Teise genotüübi näitel on genotüüp stabiilne, kuna madala saagikuse kaalu korral saab ta madala astaku. Samas, kui stabiilsuse kaalu vähemaks võtta ja saagikuse kaalu suurendada, siis tõuseb selle genotüübi astak tunduvalt, mis tähendab, et see genotüüp on küll stabiilne, kuid madala keskmise saagikusega.



Joonis 8: Esimese 50 genotüübi WAASBY astakute muutumine olenevalt WAASBY kaaludele.

Sõltuvalt sellest, mis on uuringu eesmärgiks, võivad erinevad kaalude jaotused huvi pakkuda.

Tabelis 10 on ära toodud 5 kõige kõrgema WAASBY indeksi saanud genotüüpi vastavalt sellele, millist kaalu saagikusel

kasutati. On selge, et universaalselt head genotüüpi, mis on kõigist teistest genotüüpidest üle nii saagikuses kui stabiilsuses, ei leidu. Samas leidub genotüüpe, mis on kõrgetel kohtadel väga erinevate kaalude korral. Näiteks genotüübid 293 ja 182 on vähemalt kolmel erineval kaalul esikolmes. Üleüldiselt head genotüübid on veel 243, 182, 97 ja 86.

Kokkuvõtteks võib WAASBY indeksite pealt öelda, et kõige stabiilsemad genotüübid on 182 ja 86 ning 97. Kõige saagikam genotüüp on vastuvaidlematult 293, millele järgnevad 243 ja 71.

Eelnevalt sai ära toodud 50 genotüübi võrdlus graafik. 300 genotüübi samaaegne kujutamine sarnasel graafikul ei ole aga võimalik ning tulemuslik. Et koondada erinevate WAASBY kaalude tulemusi suuremal arvul genotüüpidel võib rakendada näiteks klasterdamise meetodeid. Selles töös kokkuvõtvaid meetodeid ei rakendata ja võrreldakse tulemusi teiste statistikutega.

4.7 Eberharti ja Russeli regressioon

Tabelis 11 on ära toodud paljude genotüüpide Eberharti ja Russeli regressiooni erinevad kordajad. Seletame järgnevalt ära, mida iga muutuja tähendab ning millised neist stabiilsuse kohapealt olulised on:

b_0 - regressiooni vabaliige ehk genotüübi keskmine üle kõigi keskkondade,

b_1 - regressioonsirge tõus, ehk genotüübi vastuvõtlikkus keskkonnast tingitud mõjudele,

$t(b_1 = 1)$ - t-testi väärtus, mis kontrollib, kas b_1 väärtus võiks olla nullhüpoteesi kohaselt võrdne 1-ga,

$pval_t$ - eelneva t-testi p-väärtus,

$s_{d_i}^2 - \sigma^2$ ehk regressioonsirge kaugus tegelikkest keskmistest,

$F(s_{d_i}^2 = 0)$ - F-test kontrollimaks, kas $s_{d_i}^2$ võiks olla nullhüpoteesi kohaselt võrdne 0-ga,

$pval_f$ - eelneva F-testi p-väärtus,

$RMSE$ - regressiooni keskmine ruutviga tegelikest väärtustest,

$R^2 - R^2$, ehk mitmese korrelatsioonikordaja ruut.

Stabiilsuse poole pealt on järgmised kolm statistikut olulised : b_1 , σ^2 ja $RMSE$.

Tabel 10: 5 kõige kõrgema WAASBY indeksiga genotüüpi vastavalt saagikuse kaalule WAASBY analüüsis.

Astak	saak10	saak20	saak30	saak40	saak50	saak60	saak70	saak80	saak90
1	182	182	182	86	293	293	293	293	293
2	183	86	86	182	243	243	243	243	243
3	86	183	97	97	86	86	71	71	71
4	97	97	183	293	97	228	260	260	76
5	33	155	143	243	182	97	291	291	219

Tabel 11: Eberharti ja Russeli regressiooni tulemused

GEN	b0	b1	t(b1=1)	pval_t	s2di	F(s2di=0)	pval_f	RMSE	R2
1	459.4458	0.8314183	-1.3368887	0.1813437	611.59678	1.2793372	0.2359513	48.33780	0.7726730
2	456.1833	1.0100288	0.0795308	0.9366149	-315.88776	0.8561460	0.5742299	39.54288	0.8822929
3	456.4217	0.9264408	-0.5833396	0.5597014	899.32888	1.4106232	0.1687353	50.75746	0.7928534
4	487.2167	0.9820824	-0.1420905	0.8870165	783.57711	1.3578081	0.1936460	49.79819	0.8171298
5	430.2200	0.9581094	-0.3322015	0.7397565	-57.85567	0.9738805	0.4639119	42.17424	0.8556886
6	449.7613	0.9792527	-0.1645306	0.8693227	2048.15661	1.9348085	0.0364092	59.44473	0.7571493

$b1$ näitab, kui palju vastava genotüübi keskmine saagikus muutub eri keskkondades olenevalt sellest, kui palju keskkonna keskmine muutub. Stabiilseks genotüübiks loetakse genotüüpe, mille $b1$ väärtus on võrdne või lähedal ühele. Sel juhul, kui keskkond muutub keskmiselt soodsamaks, siis genotüübi keskmine saagikus suureneb täpselt keskkonna keskmiste vahe võrra. Suuremate $b1$ väärtuste korral tähendab soodsam keskkond suuremat saagikuse suurenemist, kuid ka ebasoodsam keskkond tähendab suuremat saagikuse langust. Madalama $b1$ väärtusega reageerib genotüüp keskkonna muutustele vähemal määral.

s_{di}^2 näitab regressioonsirge kaugust kokku genotüübi keskmistest. Kõrge väärtus siin näitab, et Eberhart ja Russeli regressioon ei suuda hästi kinni püüda GxE efekti, ehk Eberharti ja Russeli regressiooni kohaselt on ta ebastabiilne (Alwala et al. (2010)).

RMSE (*Root Mean Square Error*) on sarnane s_{di}^2 statistikule, aga kauguse asemel keskmisest vaadatakse keskmist ruut-kaugust tegelikust väärtusest. Sarnaselt eelmisega tähendab suur väärtus siin, et mudel ei püüa kinni väga hästi GxE efekti, ehk tuleks teiste mudelite poole vaadata.

Tabelist 11 vaadates on näha, et väga paljudel genotüüpidel on nii F-testi kui t-testi väärtused madalad. Madala valimimahuga on testidel väga laiad usaldusintervallid, seega ei tasu ilmingimata vaadata testide p-väärtuseid, vaid regressiooni parameetreid endid tuleks võrrelda.

Tavaliselt võrreldakse genotüüpe eraldi nii $b1$ kui ka s_{di}^2 poolest. Neid kokku ei võeta, kuna mõlemad mõõdavad stabiilsust eri aspektist.

Kui vaadata tabeleid (12,13 ja 14), siis $b1$ järgi kõige stabiilsemad genotüübid on 92, 287, 196, 275 ja 239. Kui vaadata s_{di}^2 väärtuseid, siis kõige stabiilsemad genotüübid on 213, 253, 169 ja 188. Üks võimalus mõlemat korraga arvestada oleks summeerida mõlema statistiku astakud, ning seejärel võrrelda astakute summasid. Kui kahe statistiku astakute summa järgi arvestada stabiilsust, siis stabiilsemad genotüübid on 269, 246, 6, 207, 204.

4.8 Lin ja Binns paremus määr

Funktsioon “superiority()” paketest “metan” (Olivoto and Lúcio 2020) jagab esialgselt keskkonnad kaheks. Keskkonnad, kus keskmine saagikus on parem kui üleüldine keskmine saagikus, nimetatakse headeks (*favourable*) keskkondadeks. Keskmisest madalamad nimetatakse halbadeks (*unfavourable*) keskkondadeks. Funktsioon arvutab statistiku eraldi

Tabel 12: b1 järgi parimad genotüübid

GEN	rankb1	ranks2di	Sumranks
92	1	95	96
287	2	218	220
196	3	276	279
275	4	248	252
239	5	159	164
269	6	17	23

Tabel 13: s2di järgi parimad genotüübid

GEN	rankb1	ranks2di	Sumranks
213	140	1	141
253	264	2	266
169	138	3	141
188	64	4	68
292	95	5	100
93	121	6	127

Tabel 14: Summeeritud astakute järgi parimad genotüübid

GEN	rankb1	ranks2di	Sumranks
269	6	17	23
246	16	11	27
6	19	15	34
207	26	24	50
204	41	10	51
4	36	23	59

välja heades keskkondades (Pi_{f,R_f}), halbades keskkondades (Pi_{u,R_u}) ja kõigis keskkondades (Pi_{a,R_a}). Siinkohal Pi tähistab statistiku enda väärtust ja R tähistab vastavalt statistiku astakut. Tabeli 15 järgi on kõige paremaks keskkonnaks 2022. aasta Eesti. Headeks keskkondadeks on veel Läti aastatel 2022, 2023, Leedu aastal 2021 ja Norra aastal 2022. Kõik ülejäänud olid keskmisest halvemad keskkonnad.

Tabelist 16 on näha, et kõigi keskkondade peale arvatud statistiku järgi parim genotüüp ei ole ilmtingimatta parim kõigis keskkondades. Näiteks genotüüp 293 on kõigi keskkondade peale esimene, kuid väiksemate gruppide järgi on ta 12. ja 1. koht vastavalt. Siiski on ta suhteliselt kõrgel kohal ka teistes indeksites. Kümne kõige väiksema indeksi väärtusega genotüübi järgi saab öelda, et hea genotüüp üldises mõistes on ka hea genotüüp halbades keskkondades ja heades keskkondades eraldi vaadatuna.

See statistik ei järjesta genotüüpe ainult keskmise saagikuse järgi. Tabelist 16 on näha, et saagikused ei ole ainult kahanevas järjekorras, vaid ka kõrge koht nõuab, et genotüüp oleks suhteliselt saagikas nii halbades keskkondades, kui ka heades keskkondades. Seega on kõrgel kohal olevad genotüübid head mitte ainult keskmise saagikuse poolest, vaid ka stabiilsuse poolest. Statistiku järgi parimad genotüübid on 293, 243, 219, 269, 251.

