

# Kausaalimalli viljapellon tuottavuudelle

Tilastotieteen pro gradu -tutkielma

Timo Pohjonen

Jyväskylän yliopisto  
Matematiikan ja tilastotieteen laitos  
Toukokuu 2022

## Tiivistelmä

Maanviljelyssä päätöksenteko on oleellinen osa liiketoimintaa. Päätöksiä tehdään muun muassa kylvöön, kasvinsuojeluun ja sadonkorjuuseen liittyvistä asioista. Näillä voidaan vaikuttaa tulevan sadon määrään ja laatuun sekä kustannuksiin. Tekeillä hyviä valintoja saadaan laadukas sato mahdollisimman pienin kustannuksin. Päätöksenteossa otetaan huomioon sekä taustatekijöitä että aiemmin tehtyjä päätöksiä, minkä seurauksena voidaankin ajatella, että näiden välillä on kausaalisuhte.

Tässä tutkielmassa tarkastellaan maanviljelyn tuottavuutta ja keskitytään kevätiljoihin. Tutkimusmenetelmänä käytetään kausaalimallinnusta ja tuottavuutta mitataan saadun sadon laadun ja määrän sekä kustannuksien avulla. Varsinainen kausaaligraafi muodostetaan kasvinviljelyn asiantuntijan avulla käyttäen menetelystapana haastattelututkimusta. Kausaalimallin avulla voidaan tehdä identifioituvuustarkasteluja kiinnostuksen kohteena oleville kausaalivaikutuksille, joita taas voitaisiin edelleen hyödyntää kausaalivaikutusten estimoinnissa.

Saatujen tulosten perusteella kausaalimallinnuksen hyödyntäminen maanviljelyn tuottavuuden arvioinnissa on mahdollista. Jokaisen päätösmuuttujan vaikutus saatuun satoon ja kustannuksiin saatiin identifioitua vähintäänkin ottamalla useampi muuttuja mukaan interventioon. Tutkielman jatkon kannalta seuraavat askeleet olisivat aineiston kerääminen ja kausaalivaikutusten estimointi.

---

**Avainsanat:** Kausaalimalli, kausaalivaikutusten identifiointi, kausaalivaikutusten estimointi, maanviljely, viljasato, tutkimushaastattelu

# Sisällys

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Kausaalipäätelyn vaiheet ja menetelmät</b>	<b>2</b>
2.1	Karsimisen hyödyntäminen ennen ID-algoritmia . . . . .	6
2.2	ID-algoritmin palauttaman tuloksen sieventäminen . . . . .	8
2.3	Kausaalivaikutusten estimointi . . . . .	10
<b>3</b>	<b>Tutkimushaastattelun toteutus</b>	<b>11</b>
<b>4</b>	<b>Kausaalimalli</b>	<b>13</b>
4.1	Muuttujien kuvailu . . . . .	15
4.2	Muuttujien väliset kausaalisuhteet . . . . .	19
<b>5</b>	<b>Kausaalivaikutusten identifiointi</b>	<b>22</b>
5.1	Päätösmuuttujien vaikutus satoon . . . . .	23
5.2	Päätösmuuttujien vaikutus kustannuksiin . . . . .	27
5.3	Päätösmuuttujien vaikutus sadon ja kustannusten yhteisjakaumaan .	27
<b>6</b>	<b>Mahdollisia datalähteitä jatkotutkimuksia varten</b>	<b>29</b>
<b>7</b>	<b>Pohdinta</b>	<b>30</b>



# 1 Johdanto

Kausaalimallinnuksessa pääpaino on asioiden välisten syy-seuraussuhteiden selvittäminen. Voidaan sanoa esimerkiksi, että lajikevalinta vaikuttaa saadun sadon määrään. Graafit ovat osa kausaalimallinnusta ja näihin kerätään kaikki ilmiöön keskeisesti liittyvät asiat. Kausaaligraafi muistuttaa ulkonäöltään ajatuskarttaa, mutta tässä viivat on korvattu nuolilla. Nuolten lähtöpää kertoo vaikuttavan asian ja loppupää vastaavasti sen, mihin vaikutetaan.

Kausaalimallinnus perustuu aineistoon ja asiantuntijatietoon tutkimusongelmasta. Saatavilla olevan aineiston määrä ei kuitenkaan rajoita kausaalimallinnusta niin paljoa kuin perinteisiä tilastollisia menetelmiä. Päinvastoin, sillä kausaalimallinnusta voidaan käyttää vaikutusten estimoinnin lisäksi myös tiedonkeruun suunnitteluun. Yleensä ei ole rajoittamattomasti resursseja hankkia aineistoa, jolloin haluttaisiin edetä mahdollisimman kustannustehokkaasti. Kuten luvusta 5.1 voidaan huomata, niin kaikista aiheeseen liittyvistä asioista ei tarvitse yksittäisen kausaalivaikutuksen laskemiseksi kerätä aineistoa. Tässä työssä esitetyt kausaalimallinnuksen perusteet ovat lähtöisin teoksista (Pearl, 2009) ja (Pearl, Glymour, & Jewell, 2016). Nämä teokset antavat kokonaiskuvan kausaalimallinnuksesta ja kertovat, miten kausaalimallinnus eroaa perinteisistä tilastollisista menetelmistä.

Maanviljely on monivaiheinen prosessi, johon liittyy vuosittain useita päätöksiä, joita maanviljelijä joutuu tekemään. Näitä ovat erityisesti kylvöön, kasvinsuojeluun ja sadonkorjuuseen liittyvät valinnat. Osa tehtävistä päätöksistä on kertaluontoisia, kuten lajikevalinta ja viljelytekniikka. Näiden lisäksi on myös pitkällä aikavälillä tehtäviä päätöksiä, kuten kasvinsuojeluun ja lisälannoitukseen liittyvät toimet. Näitä voidaan tehdä tarvittaessa useampaan kertaan ja eri aikoina.

Suomessa maanviljelyä tutkii muun muassa luonnonvarakeskus Luke. Suomessa viljellään pääasiassa ohraa, kauraa, vehnää ja ruista. Näistä suurin osa päätyy rehuksi ja vain pieni päätyy juomiksi tai viljatuotteiksi (Partala, Sieviläinen, & Mäitälä, 2022). Huonot satovuodet ja maailmanlaajuiset kriisit voivat vaikuttaa maatalojen kannattavuuteen merkittävästi. Esimerkiksi jos polttoöljyn tuonti Venäjältä loppuisi, niin tällöin kustannukset tulisivat nousemaan (Niemi & Huuskonen, 2022). Kuitenkin yhteiskunnan tukitoimet ja tuottajanhintojen nousu avittavat kannattavuuden säilymistä, vaikka energian ja lannoitteiden hinnat nousevat (Luke, 2022b).

Kuten yritystoiminnassa yleensäkin, myös maanviljelyssä pyritään mahdollisimman hyvään tuottavuuteen. Kiinnostuksena olisikin selvittää, voitaisiinko päätöksentekoa parantaa kausaalipäätelyllä ja tehdä sitä kautta maanviljelystä kannattavampaa. Kausaalimallinnuksen soveltaminen maanviljelyyn on sinänsä kiinnostavaa, sillä tätä ei ole tiedettävästi ennen tutkittu. Täten tutkimusongelmaa lähdetään tar-

kastelemaan uudesta näkökulmasta ja tämä voi tuottaa kiinnostavaa tietoa ja mahdollistaa laajentamisen muihin sovelluksiin. Tässä tutkielmassa rajoitutaan tarkastelemaan vain kevätiljoja, mutta ideaa voisi soveltaa myös syysviljoille sekä muille viljelykasveille. Tällöin ainoa olennainen muutos olisi kausaaligraafiin säätäminen tilanteeseen sopivaksi.

Tutkielmaan tarvittiin maanviljelyn asiantuntija, joka osaisi kertoa maanviljelyn perusteista ja erityisesti viljakasvien viljelystä. Tätä varten tutkielmaan pyydettiin mukaan Anssi Mennala, joka omaa pitkän taustan maanviljelystä. Hän omistaa kasvinviljelyyn erikoistuneen perhetilan Jämsässä ja pyrkii kehittämään toimintaansa aktiivisesti (Mennala, 2022). Tutkimusongelmaa lähdetään tarkastelemaan yksityisyrittäjän näkökulmasta perustuen Mennalan näkemyksiin.

Tässä tutkielmassa esitetään tutkimusongelmaan liittyvä kausaalimalli. Tätä varten kerättiin tietoa haastattelemalla Mennalaa. Häneltä haastateltiin muun muassa aiheeseen liittyvistä tekijöistä, niiden välisistä vaikutuksista sekä tilalla kerätyistä aineistosta. Koska tilalla ei ollut kerättynä valmista aineistoa, niin pohdittiin, miten aineistoa voitaisiin kerätä ja mitä tämä käytännössä vaatisi. Tämän lisäksi mietittiin, mitä muita aineistoja tutkielmassa voitaisiin hyödyntää. Näitä voisivat olla esimerkiksi säätilastot ja viljojen hintatilastot. Tämän työn tavoitteena on luoda kausaalimalli viljapellon tuottavuudelle ja miettiä, minkälaista aineistoa kausaalivaikutusten estimointiin vaatisi. Kausaalivaikutusten estimointi ei kuulunut tämän työn tavoitteisiin.

Tutkielman aluksi esitellään kausaalipäätelyn vaiheita ja käytettäviä menetelmiä. Tämän jälkeen kerrotaan, kuinka tutkimushaastattelu toteutettiin ja lyhyesti tutkimuksen lähtökohdista. Kappaleessa 4 esitellään varsinaiseen tutkimuskysymykseen liittyvä kausaaligraafi ja kerrotaan sen sisältämistä muuttujista. Sen jälkeen tehdään identifioituvuustarkasteluja päätösmuuttujien vaikutuksista satoon ja kustannuksiin. Kappaleessa 6 kerrotaan, mitä aineistoja olisi saatavilla ja miten niitä voitaisiin käyttää jatkotutkimuksissa. Lopuksi pohditaan saatuja tuloksia sekä mietitään, miten tätä tutkielmaa voitaisiin jatkaa.

## 2 Kausaalipäätelyn vaiheet ja menetelmät

Tässä luvussa kerrotaan teoriaa kausaalipäätelyn vaiheista ja menetelmistä. Teoria pohjautuu pääosin teoksiin (Pearl, 2009), (Pearl et al., 2016) ja (Koller & Friedman, 2009). Ensiksi käsitellään graafiteoriaa sekä rakenneyhtälöitä, joiden avulla määritellään kausaalimalli. Tämän jälkeen puhutaan kausaalivaikutuksista sekä niiden identifioinnista. Lopuksi esitellään vielä teoriaa kausaalivaikutusten estimointiseksi.

Muuttujia merkitään isoilla kirjaimilla ( $X_i$ ), näiden saamia arvoja pienillä kirjaimilla ( $x_i$ ) ja joukkoja isoilla lihavoiduilla kirjaimilla ( $\mathbf{X} = \{X_1, \dots, X_n\}$ ).

Kausaalipäätely mahdollistaa asioiden välisten syy-seuraussuhteiden mallintamisen. Maanviljelyssä voitaisiin esimerkiksi olla kiinnostuneita lajikevalinnan ( $X$ ) vaikutuksesta saatuun satoon ( $Y$ ). Kausaalipäätely mahdollistaa tämän estimoinnin ja saatua tulosta voidaan käyttää hyväksi päätöksenteossa. Jos kysymykseen lähdetäisiin vastaamaan pelkän aineiston pohjalta, niin tehdyt päätelmät voisivat olla virheellisiä. Tämä johtuu siitä, että vahva korrelaatio kahden muuttujan välillä ei tarkoita kausaalisuutta (Pearl et al., 2016). Korrelaatio voi nimittäin olla seurausta esimerkiksi latenteista sekoittavista tekijöistä tai valintaharhasta. Latenteilla tekijöillä tarkoitetaan sellaisia muuttujia, joiden arvoja ei havaita. Sovelluksesta riippuen latenteja muuttujia voi olla suuriakin määriä ja tutkijan vastuulle jääkin löytää näistä oleellimmat. Esimerkiksi otettaessa sekoittava tekijä  $Z$  mukaan tarkasteluun, voidaan saada päinvastainen tulos muuttujan  $X$  vaikutuksesta muuttujaan  $Y$  kuin alun perin. Tällainen ongelma tunnetaan yleisemmin nimellä Simpsonin paradoksi (Simpson, 1951). Aineiston lisäksi tarvitaankin siis tietämystä asioiden välisistä suhteista, jos halutaan päästä oikeisiin johtopäätöksiin (Pearl et al., 2016).

Ensimmäinen vaihe kausaalipäätelyssä on suunnatun asyklisen graafin (Directed acyclic graph, DAG) piirtäminen. Tämä koostuu muuttujista  $X_i$  ( $i=1, \dots, n$ ) ja niiden välisistä funktionaalisista riippuvuuksista. Graafissa muuttujia kutsutaan solmuiksi ja näiden välisiä suhteita joko särmiksi tai nuoliksi. Graafi on suunnattu, jos jokaisella nuolella on suunta ja asyklinen, jos se ei sisällä silmukkaa. Silmukalla tarkoitetaan solmujonoa, jossa lähtösolmusta päästään takaisin lähtösolmuun kulkemalla ainoastaan nuolten suuntaisesti. Mikäli graafi on suunnattu ja asyklinen, niin sitä kutsutaan suunnatuksi asykliseksi graafiksi (Pearl, 2009). DAG:n avulla voidaan tehdä päätelmiä muuttujien välisistä riippuvuussuhteista ja sitä voidaan käyttää myös kausaalivaikutusten identifioituvuuden tarkasteluun, josta puhutaan myöhemmin (Pearl et al., 2016).

Formaalisti kausaaligraafi esitetään parina  $G = \langle \mathbf{V}, \mathbf{E} \rangle$ , missä  $\mathbf{V}$  sisältää kaikki solmut ja  $\mathbf{E}$  sisältää näiden väliset nuolet. Tarkemmin  $\mathbf{E}$  sisältää listan pareja  $(X, Y)$ ,  $X, Y \subseteq \mathbf{V}$ ,  $X \neq Y$ , missä  $X$  kuvaa nuolen lähtösolmua ja  $Y$  päätesolmua.

