

Konstruktion eines Planungsmodells für Aussaatzeiten unter Unsicherheit des tatsächlichen Erntezeitpunkts

Jan van Grimbergen

Bachelorarbeit



Algorithmische Bioinformatik
Heinrich-Heine-Universität Düsseldorf
Deutschland

Abgabedatum: 17. März 2021
Erstgutachter: Prof. Dr. Gunnar Klau
Zweitgutachter: Prof. Dr. Egon Wanke

Zusammenfassung

In dieser Arbeit beschreiben wir, angelehnt an die Syngenta Crop Challenge 2021, ein von uns entwickeltes Verfahren in verschiedenen Varianten, um die Aussaatzeiten einer Menge an Saatpopulationen so zu bestimmen, dass diese zu möglichst konstanten und gleichbleibenden Ernteerträgen führen. Im Rahmen der Challenge werden zwei Datensätze zur Verfügung gestellt:

Der erste Datensatz beinhaltet die Saatpopulationen mit ihren Ernteerträgen, möglichen Aussaatzeitfenstern und einem Bedarf an Wachstumsgrad-Werten (GDUs) für jede Population. Sobald nach dem Aussaatzeitpunkt genug GDU-Werte angesammelt werden, sodass der GDU-Bedarf erreicht ist, wird die entsprechende Population geerntet. Der andere Datensatz umfasst die GDU-Werte der Jahre 2009 bis 2019.

Es gibt zwei verschiedene Szenarien: In Szenario 1 ist eine Erntekapazität vorgegeben, in Szenario 2 muss diese berechnet werden. In beiden Szenarien soll die maximale Abweichung und der Median dieser Abweichungen von der wöchentlichen Erntemenge und der Kapazität minimal sein.

Für beide Szenarien stellen wir das von uns entwickelte, dreistufige, heuristische Verfahren vor. Im ersten Schritt berechnen wir die möglichen Erntewochen jeder Population. Schritt 2 verwendet ein gemischt-ganzzahliges lineares Programm, um eine Aufteilung der Populationen auf die möglichen Erntewochen zu berechnen. Schritt 2 im Szenario 2 bestimmt zusätzlich die Kapazität. Wir konnten zeigen, dass das Problem, welches wir im zweiten Schritt lösen, NP-vollständig ist. Im dritten und letzten Schritt wird dann für jede Population ein Aussaatzeitpunkt bestimmt, der zu einer Ernte in der berechneten Woche führen soll.

Unter den verschiedenen Modellvarianten, die wir verglichen haben, liefert diejenige die besten Ergebnisse, welche im ersten Schritt die möglichen Erntewochen mithilfe einer statistischen Methode berechnet und im zweiten Schritt die Stabilität einer Erntewoche mit berücksichtigt hat. Als Stabilität einer Erntewoche bezeichnen wir die Wahrscheinlichkeit, dass für eine Woche und eine Menge an Saatpopulationen, die in dieser Woche geerntet werden soll, jede Population in eben jener Woche geerntet wird.

Inhaltsverzeichnis

1	Einleitung	1
2	Problemdefinition	3
2.1	Deterministische Problemformulierung	4
2.2	Stochastische Problemformulierung	5
2.3	NP-Vollständigkeit des Problems	5
3	Theoretische Grundlagen	6
3.1	Lineare Programmierung	6
3.2	Gemischt-ganzzahlige lineare Programmierung	7
3.3	Der Median in der linearen Programmierung	7
3.4	Support-Vektor-Regression (SVR)	8
3.5	Permutationstest	8
4	Methoden	9
4.1	Die Unsicherheit in den GDUs	10
4.2	Modellbeschreibung	11
4.3	Schritt 1: Berechnung der möglichen Erntewochen	11
4.4	Schritt 2: Populationsaufteilung auf die Erntewochen für SHP	11
4.5	Schritt 1b für SHPNOC	14
4.6	Schritt 2b für SHPNOC	14
4.7	Schritt 2' für SHPNOC	14
4.8	Schritt 3: Ermittlung der Aussaatage	14
5	Ergebnisse	15
5.1	SHP, Standort 0	16
5.2	SHP, Standort 1	19
5.3	SHPNOC, Standort 0	22
5.4	SHPNOC, Standort 1	25
6	Diskussion	28
6.1	Kritik an der Problemstellung	28
6.2	Minimierung des Erwartungswertes	29
6.3	Kritik an der Evaluation und an der Zufallsvariable \mathbb{G}	29
7	Resümee	30

1 Einleitung

Aufgrund des Klimawandels und der rasant wachsenden Weltbevölkerung gewinnt die globale Ernährungssicherung zunehmend an Relevanz. In letzter Zeit haben innovative und neuartige Technologien die Zeit verkürzt, die benötigt wird, um den Landwirten schnellere, ertragreichere und besser angepasste Saatgutoptionen anzubieten [1]. Damit treten allerdings neue Herausforderungen auf, denn bei unzureichender Planung der Aussaatzeiten kann ein höherer Ernteertrag Speicherengpässe verursachen, und verschieden schnell reifende Saatpopulationen können zu Schwankungen in den Erträgen führen. So könnte es vorkommen, dass in einer Woche zu wenig Ernteerträge erzielt werden und daher eine Ernte unwirtschaftlich wäre oder dass große Erntemengen vernichtet werden müssen, da die Speicherkapazitäten ausgereizt sind. Die Planung von Aussaatzeiten spielt also eine wichtige Rolle, wenn man auch in Zukunft die Ernährung der Weltbevölkerung gewährleisten möchte. Mit der Frage der Planung befasst sich auch die *Crop Challenge in Analytics* der Firma Syngenta [1]. Im Rahmen dieses Wettbewerbs wollen wir das Thema näher beleuchten und nach einer Lösung der Problematik suchen.

Unser Forschungsziel ist die Entwicklung eines Planungsmodells zur Berechnung von Aussaatzeitpunkten, das für eine Menge an Saatpopulationen die Aussaatzeitpunkte so bestimmt, dass die Abweichungen von der wöchentlichen Erntekapazität möglichst gering sind.

Das heißt, das Ziel ist die Entwicklung eines Planungsmodells, das es ermöglicht, konstante Ernteerträge einzufahren, welche die wöchentliche Erntekapazität bestmöglich auslasten. Dabei gibt es drei Fehlergrößen, die zu beachten sind. Sowohl der maximale Betrag als auch der Median des Betrags der Differenz aus wöchentlicher Kapazität und geernteten Einheiten in einer Erntewoche sollen minimiert werden. Auch die Anzahl der Wochen, in denen eine Ernte eingefahren wird, soll möglichst klein gehalten werden.

Für die Berechnungen der Aussaatzeitpunkte müssen wir auf die *Wachstumsgrad-Werte* (GDUs) der letzten Jahre zurückgreifen. Jedem Tag und Ort kann ein GDU-Wert zugeordnet werden. Dieser beschreibt, wie groß der Einfluss des Tages und Orts auf das Wachstum von Saatpopulationen ist. Zur Vereinfachung wird davon ausgegangen, dass jede Saatpopulation einen gewissen GDU-Schwellwert bzw. GDU-Bedarf besitzt. Ab dem Zeitpunkt ihrer Aussaat werden alle GDUs der Folgetage aufaddiert, bis der Wert größer oder gleich dem Schwellwert ist. Sobald der Schwellwert erreicht ist, wird die gesamte Population geerntet. Jede Saatpopulation besitzt zudem noch ein bestimmtes Zeitfenster, in dem sie angebaut werden kann. Das Flussdiagramm in Abbildung 1 verdeutlicht diesen Vorgang.

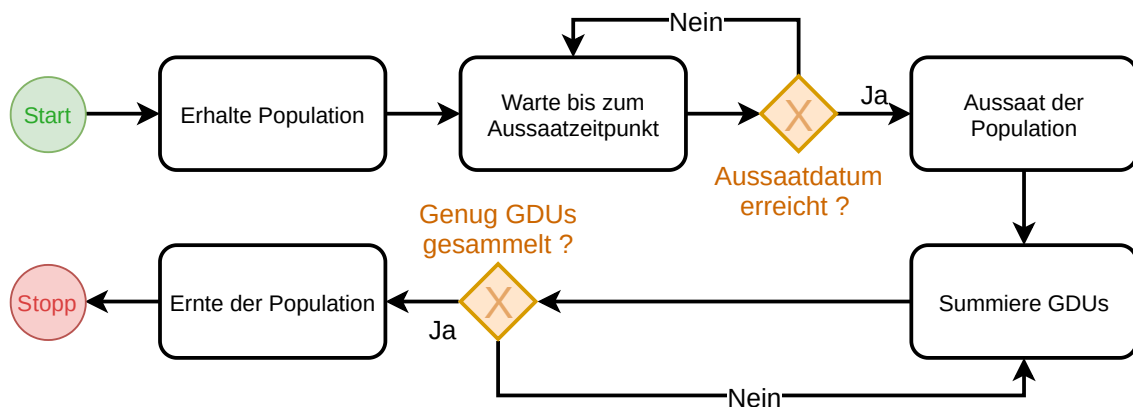


Abbildung 1: Ablaufdiagramm für den Aussaat-Ernteprozess

Die GDUs des Planungsjahres sind nicht bekannt. Eine Voraussetzung an das Modell ist, dass die Planung gute Ergebnisse unter Unsicherheit der GDUs des Planungsjahres liefert. Für die Challenge sind zwei Szenarien vorgegeben. In *Szenario 1* ist die wöchentliche Erntekapazität bekannt. In *Szenario 2* soll eine gute Erntekapazität ermittelt werden. Entsprechend der zwei Szenarien wollen wir auch zwei Modelle auf der Grundlage von gemischt-ganzzahligen linearen Programmen entwickeln.

Das Modell zu Szenario 1 besteht aus drei Stufen: die Berechnung der möglichen Erntewochen einer Saatpopulation, die Berechnung einer bestmöglichen Aufteilung der Saatpopulationen auf die möglichen Erntewochen und die Berechnung des besten Aussaatzeitpunkts für eine Population. Wir konnten zeigen, dass das Finden der bestmöglichen Aufteilung der Saatpopulationen auf die möglichen Erntewochen, d. h. die Wochen so aufzuteilen, dass der maximale Betrag und der Median des Betrags der Differenzen zwischen Kapazität und nicht leerer Erntewoche minimal ist, ein NP-vollständiges Problem ist.

Für Szenario 2 stellen wir zwei Modelle vor, die Modifikationen des Modells für Szenario 1 sind. Die eine Variante verwendet einen Ansatz, bei der die Kapazität und die Aufteilung nacheinander berechnet werden und sich gegenseitig updaten. Die andere Variante ermittelt die Aufteilung und die Kapazität direkt.

Die Arbeit ist wie folgt aufgebaut : In Kapitel 2 beschreiben wir das Problem sowie unsere Zielsetzung genauer und stellen dar, wie die Daten beschaffen sind, die uns von Syngenta im Rahmen der Challenge zur Verfügung gestellt wurden. Zudem definieren wir in diesem Kapitel die Zielfunktion und setzen damit unser Evaluationskriterium fest. Ebenfalls beweisen wir die NP-Vollständigkeit der Aufteilung der Populationen auf die möglichen Erntewochen.

In Kapitel 3 geben wir einen Überblick der theoretischen Grundlagen dieser Bachelorarbeit, welcher unter anderem benötigt wird, um das Thema umfassend zu verstehen. Kapitel 4 beschreibt die von uns entwickelten mehrstufigen Modelle für Szenario 1 und Szenario 2.

Die von unseren Modellen erzielten Ergebnisse werden in Kapitel 5 vorgestellt. Dazu vergleichen wir bei Szenario 1 und 2 die zu erwartende Fehlergröße der verschiedenen Modellvarianten untereinander. Unsere Forschung, die verwendeten Methoden und ihre Ergebnisse reflektieren wir kritisch in Kapitel 6.

Im letzten Abschnitt fassen wir die von uns erzielten Ergebnisse zusammen und geben einen Ausblick auf mögliche, zukünftige Arbeiten.