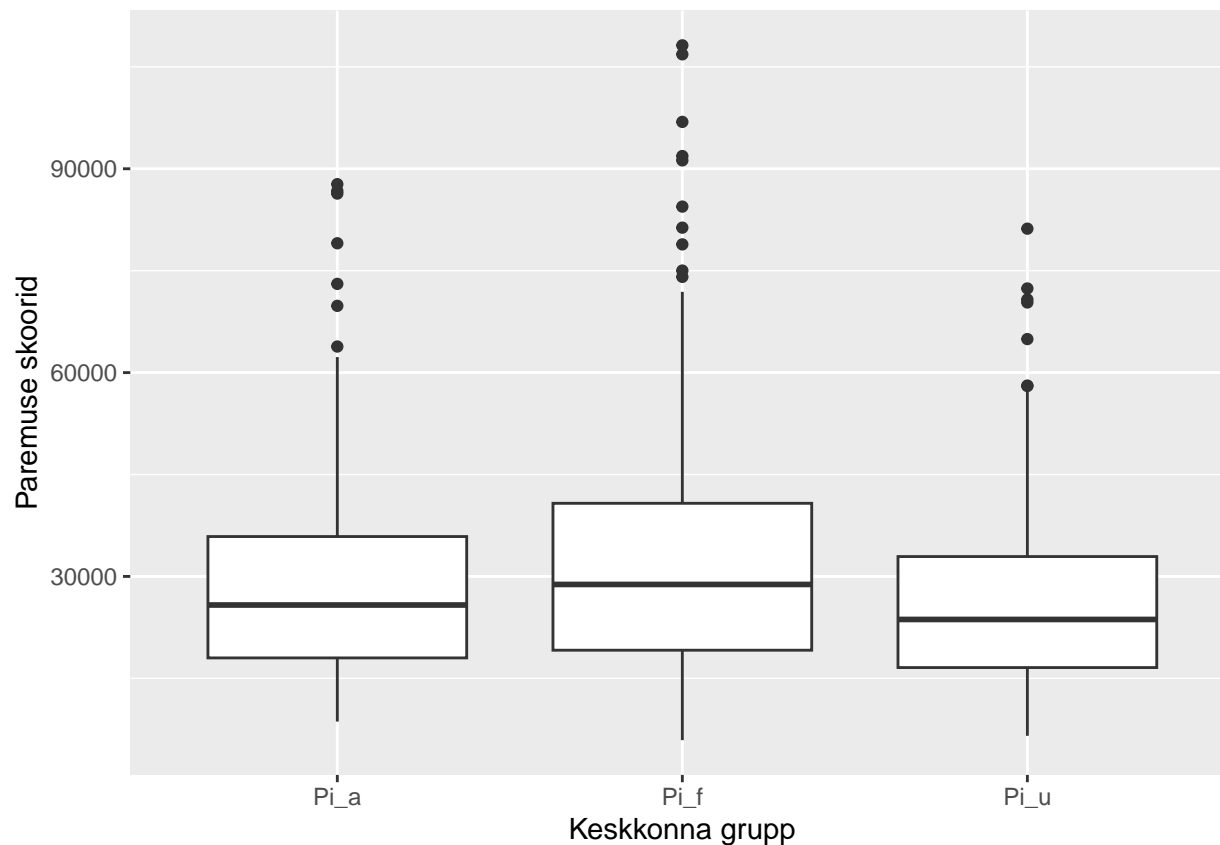
Joonisel 9 on näha erineva keskkonnagrupi indeksi jaotust karpdiagrammi kujul. Ei ole teada, mis jaotusega see statistik olema peaks, aga eelnevalt graafikult on näha, et kõikides keskkonna alamgruppides on ta enamvähem samasuguse jaotusega.

Tabel 15: Keskkondade headus ja nende efektid saagikusele

ENV	Y	index	class
1	437.3868	-71.422291	unfavorable
2	717.5422	208.733042	favorable
3	364.1767	-144.632458	unfavorable
4	503.7333	-5.075791	unfavorable
5	623.0492	114.240042	favorable
6	550.5163	41.707175	favorable
7	528.7022	19.893042	favorable
8	480.1918	-28.617291	unfavorable
9	341.1985	-167.610608	unfavorable
10	507.8702	-0.938958	unfavorable
11	634.2650	125.455875	favorable
12	417.0773	-91.731779	unfavorable

Tabel 16: Lin ja Bin paremuse määra järgi 10 parimat genotüüpi kõigi keskkondade järgi

GEN	Y	Pi_a	R_a	Pi_f	R_f	Pi_u	R_u
293	615.8250	8641.053	1	11581.252	12	6540.911	1
243	605.1817	8803.738	2	10660.346	7	7477.590	4
219	606.1742	9836.611	3	5899.083	1	12649.131	30
260	597.4071	9984.400	4	11379.068	9	8988.209	7
251	606.2117	10246.168	5	14904.222	37	6918.986	2
291	590.8608	10472.848	6	8799.904	4	11667.808	23
159	590.7958	11053.675	7	10429.654	6	11499.404	21
76	603.5029	11099.286	8	13365.161	26	9480.804	8
71	604.9800	11309.484	9	10109.901	5	12166.329	29
250	592.6387	11496.109	10	13096.584	24	10352.913	12



Joonis 9: Paremuse määrade karpdiagramm erinevate keskkonna gruppide lõikes

4.9 Ökovalents

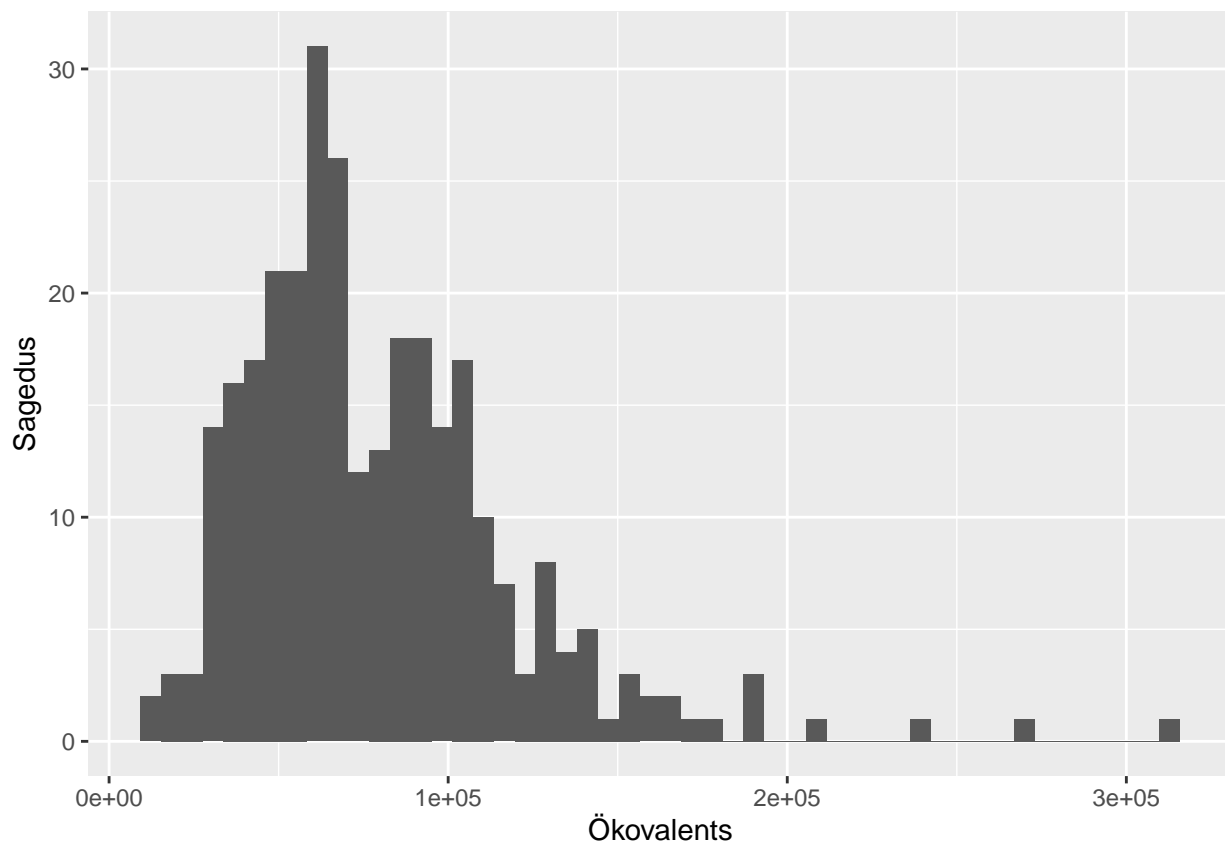
Tabelis 17 on ära toodud kümne väikseima ökovalentsi ehk kõrgeima stabiilsusega genotüüpi. Tabeli esimene tulp näitab genotüüpi, teine tulp ökovalentsi väärtust, kolmas tulp näitab, kui suure osa kogu GxE efektist vastav genotüüp annab. Suur osa GxE efekti seletamisest tähendab, et enamus GxE efektist on just selle genotüübi poolt indutseeritud. Kui genotüüp annab GxE-le väikese osa efektist, siis see tähendab, et genotüüp on lähedal oma keskmisele igas keskkonnas, mis näitab tema väikest varieeruvust keskmisest. Ökovalentsi väärtused summeeruvad kokku GxE koosmõju efekti ruutude summaks (13572726) ning osakaalud summeeruvad kokku üheks.

Tabelist on näha, et stabiilseimad genotüübid on 33, 182, 97, 86, 143.