Vaihtoehtoinen tapa kuvata kausaaligraafin sisältämä informaatio on määrittellä muuttujien väliset funktionaaliset riippuvuudet, joita kutsutaan rakenneyhtälöiksi. Näiden ja kausaaligraafin (DAG) avulla voidaan muodostaa varsinainen kausaalimalli. Merkitään kausaaligraafin muuttujia  $\mathbf{X} = (X_1, \dots, X_n)$ . Tällöin rakenneyhtälö kullekin muuttujalle  $X \subseteq \mathbf{X}$  kir-

jojetaan yleensä muodossa

$$x = f(pa(x)), \text{ missä}$$

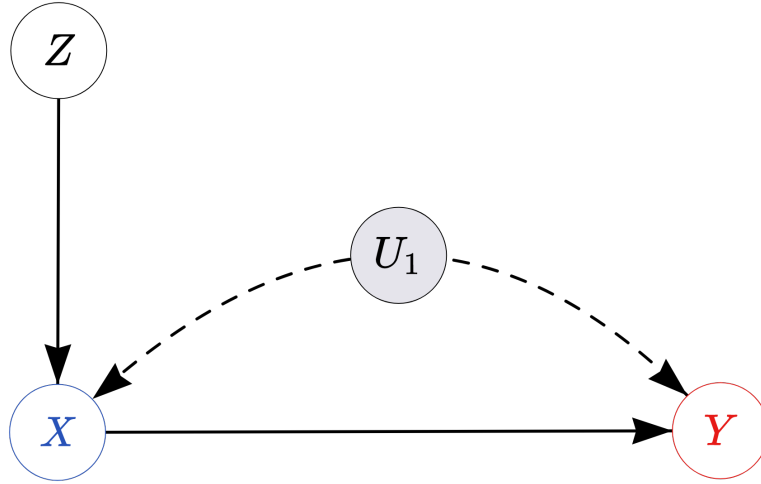
$pa(x) \subseteq \mathbf{X} \setminus X$  tarkoittaa muuttujan  $X$  vanhempia. Muuttujan  $X$  vanhemmiksi kutsutaan niitä solmuja kausaaligraafissa  $G$ , joista lähtee nuoli muuttujaan  $X$ . Lisäksi funktio  $f$  voi olla eri jokaiselle muuttujalle. Tällaisessa muodossa esitetyt rakenneyhtälöt sisältävät täysin saman informaation kuin kausaaligraafi. (Pearl, 2009).

Seuraava vaihe kausaalipäätelyssä on määrittellä kiinnostuksen kohteena olevat kausaalivaikutukset ja tarkastella näiden identifioituvuutta, kun käytössä on muuttujien yhteisjakauma. Identifioituvuutta pyritään tässä tutkimaan epäparametrisesti ilman oikeaa aineistoa, jakaumaoletuksia ja oletuksia rakenneyhtälöiden funktionaalista muodoista. Muuttujan  $X$  kausaalivaikutusta muuttujaan  $Y$  merkitään ehdollisena jakaumana  $P(Y|\text{do}(X))$ . Tässä tehdään siis muuttujalle  $X$  interventio (merkitään  $\text{do}(\cdot)$ ), jossa tutkija kiinnittää muuttujalle  $X$  arvon  $x$ . Kausaaligraafissa tämä poistaa kaikki muuttujaan  $X$  tulevat nuolet, jolloin ajatellaan, että interventiomuuttujan arvoon eivät muut muuttujat vaikuta. Lisäksi rakenneyhtälöissä muuttujan  $X$  paikalle sijoitetaan arvo  $x$ . Jotta kausaalivaikutuksia voidaan estimoida, niin niiden täytyy olla identifioituvia. Identifioinnin voivat estää havaitsemattomat tekijät, joita merkitään graafissa kaksisuuntaisella katkoviivalla kahden solmun välillä. Jos kausaalivaikutus ei identifioitu, niin tarvitaan joko vahvempia oletuksia tai aineistoa havaitsemattomasta tekijästä. Interventio voidaan tehdä myös muuttujajoukolle ja tässä tapauksessa toimitaan samalla tavalla kuin yhden muuttujan tapauksessa (Pearl, 2009).

Identifioituvuutta voidaan tarkastella graafisesti joissain erikoistapauksissa ja annetaan tästä seuraavaksi esimerkki. Olkoon kiinnostuksena identifioida kausaalivaikutus  $P(Y|\text{do}(X))$  kuvion 1 mukaisessa kausaaligraafissa. Kuviossa 1 taustamuuttuja  $Z$  vaikuttaa interventiomuuttujaan  $X$ , joka puolestaan vaikuttaa vastemuuttujaan  $Y$ . Lisäksi muuttujien  $X$  ja  $Y$  välillä on latentti tekijä  $U_1$ . Nyt kausaalivaikutus ei ole identifioituva, sillä mikä vain osa muuttujan  $X$  havaitusta vaikutuksesta muuttujaan  $Y$  voi johtua sekoittavasta latentista tekijästä  $U_1$ . Poikkeuksen edelliseen tekevät lineaariset mallit, joissa kausaalivaikutus saataisiin identifioitua käyttäen muuttujaa  $Z$  instrumenttimuuttujana (Pearl, 2009). Muiden mallien tapauksessa kyseinen kausaalivaikutus ei identifioitu, vaikka kuvion 1 kausaaligraafi olisi osa isompaa kausaaligraafia. (Pearl, 2009).

Yleisesti identifioitumista tarkastellaan erilaisten päätelysääntöjen, todennäköisyyslaskennan sekä algoritmien avulla. Näissä pyritään esittämään kausaalivaikutus havaittujen jakaumien avulla, jolloin lopullisessa lausekkeessa ei olisi enää interven-





Kuvio 1: Esimerkki kausaaligraafista

tioita jäljellä. Päättelysäännöissä tämä tapahtuu tutkimalla muuttujien riippuvuus-rakenteita alkuperäisen graafin aligraafeissa ja tämän avulla lisätään tai poistetaan interventioita/muuttujia kausaalivaikutuksesta. Lisäksi käytetään apuna todennäköisyyden laskusääntöjä, kuten laajennusta summaksi, ehdollistamista sekä faktointia (Pearl, 2009). Algoritmisesti identifioituvuutta voidaan tutkia ID-algoritmin (Shpitser & Pearl, 2006) avulla. Tämä on rekursiivinen algoritmi, joka palauttaa vain havaittuja jakaumia sisältävän lausekkeen tai tiedon, että kausaalivaikutus ei ole identifioituva.

Yleensä yksilötason kausaalivaikutuksia ei pystytä estimoimaan, jolloin estimointi toteutetaan populaatiotasolla. Tällöin lasketaan keskimääräinen kausaalivaikutus (Average treatment effect, ATE). Nyt käsittelyn  $x_1$  keskimääräinen kausaalivaikutus muuttujaan  $Y$  verrattuna käsittelyyn  $x_0$  lasketaan kaavalla (Pearl et al., 2016)

$$T = E(Y|do(X = x_1)) - E(Y|do(X = x_0)).$$

Tässä tutkielmassa käytetään menetelmien soveltamiseen R-ohjelmointikieltä (R Core Team, 2021) ja tälle tehtyjä paketteja `igraph` (Csardi & Nepusz, 2006) ja `causaleffect` (Tikka & Karvanen, 2017a) hyödynnetään identifioituvuustarkasteluissa. Näiden lisäksi hyödynnetään kausaaligraafin visualisoinnissa `yEd`-editoria (yWorks GmbH).

Kun halutaan identifoida kausaalivaikutus  $P(\mathbf{Y}|do(\mathbf{X}))$  ID-algoritmeilla (Shpitser & Pearl, 2006), annetaan sille syötteenä muuttujajoukot  $\mathbf{X}$  ja  $\mathbf{Y}$ , kausaaligraafi  $G = \langle \mathbf{V}, \mathbf{E} \rangle$  ja yhteisjakauma  $P(\mathbf{V})$ . Lisäksi tarvitaan jokin solmujen välinen topologinen järjestys  $\pi$ , joita yleensä on useita. Tällä tarkoitetaan järjestystä, jossa solmun esivanhempi esiintyy aina ennen jälkeläistään. Tämä vastaa topologisen jär-

jestyksen määritelmää: Jos  $X$  on  $Y$ :n esivanhempi graafissa  $G$ , niin  $X < Y$  topologi-  
sessa järjestyksessä  $\pi$ . Huomattavaa on, että  $P(\mathbf{V})$  sisältää havaittujen muuttujien  
yhteisjakauman ja ID-algoritmia ei voidakaan käyttää tilanteessa, jossa muuttujien  
yhteisjakaumaa ei ole saatavilla.

ID-algoritmi on sinänsä yksinkertainen, sillä se palauttaa jokaisella iteraatiokier-  
roksella vain yhden asian. Se joko kutsuu itseään uudelleen tai palauttaa identifioin-  
nin tuloksen. Mahdolliset palautetut tulokset ovat vain havaittuja jakaumia sisältä-  
vä lauseke tai tieto ei-identifioituvuudesta. ID-algoritmi on toteutettu causaleffect  
paketin funktiossa causaleffect. Tämä toteutus tarvitsee syötteenä edellä mainitut  
joukot  $\mathbf{X}$  ja  $\mathbf{Y}$  sekä paketin `igraph` avulla luodun graafiojektin.

ID-algoritmin palauttama lauseke on kuitenkin usein monimutkainen ja tällöin  
haluttaisiinkin löytää ekvivalentti yksinkertaisempi lauseke. Ensinnäkin tämä hel-  
pottaa saadun lausekkeen ymmärtämistä, mutta siitä voi olla muitakin hyötyjä, ku-  
ten esimerkiksi laskennallisia hyötyjä estimoinnissa ja puuttuvan tiedon tapauksessa.  
Tätä varten Karvanen ja Tikka ovat kehittäneet menetelmiä, jotka on implementoi-  
tu funktioon `causaleffect` parametreilla `prune` (Tikka & Karvanen, 2017b) ja `simplify`  
(Tikka & Karvanen, 2017c).

## 2.1 Karsimisen hyödyntäminen ennen ID-algoritmia

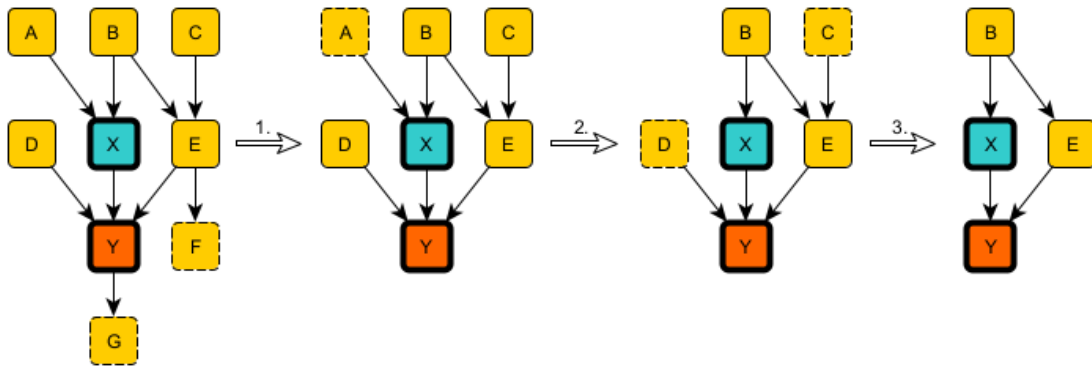
Seuraavat kappaleet perustuvat artikkeliin (Tikka & Karvanen, 2017b). Karsimisessa  
(pruning) poistetaan graafista tietyt solmut ja identifiointi tehdään tästä saadun  
aligraafin avulla. Kausaalivaikutuksen  $P(\mathbf{Y}|\text{do}(\mathbf{X}))$  tapauksessa yritetään poistaa  
graafin solmuja (paitsi solmuja  $\mathbf{Y}$  ja  $\mathbf{X}$ ) seuraavalla algoritmilla:

1. Poistetaan kaikki solmut, jotka eivät ole solmujoukon  $\mathbf{Y}$  esivanhempia.
2. Poistetaan solmujoukon  $\mathbf{X}$  esivanhemmat  $\mathbf{Z}$ , joista on yhteys solmujoukkoon  
 $\mathbf{Y}$  vain solmujoukon  $\mathbf{X}$  kautta ehdolla, että graafin latentti projektio on sama  
kuin sen aligraafi, josta on poistettu solmut  $\mathbf{Z}$ .
3. Poistetaan solmut, joista on yhteys graafin muihin solmuihin vain yhden sol-  
mun kautta.

Karsintaan kuuluu näiden lisäksi vielä neljäs askel, josta kerrotaan myöhemmin. En-  
simmäinen askel on peräisin teoksesta (Shpitser & Pearl, 2006) ja se onkin toteutettu  
jo ID-algoritmissa. Askeleet 2-3 ovat sen sijaan uusia ja peräisin artikkelista (Tikka  
& Karvanen, 2017b). Toinen askel on hieman monimutkaisempi ja siihen tarvitta-  
van latentin projektion määritelmä kerrotaan myöhemmin ja sitä avataan esimer-  
kin kautta. Kuviossa 2 on esitetty askeleiden 1-3 havainnollistus, kun identifioidaan

kausaalivaikutusta  $P(Y|\text{do}(X))$ . Vasemmassa laidassa on alkuperäinen graafi, josta lähdetään karsimaan solmuja pois. Sinisellä värillä on kuvattu interventiomuuttujaa, punaisella vastemuuttujaa ja keltaisella muita muuttujia. Lisäksi katkoviivalla on kehystetty muuttujia, jotka poistetaan kyseisessä askeleessa. Nuolien päällä olevat numerot vastaavat edellä määriteltyjen askeleiden numerointia.

Kuviossa 2 ensimmäinen askel poisti solmut F ja G. Tämä seuraa siitä, että näistä muuttujista ei pääse kulkemalla nuolien suuntaisesti vastemuuttujaan. Toinen askel vastaavasti poisti solmun A. Tässä tapauksessa  $L=\{A\}$  ja latentti projektio ei lisännyt särmiä jäljelle jääneiden solmujen välille. Tämä johtuu siitä, että solmun A kautta ei kulkenut polkuja, joka on pääehto särmien lisäykselle. Polulla tarkoitetaan solmujonoa, jossa kukin solmu esiintyy enintään kerran ja nämä on yhdistetty särmillä. Kolmas askel poisti solmut C ja D. Tämä johtuu siitä, että näihin muuttujiin tulee/lähtee vain yksi nuoli.



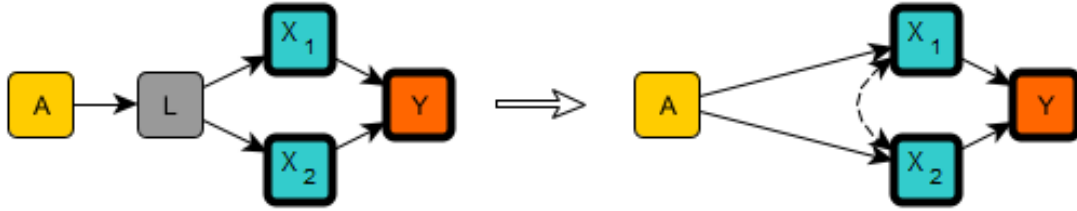
Kuvio 2: Esimerkki karsinnan käytöstä kausaalivaikutukselle  $P(Y|\text{do}(X))$

Latentin projektion muodostaminen on osa toista askelta ja se voidaan tiivistää seuraavasti: Poistetaan ensiksi graafista  $G$  tietyt solmut, jonka jälkeen lisätään tarvittaessa yksi- tai kaksisuuntaisia särmiä jäljelle jääneiden solmujen välille. Seuraavaksi annetaan formaali esitys latentille projektioille.