2 Problemdefinition

Zunächst wollen wir zusammenfassen, welche Daten uns Syngenta zur Verfügung stellt. Anschließend werden wir eine deterministische und daraus ableitend eine stochastische Definition des Problems formulieren. Syngenta stellt uns ein Planungsjahr, eine Anzahl an Erntewochen und zwei Datensätze zur Verfügung:

- Mithilfe des Planungsjahres, welches bei uns 2020 ist, und der Anzahl der Erntewochen lässt sich der Planungszeitraum bestimmen. Die erste Woche eines Planungsjahres soll diejenige sein, die den ersten Januar des Planungsjahres beinhaltet. Es soll davon ausgegangen werden, dass Sonntag der erste Tag einer Woche ist. Daraus lässt sich der Planungszeitraum vom 1. Januar 2020 bis zum 1. Mai 2021 bestimmen. Dieser umfasst 487 Tage. Wir definieren:
 - die Menge der Erntewochen w des Planungsjahres bezeichnen wir als W ,
 - die Daten einer Woche als das Intervall $\mathcal{D}_w = [f_w, l_w]$, wobei f_w der erste und l_w den letzten Wochentag darstellt,
 - die Menge aller Tage des Planungszeitraums als $\mathcal{D}_y = \{d_1, \dots, d_m\}$ mit $m \in \mathbb{N}$. In unserem konkreten Fall ist $\mathcal{D}_y = \{d_1, \dots, d_{487}\}$.
- Ein Datensatz umfasst die historischen GDU-Werte für die gegebenen Standorte 0 und 1 im Zeitraum 2009 bis 2019. Wir halten unsere Formulierung unabhängig vom Standort, weshalb im Folgenden nicht zwischen den Orten unterschieden wird. Des Weiteren bekommen wir für jeden Standort eine Erntekapazität c , welche für die Problemformulierung in Szenario 1 relevant ist. Bei Standort 0 ist $c = 7000$, bei Standort 1 ist $c = 6000$.
 - $\text{GDU}_y : \mathcal{D}_y \mapsto \mathbb{R}$ ordnet jedem Tag aus \mathcal{D}_y einen GDU-Wert zu.
 - Wir schreiben $G = (G_1, \dots, G_m) = (\text{GDU}_y(d_1), \dots, \text{GDU}_y(d_m))$ für die GDU-Werte des Planungszeitraums. In unserem konkreten Fall ist $G = (G_1, \dots, G_{487})$.

Abbildung 2 zeigt ein Beispiel für die GDU Verteilung an den beiden Standorten 0 und 1. Man erkennt, dass die GDU-Daten an dem Standort 0 innerhalb eines Jahres wesentlich größeren Schwankungen unterliegen als die GDU-Daten an Standort 1. Beide sind periodisch und unterliegen einem Rauschen. Es sei angemerkt, dass es sich bei den im Plot dargestellten Daten aus Datenschutzgründen nicht um die Originaldaten handelt, sondern um simulierte, die ähnliche Eigenschaften aufweisen wie die Originalen.

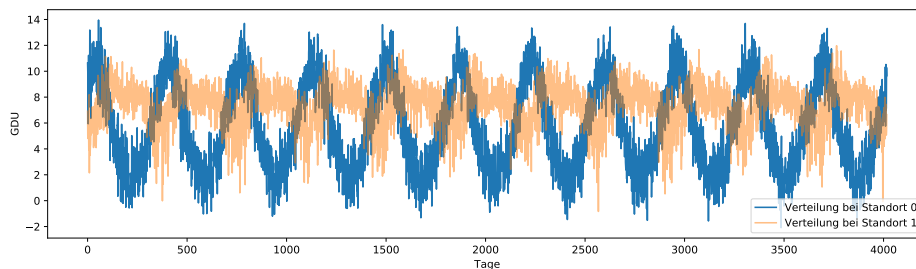


Abbildung 2: Simulierte GDU-Verteilung über mehrere Jahre für die Standorte

- Der zweite Datensatz, den wir von Syngenta erhalten, umfasst die Informationen zu den verschiedenen Saatpopulationen. Zu jeder Saatpopulation bekommen wir den frühest- und spätestmöglichen Aussaatzeitpunkt, wie auch einen ursprünglich verwendeten Aussaatzeitpunkt. Wir erhalten

die Information, wie groß der Ernteertrag einer Population und wie hoch ihr GDU-Bedarf ist. Genau so erfahren wir, an welchem Ort (Standort 0 oder Standort 1) und in welchem Szenario sie ausgesät werden soll.

- Sei \mathcal{P} die Menge der Saatpopulationen $p \in \mathcal{P}$, die ausgesät werden sollen.
- Als q_p bezeichnen wir die Erntemenge einer Population.
- Das Intervall $\hat{d}_p = [e_p, l_p]$ beinhaltet alle möglichen Aussaatdaten der Population p wobei e_p der früheste Aussaatzeitpunkt und l_p der späteste ist.
- Das Datum o_p repräsentiert den ursprünglichen Aussaatzeitpunkt der Population.
- Die Variable r_p beschreibt, wie viele GDU-Werte vom Aussaatzeitpunkt an akkumuliert werden müssen, bis die Population geerntet wird. Die Population muss auch in derselben Woche geerntet werden, in der dieser Wert erreicht wird.

2.1 Deterministische Problemformulierung

Für die deterministische Formulierung des Problems gehen wir davon aus, dass die GDU-Werte $G = (G_1, \dots, G_m)$ des Planungszeitraums bekannt sind. In Problem 1 beschreiben wir das deterministische Problem zu Szenario 1, in Problem 2 das zu Szenario 2.

Problem 1 (DETERMINISTISCHE ERNTEPLANUNG, DHP). Anders als bei der späteren Problemdefinition von Szenario 2 haben wir in Szenario 1 eine Kapazität c gegeben. Das Problem besteht darin, für eine Menge an Saatpopulationen $\mathcal{P} = \{p_1, \dots, p_n\}$ und die gegebenen GDU-Werte $G = (G_1, \dots, G_m)$ im Planungszeitraum ein Tupel an Aussaatagen

$$d^{\mathcal{P}} = (d_1, \dots, d_n) \quad (1)$$

zu bestimmen, wobei d_i der Aussaatag für p_i ist. Das Ziel dabei ist, dass d die Bewertungsfunktion $f(d, G, c)$ minimiert. Die Wahl eines Aussaatdatums legt im deterministischen Modell fest, wann eine Population geerntet wird. Sei $Q_w(d, G)$ für $w \in W$ die Erntemenge, die in der Woche w erzielt wird, wenn man die GDU-Werte G und Aussaatdaten d annimmt. Die Funktion f soll das Maximum und den Median der absoluten Differenzen aus wöchentlicher Ernte $Q_w(d, G)$ und Kapazität c , aber auch die Anzahl der tatsächlichen Erntewochen $|Q_w(d, G)| > 0$ berücksichtigen. Dabei sollen auch bei der maximalen Differenz und dem Median der Differenzen nur die Wochen berücksichtigt werden, in denen die Erntemenge nicht leer ist. Wir definieren dafür die Konsistenzabweichung einer Woche w als

$$\Delta_C(w, c, d, G) = \begin{cases} 0 & \text{falls } Q_w(d, G) = 0, \\ |c - Q_w(d, G)| & \text{sonst} \end{cases} \quad (2)$$

Die Funktion f wählen wir dann als

$$f(d, G, c) = \alpha \max_{w=1}^{70} \Delta_C(w, c, d, G) + (1 - \alpha) \max_{w=1 | Q_w(d, G) \neq 0}^{70} \Delta_C(w, c, d, G) + \beta |\{Q_w(d, G) > 0 \mid w = 1, \dots, 70\}| \quad (3)$$

Das Gewicht $\alpha \in [0, 1]$ legt fest, wie stark der maximale Fehler gegenüber dem Median der Fehler gewichtet wird. Der Parameter β legt fest, wie die Anzahl der tatsächlichen Erntewochen gegen die Fehler gewichtet wird. Ausgehend von der Crop Challenge wählen wir $\alpha = \frac{1}{2}$ und $\beta = \frac{1}{|W|}$, was bei uns im Konkreten $\frac{1}{70}$ entspricht. Wir wählen β auf diese Weise, da $|W|$ eine obere Schranke für die Anzahl der möglichen Erntewochen ist und durch diese Wahl vorrangig der Fehler

$$\alpha \max_{w=1}^{70} \Delta_C(w, c, d, G) + (1 - \alpha) \max_{w=1}^{70} \Delta_C(w, c, d, G)$$

minimiert wird und erst zweitrangig die Anzahl der tatsächlichen Erntewochen.

Problem 2 (DETERMINISTISCHE ERNTEPLANUNG MIT UNBEKANNTER KAPAZITÄT, DHPNOC). Seien die GDU-Werte $G = (G_1, \dots, G_m)$ für den Planungszeitraum und eine Menge an Saatpopulationen $\mathcal{P} = \{p_1, \dots, p_n\}$ gegeben. Anders als Problem 1 ist in dieser Variante des Problems die Kapazität c nicht bekannt. Deshalb ist das Ziel, ein Tupel an Aussaatdaten $d^{\mathcal{P}}$ und eine Kapazität c zu bestimmen, sodass $f(d, G, c)$ minimal ist. Die Formulierungen sind dabei dieselben wie aus Problem 1.

2.2 Stochastische Problemformulierung

Auch wenn die Vorhersage des Wetters immer genauer wird, hängt es von unzähligen Prozessen ab und lässt sich für einen längeren Zeitraum nicht vorhersagen, weshalb wir die Werte $G = (G_1, \dots, G_m)$, anders als in Abschnitt 2.1 angenommen, nicht kennen. Aus diesem Grund nehmen wir an, dass sich die GDU-Werte des Planungszeitraums durch die Zufallsvariable $\mathbb{G} = (\mathbb{G}_1, \dots, \mathbb{G}_m)$ beschreiben lassen. Weiterhin gehen wir zur Vereinfachung davon aus, dass \mathbb{G}_i unabhängig von \mathbb{G}_j ist für $i \neq j$. Den Erwartungswert einer Zufallsvariable \mathbb{X} bezeichnen wir mit $\mathbb{E}[\mathbb{X}]$. In Problem 3 beschreiben wir das stochastische Problem zu Szenario 1, in Problem 4 das zu Szenario 2.

Problem 3 (STOCHASTISCHE ERNTEPLANUNG, SHP). Da wir die GDU-Werte G des Planungszeitraums nicht kennen, wollen wir anstelle der Aussaatdaten $d^{\mathcal{P}} = (d_1, \dots, d_n)$, die $f(d^{\mathcal{P}}, G, c)$ minimieren, die Aussaatdaten $d^{\mathcal{P}'} = (d'_1, \dots, d'_n)$ finden, die den Erwartungswert $\mathbb{E}[f(d^{\mathcal{P}'}, \mathbb{G}, c)]$ minimieren.

Problem 4 (STOCHASTISCHE ERNTEPLANUNG MIT UNBEKANNTER KAPAZITÄT, SHPNOC). Analog zu Problem 2 und Problem 3 suchen wir hier die Aussaatdaten $d' = (d'_1, \dots, d'_n)$ und die Kapazität, die $\mathbb{E}[f(d', \mathbb{G}, c)]$ minimieren.

2.3 NP-Vollständigkeit des Problems

Wir zeigen, dass das zu dem Optimierungsproblem DHP gehörige Entscheidungsproblem NP-vollständig ist. Dazu geben wir an, wie eine Lösung in polynomieller Zeit überprüft werden kann und reduzieren das NP-vollständige Partitionsproblem [2] auf unser Problem. Unter Kenntnis der GDU-Werte des Planungszeitraums lässt sich leicht berechnen, in welchen Wochen man die Population $p \in \mathcal{P}$ ernten kann. Das Problem DHP ist daher äquivalent zu dem folgenden Optimierungsproblem:

Problem 5. INSTANZ: Gegeben seien Populationen $\mathcal{P} = (p_1, \dots, p_n)$, Wochen $W = (w_1, \dots, w_u)$, die Wochen \mathcal{H}_p , in denen man eine Population p ernten kann, die Erntemengen q_p für jede Population p und eine Kapazität c .

ZIEL: Sei h_i die Erntemenge in der Woche i . Minimiere die Funktion

$$\alpha \max_{i: h_i \neq 0} |C - h_i| + (1 - \alpha) \max_{i: h_i \neq 0} |C - h_i| + \beta |\{h_i | h_i \neq 0\}|. \quad (4)$$

Wir setzen wie im Problem DHP $\alpha = \frac{1}{2}$ und $\beta = \frac{1}{u}$.

Das zu dem Optimierungsproblem gehörige Entscheidungsproblem ergibt sich dann wie folgt:

Problem 6 (PARTITION SAAT PACKPROBLEM, PSP). INSTANZ: Gegeben seien Populationen $\mathcal{P} = (p_1, \dots, p_n)$, Wochen $W = (w_1, \dots, w_u)$, die Wochen \mathcal{H}_p in denen man eine Population p ernten kann, die Erntemengen q_p für jede Population p eine Kapazität c und ein $k \in \mathbb{N}$.