Jooniselt 10 on näha, et Ökovalentsid sarnanevad veidi Gamma-jaotusega. Jaotus sarnaneb normaaljaotusega, kus vasak saba on kerge ja parem raskem. Ei ole teada, kas see on oodatav ökovalentsi jaotus graafik või kas ökovalentsi statistikul üldse kipub olema mingi kindel jaotus.

Tabel 17: Ökovalentsi väärtused kümnel stabiilsemal genotüübil.

GEN	Ecoval	Ecov_perc	rank
33	10786.81	0.0004538	1
182	10798.20	0.0004543	2
97	16877.36	0.0007100	3
86	19070.99	0.0008023	4
143	21069.98	0.0008864	5
188	23136.95	0.0009734	6
155	23441.10	0.0009862	7
183	25322.73	0.0010653	8
101	27820.80	0.0011704	9
121	27942.66	0.0011755	10



Joonis 10: Ökovalentsi väärtuste esinemissagedused

4.10 Stabiilsuse dispersioon

Kasutatud on funktsiooni `stability.par()` paketest “agricolae” (de Mendiburu 2021). Käsuga `anova_joint()` (paketest “metan” (Olivoto and Lúcio 2020)) on arvatud keskmine viga mida käsus `stability.par()` vaja on.

Tabelis 18 on kajastatud 10 kõige väiksema stabiilsuse dispersiooniga genotüüpi, ehk 10 kõige stabiilsemat genotüüpi. Näeme, et järjestus on täpselt sama, mis ökovalentsi puhul tabelis 17. Meetodite peatükis on seletatud, miks ökovalents ja stabiilsuse dispersioon omavahel tugevalt seotud on.

4.11 Hühni mitteparameetrilised statistikud

Statistikute arvutamiseks kasutame käsku `Huen()` pakettist `metan` (Olivoto and Lúcio (2020)). See käsk arvutab välja mitte-parameetrilised statistikud $S_i^{(1)}$, $S_i^{(2)}$, $S_i^{(3)}$, $S_i^{(6)}$ ja lihtsalt keskmise saagikuse (Y) iga genotüübi jaoks, ning arvutab iga statistiku jaoks ka vastava astaku. Märkusena on välja toodud, et käsk `Huen()` arvutab statistikut $S_i^{(1)}$ valesti. Statistiku $S_i^{(1)}$ arvutati eraldi välja, ning lisati tabelisse käsitsi.

Tabeli 19 vaatamisel meeldetuletuseks on ära toodud, et $S_i^{(1)}$ ja $S_i^{(2)}$ mõõdavad vaid genotüüpide stabiilsust ning $S_i^{(3)}$ ja $S_i^{(6)}$ mõõdavad genotüüpide saagikust ja stabiilsust. Esimese kuue genotüübi statistiku vaadates näeme, et statistikute $S_i^{(1)}$ ja $S_i^{(2)}$ astakud kipuvad sarnased olema genotüübi lõikes, ning $S_i^{(3)}$ ja $S_i^{(6)}$ kipuvad samamoodi sarnased olema. See on kooskõlas teooriaga, kus esimesed kaks parameetrit mõõdavad vaid stabiilsust, ning viimased kaks mõõdavad saagikust ja stabiilsust korraga.

4.12 Thennarasu mitteparameetrilised statistikud

Meeldetuletuseks on ära toodud, et kõik Thennarasu statistikud mõõdavad ainult stabiilsust, sest Thennarasu kasutab keskmistatud genotüüpide väärtusi. Thennarasu statistikud on arvatud kasutades käsku `Thennarasu()` pakettist “metan” (Olivoto and Lúcio 2020).

Thennarasu statistikud on ära toodud tabelis 20. Esimese 6 genotüübi pealt on näha, et $N_i^{(1)}$ ja $N_i^{(4)}$ astakud kipuvad

Tabel 18: Stabiilsuse dispersiooni statistikud 10 stabiilseima genotüübi jaoks

	Yield	Stab.var
33	405.5246	963.5142
182	533.3929	964.7497
97	556.6429	1520.1398
86	558.9104	1721.3804
143	553.8238	1904.2862
188	404.5833	2094.0513
155	514.6788	2119.5371
183	523.9179	2291.5162
101	531.3571	2523.1142
121	473.9092	2531.0915

Tabel 19: Hühni statistikud S1, S2, S3, S6 ja üldkeskmine (Y) ning iga statistiku astakud.

GEN	Y	Y_R	S1	S1_R	S2	S2_R	S3	S3_R	S6	S6_R
1	459.4458	240	72.69697	160	8322.992	173	569.7507	273	7.2384	264
2	456.1833	250	52.91667	53	4703.901	18	307.2757	159	5.2678	205
3	456.4217	249	66.28788	117	9635.727	233	384.8283	209	6.2491	241
4	487.2167	210	77.91667	196	8446.932	180	402.2839	221	5.8345	228
5	430.2200	265	45.87121	35	6951.174	97	301.2527	153	7.0503	262
6	449.7613	257	69.75000	139	8230.750	162	491.5697	257	7.0249	260

sarnased olema, ning samamoodi $N_i^{(2)}$ ja $N_i^{(3)}$ astakud kipuvad sarnanema. Kui paare omavahel võrrelda, siis on näha, et madalamad $N_i^{(1)}$ ja $N_i^{(4)}$ väärtused küll näitavad ka madalamaid $N_i^{(2)}$ ja $N_i^{(3)}$ väärtuseid, kuid lineaarset seost astakute vahel ei ole.

Tabel 20: Thennarasu statistkud ja üldkeskmine (Y), ning nende astakud kuuel genotüübil.

GEN	Y	Y_R	N1	N1_R	N2	N2_R	N3	N3_R	N4	N4_R
1	459.4458	240	68.9167	125.0	0.2877523	42	0.4164	70	0.0108	166.5
2	456.1833	250	53.2500	19.5	0.2315217	14	0.3002	19	0.0048	78.5
3	456.4217	249	82.1667	227.5	0.3866667	102	0.4419	85	0.0164	224.5
4	487.2167	210	81.5833	224.0	0.4058872	116	0.4885	106	0.0169	231.0
5	430.2200	265	70.5833	145.0	0.2719974	34	0.3273	28	0.0054	91.5
6	449.7613	257	73.9167	169.0	0.3083073	55	0.4030	66	0.0127	188.0

5 Erinevad statistikud ja nende omavaheline võrdlus

Eelnevalt arvatati andmete peal palju erinevaid genotüüpide stabiilsust mõõtvaid statistikke. Kuigi kõigi stabiilsust mõõtvate statistike vahel on analüütilisel teel võimalik täpsed seosed välja arvutada, siis selle töö raamesse kõikide erinevate statistike paariviisiline analüütiline võrdlus ei mahu.

Kõik statistikud on mõeldud uurima genotüüpide stabiilsust, ehk genotüübi ja keskkonna koosmõju suurust. Teoorias võiks siis päriselt stabiilne genotüüp olla kõrgetel kohtadel ka enamustel stabiilsust kirjeldavatel statistikutel. Et vähendada väga sarnaste statistike (näiteks ökovalents ja stabiilsuse dispersioon, mille korrelatsioon on 1) liiga suurt mõju, tuleks eelnevalt läbi viia korrelatsioonanalüüs. Astakute erinevuste võrdlusel kasutatakse Pearsoni korrelatsiooni-kordajaid.

Kõikide statistike astakud on teisendatud sellisele kujule, kus madal astaku väärtus tähendab saagikat/stabiilset/mõlemat genotüüpi. Statistike tähistused on ära toodud tabelis 3.

Tabelites 21 ja 22 on ära toodud kõigi statistike omavahelised korrelatsioonid astakutele. Mõned tähelepanekud:

- Ökovalentsi ja stabiilsuse dispersiooni vahel on korrelatsioon 1 (stabiilsuse dispersiooni peatükis on põhjendatud, miks nii on).
- Statistid, mis mõõdavad vaid stabiilsust, on suuremas korrelatsioonis teiste statistikutega, mis mõõdavad vaid stabiilsust ning üldiselt väiksemas korrelatsioonis statistikutega, mis kombineerivad saagikust ja stabiilsust.
- Kõik peakomponent analüüsi sisaldavad statistid (ASI,ASV,GSI,WAASB) on keskmisest kõrgemalt korreleeritud üksteisega.
- Eberharti b1 ei ole mitte millegagi korreleeritud, kuigi Eberharti s2di on stabiilsuse statistikutega korreleeritud.
- Paremuse määr on tugevalt korreleeritud saagikusega ja nõrgalt korreleeritud stabiilsuse või mõlemat mõõtvate statistikutega.
- On näha, et Thennarasu statistikul $N_i^{(2)}, N_i^{(3)}, N_i^{(4)}$ on üldiselt negatiivsed kordajad kõigi teiste statistikutega. Lisaks on näha, et nendel on tugev negatiivne korrelatsioon üldkeskmisega (Y). Eelnevaid asjaolusid arvestades soovitaks mitte kasutada neid statistikke genotüüpide valikul. Sellele järeldusele jõudsid ka teised uurimused (Sabaghnia, Deghani, and Sabaghpour 2006, lk.1105).

5.1 Peakomponentanalüüs

Et paremini uurida erinevate statistike vahelisi korrelatsioone ja nende vahelisi sarnasusi, viiakse läbi peakomponentanalüüs korrelatsiooni-maatriksil (tabelid 21 ja 22). Tabelites olevate statistiliste meetodite lühendid on välja toodud töö alguses (tabelis 3).

Kaks esimest peakomponenti seletavad 94,1% varieeruvusest maatriksis. Kolmas peakomponent seletab kõigest 2,1% varieeruvusest. Kõik otsustusreeglid viitavad kahe peakomponenti kasutamisele. Et paremini visualiseerida statistike vahelisi seoseid, kujutame graafiliselt statistike esimese ja teise peakomponenti väärtuseid.