Olkoon  $G = \langle \mathbf{V} \cup \mathbf{L}, \mathbf{E} \rangle$  DAG, missä solmut joukossa  $\mathbf{V}$  ovat havaittuja ja joukossa  $\mathbf{L}$  latentteja. Latentti projektio  $L(G, \mathbf{V})$  on DAG  $G = \langle \mathbf{V}, \mathbf{E}_L \rangle$ , missä jokaiselle parille  $Z, W \subseteq \mathbf{V}$  pätee:

1.  $L(G, \mathbf{V})$  sisältää särmän  $Z \rightarrow W$ , jos  $G$  sisältää polun  $Z \rightarrow \dots \rightarrow W$ , missä muut solmut paitsi  $Z$  ja  $W$  kuuluvat joukkoon  $\mathbf{L}$
2.  $L(G, \mathbf{V})$  sisältää särmän  $Z \leftrightarrow W$ , jos  $G$  sisältää polun  $Z \leftarrow \dots \rightarrow W$ , missä muut solmut paitsi  $Z$  ja  $W$  kuuluvat joukkoon  $\mathbf{L}$  ja lisäksi polku ei sisällä käänteistä haarukkaa ( $A \rightarrow B \leftarrow C$ ).

Kuviossa 3 on havainnollistettu latentin projektion toimintaa. Siinä olevista soluista ainoastaan L on latentti ja muut havaittuja. Vasemmalla puolella on alkuperäinen graafi G ja oikealla siitä saatu latentti projektio. Graafi G sisältää polut  $A \rightarrow L \rightarrow X_1$  ja  $A \rightarrow L \rightarrow X_2$ , joten ensimmäisen kohdan perusteella lisätään latenttiin projektioon nuolet  $A \rightarrow X_1$  ja  $A \rightarrow X_2$ . Lisäksi G sisältää polun  $X_1 \leftarrow L \rightarrow X_2$ , joten latenttiin projektioon lisätään myös kaksisuuntainen särmä  $X_1 \leftrightarrow X_2$ .



Kuvio 3: Esimerkki latentista projektioista

Karsimis-algoritmin toisessa askeleessa ehtona on, että  $G[V \setminus Z] = L(G, V \setminus Z)$ , missä Z kuvaa poistettavia solmuja. Tämä tarkoittaa, että latentissa projektiossa täytyisi olla samat särmät kuin graafissa G, josta on poistettu solmut Z.

Karsintaan kuuluu vielä neljäs askel, jossa edellisistä askeleista poiketen ei poisteta solmuja, vaan muutetaan ne latenteiksi, jonka jälkeen identifiointi voidaan tehdä latentin projektion avulla. Latentiksi solmuksi voidaan muuttaa mikä vain solmujoukon  $\mathbf{Y}$  esivanhemmista, lukuun ottamatta solmuja  $\{X, \mathbf{Y}\}$ , kun seuraavat ehdot täyttyvät:  $\mathbf{X} = \{X\}$  saa sisältää vain yhden solmun ja solmun X ja sen lapsien välillä ei saa olla kaksisuuntaista särmää. Karsiminen voidaan toteuttaa myös rekursiivisesti osana ID-algoritmia, jolloin voidaan saada lisähyötyä. Funktion causaleffect parametri prune toteuttaakin tämän. Karsiminen tapahtuu lisäksi pääosin ennen ID-algoritmin soveltamista.

## 2.2 ID-algoritmin palauttaman tuloksen sieventäminen

Seuraavat kappaleet perustuvat teokseen (Tikka & Karvanen, 2017b). ID-algoritmin palauttamaa tulosta voidaan yksinkertaistaa karsimisen lisäksi hyödyntämällä sievennystä. Näiden erona on, että karsiminen tehdään ennen ID-algoritmia tai sen aikana, kun taas sievennys tehdään vasta ID-algoritmin palauttamalle tulokselle. Sievennyksen ja karsinnan yhdistelmästä voidaan joissakin tapauksissa saada lisähyötyä pelkkään karsintaan verrattuna. Samoin kuin karsinta, myös sievennys perustuu kausaaligraafin sisältämän riippuvuusrakenteen hyödyntämiseen.

Karvanen & Tikka esittivät sievennysalgoritmeja erilaisiin tilanteisiin. Nämä eroavat toisistaan siten, että niissä sievennettävä lauseke on eri muodossa. Ensimmä-

mäinen sieventämisalgoritmi ottaa syötteen ID-algoritmin palauttaman todennäköisyyslausekkeen, alkuperäisen kausaaligraafin  $G$  ja muuttujien välisen topologisen järjestyksen  $\pi$ . Lisäksi todennäköisyyslausekkeen täytyy olla muotoa

$$\sum_{\mathbf{S}} \prod_{i=1}^n P(V_i | \mathbf{C}_i) = \sum_{\mathbf{S}} P(V_1 | \mathbf{C}_1) \dots P(V_n | \mathbf{C}_n), \quad (1)$$

missä  $\mathbf{S} \subseteq \{V_1, \dots, V_n\}$  ja  $\{V_1, \dots, V_n, \mathbf{C}_1, \dots, \mathbf{C}_n\}$  ovat graafin  $G$  solmujoukkoja. Lisäksi  $V_i \notin \mathbf{C}_i$  ja  $V_i \neq V_j$  kaikilla  $i \neq j$ , Vastaavasti jatkuvien muuttujien tapauksessa summat voidaan korvata integraaleilla. Lausekkeen 1 täytyy myös olla topologisessa järjestyksessä, eli muuttujan  $V_i$  esivanhemmat  $(An(V_i)_G) \subseteq \mathbf{C}_i \subseteq V_i^\pi$  kaikilla  $i = 1, \dots, n$ . Merkinnällä  $V_i^\pi$  tarkoitetaan niitä graafin  $G$  solmuja, jotka ovat pienempiä kuin solmu  $V_i$  topologisessa järjestyksessä  $\pi$ . Tähän joukkoon eivät kuulu solmun  $V_i$  lapset, sillä nämä ovat suurempia topologisessa järjestyksessä kuin solmu  $V_i$ . Tämä seuraa, että  $\mathbf{C}_i$  sisältää kaikki muuttujan  $V_i$  esivanhemmat ja mahdollisesti myös muita muuttujia, mutta ei kuitenkaan muuttujan  $V_i$  lapsia.

Toimintaperiaatteena ensimmäisessä algoritmossa on eliminoida muuttujia, joiden yli summataan. Kaavassa 1 näitä muuttujia merkitään joukolla  $\mathbf{S}$ . Algoritmisesti sievennys tapahtuu rekursiivisesti käymällä yksi kerrallaan joukon  $\mathbf{S}$  muuttujia läpi ja näille tehdään seuraavat askeleet:

1. Etsitään kyseistä muuttujaa vastaavat sievennysjoukot (Tikka & Karvanen, 2017c), joiden avulla sievennettävä lauseke saadaan kirjoitettua muuttujien ehdollisen yhteisjakauman avulla.
2. Palautetaan saatu yhteisjakauma takaisin tulomuotoon, missä kyseinen muuttuja on poistettu ja asetetaan tämä uudeksi sievennettäväksi lausekkeeksi.

Aina ei kuitenkaan löydetä sopivia sievennysjoukkoja, joilla saataisiin askeleessa 1 muodostettua yhteisjakauma. Tällöin siirrytään seuraavaan muuttujaan joukossa  $\mathbf{S}$  ja yritetään uudelleen. Muuttujaa  $V_i$  vastaavia sievennysjoukkoja etsittäessä tarvitsee tarkastella lausekkeesta vain termejä, jotka esiintyvät lausekkeessa ennen termiä  $P(V_i | \mathbf{C}_i)$  (mukaan lukien termi itse), sillä termit ovat topologisessa järjestyksessä ja muista termeistä ei ole hyötyä tämän sievennyksessä. Tarkemmat yksityiskohdat muuttujaa  $V_j$  vastaavista sievennysjoukoista löytyvät teoksesta (Tikka & Karvanen, 2017c) sivulta 6.

Tämän lisäksi Karvanen & Tikka esittelivät myös algoritmeja, joilla pystytään sieventämään sisäkkäisiä kaavan 1 mukaisia lausekkeitä sekä tällaisten lausekkeiden osamäärän. Näiden toimintaperiaate on samantyylinen kuin edellä esitetty, sillä

näissä sieventäminen tehdään jokaiselle lausekkeen sisäkkäiselle termille erikseen ja lopuksi tulokset yhdistetään.

## 2.3 Kausaalivaikutusten estimointi

Tässä luvussa esitetään, kuinka kausaalivaikutuksia voidaan estimoida identifioituvien vaikutusten tapauksissa. Identifiointia varten ei tarvinnut tehdä vielä yksityiskohtaisia oletuksia muuttujista ja niiden jakaumista, mutta estimointia varten tarvitaan konkreettinen aineisto. Seuraavaksi esitetään eräs tapa, jolla kausaalivaikutuksia voidaan estimoida perustuen teokseen (Helske, Tikka, & Karvanen, 2021). Pääperiaatteena siinä esitetyssä estimoinnissa on määrätä jokaiselle estimoitavan lausekkeen termille jokin tilastollinen malli ja lopuksi yhdistää näistä saadut tulokset keskenään.

Tämä tapahtuu käyttämällä Bayes-laskentaa ja esimerkiksi MCMC-algoritmia. Tällöin poimitaan otoksia marginaali- ja ehdollisista jakaumista, joita esiintyy estimoitavassa lausekkeessa. Tämän avulla saadaan otos kiinnostavan vaikutuksen jakaumasta ja sen avulla voidaan estimoida esimerkiksi sen odotusarvoa.

Jos haluttaisiin estimoida esimerkiksi luvussa 5.1 esitetty peruslannoituksen ( $X_2$ ) vaikutus satoon ( $Y_1$ ) eli

$$P(Y_1|do(X_2)) = \sum_{Z_2, Z_1, X_1, X_4, X_3} P(Y_1|Z_2, Z_1, X_1, X_4, X_3, X_2)P(X_4|Z_2, Z_1, X_1)P(X_3|Z_2, Z_1, X_1, X_4)P(X_1|Z_2, Z_1)P(Z_1|Z_2)P(Z_2),$$

niin se voitaisiin tehdä seuraavalla algoritmilla, joka laskee N otosta kausaalivaikutuksesta  $P(Y_1|do(X_2))$ :

1. Jokaisella  $i=1, \dots, N$ :
2. Poimi  $Z_2^i \sim P_{\hat{\theta}}(Z_2)$
3. Poimi  $Z_1^i \sim P_{\hat{\theta}}(Z_1|Z_2^i)$
4. Poimi  $X_1^i \sim P_{\hat{\theta}}(X_1|Z_2^i, Z_1^i)$
5. Poimi  $X_4^i \sim P_{\hat{\theta}}(X_4|Z_2^i, Z_1^i, X_1^i)$
6. Poimi  $X_3^i \sim P_{\hat{\theta}}(X_3|Z_2^i, Z_1^i, X_1^i, X_4^i)$
7. Poimi  $Y_1^i \sim P_{\hat{\theta}}(Y_1|X_2, Z_2^i, Z_1^i, X_1^i, X_4^i, X_3^i)$ .

Näistä voidaan laskea odotusarvo  $\hat{E}_{\hat{\theta}}(Y_1|do(X_2)) = \frac{1}{N} \sum_{i=1}^N Y_1^i$ .

Edellä mainitussa menettelyssä tarvitaan kuitenkin tehdä useita jakaumaoletuksia ja termien lisääntyessä tämä on työlästä. Vaihtoehtoinen tapa tehdä estimointia on hyödyntää kovariaattien yhteisjakaumaa, jolloin tarvitaan jakaumaoletus vain vastemuuttujalle. Tämä menetelmä on esitetty teoksessa (Hernán & Robins, 2020). Edellinen kausaalivaikutus voidaan kirjoittaa ehdollisen todennäköisyyden kaavan avulla seuraavassa muodossa:

$$P(Y_1|do(X_2)) = \sum_{Z_2, Z_1, X_1, X_4, X_3} P(Y_1|X_2, Z_2, Z_1, X_1, X_4, X_3)P(X_3, Z_2, Z_1, X_1, X_4).$$

Tähän muotoon päästiin ensin yhdistämällä termit  $P(Z_2)$  ja  $P(Z_1|Z_2)$  niiden yhteisjakaumaksi  $P(Z_1, Z_2)$ , joka taas voitiin yhdistää seuraavan termin  $P(X_1|Z_2, Z_1)$  kanssa. Tätä jatkamalla saadaan lopulta kovariaattien yhteisjaukauma. Tästä voidaan taas estimoida keskimääräinen kausaalivaikutus (kun aineistossa on  $n$  riviä)

$$E(Y_1|do(X_2)) = \frac{1}{n} \sum_{i=1}^n E(Y_1|X_2, Z_2^i, Z_1^i, X_1^i, X_4^i, X_3^i).$$

Käytännössä tämän laskeminen tapahtuu sovittamalla ensin malli koko aineistolla vastemuuttujalle  $Y_1$ , jota selitetään kovariaateilla  $\{X_2, Z_2, Z_1, X_1, X_4, X_3\}$ . Tämän jälkeen tehdään aineistossa interventio muuttujaan  $X_2$  eli kiinnitetään tämän sarakkeen arvo aineistossa. Lopuksi ennustetaan tällä uudella aineistolla keskimääräinen kausaalivaikutus. Tässä luvussa esitetyt tulokset pätevät myös, jos interventio tai vastemuuttuja on moniulotteinen. Huomattavaa tässä jälkimmäisessä menetelmässä on, että aina ID-algoritmin palauttamaa lauseketta ei saada muokattua tämän vaatimaan muotoon, jolloin sitä ei voitaisi käyttää. Tässä työssä identifioidut yksittäisvaikutukset kuitenkin läpäisevät niin sanotun takaovikriteerin (Pearl et al., 2016), jolloin tätä ongelmaa ei synny. Mallintaminen tässä voisi tapahtua esimerkiksi yleistettyjen lineaaristen mallien avulla (McCullagh, 1989).

### 3 Tutkimushaastattelun toteutus

Tässä luvussa kerrotaan, miten tietoa on haettu tutkimusongelmasta. Tarkalleen ottaen haettu tieto keskittyy yleiskäsitykseen maanviljelystä. Tarkoituksena oli saada tutkielman tekijän käsitys riittävälle tasolle maanviljelystä. Tämä tapahtui ensin perehtymällä haastatteluteknikoihin ja työssä käytetystä haastattelumenetelmästä kerrotaan seuraavaksi.

Tutkimushaastattelussa pyritään saamaan tietoa tutkimusongelmasta. Perusrakenteena tässä toimii tyypillisesti kysymys-vastaus-kuittaustyyppinen toiminta. Alussa haastattelija esittää kysymyksen, jonka jälkeen haastateltava vastaa siihen avoimesti ja lopuksi haastattelija kiittää kuulleensa tai ymmärtäneensä vastauksen. Vastaukseen perustuen esitetään yleensä myös jatkokysymyksiä ennen seuraavaan kysymykseen siirtymistä valmiiksi laaditussa haastattelurungossa. Monien haastatteluoppaiden mukaan haastattelu kannattaa aloittaa laajemmilla kysymyksillä ja siirtyä haastattelun edetessä yksityiskohtaisempaan suuntaan (Ruusuvuori & Tiitula, 2005).