FRAGE: Seien h_i die Erntemenge in der Woche i die aus der Aufteilung der Populationen auf die Erntewochen resultieren. Gibt es eine Aufteilung der Populationen auf die Erntewochen, sodass

$$\alpha \max_{i: h_i \neq 0} |C - h_i| + (1 - \alpha) \max_{i: h_i \neq 0} |C - h_i| + \beta |\{h_i | h_i \neq 0\}| < k.$$

Wir setzen wieder wie im Problem DHP $\alpha = \frac{1}{2}$ und $\beta = \frac{1}{W}$.

Wir betrachten jetzt das Partitionsproblem, welches folgendermaßen definiert ist:

Problem 7 (PARTITION). INSTANZ: Gegeben seien Populationen $\mathcal{P} = (p_1, \dots, p_n)$ und eine Erntemenge $q_p \in \mathbb{N}$ für jede Population.

FRAGE: Gibt es eine Teilmenge $\mathcal{P}' \subset \mathcal{P}$ die erfüllt, dass:

$$\sum_{p \in \mathcal{P}'} q_p = \sum_{p \in \mathcal{P} \setminus \mathcal{P}'} q_p. \quad (5)$$

Satz 2.1. Das Problem PSP ist NP-vollständig.

Beweis. Zunächst zeigen wir, dass das Problem in NP liegt, indem wir angeben, wie man für eine gegebene Instanz eine Lösung in polynomieller Zeit überprüfen kann. Sei also eine Partition $\mathcal{P}^i \subseteq \mathcal{P}$ für $i = 1, \dots, W$ gegeben mit $\cup_{i=1}^W \mathcal{P}^i = \mathcal{P}$ und mit $\mathcal{P}^i \cap \mathcal{P}^j = \emptyset$ für $i \neq j$ gegeben. Die Erntemenge h_i der Woche w_i berechnet sich dann als

$$h_i = \sum_{p \in \mathcal{P}^i} q_p.$$

Der Wert der Zielfunktion (4) kann dann in polynomieller Zeit bestimmt werden, da Maximum und Median in polynomieller Zeit ermittelt werden können. Also kann auch in polynomieller Zeit überprüft werden, ob

$$\alpha \max_{i: h_i \neq 0} |c - h_i| + (1 - \alpha) \text{median}_{i: h_i \neq 0} |c - h_i| + \beta |\{h_i \mid h_i \neq 0\}| < k.$$

Das Problem PSP liegt also in NP. Wir zeigen nun noch, dass PSP auch NP-schwer ist, indem wir das Partitionsproblem auf PSP reduzieren. Sei eine Instanz des Partitionsproblems, bestehend aus \mathcal{P} und Erntemengen q_p , gegeben. Wir bauen eine Instanz für PSP. Sei $W = (w_1, w_2, w_3)$. Sei $\mathcal{H}\mathcal{W}_p = (w_1, w_2)$ für alle Populationen p . Als Kapazität setzen wir $c = \sum_{p \in \mathcal{P}} q_p / 2$. Dabei können wir annehmen, dass $\sum_{p \in \mathcal{P}} q_p$ gerade ist, da es sich sonst bei der, für das Partitionsproblem gegebenen Instanz, um eine Nein-Instanz handelt. Wir setzen $k = 1$. Dann löst PSP das Partitionsproblem, denn die Zielfunktion (4) ist genau dann kleiner als 1, wenn

$$\max_{i: h_i \neq 0} |c - h_i| + \text{median}_{i: h_i \neq 0} |c - h_i| = 0.$$

Dieser Ausdruck ist äquivalent dazu, dass die Populationen genau so auf die Wochen w_1 und w_2 aufgeteilt wurden, dass die Erntemenge in den Wochen genau der Kapazität entspricht, also sind die Erntemengen in den beiden Wochen gleich. \square

3 Theoretische Grundlagen

3.1 Lineare Programmierung

Optimierungsprobleme gibt es überall, sie sind wichtig für die Landwirtschaft, für die Gewinnmaximierung oder Berechnung von Transportwegen. Wir wollen an dieser Stelle lineare Programme (LP) vorstellen. Viele Anwendungsprobleme lassen sich als lineares Programm darstellen [3].

Definition 3.1. Sei $c, x \in \mathbb{R}^n$ für $n \in \mathbb{N}$. Sei $A \in \mathbb{R}^{n \times m}$ mit $m \in \mathbb{N}$ eine Matrix und $b \in \mathbb{R}^m$. Ein LP besteht aus einer Zielfunktion für $x \in \mathbb{R}^n$, welche durch $c^T x$ beschrieben wird und aus Bedingungen an die Lösung, welche durch $Ax \leq b$ definiert werden. Die Form eines linearen Programms ist durch

$$\min\{c^T x \mid Ax \leq b, x \geq 0\} \quad (6)$$

gegeben. Die Menge der möglichen Lösungen bezeichnen wir mit

$$LP := \{x \mid Ax \leq b, x \geq 0\}. \quad (7)$$

Es gibt mehrere Möglichkeiten ein lineares Programm anzugeben, jedoch sind alle Formen äquivalent und können ineinander überführt werden. Die Ungleichung $x \geq 0$ müssen wir dabei komponentenweise verstehen d. h. für alle $i \in \{1, \dots, n\}$ gilt $x_i \geq 0$. Genauso verstehen wir die Ungleichung $Ax \leq b$ komponentenweise, d. h. für alle $i \in \{1, \dots, n\}$ gilt $(Ax)_i \leq b_i$.

Lineare Programme lassen sich exakt mithilfe des Simplex-Verfahrens oder anderer Verfahren lösen [4].

3.2 Gemischt-ganzzahlige lineare Programmierung

Wir stellen die gemischt-ganzzahligen Programme vor, mit der viele Optimierungsprobleme modelliert und gelöst werden können [5].

Definition 3.2. Sei $c \in \mathbb{R}^N$, seien $N_1, N_2, N, W \in \mathbb{N}$ mit $N_1 + N_2 = N$. Sei $A \in \mathbb{R}^{W \times N}$ und $b \in \mathbb{R}^W$. Ein MILP besteht aus einer Zielfunktion für $x \in \mathbb{R}^{N_1} \times \mathbb{N}^{N_2}$, welche durch $c^T x$ beschrieben wird und aus Bedingungen an die Lösung, welche durch $Ax \leq b$ definiert werden. Das Problem ist dann definiert durch:

$$\min\{c^T x \mid Ax \leq b, x \in \mathbb{R}^{N_1} \times \mathbb{N}^{N_2}\}. \quad (8)$$

Der Lösungsraum wird beschrieben durch

$$LP = \{x \in \mathbb{R}^{N_1} \times \mathbb{N}^{N_2} \mid Ax \leq b\}. \quad (9)$$

Es muss angemerkt werden, dass die Ungleichung $Ax \leq b$, genau wie in Definition 3.1, komponentenweise zu verstehen ist. Es gibt viele Möglichkeiten, ein MILP anzugeben, und alle sind äquivalent.

Es ist deutlich schwieriger, ein MILP zu lösen, als ein lineares Programm. Hierzu verwendet man eine *Branch-And-Bound*-Methode, welche intern viele lineare Programme löst. Dazu kann das Simplex-Verfahren verwendet werden. Ein bekannter Algorithmus, der MILPs exakt löst, ist zum Beispiel das DAKIN-Verfahren [5].

3.3 Der Median in der linearen Programmierung

Auch der Median einer Menge von beschränkten Werten lässt sich als Optimierungsproblem beschreiben, wie der folgende Satz zeigt.

Satz 3.3. Sei eine Menge von beschränkten Werten p_1, \dots, p_N gegeben mit $0 \leq p_i \leq M$.

Es sind $M, N, p_1, \dots, p_N \in \mathbb{N}$ und N ungerade. Dann entspricht

$$\min_{z \in \{0,1\}^N \text{ mit } \sum z_i = \lfloor N/2 \rfloor} \max\{p_i - Mz_i : i = 1, \dots, N\} \quad (10)$$

dem Median der Werte p_1, \dots, p_N .

Beweis. Sei q_{j_1}, \dots, q_{j_N} mit $j_k = j_g$ für $g \neq k$ und $j_k \in 1, \dots, N$ eine absteigend sortierte Folge der Werte q_1, \dots, q_N . Das Maximum $\max\{q_i - Mz_i : i = 1, \dots, N\}$ entspricht dem maximalen Wert aus $\{q_i : z_i = 0\}$. Dieser maximale Wert wird minimal, falls $z_{j_1}, \dots, z_{j_{\lfloor \frac{N}{2} \rfloor}} = 1$. Daraus folgt $q_{j_k} - Mz_{j_k} \leq 0$ für $k = 1, \dots, \lfloor \frac{N}{2} \rfloor$, weiterhin gilt $z_{j_{\lfloor \frac{N}{2} \rfloor + 1}} = 0$ weshalb $q_{j_{\lfloor \frac{N}{2} \rfloor + 1}}$ dem Maximum von $\max\{q_i - Mz_i : i = 1, \dots, N\}$ entspricht. $q_{j_{\lfloor \frac{N}{2} \rfloor + 1}} = q_{j_{\lfloor \frac{N}{2} \rfloor}}$ entspricht genau dem Median von q_1, \dots, q_N . \square

Für eine gerade Anzahl an Werten, über die der Median minimiert werden soll, begnügen wir uns damit den Wert $q_{j, \frac{N}{2}+1}$ als Median zu wählen.

Bemerkung 3.4. Sei nun ein Problem gegeben mit Bedingungen an die Werte p_1, \dots, p_N und mit den Werten $N \in \mathbb{N}$ und M , eine obere Schranke für die p_i . Der minimale Median der Werte p_1, \dots, p_N entspricht dann:

$$\min_{v \in \mathbb{R}} v \quad (11)$$

$$\text{u.d.N. : } p_i - Mz_i \leq v \quad \forall i \in \{1, \dots, N\} \quad (12)$$

$$\sum_{i=1}^N z_i = \lfloor \frac{N}{2} \rfloor \quad (13)$$

$$\text{weitere Bedingungen an die } p_1, \dots, p_N \quad (14)$$

$$z \in \{0, 1\}^N \quad (15)$$

$$0 \leq p_i \leq M \quad \forall i \in \{1, \dots, N\} \quad (16)$$

3.4 Support-Vektor-Regression (SVR)

Support-Vektor-Maschinen sind eine wichtige Methode aus dem Bereich des maschinellen Lernens. Zur Einordnung von Datenpunkten in Klassen konstruiert bzw. trainiert man ein mathematisches Modell mithilfe von Datenpunkten und deren zugeordneten Klassen. Anschließend kann dann einem beliebigen Datenpunkt eine Klasse, basierend auf dem trainierten Modell, zugeordnet werden. SVMs kann man auch nutzen, um Datenpunkten Werte zuzuordnen. Das nennt sich Support-Vektor-Regression [6].

3.5 Permutationstest

Ein Permutationstest ist eine spezielle Methode der Statistik, um den Mittelwert zweier Stichproben D_x und D_y beliebiger Größe und mit unbekannten Verteilungen F_x und F_y zu einem Signifikanzniveau $\alpha \in [0, 1]$ vergleichen zu können [7]. Dazu verwendet man eine, von der Reihenfolge unabhängige Teststatistik $t(D_x, D_y)$, wie zum Beispiel

$$t(D_x, D_y) = \sum_{a \in D_x} \frac{a}{|D_x|} - \sum_{b \in D_y} \frac{b}{|D_y|}. \quad (17)$$

Die Idee hinter dieser Testmethode ist, den Wert der Teststatistik für die Stichproben D_x und D_y in Relation zu der Verteilung von t zu setzen, unter Annahme der Nullhypothese, dass D_x und D_y Stichproben derselben Zufallsvariable sind. Die alternative Hypothese ist dann, dass sich der Erwartungswert von F_x und F_y unterscheidet, also ist die Alternativhypothese:

1. $\mathbb{E}[F_x] \neq \mathbb{E}[F_y]$ oder
2. $\mathbb{E}[F_x] > \mathbb{E}[F_y]$ oder
3. $\mathbb{E}[F_x] < \mathbb{E}[F_y]$.

Um die Verteilung der Teststatistik unter der Annahme der Nullhypothese zu bauen, berechnet man für alle möglichen Permutationen von $D_x \cup D_y$ in zwei Stichproben D'_x und D'_y der Größe D_x und D_y , den Wert der Teststatistik $t^i := t(D'_x, D'_y)$, wobei $i = 1, \dots, \binom{|D_x|+|D_y|}{|D_x|}$. Da $R := \binom{|D_x|+|D_y|}{|D_x|}$ meistens sehr große Werte annimmt, wird in der Praxis oft anstelle aller Permutationen eine große Anzahl an Permutationen betrachtet, welche zufällig gebildet werden, zum Beispiel durch Ziehen ohne Zurücklegen. Unter der Annahme, dass die Nullhypothese wahr ist, sind alle so gebildeten Permutationen gleich wahrscheinlich.