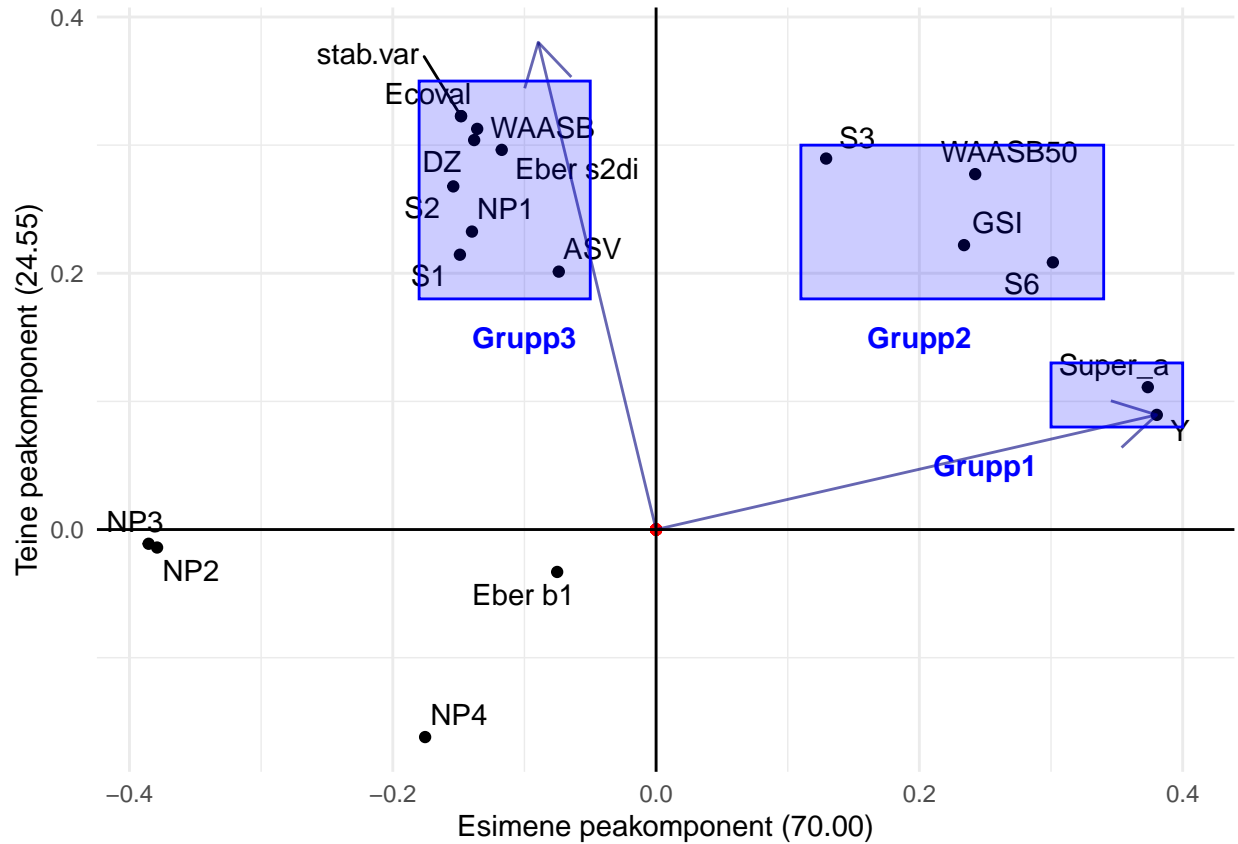
Tabel 21: Korrelatsioonid erinevate statistikute antud astakute vahel.

	Y	DZ	ASV	GSI	WAASB	WAASB50	Eber b1	Eber s2di	Super_a
Y	1.0000	-0.1737	-0.1419	0.6550	-0.1931	0.7324	-0.1402	-0.1538	0.9927
DZ	-0.1737	1.0000	0.4223	0.1898	0.8567	0.4088	0.2437	0.7271	-0.1144
ASV	-0.1419	0.4223	1.0000	0.6550	0.6866	0.3296	0.1442	0.4525	-0.1009
GSI	0.6550	0.1898	0.6550	1.0000	0.3767	0.8107	0.0031	0.2280	0.6808
WAASB	-0.1931	0.8567	0.6866	0.3767	1.0000	0.4731	0.2031	0.7487	-0.1320
WAASB50	0.7324	0.4088	0.3296	0.8107	0.4731	1.0000	0.0200	0.3338	0.7739
Eber b1	-0.1402	0.2437	0.1442	0.0031	0.2031	0.0200	1.0000	0.0026	-0.1307
Eber s2di	-0.1538	0.7271	0.4525	0.2280	0.7487	0.3338	0.0026	1.0000	-0.1052
Super_a	0.9927	-0.1144	-0.1009	0.6808	-0.1320	0.7739	-0.1307	-0.1052	1.0000
Ecoval	-0.1986	0.9179	0.5749	0.2873	0.9325	0.4319	0.2202	0.8081	-0.1342
stab.var	-0.1987	0.9180	0.5749	0.2871	0.9325	0.4318	0.2202	0.8081	-0.1343
S1	-0.2436	0.6130	0.4053	0.1235	0.6183	0.1592	0.1806	0.5176	-0.1993
S2	-0.2007	0.7610	0.4358	0.1795	0.7453	0.3168	0.2021	0.6214	-0.1443
S3	0.4622	0.5023	0.2915	0.5753	0.4953	0.6903	0.0196	0.4142	0.4993
S6	0.8669	0.1746	0.0719	0.7166	0.1562	0.8726	-0.0361	0.1196	0.8874
NP1	-0.1734	0.6616	0.3830	0.1600	0.6429	0.2807	0.1868	0.5196	-0.1221
NP2	-0.9305	0.3590	0.2435	-0.5244	0.3801	-0.5431	0.1696	0.2806	-0.9046
NP3	-0.9522	0.3743	0.2524	-0.5342	0.3892	-0.5599	0.1828	0.2930	-0.9268
NP4	-0.4115	0.0146	0.0310	-0.2905	0.0403	-0.3304	0.1720	-0.0385	-0.4130

Tabel 22: Korrelatsioonid erinevate statistikute antud astakute vahel.

	Ecoval	stab.var	S1	S2	S3	S6	NP1	NP2	NP3	NP4
Y	-0.1986	-0.1987	-0.2436	-0.2007	0.4622	0.8669	-0.1734	-0.9305	-0.9522	-0.4115
DZ	0.9179	0.9180	0.6130	0.7610	0.5023	0.1746	0.6616	0.3590	0.3743	0.0146
ASV	0.5749	0.5749	0.4053	0.4358	0.2915	0.0719	0.3830	0.2435	0.2524	0.0310
GSI	0.2873	0.2871	0.1235	0.1795	0.5753	0.7166	0.1600	-0.5244	-0.5342	-0.2905
WAASB	0.9325	0.9325	0.6183	0.7453	0.4953	0.1562	0.6429	0.3801	0.3892	0.0403
WAASB50	0.4319	0.4318	0.1592	0.3168	0.6903	0.8726	0.2807	-0.5431	-0.5599	-0.3304
Eber b1	0.2202	0.2202	0.1806	0.2021	0.0196	-0.0361	0.1868	0.1696	0.1828	0.1720
Eber s2di	0.8081	0.8081	0.5176	0.6214	0.4142	0.1196	0.5196	0.2806	0.2930	-0.0385
Super_a	-0.1342	-0.1343	-0.1993	-0.1443	0.4993	0.8874	-0.1221	-0.9046	-0.9268	-0.4130
Ecoval	1.0000	1.0000	0.6621	0.8263	0.5404	0.1831	0.7170	0.3933	0.4114	0.0563
stab.var	1.0000	1.0000	0.6621	0.8264	0.5402	0.1829	0.7171	0.3934	0.4116	0.0562
S1	0.6621	0.6621	1.0000	0.6929	0.6215	0.1279	0.6410	0.4132	0.4090	0.1653
S2	0.8263	0.8264	0.6929	1.0000	0.5092	0.2315	0.9510	0.4520	0.4524	0.1124
S3	0.5404	0.5402	0.6215	0.5092	1.0000	0.7523	0.4641	-0.2669	-0.2743	-0.1727
S6	0.1831	0.1829	0.1279	0.2315	0.7523	1.0000	0.2474	-0.6876	-0.7096	-0.3313
NP1	0.7170	0.7171	0.6410	0.9510	0.4641	0.2474	1.0000	0.4400	0.4145	0.1288
NP2	0.3933	0.3934	0.4132	0.4520	-0.2669	-0.6876	0.4400	1.0000	0.9814	0.4073
NP3	0.4114	0.4116	0.4090	0.4524	-0.2743	-0.7096	0.4145	0.9814	1.0000	0.4153
NP4	0.0563	0.0562	0.1653	0.1124	-0.1727	-0.3313	0.1288	0.4073	0.4153	1.0000

Joonisel 11 on näha, et keskmist saagikust mõõtev tunnus Y (keskmise saagikuse astak) ja sellele lähedal olevad statistikud (Grupp 1) peaaegu ühtivad esimese peakomponendi suunaga. Keskmise saagikuse astak on peaaegu samasuunaline esimese peakomponendiga. See tähendab, et esimene peakomponent mõõdab genotüübi üldist saagikuse määra.



Joonis 11: Statistike peakomponentanalüüsi esimesed kaks peakomponenti

Joonisel 11 on ka näha, et leidub suur grupp (Grupp 3) statistikuid, mis jäävad täpselt risti saagikust mõõtva peakomponendi suunaga, ehk samasuunaliselt teise peakomponendiga. Sinna gruppi kuuluvad enamasti stabiilsust mõõtvaid statistikuid ($DZ, Ecoval, WAASB, ASV, eber S_{di}^2, S_i^{(2)}, N_i^{(1)}$). Sellest võib järeldada, et teine peakomponent on mõeldud genotüüpide stabiilsuse mõõtmiseks.