Haastattelun runko laadittiin edellä esitettyjä toimintatapoja noudattaen. Esi-merkkikysymys tähän oli, että miten syys- ja kevätviljojen viljely eroavat toisistaan ja miten kausaaligraafin tapauksessa tämä ilmaistaisiin. Vastauksena tähän saatiin, että huomioon pitäisi ainakin ottaa syysviljojen tapauksessa lisäksi syksyn sää. Lisäksi tässä tutkielmassa tehdyt haastattelut nauhoitettiin. Tällöin saatiin haastateluista kaikki tieto talteen ja mahdollistettiin tähän palaaminen. Ennen haastattelun nauhoittamista kysyttiin suostumus haastateltavalta tähän. Ensimmäinen haastattelu toteutettiin Jämsässä lokakuussa 2021. Tämä mahdollisti maatalaan tutustumisen sekä helpotti vuorovaikutusta. Loput tarkentavat haastattelut toteutettiin puhe-  
luiden avulla myöhäisempinä ajankohtina. Kuviossa 4 on paikan päältä otettu kuva eräästä Mennalan kynnetystä viljapelloista.



Kuvio 4: Syksyllä kynnetty viljapello (Lokakuu 2021)

Mennala ja Juha Karvanen (tilastotieteen professori Jyväskylän yliopistossa) ovat keskustelleet tutkimusongelmasta jo aiemmin syksyllä 2020 ja tehneet tällöin



alustavan kausaaligraafin. Ensimmäisen haastattelun päätavoitteena oli tarkentaa tätä kausaaligraafia ja antaa siinä oleville muuttujille konkreettinen esitys. Päätösmuuttujien kohdalla tavoitteena oli selvittää, mitä vaihtoehtoja päätöksenteossa on ja minkä perusteella päätökset tehdään. Muiden muuttujien kohdalla selvitettiin, että mitä asioita näihin sisältyy ja millä tavalla ne vaikuttavat. Lopuksi vielä keskusteltiin tilalla kerätystä aineistosta sekä siitä, että mitä aineistoa voitaisiin vielä kerätä ja mitkä olisivat tämän vaatimukset.

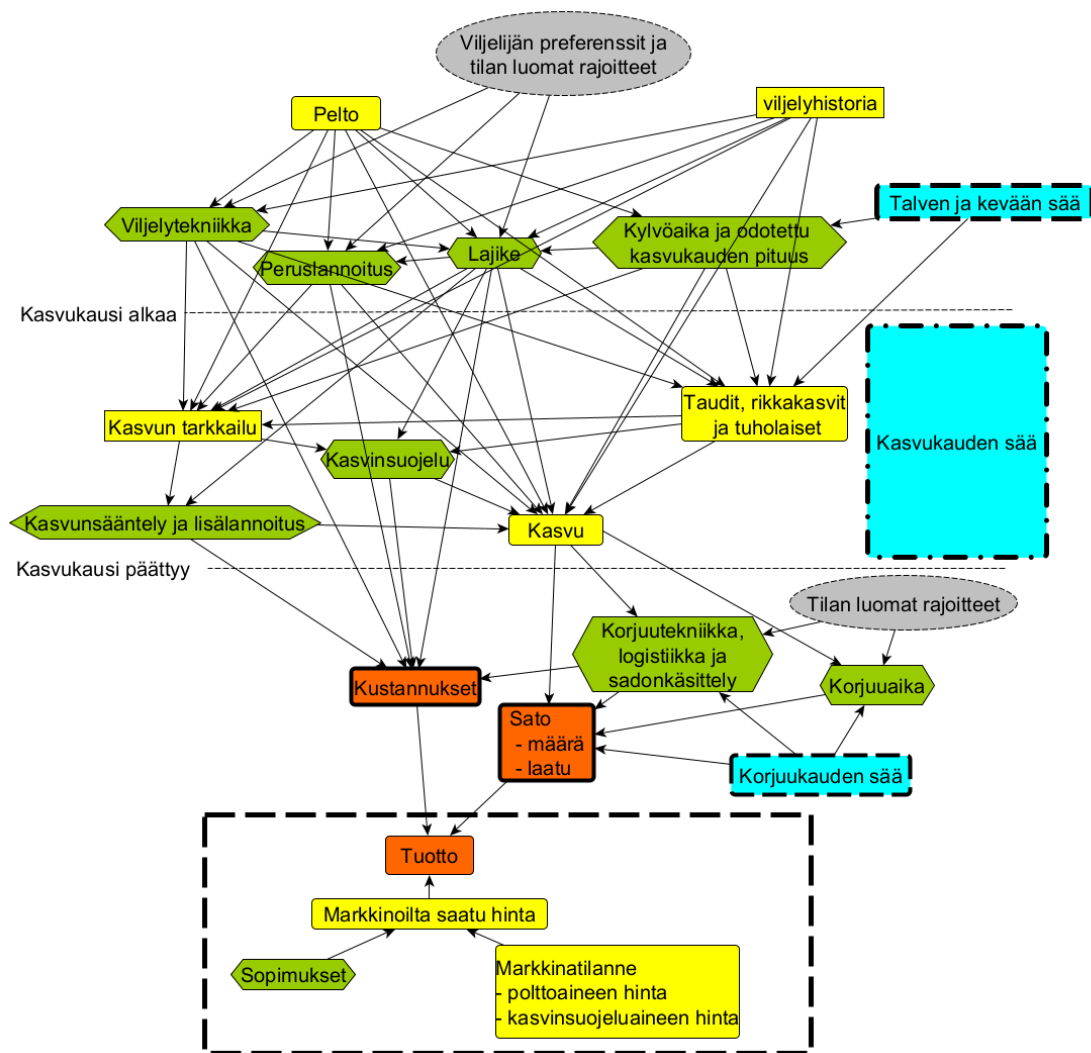
## 4 Kausaalimalli

Maanviljely koostuu hyvin monesta asiasta ja tässä luvussa luodaan kokonaiskuva tuottoon vaikuttavista tekijöistä. Luodussa kausaaligraafissa on yhdistetty tilanteeseen liittyviä tekijöitä kokonaisuuksiksi, jolloin se pystytään esittämään kompaktisti. Kuviossa 5 on esitettynä kausaaligraafi viljapellon tuottavuudelle. Siinä voidaan ajan ajatella juoksevan ylhäältä alhaalle päin. Vaakasuunnassa olevat katkoviivat jakavat viljelykauden kolmeen osaan: Ennen kasvukautta tapahtuvat asiat, kasvukaudella tapahtuvat ja kasvukauden jälkeen tapahtuvat.

Kuviossa 5 on värjätty vihreällä kiinnostuksen kohteena olevat päätösmuuttujat, keltaisella taustatekijät ja punaisella kannattavuutta mittaavat muuttujat. Lisäksi sinisellä on kuvattu säämuuttujia ja harmaalla latentteja tekijöitä. Säämuuttujat eroteltiin taustamuuttujista, koska haluttiin kokeilla, voitaisiinko ne joidenkin kausaalivaikutusten tapauksissa muuttaa latenteiksi. Tästä olisi hyötyä silloin, jos säätietoja ei olisi saatavilla. Kasvukauden sään ajatellaan vaikuttavan kaikkiin kasvukauden muuttujiin (ensimmäisen ja toisen katkoviivan välissä olevat muuttujat), mutta siitä on selvyuden takia jätetty piirtämättä nuolet. Kausaaligraafissa nuolet kertovat muuttujien välisistä vaikutussuhteista. Tässä on ajateltu kevätiljoja ja syysviljoille graafi on muuten sama, paitsi sinne lisättäisiin vielä syksyn sää.

Päätösmuuttujiin kohdistuu jonkinlainen päätös, jonka viljelijä joutuu tekemään. Päätösmuuttujiin viljelijä pystyy vaikuttamaan suoraan omien valintojensa kautta ja nämä ovat kiinnostuksen kohteena. Näiden päätösten hyvyttä mitataan tässä tutkielmassa saadun sadon ja kustannuksien avulla, mutta myös muita mahdollisia lähestymistapoja on. Taustamuuttujiin ei sen sijaan kohdistu mitään päätöstä, vaan niiden arvot määräytyvätkin aiemmin tehtyjen päätösten ja reaalimaailman ilmiöiden perusteella. Esimerkiksi maanviljelijä voi valita lajikkeeksi kauran, mutta ei tuholaisten määrää nollaksi. Näistä jälkimmäinen voitaisiin kuitenkin epäsuorasti saada haluttuun arvoon käyttämällä torjuntakeinoja.

Kannattavuutta mittaavista muuttujista kustannuksia ja tuottoa mitataan ra-



Kuvio 5: Kausaaligraafi viljapellon tuottavuudelle

hallisesti. Sato sen sijaan koostuu määrästä ja laadusta. Määrää mitataan tonneina hehtaarilta ja laatua voidaan mitata vilja-analyysin avulla. Vilja-analyysi koostuu tekijöistä, kuten viljalaji (esim. kaura), lajike (esim. Avetron), kosteus (viljan kosteus kuivauksen jälkeen, mitataan %), hlp (hehtolitrapaino eli kuinka paljon 100 litraa viljaa painaa kiloissa), DON (homepitoisuus,  $\mu g/kg$ ), valkuaispitoisuus (%), lajittelu (%), roskat (%) ja vieraiden viljalajien osuus (%). Näistä hehtolitrapainossa ja valkuaispitoisuudessa suurempi arvo on parempi, kun muissa taas pienempi arvo on parempi.

Tuotto lasketaan kustannuksien, sadon sekä markkinoilta saadun hinnan avulla. Markkinatilanne vaikuttaa viljojen hintatasoon ja tähän viljelijä ei pysty vaikuttamaan. Kuitenkin etukäteen laadittavilla sopimuksella viljelijä pystyy kiinnittämään myymänsä viljan hinnan. Esimerkiksi talvella voitaisiin sopia, että viljelijä myy seuraavan satonsa tiettyyn hintaan. Tämä olisi hyvä, jos viljojen hinnat olisivat

laskussa, jolloin saataisiin parempi hinta viljasta. Tässä työssä jätetään kuitenkin markkinatilanteeseen ja tuottoon liittyvät asiat tälle tasolle ja keskitytään muihin muuttujiin.

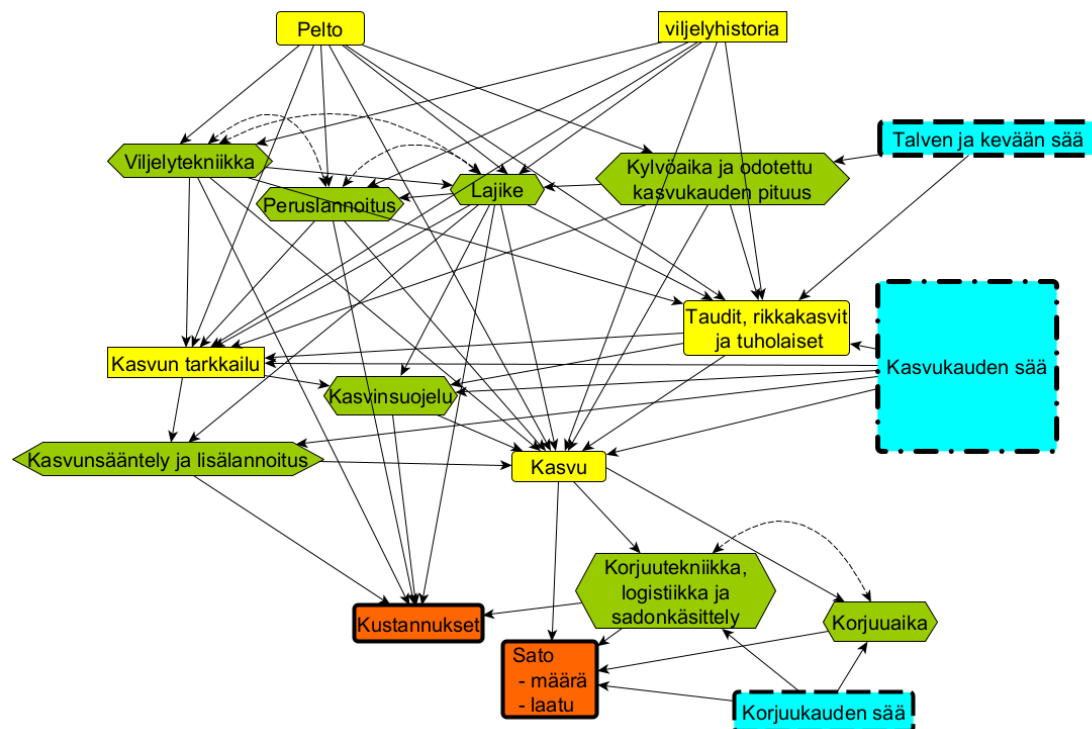
Latentit muuttajat kuvaavat maanviljelijän preferenssejä ja tilan luomia rajoitteita. Näitä rajoitteita voivat olla muun muassa koneiston puute ja työntekijöiden sairastuminen. Lisäksi maanviljelijällä voi olla tiettyjä toimintatapoja, jotka vaikuttavat päätöksentekoon. Ennen kasvukautta tehtävissä päätöksistä näiden ajatellaan vaikuttavan viljelytekniikkaan, peruslannoitukseen ja lajikkeeseen, mutta ei kylvöaikaa. Lisäksi korjuukaudella tehtävien päätöksien välillä ajatellaan olevan vastaavat latentit muuttajat.

## 4.1 Muuttujien kuvailu

Kuviossa 6 on esitettynä edellisen luvun kausaaligraafi, jossa latentit muuttajat on korvattu kaksisuuntaisilla nuolilla ja tuotto on rajattu pois. Tässä alaluvussa käydään läpi kaikki muuttajat (vastemuuttujia lukuun ottamatta) ja kerrotaan, mitä ne sisältävät. Luotu kausaaligraafi heijastaa Mennalan näkemyksiä asiasta. Siinä olevien solmujen voidaan kuitenkin ajatella olevan kiinnitettyjä eli muiden tilojen vastaavissa graafeissa esiintyisivät samat solmut. Tähän olisi mahdollista lisätä uusia solmuja, jos toisella tilalla päätöksentekoon liittyisi muita tekijöitä. Yleisesti solmujen sisältämistä muuttujista voi esittää useitakin näkemyksiä perustuen käytössä olevaan tietämykseen ja menetelmiin. Tulevaisuutta ajatellen muuttujia voitaisiin muuttaa siten, että aineiston kerääminen niistä olisi mahdollisimman helppoa. Tämä tapahtuisi tilakohtaisesti käytössä olevia menetelmiä mukaillen. Seuraavaksi annetussa esityksessä on ajateltu, että teknisiä rajoituksia ei tulisi vastaan.