Der Wert von $t(D_x, D_y)$ wird dann mit den Werten t^i verglichen. Falls wir die alternative Hypothese $\mathbb{E}[F_x] \neq \mathbb{E}[F_y]$ gewählt haben, müssen wir die Nullhypothese verwerfen, falls $t(D_x, D_y)$ größer als mindestens oder kleiner als maximal $100(1 - \frac{\alpha}{2})\%$ der Werte t^i ist und nehmen die Alternativhypothese an. Falls wir die Alternativhypothese $\mathbb{E}[F_x] > \mathbb{E}[F_y]$ wählen, verwerfen wir die Nullhypothese, falls $t(D_x, D_y)$ größer als mindestens $100(1 - \alpha)\%$ der Werte t^i ist. Analog wird die Nullhypothese bei der Alternativhypothese $\mathbb{E}[F_x] < \mathbb{E}[F_y]$ verworfen und die Alternativhypothese angenommen.

4 Methoden

In diesem Kapitel wollen wir die von uns entwickelten Planungsmodelle und ihre Funktionsweisen näher erläutern. Zunächst betrachten wir DHP (Problem 1) und zeigen eine Möglichkeit, dieses exakt zu lösen. Daraus entwickeln wir ein dreistufiges Planungsmodell für SHP (Problem 3). Aus den Überlegungen ergeben sich dann weiter zwei Planungsmodelle für SHPNOC (Problem 4).

Bemerkung 4.1. Angenommen man kennt die GDUs $G = (G_1, \dots, G_m)$ des Planungsjahres, dann kann man für jede Saatpopulation $p \in \mathcal{P}$ die Menge der möglichen Erntewochen $\mathcal{HW}_p \subseteq W$ einfach bestimmen, indem man für die Population p und für jeden ihrer Aussaatage $d \in [e_p, l_p]$ den Tag d^H ermittelt, sodass

$$\sum_{t=d}^{d^H} G_t \geq r_p \wedge \sum_{t=d}^{d^H-1} G_t < r_p. \quad (18)$$

Die Erntewoche $w \in W$, in der d^H liegt, wird anschließend zur Menge \mathcal{HW}_p hinzugefügt. Kennt man die Mengen \mathcal{HW}_p aller so berechneten möglichen Erntewochen jeder Population $p \in \mathcal{P}$, dann lässt sich DHP exakt lösen. Zunächst berechnet man die bestmögliche Aufteilung der Saatpopulationen \mathcal{P}_y auf die Erntewochen \mathcal{W}_y . Das setzen wir mithilfe eines ganzzahligen linearen Programms um. Anschließend kann man auch die Aussaatage berechnen.

Sind die GDU-Werte eines Planungsjahres bekannt, lässt sich mit dem vorgestellten Vorgehen eine perfekte Planung der Erntetage in Szenario 1 erzeugen.

Bemerkung 4.2. Wie bereits erwähnt, kennen wir die GDUs des Planungsjahres nicht, wollen diese Vorgehensweise im Wesentlichen dennoch für unser Planungsmodell verwenden. Wir werden unser Modell für SHP deshalb bestehend aus den folgenden drei Hauptschritten konstruieren:

1. Berechnung der möglichen Erntewochen \mathcal{HW}_p einer Population $p \in \mathcal{P}$.
2. Berechnung der bestmöglichen Aufteilung der Populationen p auf die möglichen Erntewochen \mathcal{HW}_p , d. h. , berechne, in welcher Woche $w \in \mathcal{HW}_p$ die Population p geerntet werden soll.
3. Berechnung des Tupels der Aussaatage $d^{\mathcal{P}} = (d_1, \dots, d_n)$ (wie in Abschnitt 2.1 beschrieben).

Anders als in SHP ist die Kapazität in SHPNOC eine Variable. Ein Modell für SHPNOC welches wir direkt aus dem für SHP ableiten, ist, dass wir eine initiale Kapazität geben und nach dem Optimierungsschritt in Schritt 2 des Modells zu SHP die Kapazität wieder anpassen. Dieser Zyklus wird so lange wiederholt, bis sich keine wesentliche Veränderung der Kapazität feststellen lässt. In unseren Berechnungen war die Veränderung der Kapazität bereits nach drei Durchläufen so klein, dass der Zyklus terminierte. Das andere Modell für SHPNOC verfolgt den Ansatz, anstelle des Lösen von DHP in Schritt 2, das Problem DHPNOC zu lösen. Abbildung 3 zeigt den Ablauf der Planungsmodelle zu den Problemen SHP und SHPNOC.

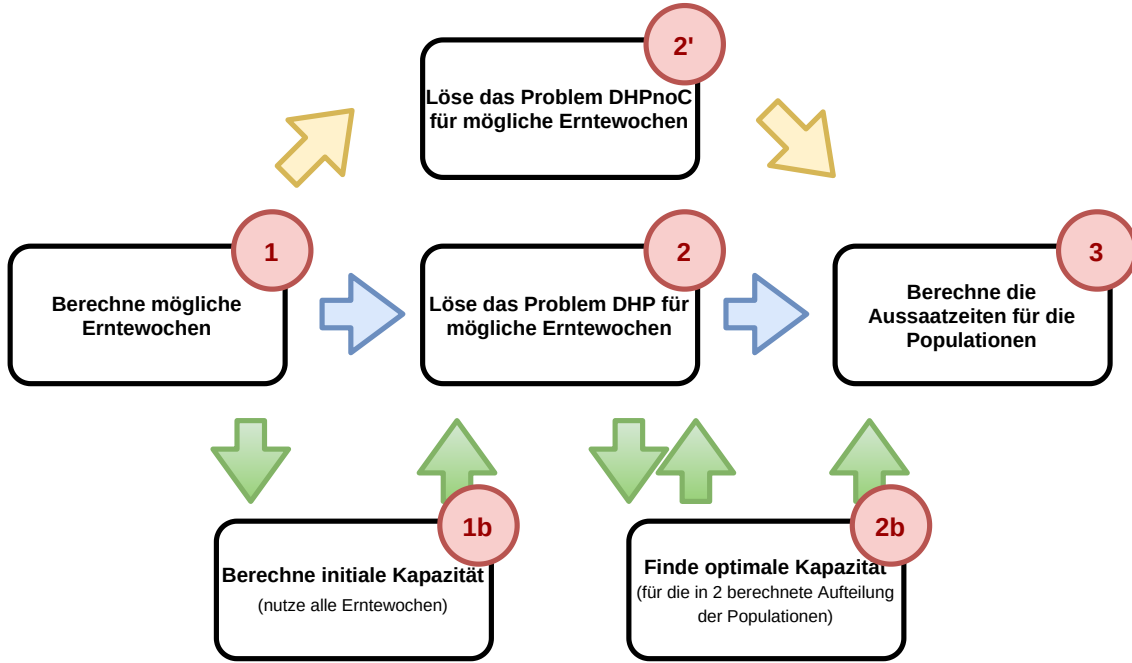


Abbildung 3: Ablaufdiagramm des Modells zu SHP (blaue Pfeile) und der beiden Modelle zu SHPNOC (gelbe und grüne Pfeile)). Für SHP durchlaufen wir die Schritte 1, 2, 3. Das eine Planungsmodell für SHPNOC durchläuft die Schritte 1, 1b, $(2, 2b)^k, 3$ für ein $k \in \mathbb{N}$. Das soll bedeuten, dass die Schritte 2 und 2b genau k mal abwechselnd durchlaufen werden. Das andere Modell zu SHPNOC durchläuft die Schritte 1, 2' und 3.

4.1 Die Unsicherheit in den GDUs

Wir wollen, dass unser Modell Pflanzdaten $d^{\mathcal{P}} = (d_1, \dots, d_n)$ bestimmt, die zu einem möglichst kleinen Erwartungswert der Zufallsvariable $f(d^{\mathcal{P}}, \mathbb{G}, c)$ führen. Allerdings kennen wir die Verteilung der Zufallsvariable $\mathbb{G} = (\mathbb{G}_1, \dots, \mathbb{G}_m)$ nicht. Um das Modell und seine Güte einordnen zu können, besitzen wir die GDU-Verteilungen von 11 Jahren und somit nur eine Stichprobe der Größe 11 von Werten der Zufallsvariable \mathbb{G} . Generell werden Modelle die Probleme SHP und SHPNOC nur näherungsweise lösen können, da wir nicht genug Informationen über die Zufallsvariable \mathbb{G} besitzen. Wenn man annimmt, dass die GDU-Werte der einzelnen Tage unabhängig voneinander sind und $\mathbb{G}_{i-1} \approx \mathbb{G}_i \approx \mathbb{G}_{i+1}$ ist, dann können wir anstelle der 11 GDU-Verteilungen 33^m GDU-Verteilungen für unsere Berechnungen nutzen. In unserem konkreten Fall also 33^{487} . Wir werden im Folgenden immer mit dieser Vergrößerung der Stichprobe arbeiten.

Sei $S = (s_1, \dots, s_{1000})$ eine Stichprobe von Werten aus \mathbb{G} . Dabei ist $s_i = (G_1^{s_i}, \dots, G_m^{s_i})$ wobei m die Anzahl der Tage im Planungszeitraum ist und $G_j^{s_i}$ ein GDU-Wert für den j -ten Tag im Zeitraum. Auf Grundlage dieser Stichprobe lässt sich der Erwartungswert der Zufallsvariable $f(d^{\mathcal{P}}, \mathbb{G}, c)$ approximieren, indem wir die folgende Summe berechnen:

$$\mathbb{E}[f(d^{\mathcal{P}}, \mathbb{G}, c)] \approx \frac{1}{|S|} \sum_{s \in S} f(d^{\mathcal{P}}, s, c) \quad (19)$$

Um zwei verschiedene Modelle bzw. Modellvarianten zu vergleichen, werden wir den Erwartungswert wie oben approximieren und einen Permutationstest zum Signifikanzniveau $\alpha = 5\%$ durchführen.

4.2 Modellbeschreibung

Wir stellen nun unser Planungsmodell und seine Varianten vor. Wie schon erwähnt wird dieses aus drei Schritten bestehen. Die Schritte aus Bemerkung 4.2 und ihre Varianten werden wir im Folgenden näher erläutern. Bei Schritt 2 gehen wir dann auch auf die Unterschiede der Modelle für SHP und SHPNOC ein.

4.3 Schritt 1: Berechnung der möglichen Erntewochen

Der erste Schritt des Modells besteht daraus, für jede Saatspopulation $p \in \mathcal{P}$ die möglichen Erntewochen \mathcal{HW}_p zu bestimmen. Um dieses Problem zu lösen, stellen wir zwei Varianten vor. Die erste Variante beruht auf einer Vorhersage der GDU-Werte des Planungszeitraums mithilfe einer Support-Vektor-Maschine. Die zweite Variante verwendet eine statistische Methode.

Bemerkung 4.3 (Berechnung mittels Vorhersage der GDUs). In dieser Variante zur Berechnung der möglichen Erntewochen einer Population versuchen wir, eine möglichst gute Vorhersage der GDUs des Planungszeitraums auf Grundlage der GDU-Werte der uns zur Verfügung stehenden Jahre zu berechnen. Diese historischen Werte stellen die Trainingsdaten für die Support-Vektor-Maschine dar, mittels der wir die Vorhersage vornehmen.

Für jedes Jahr y und jeden Tag d , über den wir historische GDU-Daten besitzen, wird die SVM dann mit Tupeln (t, G_d^y) trainiert, wobei $t \in \{1, \dots, 365\}$ die Tagesnummer in einem Jahr entspricht und G_t^y dem zugehörigen GDU-Wert aus dem Jahr y .

Für eine gegebene Verteilung der GDU-Werte lässt sich, wie wir gesehen haben, leicht berechnen, in welchen Erntewochen man eine bestimmte Population p ernten kann. Allerdings sind GDU-Verteilungen Zufallsereignisse, weshalb wir solche Wochen ausschließen wollen, in denen nur in den wenigsten Fällen eine Ernte der Population p möglich ist. Diese Überlegung führt zur nächsten Variante des ersten Schrittes unseres Planungsmodells.