Grupp 2 sisaldab statistikuid, mis on mõeldud genotüüpide valikuks nii stabiilsuse aspekti, kui ka keskmist saagikust arvestades. On näha, et need jäävad diagonaalile saagikust mõõtvate statistike ja stabiilsust mõõtvate statistike vahele, mis kinnitaks esialgset hüpoteesi, et esimene peakomponent tõesti mõõdab saagikust ja teine peakomponent genotüübi stabiilsust. Kaugus kolmandast või esimesest grupist näitab, kui suurt rõhku statistik annab kas stabiilsusele või saagikusele. Statistiku $S_i^{(3)}$ paneb rohkem rõhku stabiilsusele, kuid $S_i^{(6)}$ annab rohkem rõhku keskmisele saagikusele.

Graafiku pealt on näha, et Eberhart-Russeli regressioonist tulenev statistik b_1 (Eber b1) on lähedal keskpunktile. See tähendab, et statistikul b_1 on üldiselt nõrgad korrelatsioonid kõigi teiste statistikutega.

Kui enamus statistikuid on jagatavad kolme gruppi peakomponent analüüsi põhjal, siis $N_i^{(2)}$, $N_i^{(3)}$ ja $N_i^{(4)}$ statistikud, mis on mõeldud mõõtma stabiilsust, ei sarnane grupi 3 statistikutega. Korrelatsioonanalüüsist oli näha, et $N_i^{(2)}$ ja $N_i^{(3)}$ olid tugevas negatiivses korrelatsioonis keskmise saagikusega, mis teeb nende statistikute arvestamise genotüübi valikul kaheldavaks. Statistiku $N_i^{(4)}$ on veidi väiksema negatiivse korrelatsiooniga saagikusele, kuid siiski tunduvalt suurema negatiivse korrelatsiooniga, kui teised grupi 3 statistikud ja sellel on nõrk korrelatsioon kõigi teiste stabiilsusstatistikutega. Ka teised stabiilsust mõõtvaid statistikuid võrdlevad uurimused (Sabaghnia, Dehghani, and Sabaghpour (2006)) on näidanud, et $N_i^{(2)}$, $N_i^{(3)}$, $N_i^{(4)}$ on negatiivses korrelatsioonis saagikusega, seega on nende statistikute arvestamine genotüübi valikul suuresti kaheldav.

5.2 Millist statistikut valida?

Eelnevas peatükis sai ära näidatud, et suur osa statistikuid on omavahel teugevalt korreleeritud. Olenevalt sellest, kas sordiaretajad rõhutavad rohkem genotüüpide stabiilsust (näiteks on oodata väga halbu ilmastikutingimusi või kasvutingimusi), siis tasub rohkem rõhku panna kolmandas grupis olevatele statistikutele (näiteks ökovalents, $S_i^{(2)}$, WAASB, ASV). Kui aretajad rõhutavad üleüldist saagikust (näiteks on oodata häid ilmastikutingimusi), siis tasub vaadata esimese grupi statistikute poole (paremuse määra, keskmine saagikus). Teises grupis on statistikud, mis mõõdavad mõlemat aspekti, seega nende statistikute järgi valitud genotüübid on võrdlemisi stabiilsed ja ka hea saagikusega (näiteks on oodata keskpäraseid kasvutingimusi või ilmastikutingimusi ei ole võimalik ette ennustada)

5.3 Statistike agregeerimine

Kui eesmärgiks on kombineerida infot kõigi statistikute peale kokku ning valida üldiselt häid genotüüpe, mida edasi aretada, siis tasuks otsida genotüüpide grupe, mis on kõigi statistikute järgi heal kohal. Samas ei saa klasterdamist viia läbi tervel astakutemaatriksil, kuna korreleeritud astakud hakkavad oma mõju tunnuste suure arvu tõttu suurendama. Autor pakkub välja idee, kus igast grupist võetakse ainult üks statistik, sest grupi statistikud on omavahel tugevalt korreleeritud. Statistikeid $N_i^{(2)}$, $N_i^{(3)}$, $N_i^{(4)}$ ei tohiks kasutada, kuna nende astakud on vastuolus kõigi teiste statistikute poolt saadud tulemustega ning statistikut b_1 tuleks käsitleda kui puhtalt stabiilsust mõõtvat statistikut.

Et võtta kokku informatsiooni erinevate statistikute kohta, kasutati erinevaid klasterdamismeetodeid. Kõigist statistikute hulgast võeti alamhulkasid, kus igast grupist võeti üks statistik, ehk igas alamgrupis oli kolm statistikut ja viidi läbi hierarhiline klasterdamine alamgrupil statistikutest. Kuna klasterdamine põhineb kaugustel, siis astakute asemel kasutati statistikute väärtuseid ning väärtused normaliseeriti lahutades statistiku väärtusest selle keskmine ja jagades see statistiku haardega. Klasterdamisel kasutati Manhattani kaugust. Meetodina kasutati “Complete linkage”-i ning klastrite arvuks valiti 10. Klasterdamine viidi läbi kasutades käsku `daisy()` paketist “cluster” (Maechler et al. 2022).

Klasterdamiseks tekitati järgnevad statistikute grupid:

- 1.grupp : WAASB, GSI, Y
- 2.grupp : ASV, S3, Super_a
- 3.grupp : Ökovalents, WAASBY50, Y
- 4.grupp : Eber S2di, S6, Super_a
- 5.grupp : NP1, WAASBY50, Y

Klasterdamine viidi läbi nii statistikute astakuid kasutades, kui ka statistikute väärtuseid kasutades. Astakute kasutamisel olid vaatlused lähestikku ja selgelt eristuvaid klastreid ei tekkinud, kuid ka tunnuse teisendatud algväärtuseid kasutades ei leitud klasterdamisel andmetel selgelt eristuvaid grupe. Näiteks keskmine silueti väärtus (mille definitsioon on ära toodud (Rousseuw 1987),lk.57), mis näitab üleüldist klasterdamisel tekkinud struktuuri headust, jäi kõigil juhtudel 0,2 ja 0,35 vahele, mis näitab üldiselt halba struktuuri klasterdamisel. Sellele vaatamata prooviti klasterdamise tulemuste järgi valida genotüüpe, mis on nii stabiilsuse ja saagikuse poole pealt teistest paremad. Selle uurimiseks vaadati kõikide genotüüpide klastritesse kuuluvusi, ning loeti kokku, mitu korda iga genotüüp sattus klastrisse, mis oli soovitud statistikute väärtusega, ehk millel stabiilsust ja keskmist saagikust näitavad statistikud olid kõige soovitud väärtustega.

Igal klasterdamisel vaadati, milline grupp on statistikute poolest kõige sobivam. See tähendab, et keskmise saagikuse näitajatel näiteks kõige suurem, ökovalentsi väärtuse juures aga kõige väiksem. Juhul, kui mitu erinevat klastrit olid parimad (näiteks üks oli parim stabiilsuse poole pealt, aga teine oli parim keskmise saagikuse poole pealt), siis loeti mõlemasse klastrisse sattunud genotüübid “võitjateks”. Lõpuks loeti kokku, mitu korda iga genotüüp “võitis”. Lisades 1 kuni 5 on ära toodud kõigi viie klasterdamise näited kahel statistikul nendest gruppidest. Näiteks 3. grupi “võitis” klaster 7, kuna sellel on keskmine saagikus kõige kõrgem ja ökovalentsi väärtus kõige madalam.

Sellise klasterdamise tulemusel “võitsid” 22 genotüüpi kõigil viiel korral. Kõigil viiel korral võitsid järgnevad genotüübid: 71, 159, 169, 203, 216, 218, 221, 223, 224, 231, 234, 243, 244, 247, 260, 270, 272, 276, 291, 293, 294, 300.

Kõik välja toodud genotüübid on nii stabiilsuse, kui ka saagikuse poole pealt teistest tunduvalt kõrgemal. Sõltuvalt sellest, kui palju parimaid genotüüpe on uuringu eesmärgiks välja selgitada, võib klastrite arvu suurendada või vähendada. Klastrite arvu suurenemisel väheneb “võitjate” arv ja klastrite arvu vähendamisel suureneb “võitjate” arv.

Teise agregeerimismeetodina prooviti võtta keskmine astak gruppide kaupa, kus grupid on defineeritud, nagu statistikute peakomponentanalüüsi peatükis. Esimene grupp mõõdab genotüüpide keskmist saagikust. Teine sisaldab statistikuid, mis mõõdavad korruga nii saagikust, kui ka genotüübi stabiilsust. Kolmas grupp sisaldab stabiilsust mõõtvaid statistikuid. Kolmandasse gruppi on ka lisatud statistik “Eber b1”, kuna peakomponentanalüüsis leiti, et see võib mõõta stabiilsust mingist muust aspektist ning see ei ole negatiivses korrelatsioonis teiste stabiilsust mõõtvate statistikutega. Lisaks arvutati keskmine astak üle kõigi kolme gruppi kuulavate statistikute astakute, et selgitada eelistatum genotüüp.

Tabelis 23 on ära toodud 20 madalaima koguastaku saanud genotüüpi. Lisaks on ära toodud tunnuses “Klastrite arv”, kui mitmesse parimasse klastrisse see genotüüp eelneva klasterdamisega sattus. Viimased tunnused näitavad keskmist

astakut vastavalt gruppides 1, 2, 3 ja kõigi statistikute peale kokku.

Kui vaadata kõiki astakuid keskmistava tunnuse järgi, siis genotüübid 293, 228 on selgelt teistest genotüüpidest ees. Genotüübil 293 on kõigist genotüüpidest kõrgeim saagikus. Järgnevad genotüübid on 143, 154, 226 ja 86.

Vaadates tabelit 23 näeme, et leidub genotüüpe, mis on kõrge keskmise saagikusega ja madalama stabiilsusega (näitena genotüübid 293, 260, 243, 244), kõrge stabiilsusega ja madalama keskmise saagikusega (näitena genotüübid 143, 86, 97, 182) või genotüüpe, mis on võrdlemisi head mõlemas (näitena genotüübid 228, 226, 86, 97).