Kuviossa 6 muuttajat ovat erityyppisiä monellakin tapaa. Osa muuttujista voidaan ajatella aikasarjana, kun toisiin taas riittää yksittäinen mittaus. Lisäksi muuttujien tyypit voivat olla jatkuvia tai luokka-asteikollisia. Vaikka mittaukset tapahtuvat eri aikaan, niin voidaan silti ajatella, että käytössä on muuttujien yhteisjakama. Esitelty kausaaligraafi on myös yksinkertaistus ongelmasta, sillä tämä ei salli kaikkia vaikutuksia. Esimerkiksi kasvinsuojelutoimenpiteiden vaikutus kasvuun huomioidaan graafissa vain päätepisteen kautta. Tällöin vaikutus kohdistuu ainoastaan kasvukauden päätteeksi olevaan kasvuun eikä viikoittaisiin kasvuhavaintoihin. Tätä varten kausaaligraafin solmuja jouduttaisiin monistamaan ja kokonaiskuvan hahmottaminen vaikeutuisi.

Pellolla tarkoitetaan tässä viljeltävää lohkoa. Pellosta täytyy kerätä tiedot viljelyominaisuuksista, jotka saadaan tekemällä viljavuustutkimus lohkoakohtaisesti (Eurofins, 2022). Nämä tehdään vähintään viiden vuoden välein, mikä on raja ympäris-



Kuvio 6: Kausaaligraafi viljapellon sadolle ja kustannuksille. Ero kuvioon 5 on, että tuotto on rajattu pois. Tätä graafia tullaan käyttämään seuraavan luvun identifioituvuustarkasteluissa.

tökorvauksen saannille. Viljavuustutkimus sisältää joukon muuttujia, jotka kuvaavat maaperän laatua. Tärkeimpiä näistä ovat pH, maalaji ja perusravinneaineet, mutta näiden lisäksi tämä sisältää muun muassa multavuuden ja johtoluvun. Viljavuustutkimuksesta on myös eri laajuuksia, joilla voisi saada lisää tietoa eri ravinnepitoisuuksista. Toinen kerättävä asia pellostä on sen sijainti. Olennaisesti jos pellot sijaitsevat toistensa läheisyydessä, niin tätä ei tarvita. Muissa tapauksissa sijainnin tarkkuuden ilmaisemiseen täytyy käyttää harkintaa. Jos kyse muutamasta eri paikkakunnasta, niin paikkakunnan nimi voisi olla sopiva. Jos taas peltoja olisi ympäri Suomea, niin nämä voitaisiin jakaa esimerkiksi Etelä-, Pohjois-, Länsi- ja Itä-Suomeen.

Viljelyhistoria koostuu siitä, miten peltoa on viime vuosina viljelty. Viljelyhistoria voidaan tiivistää muuttujiin {kasvijäte, lajike, maanmuokkaus}. Puinnissa puumuri jättää kasvien korret ynnä muun jätteen pellolle. Kasvijätteeseen kuuluu valinta sen jättämisestä pellolle tai keräämisestä pois. Lajikkeeseen kuuluu tieto viljelystä lajista sekä lajikkeesta. Maanmuokkaustekniikkaan kuuluu se, miten maata on muokattu. Maata muokataan, jotta kasvijäte saadaan hävitettyä pellon pinnalta ja uusi kasvi voidaan istuttaa tilalle. Maanmuokkaus voidaan jakaa kyntöön ja muihin menetelmiin. Tulevia päätöksiä tehdessä tulisia tarkastella viljelyhistoriaa edellisvuosilta ja tähän sopiva aikaväli voisi olla kolmesta viiteen vuotta, mikä perustuu

tautien selviytymiseen pellossa kasvijätteen seassa.

Viljelytekniikka kuvaa toimenpiteitä, jotka tehdään siemenen kylvämiseksi maahan. Viljelytekniikoita on useita ja vaihtoehdot määräytyvätkin maanviljelijän ja käytössä olevan koneiston perusteella. Yksinkertaistettuna viljelytekniikat voidaan jakaa kyntöön, perusmuokkaukseen ja suorakylvöön. Kyntö ja perusmuokkaus ovat maanmuokkaustekniikoita ja suorakylvössä siemen puolestaan kylvetään muokkaamattomaan maahan. Maanmuokkaustekniikoihin liittyen voitaisiin huomioida myös muita tekijöitä, kuten muokkaussyvyys eli kuinka syvältä maata käännetään ympäri.

Peruslannoituksella kuvataan lannoitusta, joka tehdään ennen kylvöä tai sen yhteydessä. Eräs tapa on tehdä lannoitus kerralla, mutta tässä työssä lannoitus on jaettu perus- ja lisälannoitukseen. Lisälannoitus tehdään kasvukauden aikana tarvittaessa. Peruslannoituksen voi ajatella koostuvan tyypistä ja määrästä. Tyyppi jaetaan keinolannoitteiksi ja liete- tai kuivalantaan. Keinolannalla tarkoitetaan kemiallisesti valmistettuja lannoitteita. Lantoihin kohdistuu tiettyjä rajoitteita, kuten että lietelantaa ei saa levittää taajamissa. Tällaiset rajoitteet täytyisi erikseen ottaa huomioon tulosten tulkinnoissa. On myös muita tapoja lannoittaa, kuten esimerkiksi viljelemällä apilaa, joka sitoo maahan typpeä. Tällaisessa menettelyssä uhrataan vuoden sato, mutta mahdollisesti parannetaan tulevien vuosien satoja.

Lajikkeeseen kuuluu viljeltävä viljalaji sekä viljalajike. Kotimaisia viljalajeja ovat vehnä, ohra, ruis ja kaura. Näiden lisäksi tyypillinen viljelykasvi on rypsi. Näistä kaikista on myös paljon alalajeja, jotka vaihtelevat ominaisuuksiltaan. Näitä ominaisuuksia ovat muun muassa laon- ja taudinkestävyys, satoluku ja siementen laatu.

Kylvöaikaan kuuluu kylvöpäivämäärä sekä odotettu kasvukauden pituus. Jälkimmäistä voidaan arvioida tarkastelemalla edellisvuoden kasvukauden päättymispäiviä, jotka löytyvät nettisivulta (Ilmatieteenlaitos, 2021). Odotettu kasvukauden pituus laskettaisiin edellisvuoden kasvukauden päättymispäivän ja tämän vuoden kylvöpäivän välisenä erotuksena. Keski-Suomen tapauksessa päättymispäivä sijoittuisi loka-marraskuulle. Odotettuun kasvukauden pituuteen ei kohdistu päätöstä, vaikka se onkin graafissa värjätty vihreäksi.

Viljan kasvua kuvaavat solmut kasvun tarkkailu ja kasvu. Kasvun tarkkailulla tarkoitetaan jatkuvaa viljan kasvun seurantaa. Tämä voisi koostua viikoittaisista mittauksista samaan tapaan kuin kasvintuhoojien tapauksessa. Mitattavana muuttujana toimisi esimerkiksi vihermassan ja lehtivihreän määrä, ja näitä arvioidaan lohkoa parhaiten edustavien kasvien avulla. Kasvu sen sijaan tarkoittaa kausaali-graafissa sitä tilannetta, millaiseksi vilja on kasvanut kasvukauden päätteeksi. Tämä olisi yksittäinen havainto ja siinä otettaisiin huomioon vähintään samat muuttujat mitä kasvun tarkkailussa.

Kasvintuhoojilla tarkoitetaan tauteja, rikkakasveja ja tuholaisia. Kasvintuhoojista tarvitsee arvioida kunkin määrät viikkokohtaisesti. Nämä voisi luokitella niiden määrän mukaan kolmeen luokkaan: Alittaa reilusti torjuntakynnyksen, lähellä torjuntakynnystä ja ylittää torjuntakynnyksen. Torjuntakynnyksen ylittyessä kasvinsuojelutoimenpiteet pitäisi aloittaa heti.

Säätitietoja voidaan kerätä joko yksityisten sääasemien kautta tarkasti tai sitten julkisten säätilastojen kautta. Näistä on löydettävissä päivittäiset keskilämpötilat ja sademäärät. Lisäksi yksityisten sääasemien tapauksessa voitaisiin tarkastella myös muita tekijöitä. Säästä voitaisiin käyttää päivittäisiä havaintoja hyödyksi, mutta seuraavaksi esitellään vaihtoehtoisia tapoja, joilla näitä havaintoja pystytään tiivistämään.

Talven sään voisi esimerkiksi tiivistää pakkaspäivien lukumäärään (päivät, jolloin keskilämpötila on pakkasella) ja tämä mitattaisiin marras-joulukuulta. kevään säätä taas tarkastellaan aikasarjana tammikuusta kevään kylvöpäivään. Näistä voidaan mitata keskilämpötila viikoittain ja koko sademäärä. Kasvukauden sää kuvaa aikaväliä kylvöpäivästä siihen, että kasvi on valmis. Kasvukauden säästä riittää mitata lämpötilasumma ja sademäärä esimerkiksi viimeiseltä 2 viikolta. Korjuukauden sää kuvaa aikaväliä siitä, kun kasvi on valmis ja sato puidaan. Korjuukauden säästä voidaan myös mitata viikoittain keskilämpötila ja sademäärä.

Kasvinsuojelu kattaa kasvintuhoojia vastaan tehdyt toimet. Kasvinsuojelusta tarvitsee kirjata ylös ajoitus ja mitä vastaan se tehtiin (rikkakasveja, tuholaisia tai kasvitauteja). Ajoituksella tarkoitetaan kulunutta aikaa kylvöpäivästä. Kasvinsuojelupäätöksiä tehdään enintään kerran kutakin haittaa vastaan. Eräs vaihtoehto olisi tehdä kustakin haitasta oma muuttujansa ja näiden arvoksi antaa ajoitus. Lisäksi jos jotakin haittaa vastaan ei tehdä toimenpiteitä, niin tällöin se saisi arvoksi kuluneen ajan kylvöstä puintiin. Tämän voisi kuvitella vastaavan tilannetta, jossa kasvinsuojelua tehtäisiin viimeisenä päivänä, jolloin sillä ei olisi vaikutusta saatuun satoon.

Kasvunsäätely erotettiin kasvinsuojelusta, sillä tähän ei kasvintuhoojien ajatella vaikuttavan. Kasvunsäätelyssä päätetään se, että annetaanko kasville kasvunsäätteitä. Tämän tärkein tehtävä on viljan lakoutumisen estämisen, jotta kasvu pysyy hyvänä ja sadon puinti on helpompaa. Lakoutumisella tarkoitetaan kasvuston kaatumista, joka voi seurata kovista sateista tai tuulista. Kasvunsäätelystä voitaisiin tehdä muuttuja, jonka arvona olisi ajoitus.

Kasvun ja sadon välistä aikaa kuvaa korjuuaika, joka on suoraan viljan valmistuspäivän ja sen korjuun välinen erotus. Korjuussa vilja puidaan pois pelloilta, kuljetetaan viljankuivaamolle ja siellä edelleen kuivataan. Sato eroaa kasvusta siten,

että tämä mitataan tarkasti korjuun jälkeen. Kasvua sen sijaan joudutaan arvioimaan jollakin tapaa. Korjuutekniikkaan, logistiikkaan ja sadonkäsittelyyn kuuluvat toimenpiteet viljan saamiseksi myyntikelpoiseksi. Näitä on huomattavasti vaikeampi muuttaa kuin muita päätösmuuttujia, sillä nämä vaativat suurempia investointeja. Esimerkiksi Puimurin, viljankeräyslavojen ja viljankuivaamon vaihto eivät käy hetkessä ja näiden optimoinniksi tarvittaisiin aineistoa eri tiloilta.

## 4.2 Muuttujien väliset kausaalisuhteet

Tässä alaluvussa kerrotaan, miten kausaaligraafissa on määritelty muuttujien väliset suhteet. Muuttujista pellon, viljelyhistorian ja säämuuttujien ajatellaan olevan muuttujia, joihin eivät kausaaligraafin muut muuttujat vaikuta. Toisin sanoen näihin ei tule nuolia muista muuttujista, mutta näistä lähtee nuolia muihin muuttujiin.

Pellossa tärkeimmät vaikuttavat tekijät ovat maalaji ja sijainti. Sijainti liittyy erityisesti lämpötilasummaan, sillä tämä vaihtelee alueittain. Pelto vaikuttaa ennen kasvukautta tehtäviin päätösmuuttujiin viljelytekniikka, peruslannoitus, lajike ja kylvöaika. Lajikkeen valinta tehdään pääasiassa maaperän perusteella. Vaatimattomille kasveille kuten kauralle kelpaa tavallinen multamaa, kun taas vaativat kasvit kuten vehnä ja rypsi tarvitsevat otolliset olosuhteet. Päätös kylvönajasta tehdään maaperän kuivuuden perusteella. Kylvö voidaan tehdä aikaisintaan sitten, kun maaperästä otettu maakimpale hajoaa siihen tökättäessä. Lannoitus tehdään pellolle tehtyyn viljavuustutkimukseen perustuen. Tämän avulla päätetään, mitä lannoitetta laitetaan ja kuinka paljon. Pellon vaikutus viljelytekniikkaan tulee maaperän kautta, sillä esimerkiksi eroosioherkkiin maihin ei voida tehdä kyntöä. Pellon vaikutus kasvintuhoojien määrään tulee sijainnin kautta.

Viljelyhistoriasta lähtevät nuolet perustuvat osittain sen kautta, että jätettiinkö viime vuoden kasvijätteet peltoon vai ei. Kasvijäte sisältää ravinteita, mikä vähentää tarvittavan lannoituksen määrää, mutta tällöin kasvintuhoojien torjuntapaine kasvaa. Kasvitaudit voivat säilyä pellossa muutaman vuoden, vaikka siellä ei mitään kylvettäisikään. Jos kylvetään samansukuisia viljakasveja, niin kasvitaudit alkavat yleistymään. Tätä ongelmaa voidaan kiertää vaihtelemalla kasvatettavia viljalajeja siten, että niillä ei ole yhteisiä kasvitauteja.

Talven ja kevään sään ajatellaan vaikuttavan kylvöaikaan ja kasvintuhoojiin. Monet tuholaiset talvehtivat pellolla kasvijätteen seassa. Lämpiminä vuosina etelästä tulevat tuholaiset yleistyvät, mutta kovat pakkastalvet hävittävät kasvintuhoojia. Keväällä on taas kyse siitä, miten kevät kehittyy. Ensinnäkin pellon täytyy olla kuiva, jotta kylvöt voidaan aloittaa. Jos kevät on sateinen, niin pellolle päästään vasta myöhemmin. Jämsän seudulla kylvötöihin voidaan päästä jo huhtikuun puolella,

mutta joinakin vuosina tämä venyy toukokuun loppupuolelle. Kevään lämpötilalla on vaikutusta erityisesti rikkakasveihin. Jos kevät on lämmin ja sateinen, niin rikkakasvit alkavat kasvamaan pellossa, ennen kuin on päästy kylvämään. Tästä taas seuraa, että rikkakasvit varjostavat viljeltävää kasvia ja kasvu heikkenee.

Viljelytekniikka vaikuttaa kasvintuhoojiin, lajikevalintaan, kustannuksiin ja kasvuun. Viljelytekniikoista kynnessä maaperä käännetään ympäri ja tämä hävittää paremmin kasvintuhoojia, mutta maksaa enemmän kuin kevytmuokkaus, jossa maanpintaa vain vähän sekoitetaan. Viljelytekniikan vaikutus lajikevalintaan vaikutus tulee lajikkeen kasvintuhoojien kestävyys kautta. Esimerkiksi jos maata ei kynnetä ja valitaan huonosti kasvintuhoojia kestävä lajike, niin tällöin joudutaan lähes varmasti käyttämään kemiallisia kasvinsuojeluaineita myöhemmin.