Bemerkung 4.4 (Berechnung auf statistischer Grundlage). Bei dieser Variante wählen wir diejenigen Wochen als mögliche Erntewochen einer Population, bei denen es eine hohe Wahrscheinlichkeit gibt, dass die Population in dieser Woche geerntet werden kann. Wir wollen dazu als mögliche Erntewochen \mathcal{HW}_p der Population p alle Wochen w auswählen, für die mindestens eine Wahrscheinlichkeit von 95 % besteht, dass p einen Aussaatzeitpunkt $d \in [e_p, l_p]$ besitzt, sodass p in Woche w geerntet wird, falls man an diesem Tag aussät. Natürlich können wir auch diese Wahrscheinlichkeiten nur approximieren. Wir ziehen eine Stichprobe $S = (s_1, \dots, s_{100})$ von 100 Werten der Zufallsvariable \mathbb{G} . Für eine Population p lässt sich dann für jede GDU-Verteilung s_i , wie in Bemerkung 4.1, eine Menge an möglichen Erntewochen $\mathcal{HW}_p^{s_i}$ der Population bestimmen. Als mögliche Erntewochen einer Population werden dann die Wochen $w \in W$ ausgewählt, die in mindestens 95 % der Mengen $\mathcal{HW}_p^{s_i}$, für $i = 1, \dots, 100$ enthalten sind.

4.4 Schritt 2: Populationsaufteilung auf die Erntewochen für SHP

Auf Grundlage der im ersten Schritt berechneten möglichen Erntewochen einer Population p , soll in diesem Schritt die bestmögliche Aufteilung der Population auf die möglichen Erntewochen \mathcal{HW}_p berechnet werden.

In diesem Abschnitt stellen wir hierzu zwei Varianten vor. Die erste Variante wird die Aufteilung berechnen, sodass für die durch die Aufteilung zu erwarteten Erntemengen h_1, \dots, h_m gilt, dass sie

$$\alpha \max_{i: h_i \neq 0} |C - h_i| + (1 - \alpha) \max_{i: h_i \neq 0} |C - h_i| + \beta |\{h_i \mid h_i \neq 0\}|.$$

minimieren. Dabei ist h_i die Erntemenge der Woche i und $\alpha = \frac{1}{2}, \beta = \frac{1}{|W|}$. Diese Zielfunktion entspricht unserer Zielfunktion f , nur dass hier die Erntemengen direkt gegeben sind und nicht erst durch die GDU-Werte und die Aussaatdaten bestimmt werden. Es ist genau die Zielfunktion aus dem zu DHP äquivalenten Problem PSP. Die erste Variante löst das Problem PSP exakt. Die zweite Variante wird daneben noch versuchen, die Populationen so auf die Erntewochen aufzuteilen, dass trotz der Unsicherheit in den GDU-Werten die Wahrscheinlichkeit maximiert wird, dass wie geplant geerntet wird. Dazu werden später die in Definition 4.5 vorgestellten Stabilitätswerte verwendet.

Variante 1 Die Eingabedaten sind die Populationen $p \in \mathcal{P}$ mit den Erntemengen q_p die möglichen Erntewochen $\mathcal{H}\mathcal{W}_p$ und eine Erntekapazität c . Wir definieren die Menge aller möglichen Erntewochen als $\mathcal{W} = \cup_{p \in \mathcal{P}} \mathcal{H}\mathcal{W}_p$. Die Variablen Δ_{\max} und Δ_{med} werden benötigt, um das Maximum und den Median der in Gleichung (2) vorgestellten Konsistenzabweichung zu beschreiben. Zur Beschreibung des MILPs werden noch einige binäre Variablen benötigt:

$$\begin{aligned} x_{pw} &= \begin{cases} 1 & \text{falls Population } p \text{ in der Woche } w \in \mathcal{H}\mathcal{W}_p \text{ geerntet wird,} \\ 0 & \text{sonst,} \end{cases} \\ y_w &= \begin{cases} 1 & \text{falls in der Woche } w \in \mathcal{W} \text{ geerntet werden soll,} \\ 0 & \text{sonst,} \end{cases} \\ z_w &= \begin{cases} 1 & \text{falls der Konsistenzwert zur Woche } w \geq \Delta_{\text{med}}, \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

An dieser Stelle sei erwähnt, dass es sich bei den Variablen z_w um die Variablen aus Abschnitt 3.3 handelt, welche zur Beschreibung des Medians in Form eines MILPs notwendig sind.

Wir formulieren unser MILP in dieser Variante für Szenario 1 dann wie folgt:

$$\begin{aligned} \min \quad & \alpha \cdot \Delta_{\max} + (1 - \alpha) \cdot \Delta_{\text{med}} + \beta \cdot \sum_{w \in W} y_w & (20) \\ \text{u.d.N.} \quad & \sum_{w \in \mathcal{W}} x_{pw} = 1 & p = 1, \dots, n & (21) \\ & x_{pw} \leq y_w & p = 1, \dots, n, w \in \mathcal{H}\mathcal{W}_p & (22) \\ & |c \cdot y_w - \sum_{\{p | w \in \mathcal{H}\mathcal{W}_p\}} x_{pw} q_p| \leq \Delta_{\max} & w \in W & (23) \\ & |c \cdot y_w - \sum_{\{p | w \in \mathcal{H}\mathcal{W}_p\}} x_{pw} q_p| \leq \Delta_{\text{med}} + M \cdot z_w & w \in W & (24) \\ & \sum_{w \in W} \frac{y_w}{2} - 1 \leq \sum_{w \in W} z_w \leq \sum_{w \in W} \frac{y_w}{2} & & (25) \\ & z_w \leq y_w & w \in W & (26) \\ & \Delta_{\max}, \Delta_{\text{med}} \in \mathbb{R}^+ & & (27) \\ & x_{pw} \in \{0, 1\} & p = 1, \dots, n, w \in \mathcal{H}\mathcal{W}_p & (28) \\ & y_w, z_w \in \{0, 1\} & w \in W & (29) \end{aligned}$$

Die in Gleichung (20) beschriebene Zielfunktion minimiert das Maximum und den Median der Konsistenzabweichung sowie die Anzahl der Erntewochen. Die Parameter $\alpha = \frac{1}{2}$ und $\beta = \frac{1}{|W|}$ setzen wir wie schon in Problem 1 beschrieben. Bedingungen (21) sorgen dafür, dass jede Population nur genau einmal gepflanzt und geerntet werden kann. Die Bedingungen (22) werden benötigt, damit keine Population in einer Woche geerntet wird, für die $y_w = 0$, also entschieden wurde, dass in der Woche w

nicht geerntet werden soll. Wir minimieren auch den Median der Konsistenzabweichung: Die Bedingungen (24), (25) und (26) sind dabei genau die Bedingungen, die es nach Abschnitt 3.3 braucht, um den Median zu minimieren. Der Wert M ist eine obere Schranke für die Beträge

$|c \cdot y_w - \sum_{\{p|w \in \mathcal{H}\mathcal{W}_p\}} x_{pw} q_p|$, diese Beträge entsprechen den Werten, für die der Median gefunden werden soll. Wir wählen $M = \sum_{p \in \mathcal{P}} q_p$. Weiter wollen wir auch die maximale Konsistenzabweichung minimieren. Die Bedingungen (23) setzen Δ_{\max} auf die maximale Konsistenzabweichung für alle Erntewochen.

Variante 2 Die zweite Variante soll versuchen, die Unsicherheit in den GDU-Werten mitzuberücksichtigen. Dafür definieren wir zunächst, was unter der Stabilität eines Aussaatzeitpunkts einer bestimmten Population bezüglich einer Erntewoche zu verstehen ist.

Definition 4.5. Für eine Saatpopulation p definieren wir den Stabilitätswert $\zeta_{p,d,w}$ des Aussaattags d bezüglich einer Erntewoche w als die Wahrscheinlichkeit, dass p in der Woche w geerntet wird, wenn man sie am Tag d anpflanzt.

Aufbauend auf der vorherigen Definition 4.5, definieren wir die Stabilität für eine Woche w und für die Populationen $\mathcal{P}'_w \subseteq \mathcal{P}$, die in w geerntet werden.

Definition 4.6. Der Stabilitätswert einer Population p bezüglich einer Erntewoche $w \in \mathcal{H}\mathcal{W}_p$ ist definiert als

$$Z_{p,w} := \begin{cases} \max\{\zeta_{p,d,w} \mid d \in [e_p, l_p]\} & \text{falls } \max\{\zeta_{p,d,w} \mid d \in [e_p, l_p]\} > 0 \\ 1 & \text{sonst.} \end{cases} \quad (30)$$

Sei w eine Erntewoche und \mathcal{P}'_w die Menge aller Populationen $p \in \mathcal{P}$, die in w geerntet werden sollen. Die Idee, die zur zweiten Variante führt, ist die, das Minimum der Stabilität der möglichen Erntewochen \mathcal{W} zu maximieren. Dieses definieren wir folgendermaßen:

$$\min_{w \in \mathcal{W}} \prod_{p \in \mathcal{P}'_w} Z_{p,w} \quad (31)$$

Das ist ohne Weiteres in einem linearen Programm nicht einfach möglich, weshalb wir das Produkt mit Hilfe des Logarithmus linearisieren.

$$\sum_{p \in \mathcal{P}'_w} \log(Z_{p,w}) \quad (32)$$

Diese Summe lässt sich dann gut in einem gemischt-ganzzahligen linearen Programm maximieren. An dieser Stelle ist dann auch klar, warum wir $Z_{p,w} := 1$ und nicht wie zu erwarten $Z_{p,w} := 0$ setzen, falls die Woche w keine mögliche Erntewoche der Population p ist. Einen Einfluss auf das gemischt-ganzzahlige Programm hat diese Wahl allerdings nicht, da p nicht in w geerntet werden kann.

Variante 2 ergibt sich durch zwei kleinere Anpassungen. Aus der Zielfunktion (21) wird

$$\min \quad \alpha \cdot \Delta_{\max} + (1 - \alpha) \cdot \Delta_{\text{med}} + \beta \cdot \sum_{w \in \mathcal{W}} y_w - \gamma \Delta_{\text{stab}}. \quad (33)$$

Wir wählen dabei $\gamma = 1$. Zudem ergänzen wir das in Variante 1 beschriebene MILP noch um eine weitere Bedingung, um die Variable Δ_{stab} auf das Minimum der Stabilitätswerte der Erntewochen zu legen.

$$\Delta_{\text{stab}} \leq \sum_{p \in \{p|w \in \mathcal{H}\mathcal{W}_p\}} \log(Z_{p,w}) x_{pw} \quad w \in \mathcal{W} \quad (34)$$

Bemerkung 4.7 (SHPNOC). Die Varianten 1 und 2 für Schritt 2 des Problems SHPNOC sind dieselben wie für das Problem SHP. Als Kapazität c wählt man dabei die zuvor im Schritt 1b oder Schritt 2b berechnete Kapazität.

4.5 Schritt 1b für SHPNOc

Anders als in Szenario 1 ist in Szenario 2 keine Kapazität gegeben. Für das MILP in Schritt 2, welches eine Aufteilung der Populationen auf die Erntewochen berechnet, wird jedoch eine Kapazität benötigt. Aus diesem Grund suchen wir eine initiale Kapazität und wollen als diese die Kapazität wählen, die unter der Annahme, dass in jeder Woche geerntet werden muss für die besten Aufteilungen der Populationen sorgt. Für die Berechnung dieser Kapazität können wir erneut das MILP zu SHP aus 4.4 modifizieren. Wir setzen dafür alle $y_w = 1$ und erhalten dadurch die Möglichkeit, die Kapazität c als weitere Variable in dem MILP zu verwenden.

4.6 Schritt 2b für SHPNOc

Nachdem man in unserem Modell bei Schritt 2 für eine Kapazität c eine gute Aufteilung der Populationen auf die Erntewochen berechnet hat, wollen wir in diesem Schritt für diese Aufteilung eine bessere Kapazität berechnen. Seien $h = (h_1, \dots, h_{|W|})$ die Erntemengen, die durch die zuvor berechneten Aufteilungen erzeugt werden, dabei ist h_i die Erntemenge in Woche i . Sei h_l der niedrigste Erntewochenenertrag und h_u der höchste Erntewochenenertrag, der durch die Planung der Erntewoche entsteht. Wir wählen als die neue Kapazität \tilde{c} diejenige Kapazität $h_l \leq c \leq h_u$, die den Wert der gewichteten Summe aus Maximum und Median der Konsistenzabweichung minimiert:

$$\alpha \max \Delta_C + (1 - \alpha) \text{median } \Delta_C. \quad (35)$$

Falls es mehrere Kapazitäten gibt, die diese Summe minimieren, wählen wir von diesen eine Kapazität \hat{c} , die zusätzlich die Summe aller Konsistenzabweichungen $\sum_{i=1}^{|W|} |\hat{c} - h_i|$ minimiert.