Tabel 23: Agreeerivate statistikute väärtused genotüüpidel esimesel 20-l genotüübil keskmise astaku järgi

	Klastrite arv	Saakkokku	Keskkokku	Stabkokku	Kokkukokku
293	5	1.000000	10.000	70.55556	27.18519
228	4	36.000000	14.625	41.11111	30.57870
143	4	70.000000	15.375	29.55556	38.31019
154	3	34.000000	31.875	53.00000	39.62500
226	4	54.000000	23.875	41.27778	39.71759
86	4	63.333333	17.375	40.66667	40.45833
97	4	67.333333	21.750	34.11111	41.06481
260	5	6.000000	23.375	102.88889	44.08796
291	5	10.666667	15.625	110.05556	45.44907
243	5	3.333333	7.125	126.94444	45.80093
244	5	15.333333	35.750	93.33333	48.13889
98	4	34.000000	36.250	74.22222	48.15741
159	5	11.666667	41.875	109.88889	54.47685
216	5	28.000000	32.250	103.33333	54.52778
182	4	113.333333	26.125	35.22222	58.22685
201	4	21.333333	53.625	99.77778	58.24537
144	4	79.333333	40.375	55.66667	58.45833
81	4	91.000000	47.250	37.66667	58.63889
276	5	50.333333	36.250	92.55556	59.71296
292	4	17.000000	43.375	118.77778	59.71759

Kokkuvõte

Rahvusvahelise NOBALwheat projekti raames viidi 2021.–2023. aastal läbi ulatuslik suvinisu uuring Eestis, Lätis, Leedus ja Norras, mille raames rajati põldkatsed 300 erineva genotüübiga nende omaduste ja tunnuste hindamiseks. Peatähelepanu oli saagikuse ja selle stabiilsuse hindamisel muutuvates ilmastikuoludes. Projekti peamiseks eesmärgiks oli leida algmaterjaliks sobivaid genotüüpe erinevates ilmastikutingimustes kõrgema ja stabiilsema saagiga suvinisu sortide aretamiseks. Jätakuuutliku nisutootmise tarbeks regioonis on vajalik kliimamuutustega paremini kohastunud sortide aretamine.

Saagikuse ja stabiilsuse hindamiseks kasutatakse palju erinevaid statistikuid, millest osad on mõeldud mõõtma ainult stabiilsust ning osad nii keskmist saagikust kui ka stabiilsust. Uurimistöösse valiti 19 sobivat statistikut. Vastavalt töö eesmärgile koostati ülevaated nende põhjal arvatud tulemuste interpreteerimiseks. Analüüsi tulemusena grupeeriti statistikud ja anti soovitused nende valikuks sõltuvalt aretuse eesmärkidest. Stabiilsemate genotüüpide väljaselgitamiseks on sobivamad ökovalents (w), Hühn-i statistik $S^{(2)}$, absoluutsete skooride kaalutud keskmine

Tabel 24: Parimad genotüübid sõltuvalt sordiaretmise eesmärgist

Saagikas genotüüp	Nimi	Stabiilne ja saagikas	Nimi	Stabiilne	Nimi
293	Polkka	228	DS-720-3-DH	143	GN03503
260	T9040	154	GN13618	182	Serenada
243	DS-627-10-DH	226	DS-674-7-DH	86	Manu

(WAASB), AMMI stabiilsusvärtus (ASV). Saagikamate genotüüpide leidmiseks sobivad Paremuse määär (P) ja keskmine saagikus. Hea saagi ja hea saagi stabiilsuse tasakaaluga genotüüpide valikuks on otstarbekas kasutada näiteks Hühn-i statistikuid $S^{(3)}$ ja $S^{(6)}$.

Analüüsi tulemusena reastati kõik genotüübid 19 erineva statistiku järgi paremuse järjekorda. Statistike vahel tehti ka lihtsamat korrelatsioon- ja peakomponentanalüüsi, et uurida, kuidas erinevad statistikud käituvad ning milliseid neist võiks stabiilsuse ja saagikuse hindamisel kasutada. Kahjuks põhjalikum statistiline analüüs siia töösse keerukuse ja statistikute rohkuse tõttu ei mahtunud. Tehtud analüüsid näitasid, et suur osa statistikuid mõõdavad stabiilsust sarnaselt. Lisaks leidis statistikuid, peamiselt Thennarasu statistikud, mis mõõtsid stabiilsust tunduvalt teistmoodi kui teised. Tugeva negatiivse korrelatsiooni tõttu keskmise saagikusega ja ka teiste stabiilsusstatistikutega on nende kasutamise otstarbekus kaheldav. Antud töös jäeti need analüüsist välja, kasutamise soovitude andmiseks vajaks need täiendavat uurimist. Erinevate statistikute põhjal parima genotüübi valimiseks prooviti kasutada nii klasterdamist, kui ka lihtsamaid agregeerivaid meetodeid. Kuna klasterdamise tulemusel ei tekkinud selgelt eristuvaid gruppe vaatlustel, siis jäädlihtsamate agregeerivate meetodite juurde ja vaadati keskmiseid astakuid erinevates huvi pakkuvates aspektides. Antud uuringu põhjal saab järeldada, et üldise saagikuse-stabiilsuse analüüsi läbiviimiseks ei ole vaja välja arvutada paljude erinevate statistikute väärtusi, sest enamus ühtivad ning mõõdavad stabiilsust sarnaselt.

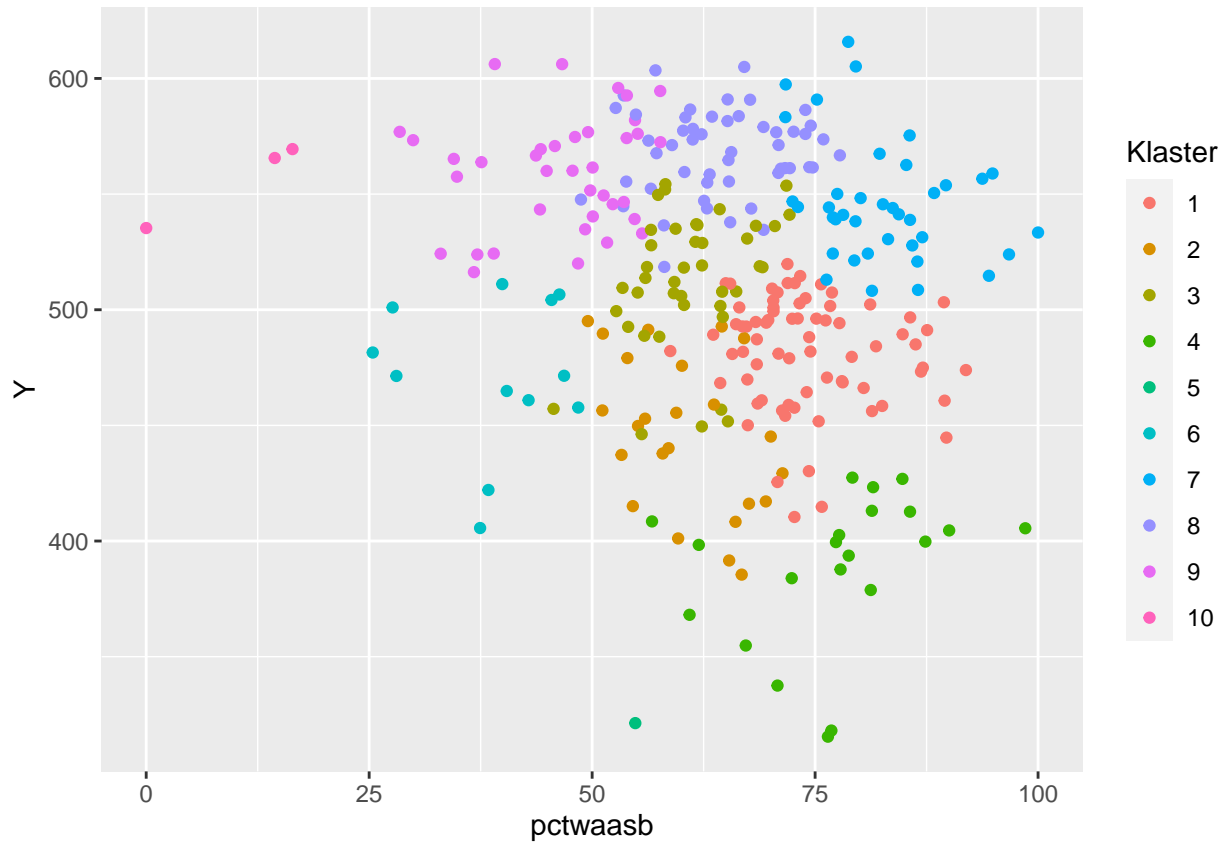
Töö teiseks eesmärgiks seatud saagikamate ja parema saagi stabiilsusega genotüüpide väljaselgitamiseks koostati järjestus kõigist genotüüpidest. Kõige paremad saagikuse, stabiilsus ja saagikuse/stabiilsuse hea tasakaaluga genotüübid on välja toodud tabelis 24. Selle töö edasi arenduseks on mitu erinevat suunda. Esiteks, ei ole siia töösse kaasatud kõiki stabiilsust mõõtvaid statistikuid, mida autor uuris. Nii pakettides “metan”, kui ka “agricolae” on kirjeldatud veel stabiilsusstatistikuid, mida saaks kasutada stabiilsuse hindamiseks. Näiteks Mohammadi and Amri (2008) on kirjeldanud geomeetrilist kohanemisindeksit. Teiseks, võiks statistikute vahel teha täpsemaid statistilisi analüüse, kui vaid astakute vahelist korrelatsioonanalüüsi, et täpsemini seletada statistikute vahelist struktuuri ja välistada statistikute dubleerimist.