Kylvöaika ja odotettu kasvukauden pituus vaikuttavat lajikkeeseen, kasvintuhoojiin ja kasvuun. Kylvö halutaan tehdä mahdollisimman aikaisin, jotta saadaan kevään kosteus hyödynnettyä eikä anneta rikkakasveille etumatkaa. Lisäksi tuholaisia ja tauteja voi alkaa ilmestymään varhaisemmassa vaiheessa kasvuun, jos kylvöä pitkitetään. Odotetun kasvukauden pituuden avulla valitaan kasvuajaltaan sopiva lajike.

Peruslannoitus vaikuttaa ainoastaan kasvuun ja kustannuksiin. Jos kasvilla on ravinnonpuutetta, niin sen lehdet ovat hailakampia ja tämä voidaan havaita kasvun tarkkailussa. Liian suuret ravinnemäärät eivät ole haitaksi kasville, mutta nämä lisäävät turhaan kustannuksia. Täten haluttaisiinkin laittaa juuri riittävä määrä lannoitetta.

Lajike vaikuttaa kasvintuhoojiin, kasvinsuojeluun, kasvunsäätelyyn, kasvuun, lannoitukseen sekä kustannuksiin. Lajikeominaisuuksia käytetään hyödyksi siten, että satoisuuden mukaan laitetaan sopivasti lannoitetta, laonkestävyys pienentää kasvunsäätelyaineiden tarvetta ja kasvintuhoojien kestävyys vähentää kasvinsuojelutoimenpiteiden tarvetta.

Kavinsuojeluun liittyvät päätökset tehdään tarkkailemalla pellolla tilannetta ja tarpeen vaatiessa aloitetaan kasvinsuojelutoimet. Jos kasvinsuojelua tarvitsee tehdä, niin se kannattaa mahdollisimman aikaisin, jotta kasvin kasvu pysyisi mahdollisimman hyvänä. Toisaalta turhaan tehdyt kasvinsuojelutoimenpiteet kasvattavat kustannuksia. Näistä seuraavat kasvinsuojelun vaikutus kasvuun ja kustannuksiin.

Myös kasvunsäätely ja lisälannoitus perustuvat viikoittaiseen havainnointiin ja ne vaikuttavat ainoastaan kasvuun ja kustannuksiin. Päätös lisälannoituksesta tehdään lehtivihreämittauksiin perustuen. Jos jossakin alueilla kasvilla näyttää olevan ravinnonpuutetta, niin silloin tänne annetaan lisälannoitusta. Tämä vaikuttaa positiivisesti kasvuun, mutta lisää kustannuksia.



Kasvukauden säästä tärkeimmät tekijät ovat lämpötila sekä sademäärä. Näiden ajatellaan vaikuttavan kasvintuhoojiin, kasvuun, kasvinsuojeluun, kasvunsäätelyyn ja lisälannoitukseen. Kasvuun tarvitaan sopivasti vettä, sillä kovat yksittäiset sadekuurot voivat huuhdella ravinneaineita pois pellostä. Lämpötila ja sademäärä vaikuttavat kasvintuhoojiin siten, että ne leviävät paremmin kostealla ja lämpimällä säällä. Kasvunsäätelyn ja lisälannoituksen tarvetta arvioitaessa täytyy menneisyyden lisäksi arvioida tulevaa säätä, sillä esimerkiksi lähitulevaisuudessa tulevien helputkien tapauksissa nämä toimenpiteet eivät avittaisi kasvua.

Taudit voidaan havaita katsomalla viljan lehteä valoa vastaan. Jos kasvitauteja on, niin lehdissä näkyy viiruja tai laikkuja. Olennaisinta on, että kuinka suurella peittoalalla tauteja on ja onko niitä ylälehdissä. Alalehdillä ei ole niinkään väliä, sillä ne kuolevat muutenkin pois ylälehtien varjostaessa niitä. Toinen olennainen asia koskee sitä, milloin tauti ilmestyy. Jos tauti ilmestyy heti kylvön yhteydessä, niin se voi tappaa koko kasvin tai vähintäänkin heikentää sen kasvua. Taas jos tauti ilmestyy vasta kun kasvin tähkä alkaa olemaan täysi, niin se ei enää haittaa. Rikkakasveja on kevätitoisia, kaksivuotisia ja monivuotisia. Monivuotiset on hävitettävä aina, kun taas kevätitoiset eivät haittaa, jos ne jäävät varjostukseen viljan alle. Rikkakasveista merkittävät tekijät ovat tyyppi, määrä ja se, että jääkö rikkakasvi varjostuksen alle. Tuholaisista taas tarkkaillaan niiden määrää ja ilmestymisajankohtaa. Nämä kaikki vaikuttavat kasvuun ja kasvinsuojeluun.

Korjuukauden sää vaikuttaa korjuutekniikkaan, logistiikkaan, sadonkäsittelyyn, korjuu aikaan ja satoon. Märkkää viljaa ei ole järkevää korjata, sillä sen kuivattaminen vaatisi paljon energiaa ja lisäisi kustannuksia. Tämän takia korjuu pyritään ajoittamaan sellaiseen väliin, jossa kasvi on valmis sekä kuiva. Kuitenkin mitä pidemmälle korjuu pitkittyy, niin sitä suurempi riski on sadon pilaantumisella. Korjuutekniikkaan ja logistiikkaan liittyvä päätös on lähinnä puimurin sekä kuljetustekniikoiden valinta lohko kohtaisesti. Pienillä tiloilla, jossa etäisyydet ovat lyhyitä ja viljeltävää alaa on vähän, puimuri- ja kuljetusvalinnoilla ei ole niin suurta merkitystä. Kuitenkin etäisyyksien ja puitavien hehtaarien kasvaessa tarvitsee tehdä optimointia. Esimerkiksi Mennalan tilalla on käytössään vaihtolavoja, joita traktori pystyy vetämään perässään kahta kerrallaan. Tällöin tyhjä lava voidaan jättää pellolle puimuria varten, kun täyttä lavaa kuljetetaan viljankuivaamolle. Tämä mahdollistaa puimurin pysymisen liikkeessä aina tyhjennyksiä lukuun ottamatta ja tehokkuus lisääntyy.

## 5 Kausaalivaikutusten identifointi

Tässä luvussa identifoidaan kausaalivaikutuksia perustuen edellisessä luvussa esitettyyn kausaaligraafiin. Aluksi identifioitiin yksi kausaalivaikutus käyttämällä pelkkää ID-algoritmia, jolloin päädyttiin monimutkaisiin lausekkeisiin. Tämän jälkeen hyödynnettiin karsintaa osana ID-algoritmia, jolloin päästiin jo huomattavasti yksinkertaisempaan muotoon. Lopuksi yritettiin vielä hyödyntää sieventämistä edellä saatuun lausekkeeseen, mikäli se onnistui järkevässä ajassa. Tässä luvussa tarkastellaan ensiksi päätösmuuttujien vaikutusta satoon ja myöhemmin näiden vaikutuksista kustannuksiin.

Päätösmuuttujien yksittäisvaikutukset satoon olivat identifioituvia lukuun ottamatta lajiketta ja viljelytekniikkaa. Ensiksi esitellään yksittäisvaikutuksia, jonka jälkeen siirrytään yhdysvaikutuksiin. Käytetään muuttujista lyhennyksiä pitkien lausekkeiden välttämiseksi seuraavasti:

$$Y_1 = \text{Sato}$$

$$Y_2 = \text{Kustannukset}$$

$$X_1 = \text{Viljelytekniikka}$$

$$X_2 = \text{Peruslannoitus}$$

$$X_3 = \text{Lajike}$$

$$X_4 = \text{Kylvöaika}$$

$$X_5 = \text{Kasvinsuojelu}$$

$$X_6 = \text{Kasvunsäätely ja lisälannoitus}$$

$$X_7 = \text{Korjuuaika}$$

$$X_8 = \text{Korjuutekniikka, logistiikka ja sadonkäsittely}$$

$$Z_1 = \text{Pelto}$$

$$Z_2 = \text{Viljelyhistoria}$$

$$Z_3 = \text{Taudit, rikkakasvit ja tuholaiset}$$

$$Z_4 = \text{Kasvun tarkkailu}$$

$$Z_5 = \text{Kasvu}$$

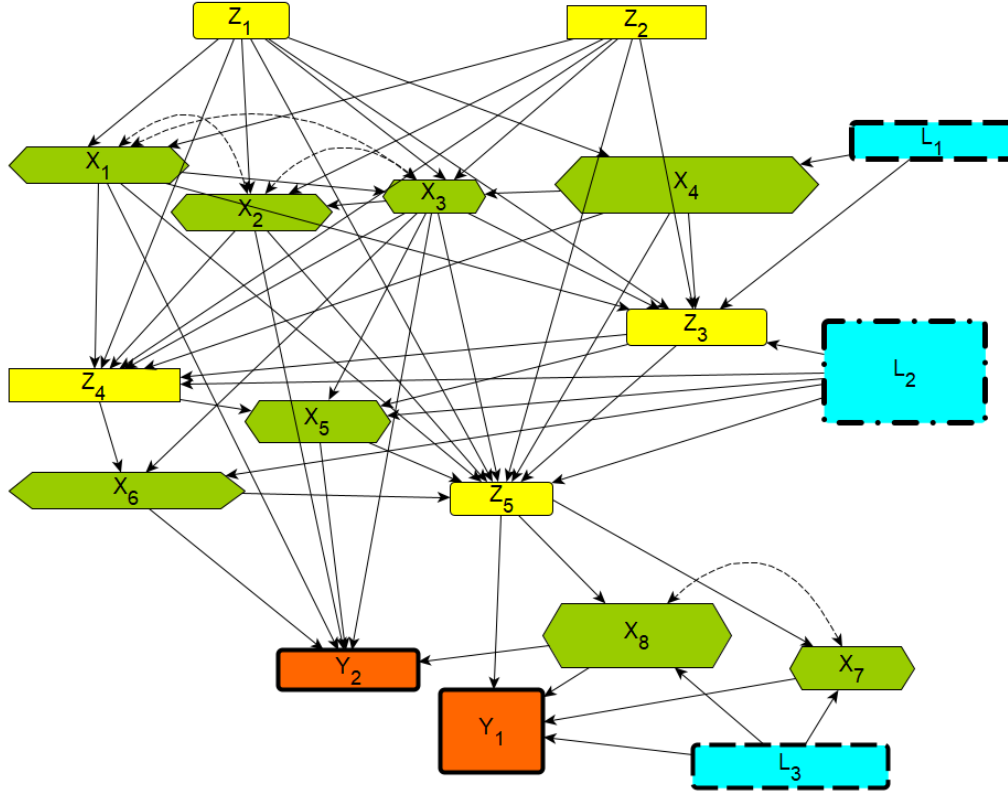
$$L_1 = \text{Talven ja kevään sää}$$

$$L_2 = \text{Kasvukauden sää}$$

$$L_3 = \text{Korjuukauden sää.}$$

Tässä  $\mathbf{Y}=\{Y_1, Y_2\}$  kuvaa vastemuuttujia,  $\mathbf{X}=\{X_1, \dots, X_8\}$  päätösmuuttujia,  $\mathbf{Z}=\{Z_1, \dots, Z_5\}$  taustatekijöitä ja  $\mathbf{L}=\{L_1, \dots, L_3\}$  säämuuttujia. Merkitään vielä, että  $\mathbf{M} = \{\mathbf{X}, \mathbf{Z}, \mathbf{L}\}$ .  $\mathbf{M}$  siis sisältää vastemuuttujia lukuun ottamatta kaikki graafin muuttujat. Säämuuttujat eroteltiin tässä taustatekijöistä, sillä haluttiin selvittää, voisiko jotkut säämuuttujat muuttaa latenteiksi ja miten tämä muuttaisi saatuja

lausekkeita. Kuviossa 7 on esitetty luvun 4 kausaaligraafi, jossa muuttujat korvataan edellisillä merkinnöillä. Lisäksi latentit muuttujat on merkitty kaksisuuntaisilla nuolilla muuttujien välillä.



Kuvio 7: Kausaaligraafi viljapellon tuottavuudelle, jossa käytetään muuttujien tilalla lyhennettyjä merkintöjä

## 5.1 Päätösmuuttujien vaikutus satoon

ID-algoritmin palauttama lauseke kausaalivaikutukselle  $P(\text{Sato}|\text{do}(\text{peruslannoitus})) = P(Y_1|\text{do}(X_2))$  on

$$\begin{aligned} & \sum_{\mathbf{M} \setminus \{X_2\}} P(X_8|\mathbf{M} \setminus \{X_8\})P(X_7|\mathbf{M} \setminus \{X_7, X_8\}) \\ & P(Y_1|\mathbf{M})P(X_3|Z_2, L_1, Z_1, X_1, X_4)P(X_1|Z_2, L_1, Z_1) \\ & P(Z_5|L_1, L_2, \mathbf{Z} \setminus Z_5, \mathbf{X} \setminus \{X_7, X_8\})P(X_6|L_1, L_2, X_1, X_4, X_3, X_2, \mathbf{Z} \setminus Z_5) \\ & P(X_5|L_1, L_2, X_1, X_4, X_3, X_2, \mathbf{Z} \setminus Z_5)P(Z_4|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3) \\ & P(Z_3|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3)P(X_4|Z_2, L_1, Z_1)P(Z_1|Z_2)P(L_3)P(L_2)P(L_1)P(Z_2). \end{aligned}$$

Tämän lausekkeen tulkitseminen olisi työlästä, sillä se sisältää kaikki kausaaligraafin muuttujat. Tästä kuitenkin saadaan paljon sievempi lauseke käyttämällä karsintaa:

$$P(Y_1|do(X_2)) = \sum_{Z_2, Z_1, X_1, X_4, X_3} P(Y_1|Z_2, Z_1, X_1, X_4, X_3, X_2)P(X_4|Z_2, Z_1, X_1)P(X_3|Z_2, Z_1, X_1, X_4)P(X_1|Z_2, Z_1)P(Z_1|Z_2)P(Z_2).$$

Huomataan, että karsinta yksinkertaisti lauseketta huomattavasti. Ensinnäkin jälkimmäisessä lausekkeessa on paljon vähemmän termejä ja toiseksi ehdollistusjoukot ovat myös pienempiä. Sieventämisen käyttäminen karsinnan lisäksi antoi tässä tapauksessa saman lausekkeen kuin pelkkä karsinta. Näiden menetelmien ajoajat sen sijaan olivat täysin eri luokissa, kun karsinta tapahtui sekunneissa ja sieventäminen vei useamman tunnin tavallisella pöytätietokoneella. Sieventämistä hyödynnettiin myös muihin tämän alaluvun yksittäisvaikutuksiin karsinnan lisäksi ja tämä palautti täysin saman lausekkeen kuin pelkkä karsinta.