4.7 Schritt 2' für SHPNOc

Das Problem DHPNOc ist sehr ähnlich wie DHP, welches wir mithilfe des MILP in Abschnitt 4.4 lösen. Dieses MILP kann einfach angepasst werden, sodass es DHPNOc lösen kann. Dazu tauschen wir im MILP die Bedingungen

$$|c \cdot y_w - \sum_{\{p|w \in \mathcal{H}^w \mathcal{W}_p\}} x_{pw} q_p| \leq \Delta_{\max}$$

gegen

$$|c - \sum_{\{p|w \in \mathcal{H}^w \mathcal{W}_p\}} x_{pw} q_p| \leq \Delta_{\max} + (1 - y_w)M \quad (36)$$

aus, wobei M eine obere Schranke für die Kapazität sein muss. Weiterhin ist c nun eine Variable und nicht länger ein Parameter. Die beiden Formulierungen sind äquivalent, jedoch besitzt (36) den großen Vorteil, dass die Ungleichungen, trotz der Einführung von c als Variable, linear bleiben.

4.8 Schritt 3: Ermittlung der Aussaatage

Der letzte Schritt befasst sich damit, für eine Saatpopulation p und eine Erntewoche w einen Aussaatag auszuwählen, sodass p in w geerntet wird. Die beste Erntewoche w für die Population p wurde bereits in Schritt 2 berechnet. Für Schritt 3 verwenden wir dann folgende Vorgehensweise, welche die in Definition 4.5 vorgestellten Stabilitätswerte verwendet. Für jede Population p und ihre entsprechende, in Schritt 2 berechnete Erntewoche w wählen wir das Aussaatdatum $d \in [e_p, l_p]$, das erfüllt, dass

$$d = \operatorname{argmax}\{\zeta_{p,d^*,w} | d^* \in [e_p, l_p]\}. \quad (37)$$

Den Wert $\zeta_{p,d^*,w}$ werden wir approximieren. Sei $S = (s_1, \dots, s_{1000})$ eine Stichprobe von \mathbb{G} . Für jedes $s_i \in S$ lässt sich dann berechnen, ob sich p in w ernten lässt, wenn die Population an d^* angepflanzt wurde. Die Funktion $h^*(p, d^*, w, s)$ nimmt den Wert 1 an, falls so eine Ernte möglich ist, und sonst den Wert 0. Dann können wir den Wert $\zeta_{p,d^*,w}$ wie folgt annähern:

$$\zeta_{p,d^*,w} \approx \sum_{i=1}^{1000} \frac{h^*(p, d^*, w, s_i)}{1000}. \quad (38)$$

5 Ergebnisse

In diesem Kapitel stellen wir die Ergebnisse vor, die unsere Modelle für SHP und SHPNOC für die Standorte 0 und 1 erzielen. Wir zeigen jeweils, welche Planung der Erntewochen sich nach Schritt 2 des Modells ergibt und welche Verteilungen der Bewertungsfunktion f und der Anzahl der genutzten Erntewochen daraus resultieren. Wir wollen auch die Ergebnisse, d. h. die Verteilungen von f , der einzelnen Modelle bzw. der Modellvarianten miteinander vergleichen. Dafür werden wir Permutationstests verwenden.

Die Verteilung der Bewertungsfunktion f für die verschiedenen Modellvarianten und Szenarien approximieren wir, indem wir zunächst eine Stichprobe $S = (s_1, \dots, s_{1000})$ aus den Werten der Zufallsvariable \mathbb{G} ziehen. Bei den, durch Schritt 3 ermittelten, Pflanzdaten d ist es dann möglich, für jedes Element s_i die Erntezeitpunkte einer Population zu bestimmen und daraus folgernd die Erntemenge jeder Woche. Daraus ergibt sich wiederum der Wert der Funktion $f(d, s_i, c)$. Auf Grundlage der Werte $f(d, s_i, c)$ lassen sich die verschiedenen Varianten vergleichen.

Tabelle 1 zeigt die Modellvarianten, die wir für das Problem SHP untersuchen wollen. Tabelle 2 zeigt die Modellvarianten, welche beim Problem SHPNOC untersucht werden sollen. Die Zeile in der eine Variante in der Tabelle steht gibt an, welche Variante des zweiten Schrittes des Modells verwendet wurde vgl. Abschnitt 4.4 bis Abschnitt 4.6. Die Spalte der Tabelle gibt an, welche Variante des ersten Schrittes verwendet wurde, vergleiche Abschnitt 4.3.

Schritt 2 \ Schritt 1	SVM	Statistisch
MILP	Var_{SVM}	Var_{Stat}
MILP mit Stabilität	$\text{Var}_{\text{SVM}}^{\text{Stabi}}$	$\text{Var}_{\text{Stat}}^{\text{Stabi}}$

Tabelle 1: Modellvarianten für das Problem SHP

Schritt 2 \ Schritt 1	SVM	Statistisch
MILP	Var_{SVM}	Var_{Stat}
MILP mit Stabilität	$\text{Var}_{\text{SVM}}^{\text{Stabi}}$	$\text{Var}_{\text{Stat}}^{\text{Stabi}}$
angepasstes MILP (Schritt 2')	$\text{Var}_{\text{SVM}}^*$	$\text{Var}_{\text{Stat}}^*$

Tabelle 2: Modellvarianten für das Problem SHPNOC

Für die Berechnungen wurde folgendes System verwendet:

- Betriebssystem: Linux,
- Kernel-Version: 5.4.0-66-generic,

- Prozessoren: 4 × Intel Core i5-5300U CPU mit 2,30 GHz,
- Speicher: 16 GiB Arbeitsspeicher.

Es ist anzumerken, dass wir die MILPs in Schritt 2 aufgrund von langen Rechenzeiten nach jeweils einer Stunde terminieren. Die Berechnungen für Schritt 1 in der statistischen Variante dauerten für die Eingabegröße (ca. 1100 bzw. 1300 Populationen und 70 Wochen) etwa 8 Stunden. In Varianten mit SVM dauerten nur wenige Sekunden. Die Berechnungen für Schritt 3 terminierten in unter einer Stunde. Zur Implementierung unserer Modelle haben wir Python 3 und Gurobi 9.1.0 verwendet. Gurobi [8] ist ein Programm zum Lösen mathematischer Programme, welches ein Python-Interface anbietet und geeignet ist um MILPs zu lösen.

5.1 SHP, Standort 0

Planung der Erntewochen. Abbildung 4 zeigt für alle 70 Erntewochen und alle Modellvarianten die Planungen aus Schritt 2, welche auf Grundlage der gegebenen Daten zu Standort 0 und dem Problem SHP durchgeführt wurden. Die Δ_{\max} -Werte bei der Planung von Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ entsprechen ungefähr

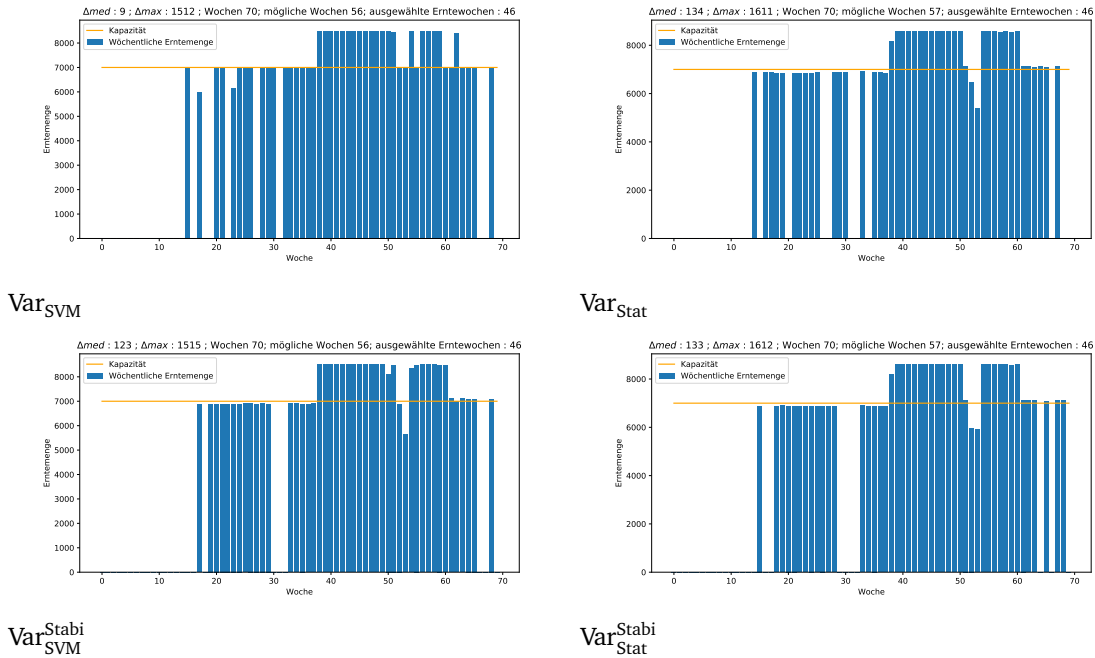


Abbildung 4: Geplante Erntemengen aller 70 Wochen für das Problem SHP für den Standort 0 und die vier verschiedenen Modellvarianten.

einander, genau wie Δ_{\max} von Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$. Die Varianten $\text{Var}_{\text{SVM}}^{\text{Stabi}}$, Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ planen mit einem Δ_{med} in gleicher Größenordnung. Dass der Wert Δ_{med} in Variante Var_{SVM} deutlich kleiner ist, als bei den übrigen Varianten, kann als Ursache haben, dass die Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ die möglichen Erntewochen in Schritt 1, mithilfe einer Vorhersage der GDUs, berechnen. Dadurch stehen der Planung in den Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ andere mögliche Erntewochen zur Verfügung, die zu einer anderen Planung in Schritt 2 führen können. Dass Δ_{med} bei Var_{SVM} auch deutlich kleiner ist als bei $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ liegt vermutlich daran, dass in Var_{SVM} in mindestens einer Woche geerntet wird, für die die meisten Stabilitätswerte der Populationen nicht gut sind. In Var_{SVM} wird, anders als bei $\text{Var}_{\text{SVM}}^{\text{Stabi}}$, nicht auf die Stabilitätswerte geachtet, weshalb in $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ nicht in einer solchen Woche geerntet wird. Das erklärt, warum Δ_{\max} und Δ_{med} aus Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ durchaus kleiner sind als für Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$. Weiterhin erklärt dies, weshalb Δ_{med} bei Var_{SVM} so viel kleiner sein kann als bei den anderen Varianten.

In allen Planungen wird eine Ernte in 46 Wochen angestrebt. In den Grafiken zu allen vier Varianten fällt besonders auf, dass sich die Erntemengen in zwei Plateaus aufteilen. Das eine Plateau entspricht dabei in seinen Erntemengen in etwa Δ_{med} , das andere Δ_{max} . Dies lässt sich dadurch erklären, dass sowohl Δ_{max} als auch Δ_{med} minimiert werden sollen. Da Δ_{med} relativ klein ist, ist schon einmal klar, dass sich die Hälfte der Erntemengen in einem Plateau um die Kapazitätsgrenze anordnen. Gleichzeitig wird auch Δ_{max} minimiert. Es wird versucht, die Erntemengen jener Wochen, welche eine Konsistenzabweichung größer als Δ_{med} besitzen, so gleichmäßig wie möglich aufzuteilen, um Δ_{max} so zu minimieren. Dabei entsteht das Δ_{max} -Plateau.

Anzahl der Erntewochen. Wir betrachten in der Tabelle 3 die Verteilungen der genutzten Erntewochen über die 1000 Werte s_i der Stichprobe aus den GDU-Verteilungen.

Var _{SVM}	Anzahl der Erntewochen	Wert	Var _{Stat}	Anzahl der Erntewochen	Wert
	46	1		46	571
	47	176		47	162
	48	215		48	192
	49	401		49	62
	50	201		50	9
Var _{SVM} ^{Stabi}	51	6		51	4
	Anzahl der Erntewochen	Wert	Var _{Stat} ^{Stabi}	Anzahl der Erntewochen	Wert
	46	73		46	772
	47	740		47	223
	48	184		48	5
	49	3			

Tabelle 3: Verteilung der Anzahl der Erntewochen für das Problem SHP am Standort 0 für die vier Modellvarianten

Was stark auffällt ist, dass die Varianten Var_{Stat}^{Stabi} und Var_{SVM}^{Stabi}, die in Schritt 2 den Stabilitätswert mit berücksichtigen, eine weniger stark gestreute Verteilung der Erntewochen aufweisen. In allen Modellvarianten wurde mit 46 Erntewochen geplant. Bei den Varianten Var_{Stat} und Var_{Stat}^{Stabi} ergeben sich in den meisten Fällen 46 Erntewochen. Die Verteilung von Var_{Stat} weist jedoch eine deutlich größere Streuung auf als die Verteilung von Var_{Stat}^{Stabi}, und auch der Anteil mit 46 Erntewochen ist bei Var_{Stat} (57,1 %) deutlich geringer als bei Var_{Stat}^{Stabi} (77,2 %). Bei der Modellvariante Var_{SVM} ergeben sich meistens 49 (40,1 %) Erntewochen anstatt der geplanten 46 (0,1 %), und auch die Verteilung der Erntewochen ist ähnlich stark gestreut wie bei Var_{Stat}. Die Variante Var_{SVM}^{Stabi} zeigt eine weniger starke Streuung als Var_{SVM}, allerdings wird hier meistens in 47 (74 %) verschiedenen Wochen geerntet.