Kasutatud Kirjandus

- Alwala, S., T. Kwolek, M. McPherson, J. Pellow, and D. Meyer. 2010. "A Comprehensive Comparison Between Eberhart and Russell Joint Regression and GGE Biplot Analyses to Identify Stable and High Yielding Maize Hybrids." *Field Crops Research* 119 (2): 225–30. <https://doi.org/10.1016/j.fcr.2010.07.010>.
- Annicchiarico, P. 1997. "Joint Regression Vs AMMI Analysis of Genotype-Environment Interactions for Cereals in Italy." *Euphytica* 94 (1): 53–62. <https://doi.org/10.1023/A:1002954824178>.
- Becker, H. C., and J. Léon. 1988. "Stability Analysis in Plant Breeding." *Plant Breeding* 101 (1): 1–23. <https://doi.org/10.1111/j.1439-0523.1988.tb00261.x>.
- Bocianowski, Jan, Janetta Niemann, and Kamila Nowosad. 2018. "Genotype-by-Environment Interaction for Seed Quality Traits in Interspecific Cross-Derived Brassica Lines Using Additive Main Effects and Multiplicative Interaction Model." *Euphytica*. <https://doi.org/10.1007/s10681-018-2328-7>.
- Bose, Lotan Kumar, Nitiprasad Namdeorao Jambhulkar, Kanailal Pande, and Onkar Nath Singh. 2014. "Use of AMMI and Other Stability Statistics in the Simultaneous Selection of Rice Genotypes for Yield and Stability Under Direct-Seeded Conditions." *CHILEAN JOURNAL OF AGRICULTURAL RESEARCH* 74 (1): 3–9. <https://doi.org/doi:10.4067/S0718-58392014000100001>.
- Bose, Nitiprasad Namdeorao AND Pande, Lotan Kumar AND Jambhulkar. 2014. "Use of AMMI and other stability statistics in the simultaneous selection of rice genotypes for yield and stability under direct-seeded conditions." *Chilean Journal of Agricultural Research* 74 (March): 3–9. http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-58392014000100001&nrm=iso.
- de Mendiburu, Felipe. 2021. *Agricolae: Statistical Procedures for Agricultural Research*. <https://CRAN.R-project.org/package=agricolae>.
- Eberhart, S. A., and W. A. Russell. 1966. "Stability Parameters for Comparing Varieties1." *Crop Science* 6 (1): [cropsci1966.0011183X000600010011x](https://doi.org/10.2135/cropsci1966.0011183X000600010011x). <https://doi.org/10.2135/cropsci1966.0011183X000600010011x>.
- Hühn, M., and R. Nassar. 1989. "On Tests of Significance for Nonparametric Measures of Phenotypic Stability." *Biometrics* 45 (3): 997–1000. <http://www.jstor.org/stable/2531698>.
- Lin, C. S., and M. R. Binns. 1988. "A SUPERIORITY MEASURE OF CULTIVAR PERFORMANCE FOR CULTIVAR × LOCATION DATA." *Canadian Journal of Plant Science* 68 (1): 193–98. <https://doi.org/10.4141/cjps88-018>.
- Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2022. *Cluster: Cluster Analysis Basics and Extensions*. <https://CRAN.R-project.org/package=cluster>.
- Mohammadi, Reza, and Ahmed Amri. 2008. "Comparison of Parametric and Non-Parametric Methods for Selecting Stable and Adapted Durum Wheat Genotypes in Variable Environments." *Euphytica* 159 (3): 419–32. <https://doi.org/10.1007/s10681-007-9600-6>.
- "NOBALwheat – Breeding Toolbox for Sustainable Food System of the NOrdic BALtic Region | EEA Grants." n.d. Accessed January 15, 2024. <https://eeagrants.org/archive/2014-2021/projects/LT-RESEARCH-0001>.
- Olivoto, Tiago, and Alessandro Dal'Col Lúcio. 2020. "Metan: An r Package for Multi-Environment Trial Analysis."

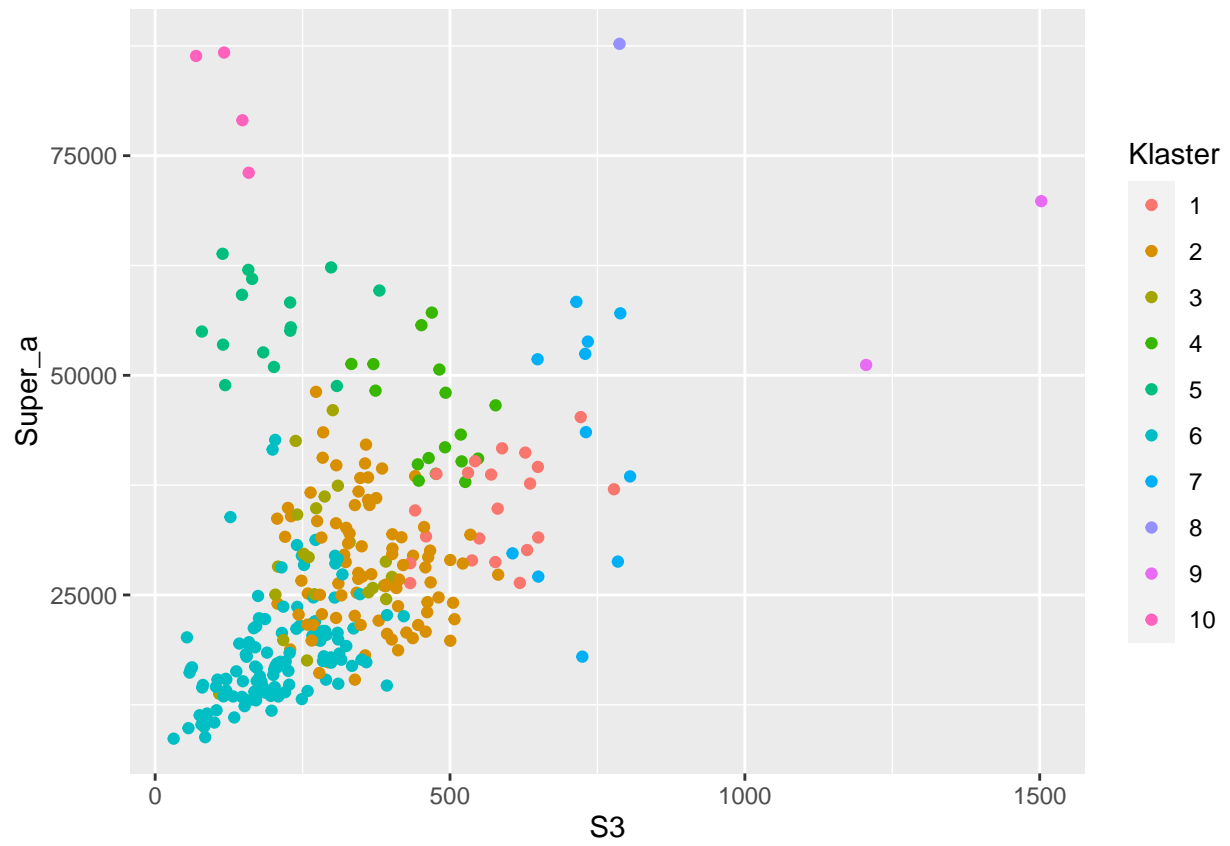
- Methods in Ecology and Evolution* 11 (6): 783–89. <https://doi.org/10.1111/2041-210X.13384>.
- Olivoto, Tiago, Alessandro Lúcio, José Silva, Volmir Marchioro, Velci Souza, and Evandro Jost. 2019. “Mean Performance and Stability in Multi-Environment Trials i: Combining Features of AMMI and BLUP Techniques.” *Agronomy Journal* 111 (July): 2951–56. <https://doi.org/10.2134/agronj2019.03.0220>.
- Purchase, J. L., Hesta Hatting, and C. S. van Deventer. 2000. “Genotype \times Environment Interaction of Winter Wheat (*Triticum Aestivum* L.) in South Africa: II. Stability Analysis of Yield Performance.” *South African Journal of Plant and Soil* 17 (3): 101–7. <https://doi.org/10.1080/02571862.2000.10634878>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rousseeuw, Peter J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Journal of Computational and Applied Mathematics* 20: 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).
- Sabaghnia, Naser, Hamid Dehghani, and Syyed Hossain Sabaghpour. 2006. “Nonparametric Methods for Interpreting Genotype \times Environment Interaction of Lentil Genotypes.” *Crop Science* 46 (3): 1100–1106. <https://doi.org/10.2135/cropsci2005.06-0122>.
- Shukla, G. K. 1972. “Some Statistical Aspects of Partitioning Genotype-Environmental Components of Variability.” *Heredity* 29 (2): 237–45. <https://doi.org/10.1038/hdy.1972.87>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wodebo, Kibreab Yosefe, Taye Tolemariam, Solomon Demeke, Weyessa Garedew, Tessema Tesfaye, Muluken Zeleke, Deribe Gemiyu, Worku Bedeke, Jane Wamatu, and Mamta Sharma. 2023. “AMMI and GGE Biplot Analyses for Mega-Environment Identification and Selection of Some High-Yielding Oat (*Avena Sativa* L.) Genotypes for Multiple Environments.” *Plants* 12 (17). <https://doi.org/10.3390/plants12173064>.
- Zobel, Richard, Madison Wright, and Hugh Gauch. 1988. “Statistical Analysis of a Yield Trial.” *Agronomy Journal - AGRON J* 80 (May). <https://doi.org/10.2134/agronj1988.00021962008000030002x>.