Tulkitaan seuraavaksi kylvöajan vaikutuksen satoon. Karsinnan ja sievennyksen avulla saatiin tälle lausekkeeksi

$$P(Y_1|do(X_4)) = \sum_{L_1, Z_1} P(Y_1|L_1, Z_1, X_4)P(Z_1|L_1)P(L_1).$$

Tämä lauseke on yksinkertainen, sillä se sisältää vähän termejä ja muuttujia. Tämän vaikutuksen estimoiminen vaatisi vain vastemuuttujan (saton= $Y_1$ ) ja interventiomuuttujan (kylvöajan= $X_4$ ) lisäksi tiedon talven ja kevään säästä ( $L_1$ ) ja pellostasta ( $Z_1$ ). Tarkastelemalla kausaaligraafia huomataan, että muuttujat  $Z_1$  ja  $L_1$  ovat riippumattomia keskenään, mutta saadussa lausekkeessa esiintyy termi  $P(Z_1|L_1)$ , jonka voisi sieventää muotoon  $P(Z_1)$ . Tästä voidaan huomata, että karsinta ja sieventäminen eivät ole täydellisiä menetelmiä ja onkin hyväksi tarkastella jälkikäteen kausaaligraafia ja saatua lauseketta yhdessä. Seuraavaksi listataan loput yksittäisvaikutukset:

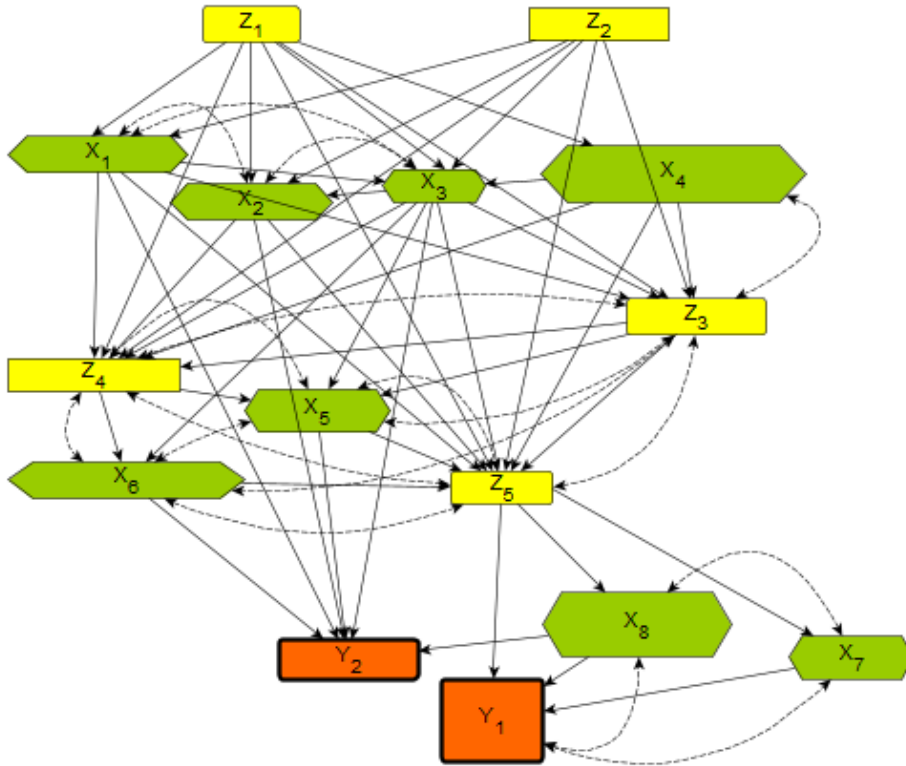
$$P(Y_1|do(X_5)) = \sum_{L_2, X_3, Z_3, Z_4} P(Y_1|L_2, X_3, Z_3, Z_4, X_5)P(Z_4|L_2, X_3, Z_3)P(Z_3|L_2, X_3)P(X_3|L_2)P(L_2),$$

$$P(Y_1|do(X_6)) = \sum_{L_2, X_3, Z_4} P(Y_1|L_2, X_3, Z_4, X_6)P(Z_4|L_2, X_3)P(X_3|L_2)P(L_2),$$

$$P(Y_1|do(X_7)) = \sum_{L_3, Z_5, X_8} P(Y_1|L_3, Z_5, X_7, X_8)P(X_8|L_3, Z_5)P(Z_5)P(L_3) \text{ ja}$$

$$P(Y_1|do(X_8)) = \sum_{L_3, Z_5, X_7} P(Y_1|L_3, Z_5, X_7, X_8)P(X_7|L_3, Z_5)P(Z_5)P(L_3).$$

Edeltävistä kausaalivaikutuksista mistään ei voitu muuttaa kaikkia säämuuttujia latenteiksi (paitsi  $P(Y_1|do(X_2))$ :n tapauksessa, sillä se ei sisältänyt säämuuttujia). Seuraavaksi esitettävissä yhdysvaikutuksissa tämä oli kuitenkin mahdollista ja tällöin säämuuttajat saatiin korvattua muilla muuttujilla. Kuviossa 8 on esitetty nä aiempi kausaaligraafi, mutta tässä säämuuttajat on muutettu latenteiksi. Tästä seuraa, että kaksisuuntaisia särmiä joudutaan lisäämään poistuneiden polkujen tilalle. Vaikka viljelytekniikan ( $X_1$ ) ja lajikkeen ( $X_3$ ) yksittäisvaikutukset satoon eivät olleet identifioituvia, niin näistä voidaan silti laskea tiettyjä yhdysvaikutuksia. Esimerkiksi lajikkeen ja peruslannoituksen kausaalivaikutukselle saadaan lausekkeeksi



Kuvio 8: Kausaaligraafi viljapellon tuottavuudelle, jossa säämuuttajat on muutettu latenteiksi

$$\begin{aligned}
& P(Y_1|do(X_2, X_3)) \\
&= \sum_{\mathbf{X} \setminus \{X_2, X_3\}, \mathbf{Z}, \mathbf{L}} P(Y_1|L_3, Z_5, X_7, X_8)P(X_8|L_3, Z_5, X_7)P(X_7|L_3, Z_5) \\
& P(Z_5|L_1, L_2, \mathbf{Z} \setminus \{Z_5\}, \mathbf{X} \setminus \{X_7, X_8\})P(X_6|L_1, L_2, X_1, X_4, X_3, X_2, \mathbf{Z} \setminus \{Z_5\}) \\
& P(X_5|L_1, L_2, X_1, X_4, X_3, X_2, \mathbf{Z} \setminus \{Z_5\})P(Z_4|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3) \\
& P(Z_3|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3)P(X_4|L_1, Z_1) \\
& P(X_1|Z_2, Z_1)P(Z_1|Z_2)P(L_3)P(L_2)P(L_1)P(Z_2).
\end{aligned}$$

Lisäksi viljetekniikan, peruslannoituksen ja lajikkeen kausaalivaikutukselle saadaan lausekkeeksi

$$\begin{aligned}
& P(Y_1|do(X_1, X_2, X_3)) \\
&= \sum_{\mathbf{X} \setminus \{X_1, X_2, X_3\}, \mathbf{Z}, \mathbf{L}} P(Y_1|L_3, Z_5, X_7, X_8)P(X_8|L_3, Z_5, X_7)P(X_7|L_3, Z_5) \\
& P(Z_5|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4, X_5, X_6) \\
& P(X_6|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4) \\
& P(X_5|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4) \\
& P(Z_4|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3, X_2, Z_3)P(Z_3|Z_2, L_1, L_2, Z_1, X_1, X_4, X_3) \\
& P(X_4|L_1, Z_1)P(Z_1|Z_2)P(L_3)P(L_2)P(L_1)P(Z_2).
\end{aligned}$$

Tässä pystyttiin muuttamaan säämuuttujat latenteiksi ja saatiin vaihtoehtoiseksi lausekkeeksi

$$\begin{aligned}
& P(Y_1|do(X_1, X_2, X_3)) \\
&= \sum_{\mathbf{X} \setminus \{X_1, X_2, X_3\}, \mathbf{Z}, \mathbf{L}} P(Y_1|Z_5, X_8, X_7)P(Z_5|Z_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4, X_5, X_6) \\
& P(X_6|Z_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4, X_5) \tag{2} \\
& P(Z_4|Z_2, Z_1, X_1, X_4, X_3, X_2, Z_3)P(Z_3|Z_2, Z_1, X_1, X_4, X_3, X_2) \\
& P(X_4|Z_2, Z_1, X_1)P(X_7|Z_5, X_8) \\
& P(X_5|Z_2, Z_1, X_1, X_4, X_3, X_2, Z_3, Z_4)P(X_8|Z_5)P(Z_1|Z_2)P(Z_2).
\end{aligned}$$

Huomataan, että summauksesta on hävinnyt säämuuttujat sekä  $X_1, X_3$  ja näiden

tilalle on tullut  $X_4$ . Tässä siis muuttujien muuttaminen latenteiksi yksinkertaisti saatua lauseketta. Tästä voidaankin huomata, että karsinta ja sieventäminen eivät välttämättä johda yksinkertaisimpaan lausekkeeseen. Tämän seurauksena kausaalimallinnuksessa kannattaakin yrittää manuaalisesti muuttaa sellaisia muuttujia latenteiksi, joista on vaikeaa kerätä aineistoa. Esimerkiksi tämän tutkielman tapauksessa, jos säätietoja ei olisi saatavilla, niin tällöin voitaisiin käyttää lauseketta (2).

## 5.2 Päätösmuuttujien vaikutus kustannuksiin

Lasketaan seuraavaksi päätösmuuttujien vaikutus kustannuksiin käyttäen karsintaa. Päätösmuuttujista ainoastaan korjuuajalla ei ole vaikutusta kustannuksiin eli  $P(Y_2|do(X_7)) = P(Y_2)$ . Tämän huomaa myös graafista, sillä liikkumalla nuolten suuntaisesti ei päästä korjuuajasta kustannuksiin. Lisäksi viljelytekniikan ja lajikkeen vaikutus ei identifioitunut.

$$P(Y_2|do(X_2)) = \sum_{Z_2, Z_1, X_1, X_4, X_3} P(Y_2|Z_2, Z_1, X_1, X_4, X_3, X_2)P(X_4|Z_2, Z_1, X_1)P(X_3|Z_2, Z_1, X_1, X_4)P(X_1|Z_2, Z_1)P(Z_1|Z_2)P(Z_2),$$

$$P(Y_2|do(X_4)) = \sum_{L_1, Z_1} P(Y_2|L_1, Z_1, X_4)P(Z_1|L_1)P(L_1),$$

$$P(Y_2|do(X_5)) = \sum_{L_2, X_3, Z_3, Z_4} P(Y_2|L_2, X_3, Z_3, Z_4, X_5)P(Z_4|L_2, X_3, Z_3)P(Z_3|L_2, X_3)P(X_3|L_2)P(L_2),$$

$$P(Y_2|do(X_6)) = \sum_{L_2, X_3, Z_4} P(Y_2|L_2, X_3, Z_4, X_6)P(Z_4|L_2, X_3)P(X_3|L_2)P(L_2),$$

$$P(Y_2|do(X_8)) = \sum_{Z_5} P(Y_2|Z_5, X_8)P(Z_5).$$

## 5.3 Päätösmuuttujien vaikutus sadon ja kustannusten yhteisjakaumaan

Lasketaan seuraavaksi päätösmuuttujien vaikutus sadon ja kustannusten yhteisjakaumaan käyttäen karsintaa. Näistä viljelytekniikan ja lajikkeen vaikutus ei identifioitunut. Causaleffect-funktion palauttama lauseke koostuu aina yksittäisistä ehdollisista jakaumista ja mallintamista varten näitä voidaan joutua käsin yhdistelemään. Esimerkiksi seuraavan kausaalivaikutuksen tapauksessa vastemuuttujien yh-

teisjakauman saa yhdistämällä ehdolliset termit  $P(Y_1|Z_2, Z_1, X_1, X_4, X_3, X_2, Y_2)$  ja  $P(Y_2|Z_2, Z_1, X_1, X_4, X_3, X_2)$  termiksi  $P(Y_1, Y_2|Z_2, Z_1, X_1, X_4, X_3, X_2)$ . Seuraavaksi listataan kaikki identifioituneet yksittäisvaikutukset:

$$\begin{aligned} P(Y_1, Y_2|do(X_2)) &= \sum_{Z_2, Z_1, X_1, X_4, X_3} P(Y_1|Z_2, Z_1, X_1, X_4, X_3, X_2, Y_2)P(Y_2|Z_2, Z_1, X_1, X_4, X_3, X_2) \\ &P(X_4|Z_2, Z_1, X_1)P(X_3|Z_2, Z_1, X_1, X_4)P(X_1|Z_2, Z_1)P(Z_1|Z_2)P(Z_2), \end{aligned}$$

$$P(Y_1, Y_2|do(X_4)) = \sum_{L_1, Z_1} P(Y_1|L_1, Z_1, X_4, Y_2)P(Y_2|L_1, Z_1, X_4)P(Z_1|L_1)P(L_1),$$

$$\begin{aligned} P(Y_1, Y_2|do(X_5)) &= \sum_{L_2, X_3, Z_3, Z_4} P(Y_1|L_2, X_3, Z_3, Z_4, X_5, Y_2)P(Y_2|L_2, X_3, Z_3, Z_4, X_5) \\ &P(Z_4|L_2, X_3, Z_3)P(Z_3|L_2, X_3)P(X_3|L_2)P(L_2), \end{aligned}$$

$$\begin{aligned} P(Y_1, Y_2|do(X_6)) &= \sum_{L_2, X_3, Z_4} P(Y_1|L_2, X_3, Z_4, X_6, Y_2)P(Y_2|L_2, X_3, Z_4, X_6) \\ &P(Z_4|L_2, X_3)P(X_3|L_2)P(L_2), \end{aligned}$$

$$\begin{aligned} P(Y_1, Y_2|do(X_7)) &= \sum_{L_3, Z_5, X_8} P(Y_1|L_3, Z_5, X_7, X_8)P(Y_2|Z_5, X_8)P(Z_5)P(X_8|L_3, Z_5)P(L_3). \end{aligned}$$

$$\begin{aligned} P(Y_1, Y_2|do(X_8)) &= \sum_{L_3, Z_5, X_7} P(Y_1|L_3, Z_5, X_7, X_8)P(Y_2|Z_5, X_8)P(Z_5)P(X_7|L_3, Z_5)P(L_3). \end{aligned}$$



## 6 Mahdollisia datalähteitä jatkotutkimuksia varten

Mennalan tilalta ei ollut kerättyä valmista aineistoa, jota päästäisiin suoraan soveltamaan. Tämä johtuu lähinnä siitä, että aineiston kerääminen on aikaa vievää ja Mennalalla ei ole ollut tälle tarvetta. Yleensäkin viljelyyn liittyvät päätökset tapahtuvat jatkuvaan havainnointiin perustuen. Tämä tarkoittaa sitä, että pellolla käydään jatkuvasti katsomassa tilannetta ja päätökset tehdään pellon sen hetkisen tilan perusteella. Tästä esimerkkinä tuholaisten torjunta: Käydään pellolla jatkuvasti tutkimassa tuholaisten määrää ja sitten kun tuholaista alkaa olemaan merkittävää haittaa, niin aloitetaan myrkyttäminen. Tällöin esimerkiksi viime viikon kirvojen määrällä ei ole merkitystä tämänhetkiseen päätöksentekoon ja sen takia siitä ei ole kerätty talteen aineistoa.