Die Verteilung der Bewertungsfunktion f . Abbildung 5 zeigt die Verteilungen der Werte Δ_{med} , Δ_{max} und die Verteilung der Zielfunktion f . Schnell erkennt man, dass Δ_{med} deutlich kleinere Werte annimmt und nicht so stark gestreut ist wie Δ_{max} .

Die Verteilungen von Δ_{max} , von Var_{SVM} und Var_{SVM}^{Stabi} sind deutlich verschieden zu den Verteilungen von Var_{Stat} und Var_{Stat}^{Stabi}. Es fällt auf, dass der in Schritt 2 (vgl. Abb. 4) berechnete Wert eine untere Schranke für die Verteilung von Δ_{max} in allen Varianten darstellt. Sowohl Var_{Stat} und Var_{Stat}^{Stabi} weisen

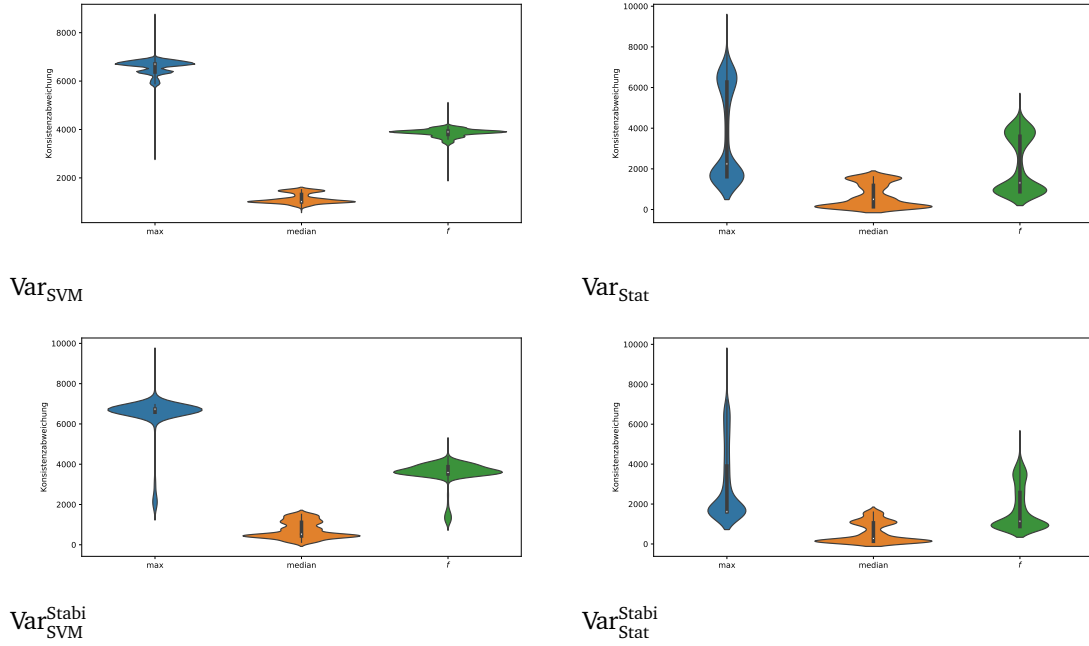


Abbildung 5: Verteilung der Zielfunktion f , sowie der Werte Δ_{\max} und Δ_{med} für das Problem SHP am Standort 0 für die vier Modellvarianten.

eine große Anzahl an Δ_{\max} -Werten auf, die zwischen ihrer unteren Schranke bei etwa 1600 und etwas über 2000 liegen. Im Vergleich zu $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ hat die Verteilung von Δ_{\max} von Var_{Stat} deutlich mehr Werte in der Höhe der Kapazität, allerdings nicht ganz so häufig wie Werte zwischen 1600 und etwas über 2000. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ hingegen nimmt häufig kleine Werte, auf dem eben beschriebenen Intervall an und ist abgesehen davon relativ gleichverteilt, mit maximalen Werten in der Größe der Kapazität. Die Verteilungen von Δ_{\max} für Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ nehmen hauptsächlich Werte in der Größenordnung der Kapazität an.

Die Verteilung von Δ_{med} wirkt in allen Varianten recht ähnlich. Es werden meistens Werte, die etwas über den in Schritt 2 berechneten Wert liegen, angenommen. Allerdings gibt es einen Ausschlag an Werten, die etwas unter 2000 liegen. Bei der Variante Δ_{\max} von Var_{SVM} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ist dieser Ausschlag am geringsten. Aus den Verteilungen von Δ_{\max} und Δ_{med} resultiert auch die Form der Verteilung von f .

Zusammenfassend kann man sagen, dass Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ zu deutlich besseren Ergebnissen für den Standort 0 führen als Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$. Wir wollen deshalb, und wegen der Ähnlichkeit der Verteilung von f in den Varianten Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$, die beiden Varianten noch mithilfe eines Permutationstests zu einem Signifikanzniveau von $\alpha = 5\%$ vergleichen. Als Teststatistik wurde, wie in Gleichung (17) beschrieben, die Funktion

$$t(D_x, D_y) = \sum_{a \in D_x} \frac{a}{|D_x|} - \sum_{b \in D_y} \frac{b}{|D_y|}.$$

gewählt. D_x und D_y sind dabei die Mengen der Zielfunktionswerte für die verschiedenen Modellvarianten. Unsere Nullhypothese ist dabei dass Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ der gleichen Verteilung unterliegen und unsere Alternativhypothese ist, dass der Erwartungswert von $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ kleiner ist als der von Var_{Stat} . Abbildung 6 zeigt diese Verteilung der Teststatistik. Die Linie t zeigt dabei den Wert der Teststatistik für die Werte der Zielfunktion f für die nicht permutierten Daten. Man erkennt sofort, dass t (Differenz der Durchschnittswerte von f bei Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$) größer ist als alle zufällig erzeugten Werte.

Wir müssen daher unsere Nullhypothese verwerfen und können die Alternativhypothese annehmen, dass der Erwartungswert der Bewertungsfunktion f bei Var_{Stat} größer ist als bei $\text{Var}_{\text{Stat}}^{\text{Stabi}}$.

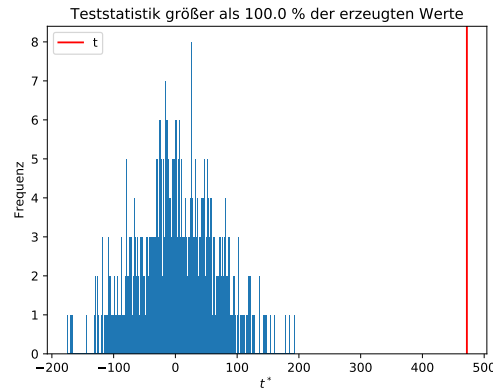


Abbildung 6: Verteilung der Teststatistik im Permutationstest für Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$

5.2 SHP, Standort 1

Planung der Erntewochen. Abbildung 7 zeigt, für alle 70 Erntewochen und alle Modellvarianten die Planungen aus Schritt 2, welche auf Grundlage der gegebenen Daten zu Standort 1 und dem Problem SHP durchgeführt wurden. Man erkennt, dass die Planung in allen Wochen sehr gleichmäßig ist, was möglicher Weise auf die nicht so stark schwankenden GDU-Werte am Standort 1 zurückzuführen ist. In Var_{Stat} wirkt die Planung minimal gleichmäßiger als in den restlichen Varianten. In allen vier Fällen wird mit 51 Wochen geplant.

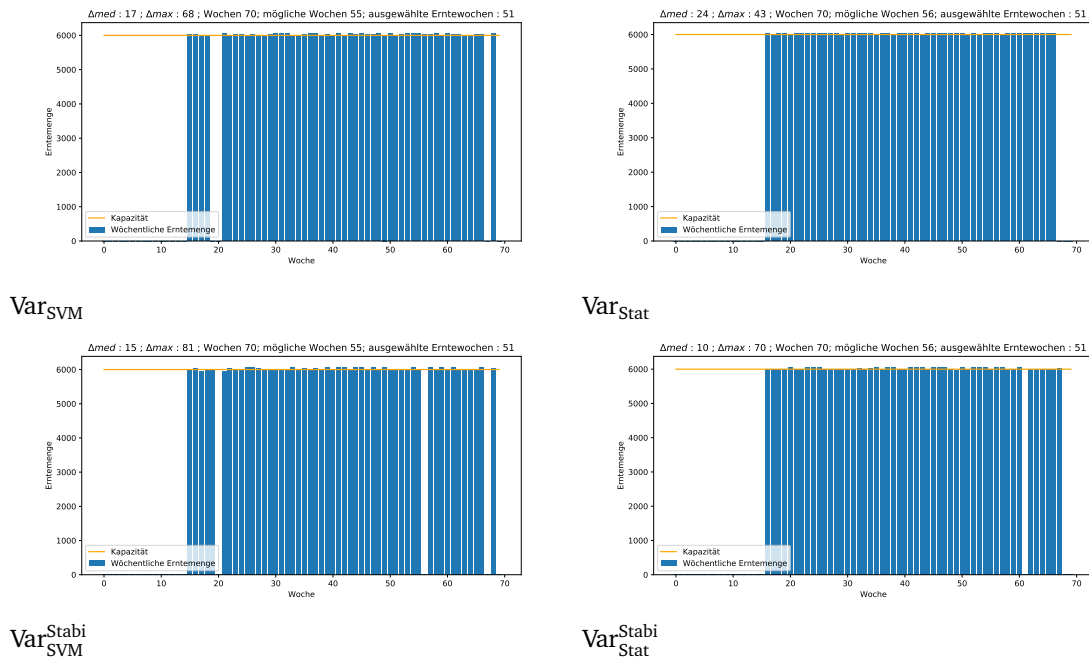


Abbildung 7: Geplante Erntemengen aller 70 Wochen für das Problem SHP für den Standort 0 und die vier verschiedenen Modellvarianten.

Anzahl der Erntewochen. Tabelle 4 zeigt die Verteilung der Anzahl der Erntewochen für die verschiedenen Stichproben s_i , welche sich aus den vier Varianten ergeben. Man erkennt, dass die Var_{Stat} eine weniger starke Streuung als die restlichen Varianten besitzt. Die Streuungen der Varianten $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ sind sehr ähnlich. Die Kurve der Verteilung von Var_{SVM} ist flacher als bei den restlichen Wochen. In dieser Variante wird in den meisten Fällen (63,8 %) in 52 Wochen geerntet anstelle der geplanten 51 Wochen (34,1 %). In den Varianten Var_{Stat} (98,6 %), $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ (93,4 %) und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ (90,8 %) wird in den meisten Fällen in den geplanten 51 Wochen geerntet.