Lisa 1



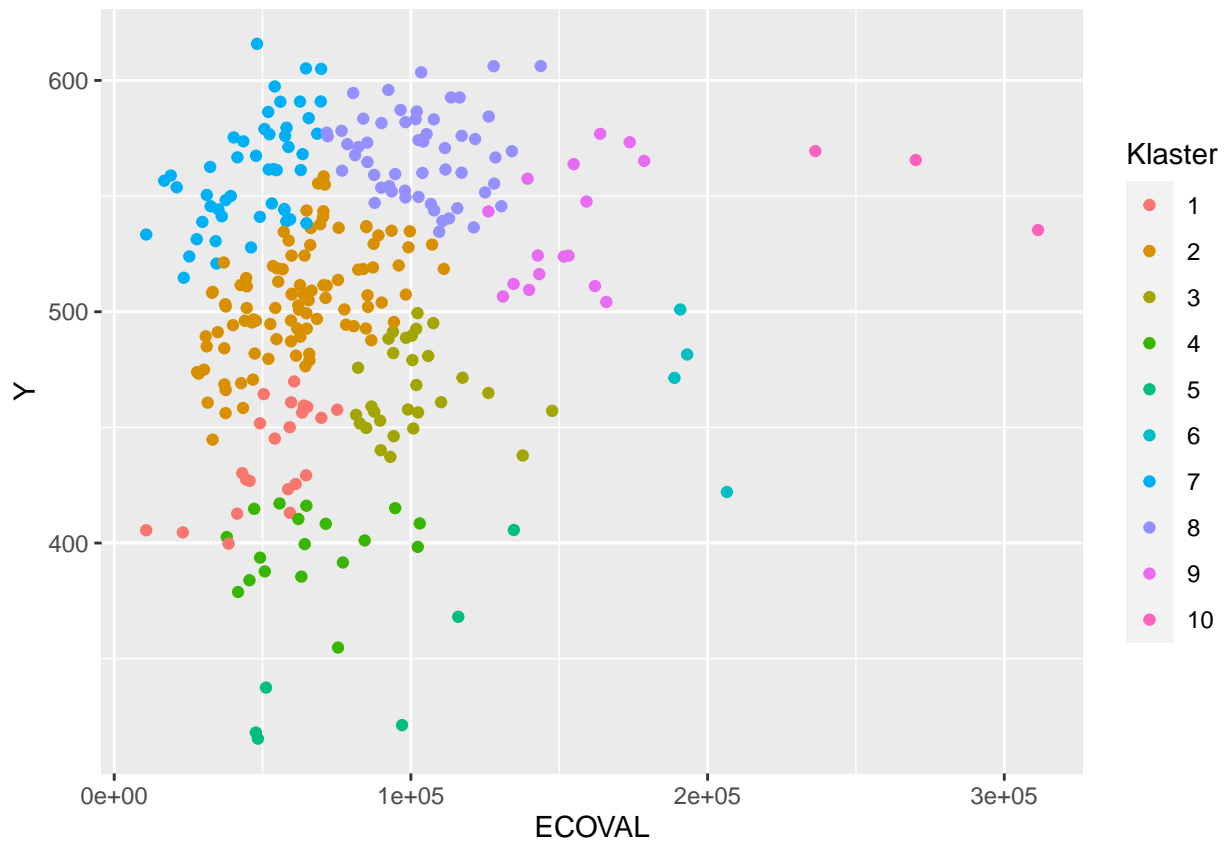
Joonis 12: 1.gruppi statistikute järgi saadud klastrid

Lisa 2



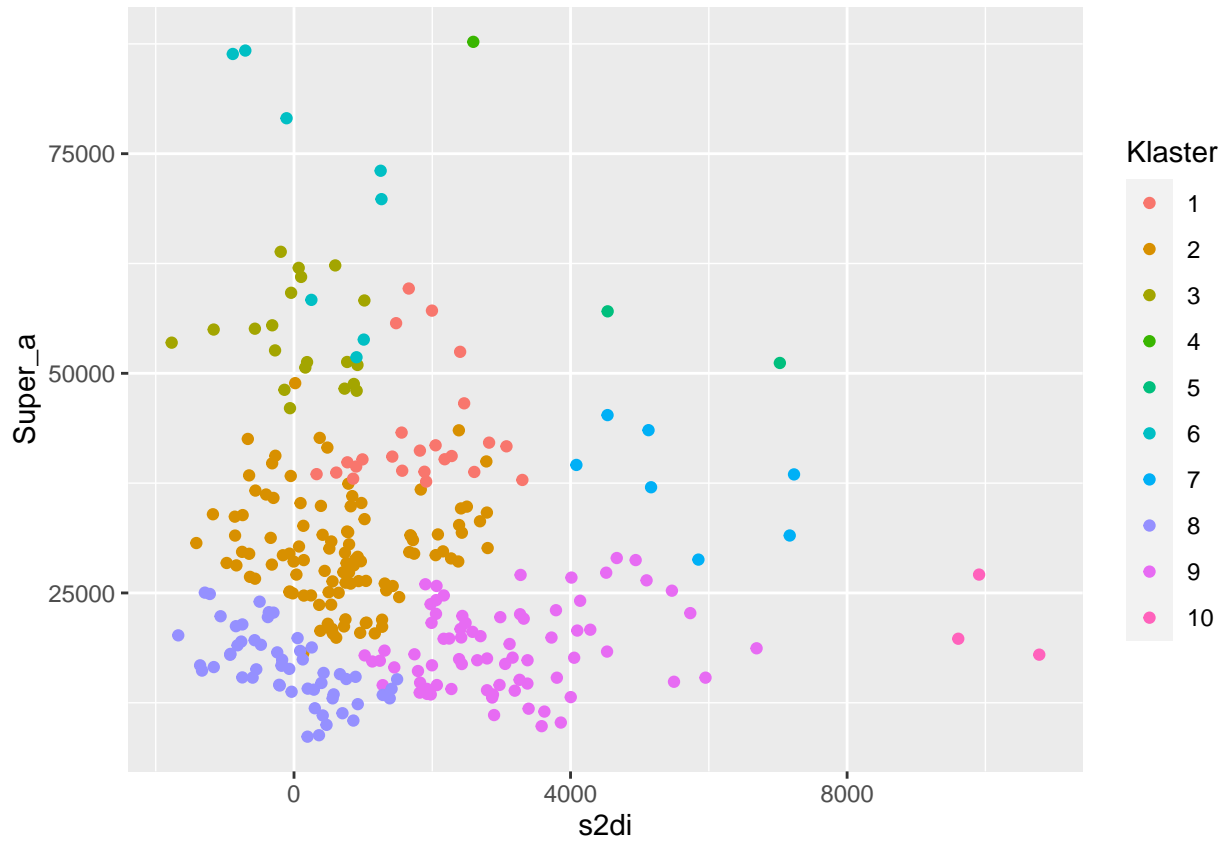
Joonis 13: 2.gruppi statistikute järgi saadud klastrid

Lisa 3



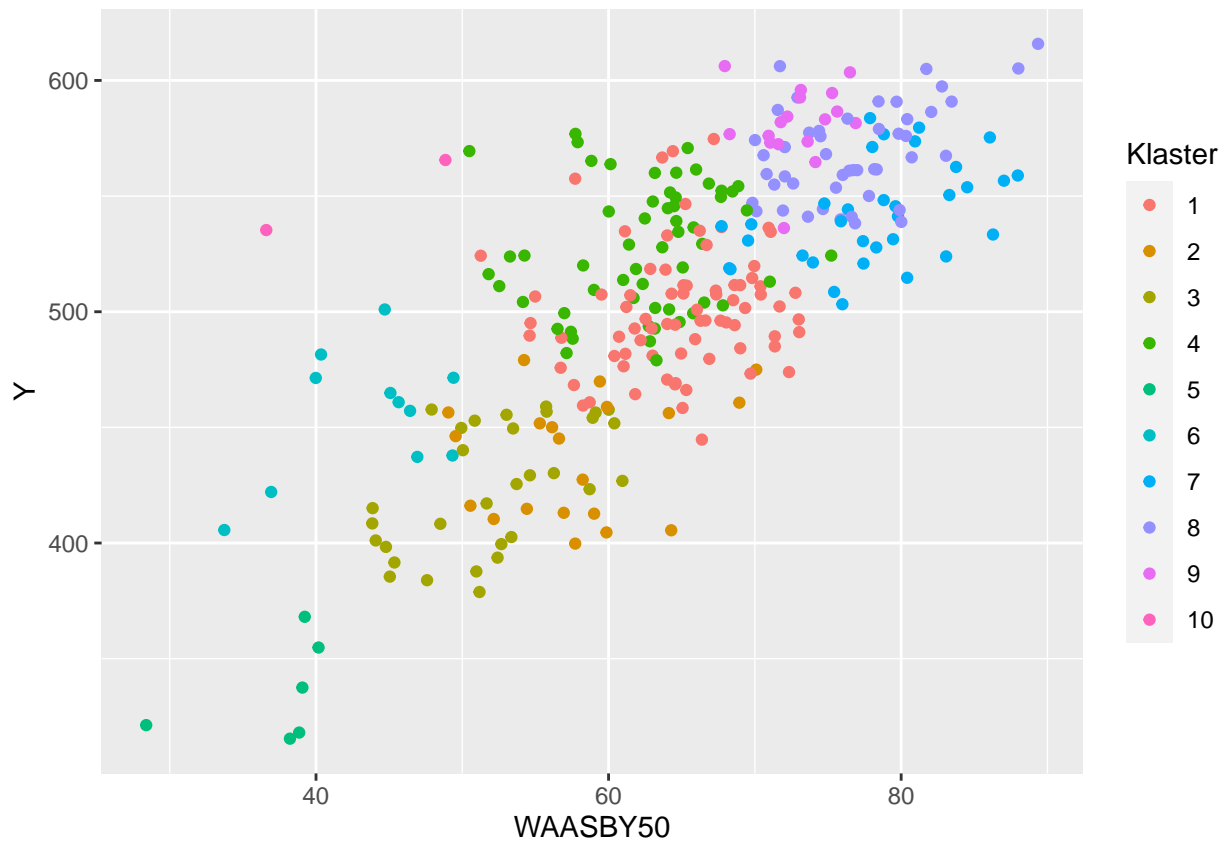
Joonis 14: 3.gruppi statistikute järgi saadud klastrid

Lisa 4



Joonis 15: 4.gruppi statistikute järgi saadud klastrid

Lisa 5



Joonis 16: 5.gruppi statistikute järgi saadud klastrid

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Joonas-Sander Tamm,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose "Genotüübi ja keskkonna vahelist koosmõju hindavate meetodite kirjeldus ja võrdlus NOBALwheat projekti andmete näitel", mille juhendajad on Tanel Kaart ja Ilmar Tamm, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Joonas-Sander Tamm

22.05.2024