Aineistoa olisi kuitenkin mahdollista kerätä kaikista edellisessä luvussa esitetyistä tekijöistä. Tämä kuitenkin vaatii tiettyjä välineistöjä, joita ei välttämättä kaikilta tiloilta löydy. Tärkein tarvittava ominaisuus olisi sadonmäärän mittausjärjestelmä puimurissa, jolla saataisiin mitattua sadon määrä lohkoittain. Tähän tarvitaan siis nykyaikainen puimuri sekä tarpeeksi mielenkiintoa viljelijältä ylläpitää mittauslaitteen käyttöä. Vastaavasti sadon laatu saadaan tarkasti lähettämällä otos vilja-analyysiin. Lisäksi kustannuksista ja päätösmuuttujista olisi myös suora- viivaista kerätä aineistoa. Viljelytoimenpiteet, kuten viljelytekniikka ja lannoitus veloitetaan tunti- tai hehtaarihinnalla ja nämä voitaisiin suhteuttaa viljeltävän lohkon kokoon nähden. Nämä voidaan löytää esimerkiksi työtehosteuran sivuilta (Työtehosteura, 2022).

Taustatekijöistä aineiston kerääminen on sen sijaan haastavampaa. Näistä kasvua ja kasvintuhoojia koskevat muuttujat jouduttaisiin arvioimaan silmämääräisesti. Tämä tapahtuisi lohkoittain valitsemalla tyypillisimmät kasvit, jotka edustavat mahdollisimman hyvin koko lohkon tilaa ja ottamalla näistä keskiarvo. Viljelyhistoriaa varten taas tarvitaan aineisto vähintään muutamalta viime vuodelta tehdyistä viljelytoimenpiteistä.

Sää tiedot voidaan löytää esimerkiksi ilmatieteenlaitoksen säähistoria-aineistoista (Ilmatieteenlaitos, 2022). Sieltä löytyy muun muassa vuorokausihavaintoja lämpötiloista ja sademääristä noin 400 havaintoasemalta. Näistä voitaisiin etsiä peltoa lähinnä oleva ja sitä kautta saada arvio sään vaikutuksesta. Sääasemia ei kuitenkaan löydy jokaiselta paikkakunnalta ja tässä tapauksessa jouduttaisiin etsimään muualta säähavaintoja.

Luken tekemää lajikevalintatyökalua voitaisiin käyttää hyväksi monellakin tapaa (Luke, 2022a). Se sisältää Luken tekemien lajikekokeiden tulokset, joita voidaan rajata eri kriteerien mukaan. Ensimmäinen valinta koskee viljalajia, joka kattaa kaikki

yleisimmät lajit. Tämän jälkeen voidaan tehdä lisärajauskia perustuen muun muassa kasvuaikaan ja taudinkestävyyteen. Lopuksi lajikkeita voidaan painottaa lisäksi maalajilla, viljelyvyöhykkeellä ja satotasolla. Heuristinen tapa hyödyntää tätä työkalua olisi rajata mahdollisimman tarkasti peltoa vastaavat ominaisuudet ja sitten valita satoisuudeltaan paras lajike. Tämän lisäksi on otettava huomioon viime vuosien viljelytoimenpiteet, jotta valitaan taudinkestävyydeltään sopiva lajike.

Lajikevalintatyökalua voitaisiin käyttää myös osana kausaalimallinnusta. Tässä työssä käytetty ID-algoritmi tarvitsee syötteenä muuttujien yhteisjakauman, jolloin useita datalähteitä ei voida käyttää. Tähän on ratkaisuksi kuitenkin kehitetty uusia menetelmiä, kuten dosearch (Tikka, Hyttinen, & Karvanen, 2021). Tämä mahdollistaa erillisten datalähteiden käytön, kunhan niissä on vähintään yksi yhteinen muuttuja. Lisäksi vain sellaisia muuttujia voidaan hyödyntää, jotka ovat osana kausaaligraafia.

## 7 Pohdinta

Tämän tutkielman tavoitteena oli selvittää, miten kausaalimalleja voitaisiin hyödyntää maanviljelyssä ja mitä tämä vaatisi. Tätä varten luotiin kausaalimalli kevätviljoille satoon ja kustannuksiin vaikuttavista tekijöistä perustuen asiantuntilausuntoihin. Kausaalimallin avulla tehtiin identifiointuustarkasteluja päätösmuuttujien vaikutuksista satoon ja kustannuksiin. Lähes kaikkien päätösmuuttujien yksittäisvaikutukset vastemuuttujiin saatiin identifioidua. Lisäksi ottamalla useampi päätösmuuttuja mukaan interventioon saatiin myös niiden päätösmuuttujien vaikutus identifioidua, joiden yksittäisvaikutus ei ollut identifioituva.

Karsinnan hyödyntäminen ID-algoritmin tukena mahdollisti paljon sievempiin lausekkeisiin pääsyn. Sieventämisen ja karsinnan yhdistämisellä ei saavutettu lisähyötyä pelkkään karsintaan verrattuna, kun interventio kohdistettiin yhteen muuttujaan. Näiden menetelmien ajoajat kuitenkin erosivat selvästi, sillä karsinta toimi sekunneissa, kun taas sievennys vei useamman tunnin tavallisella PC-koneella. Isompien graafien tapauksessa kannattaakin välttää sieventämisen käyttöä, sillä tämän aikavaativuus kasvaa eksponentiaalisesti muuttujien määrän mukaan.

Kausaalivaikutuksia ei tässä työssä estimoitu, sillä käytettävissä ei ollut tarpeeksi aineistoa. Haastatellulla maanviljelijällä ei ollut kerättynä lohkokohtaista aineistoa saadusta sadosta, joka olisi välttämätön tieto, kun halutaan selvittää päätösmuuttujien vaikutuksia satoon. Tulevaisuudessa maatiloilla tullaan luultavasti keräämään entistä enemmän aineistoa, jolloin tästä tutkielmassa saatuja tuloksia voitaisiin hyödyntää monella tavalla. Ensinnäkin tässä tutkielmassa on pohdittu, minkälaisista ke-

rättävän aineiston tulisi olla. Tämän avulla maataloilla voitaisiin suunnitella aineiston keruuta kattavammaksi. Toiseksi identifioinnista saadut tulokset näyttivät, että kaikista tekijöistä ei tarvitse kerätä aineistoa, jos ollaan kiinnostuneita vain tiettyjen päätösten vaikutuksista.

Tässä työssä luodussa kausaaligraafissa on esitetty kompaktissa muodossa päätekijät, joilla arvellaan olevan vaikutusta saatuun satoon. Kausaaligraafi on kuitenkin toteutettu haastatteleamalla vain yhtä asiantuntijaa, joten myös eriäviä näkemyksiä löytyy varmasti toisilta asiantuntijoilta. Yleensäkin kausaaligraafeja voidaan luoda samasta ilmiöstä useampia, joten tämä ei ole sinänsä ongelma. Kausaaligraafia tulisikin muokata aina sovellettavaa tilaa vastaavaksi, perustuen tilanomistajien näkemyksiin. Erityisesti päätöksiä voi olla tehtävissä rajallinen tai suurempi määrä verrattuna haastatellulla tilalla tehtyihin päätöksiin. Esimerkiksi tilan kalusto voisi olla rajallinen, jolloin päätöksiä ei voitaisi tehdä perustuen vaikkapa siihen, millaisella koneistolla viljasato puidaan, kuljetetaan taikka kuivataan. Tällöin kyseisiin muuttujiin tulevat nuolet voitaisiin poistaa kausaaligraafista, ja jotkut identifioituvat kausaalivaikutukset muuttuisivat sievemmiksi.

Koska aineistoa ei ollut kerätty, niin kausaalimallia lähdettiin rakentamaan pääasiassa perustuen haastatellun maanviljelijän näkemyksiin. Lisäksi kausaaligraafi ei valmistunut hetkessä, vaan se vaati useita haastattelukierroksia. Tämän onnistuminen ei ollut itsestäänselvyys, vaan se vaatikin yhteistyökykyä tilastotieteilijän ja maanviljelijän välillä. Ensinnäkin maanviljelijän täytyi osata selittää asioita ymmärrettävällä tasolla, jotta tilastotieteilijä sai peruskäsityksen sovelluksesta. Jos lähdettäisiin yksityiskohdista liikkeelle, niin kokonaiskuvan laatiminen olisi hyvin hankalaa tilastotieteilijälle. Tilastotieteilijältä puolestaan työ vaatii muuttujakohtaista ajattelua ja kausaaligraafin hienosäätöä.

Tosimaailman tapahtumiin liittyy paljon asioita, joita ei havaita. tämän seurauksena on mahdollista, että kausaalimallista puuttuu vielä joitakin ulkopuolisia tekijöitä, joita ei ole keksitty. Näiden olemukseen kannattaisi vähintään tulosten pohdinnassa kiinnittää huomiota. Esimerkiksi jos aineistoa on kerätty eri vuosilta, niin vuoteen liittyviä havaitsemattomia taustatekijöitä on runsaasti.

Tutkimusongelmaa varten selvitettiin, miten kausaalimalli voitaisiin esittää muuttujien avulla ja mitä aineiston kerääminen tästä vaatisi. Saatujen tulosten perusteella kausaalimallinnuksen toteuttaminen olisi mahdollista. Seuraavana vaiheena tämän työn kannalta olisi aineiston kerääminen perustuen luotuun kausaaligraafiin. Aineiston ja tehtyjen identifioituvuustarkastelujen avulla voitaisiin estimoida kiinnostavia kausaalivaikutuksia. Näistä saatuja tuloksia voitaisiin hyödyntää edelleen päätöksenteon apuna tulevaisuuden maanviljelyssä.

Tämän tutkielman päätarkoituksena oli optimoida päätöksentekoa ottamalla huomioon saatu sato ja kustannukset. Nämä eivät kuitenkaan ole suinkaan ainoat mittarit, jotka kuvaavat päätösten hyvyttä. Nimittäin pitemmällä aikavälillä kannattaisi erityisesti huomioida päätösten vaikutus maan kasvukuntoon. Tämä koostuu erilaisista fysikaalisista, biologisista ja kemiallisista tekijöistä. Näihin kuuluvat muun muassa pieneliöiden määrät, maan rakenne ja pH-arvo. Kasvukunnolla on suuri merkitys tulevien satojen määrään ja laatuun. Tämän seurauksena ei välttämättä kannattaisikaan tehdä sellaista päätöstä, joka antaa vain vähän paremman tuoton, mutta heikentää merkittävästi enemmän pellon kasvukuntoa kuin toinen päätös.

## Viitteet

- Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal, Complex Systems*:1695, 2006. URL <https://igraph.org>.
- Eurofins. Viljavuustutkimukset maasta. 2022. URL <https://www.eurofins.fi/agro/analyysit/viljavuustutkimukset-maasta>. Luettu: 13.5.2022.
- Helske, J., Tikka, S., & Karvanen, J. Estimation of causal effects with small data in the presence of trapdoor variables. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 184(3):1030–1051, 2021. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12699>.
- Hernán, M. & Robins, J. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.
- Ilmatieteenlaitos. kasvukausi-2021. 2021. URL <https://www.ilmatieteenlaitos.fi/kasvukausi-2021>. Luettu: 13.5.2022.
- Ilmatieteenlaitos. havaintojen-lataus. 2022. URL <https://www.ilmatieteenlaitos.fi/havaintojen-lataus>. Luettu: 13.5.2022.
- Koller, D. & Friedman, N. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- Luke. Lajikevalinta-työkalu. 2022a. URL <https://maatalousinfo.luke.fi/fi/lajikevalinta>. Luettu: 4.5.2022.
- Luke. Luke ennustaa maatalouden kannattavuuden säilyvän edellisen vuoden tasolla. 2022b. URL <https://www.luke.fi/fi/uutiset/luke-ennustaa-maatalouden-kannattavuuden-sailyvan-edellisen-vuoden-tasolla>. Luettu: 13.5.2022.
- McCullagh, P. *Generalized linear models*. Monographs on statistics and applied probability. Chapman and Hall, London, 2. ed edition, 1989.
- Mennala, A. Yleistä. 2022. URL <http://www.mennala.fi/index.php>. Luettu: 20.4.2022.
- Niemi, J. & Huuskonen, H. Maatilat ovat erityisen riippuvaisia polttoöljystä – miten kävisi, jos tuonti venäjältä katkeaisi? 2022. URL <https://www.luke.fi/fi/blogit/maatilat-ovat-erityisen-riippuvaisia-polttooljysta-miten-kavisi-jos-tuonti-venajalta-katkeaisi>. Luettu: 13.5.2022.

- Partala, A., Sieviläinen, E., & Mäittälä, A. Mihin-me-viljan-kaytamme. 2022. URL <https://www.luke.fi/fi/blogit/mihin-me-viljan-kaytamme>. Luettu: 13.5.2022.
- Pearl, J. *Causality: Models, Reasoning, and Inference*. Cambridge university press, 2009.
- Pearl, J., Glymour, M., & Jewell, N. P. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL <https://www.R-project.org/>.
- Ruusuvuori, J. & Tiittula, L. *Haastattelu. Tutkimus, tilanteet ja vuorovaikutus*. Vastapaino, Tampere, 2005.
- Shpitser, I. & Pearl, J. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1219. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.
- Simpson, E. H. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2):238–241, 1951.
- Tikka, S. & Karvanen, J. Identifying causal effects with the R package causaleffect. *Journal of Statistical Software*, 76(12):1–30, 2017a. URL <https://doi.org/10.18637/jss.v076.i12>.
- Tikka, S. & Karvanen, J. Enhancing identification of causal effects by pruning. *Journal of Machine Learning Research*, 18:194–1, 2017b. URL <https://www.jmlr.org/papers/volume18/17-563/17-563.pdf>.
- Tikka, S. & Karvanen, J. Simplifying probabilistic expressions in causal inference. *Journal of Machine Learning Research*, 18(1):1203–1232, 2017c. URL <https://jmlr.csail.mit.edu/papers/volume18/16-166/16-166.pdf>.
- Tikka, S., Hyttinen, A., & Karvanen, J. Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software*, 99(5):1–40, 2021. URL <https://www.jstatsoft.org/article/view/v099i05>.

Työtehoseura, T. Konetyön kustannukset ja tilastolliset urakointihinnat. 2022. URL [https://www.tts.fi/files/4415/Konetyon\\_kustannukset\\_ja\\_tilastolliset\\_urakointihinnat\\_2021.pdf](https://www.tts.fi/files/4415/Konetyon_kustannukset_ja_tilastolliset_urakointihinnat_2021.pdf). Luettu: 24.5.2022.

yWorks GmbH. yed. URL <https://www.yworks.com/products/yed>.