Var_{SVM}

Anzahl der Erntewochen	Wert
51	341
52	638
53	19
54	2

Var_{Stat}

Anzahl der Erntewochen	Wert
51	986
52	014

$\text{Var}_{\text{SVM}}^{\text{Stabi}}$

Anzahl der Erntewochen	Wert
50	1
51	908
52	88
53	2
54	1

$\text{Var}_{\text{Stat}}^{\text{Stabi}}$

Anzahl der Erntewochen	Wert
50	1
51	934
52	61
53	4

Tabelle 4: Verteilung der Anzahl der Erntewochen für das Problem SHP am Standort 1 für die vier Modellvarianten

Die Verteilung der Bewertungsfunktion f . Abbildung 8 zeigt die Verteilungen der Werte Δ_{med} , Δ_{max} und die Verteilung der Zielfunktion f . Man erkennt, dass die Verteilung von f in Var_{SVM} deutlich häufi-

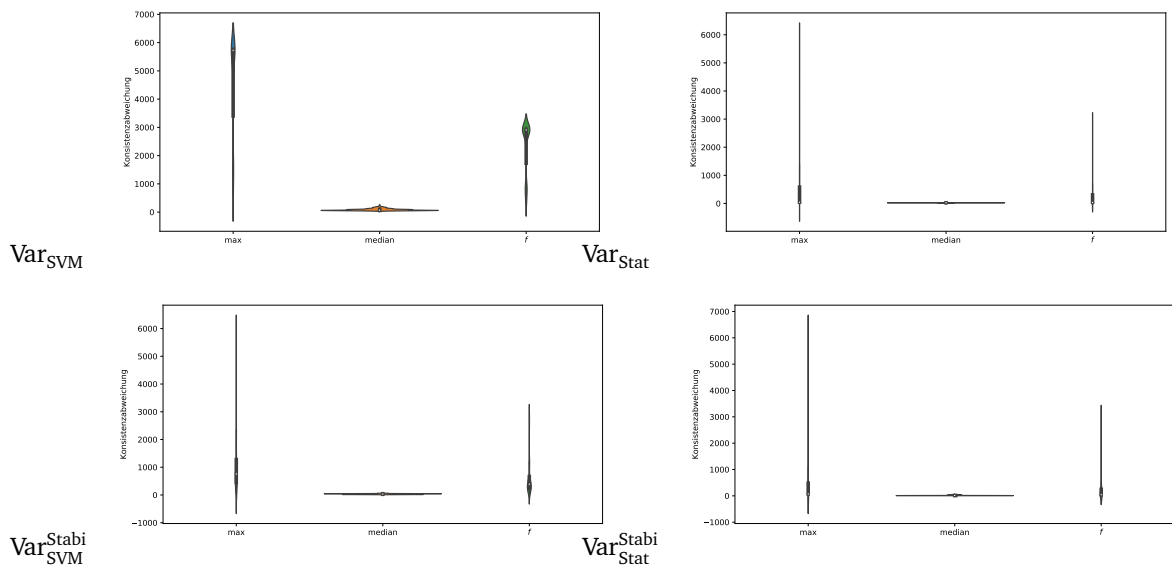


Abbildung 8: Verteilung der Zielfunktion f , sowie der Werte Δ_{max} und Δ_{med} für das Problem SHP am Standort 1 für die vier Modellvarianten

ger große Werte annimmt als in den anderen Varianten. In den restlichen Varianten scheinen die Verteilungen recht gleich. $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ hat einen höheren Mittelwert bezüglich Δ_{max} , und auch f wirkt schlechter verteilt. Die Werte, die Δ_{med} annimmt, liegen in allen Varianten nahe null und sind nicht stark gestreut.

Die Werte, die Δ_{\max} annimmt, sind im Minimum sehr klein und liegen unter hundert. Maximal nehmen sie Werte, die in etwa in der Höhe der Kapazität liegen, an. In den Varianten Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ist der Mittelwert bezüglich Δ_{\max} deutlich kleiner als der Mittelwert bei Variante Var_{SVM} oder Variante $\text{Var}_{\text{SVM}}^{\text{Stabi}}$.

Wir werden Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ mit $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ in einem Permutationstest zum Signifikanzniveau $\alpha = 5\%$ vergleichen, um sicher sagen zu können, dass $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ schlechter ist als die anderen beiden Varianten. Anschließend wollen wir auch Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ mit einem Permutationstest vergleichen. Die Nullhypothese wird jeweils sein, dass die Werte der Funktion f in beiden Varianten Stichproben der gleichen Verteilung sind. Die Alternativhypothese ist, dass der Erwartungswert der Funktion f in $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ größer ist als in den Varianten Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$.

Die Abbildungen 9 und 10 zeigen die Ergebnisse der Permutationstests. Man sieht die Verteilung der Teststatistik unter der Annahme, dass die Nullhypothese wahr ist. Die Linie t zeigt den Wert der Teststatistik für die nicht permutierten Daten. In Abbildung 9 erkennt man deutlich, dass der Unterschied zwischen Var_{Stat} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ bzw. zwischen $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ signifikant ist. Also ist der Erwartungswert von f bei Variante Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ signifikant kleiner als bei $\text{Var}_{\text{SVM}}^{\text{Stabi}}$. Ein solcher signifikanter Unterschied kann, wie man in Abbildung 10 erkennt, bei dem Vergleich zwischen Variante Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ nicht festgestellt werden. Deshalb müssen wir abschließend davon ausgehen, dass Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ für den Standort 1 gleich gute und die beste Ergebnisse liefern.

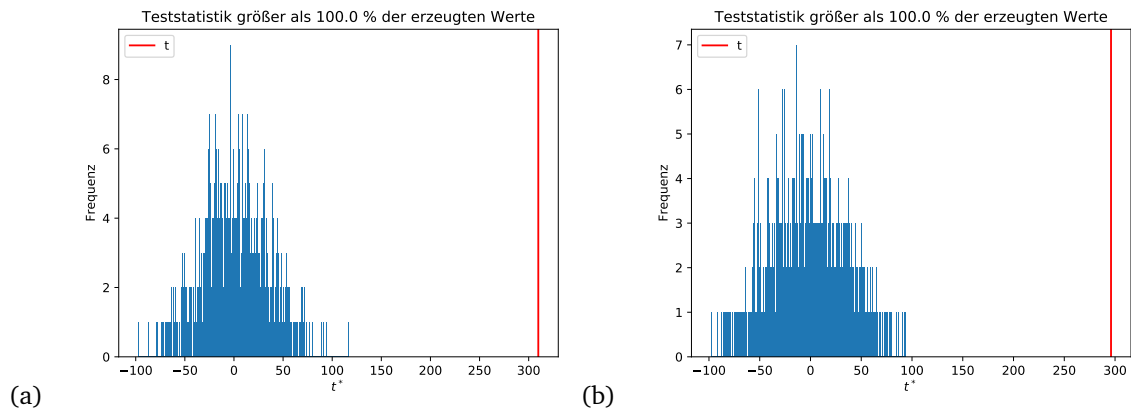


Abbildung 9: Verteilung der Teststatistiken bei den Permutationstest für $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ vs. Var_{Stat} (a) und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ (b)

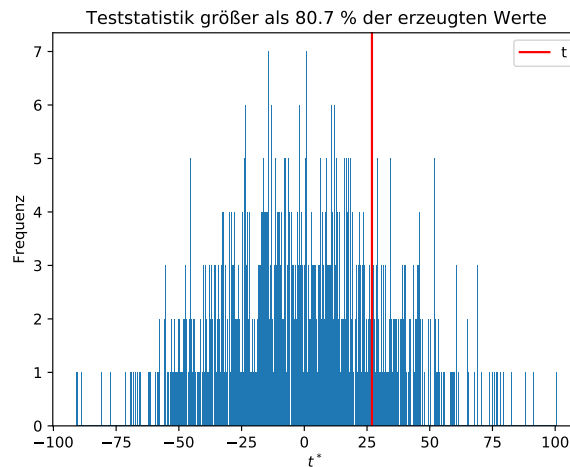


Abbildung 10: Verteilung der Teststatistik im Permutationstest für Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$

5.3 SHPnOC, Standort 0

Planung der Erntewochen Abbildung 11 zeigt für alle 70 Erntewochen und alle Modellvarianten die Planungen aus Schritt 2, welche auf Grundlage der gegebenen Daten zu Standort 0 und dem Problem SHPnOC durchgeführt wurden. Was sofort auffällt ist, dass die Varianten, welche das angepasste MILP

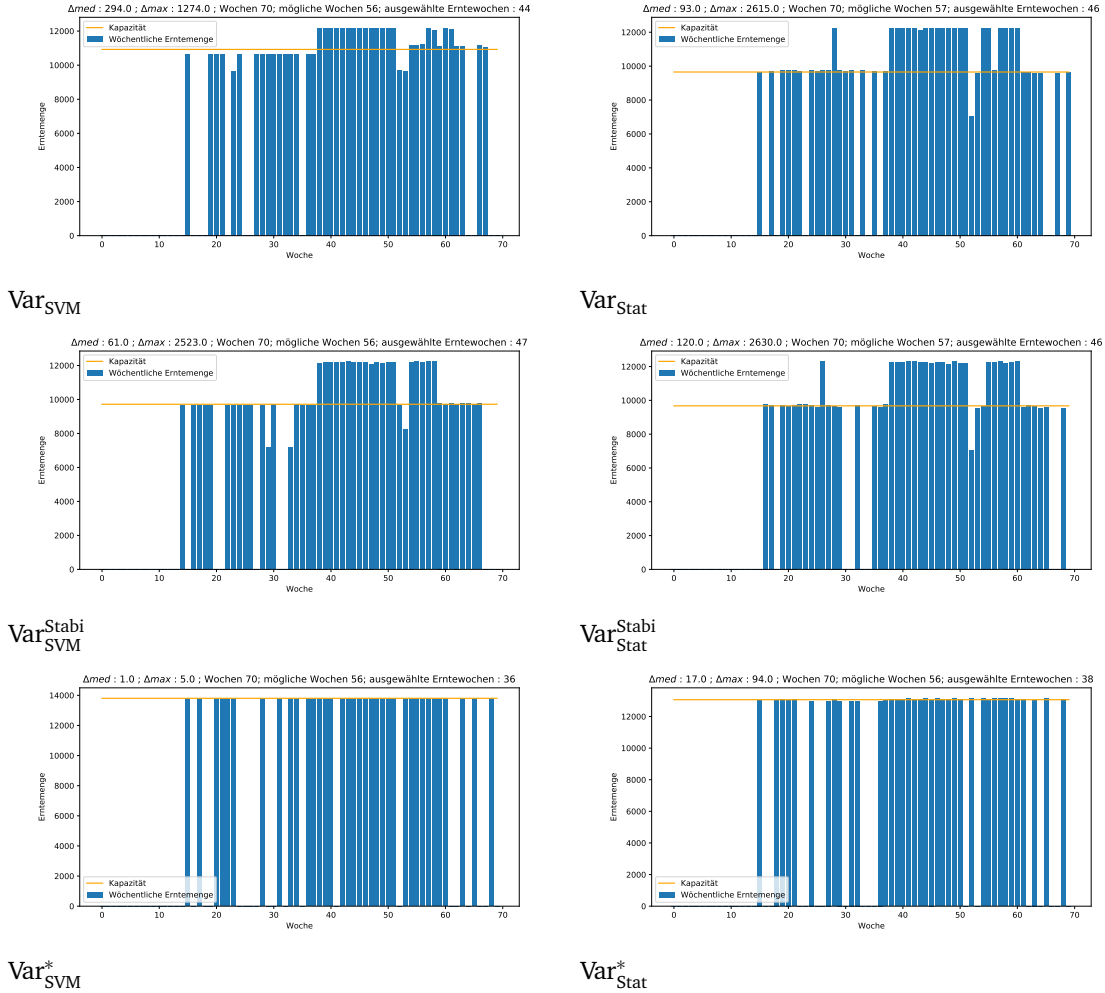


Abbildung 11: Geplante Erntemengen aller 70 Wochen für das Problem SHP für den Standort 0 und die vier verschiedenen Modellvarianten.

in Schritt 2 verwenden, eine deutlich gleichmäßigere Ernteplanung erzeugen. Die anderen Varianten ordnen sich in zwei Plateaus an, welche auf die gleiche Weise zu erklären sind wie bei den Ergebnissen zu SHP am Standort 0. Auch ist deutlich zu erkennen, dass die Planung in $\text{Var}^*_{\text{SVM}}$ und $\text{Var}^*_{\text{Stat}}$ nicht nur gleichmäßiger ist, sondern auch zu einem wesentlich kleineren Wert für Δ_{\max} und Δ_{med} führt. Bei den Varianten Var_{SVM} , Var_{Stat} , $\text{Var}^{\text{Stabi}}_{\text{SVM}}$ und $\text{Var}^{\text{Stabi}}_{\text{Stat}}$ liegt der Wert für Δ_{med} zwischen 61 und 294, der Wert für Δ_{\max} liegt zwischen 1274 und 2630. Bei der Variante $\text{Var}^*_{\text{SVM}}$ liegt der Wert für Δ_{\max} bei 5 und der Wert für Δ_{med} bei 1. Dieser große Unterschied zwischen den Varianten, die nicht das angepasste MILP verwenden, und den Varianten $\text{Var}^*_{\text{SVM}}$ / $\text{Var}^*_{\text{Stat}}$ ist dadurch zu erklären, dass ein anderes MILP in Schritt 2 verwendet wird. Die Varianten $\text{Var}^*_{\text{SVM}}$ und $\text{Var}^*_{\text{Stat}}$ suchen ein globales Minimum für die Zielfunktion, wo hingegen die restlichen Varianten lediglich ein lokales Minimum suchen. Man sieht, dass die Varianten Var_{SVM} , Var_{Stat} , $\text{Var}^{\text{Stabi}}_{\text{SVM}}$ und $\text{Var}^{\text{Stabi}}_{\text{Stat}}$ eine kleinere die Kapazität berechnen als in den beiden anderen. *Anzahl der Erntewochen.* Tabelle 5 zeigt die Verteilung der Anzahl der Erntewochen für

die verschiedenen Stichproben s_i , welche sich aus den sechs Varianten ergeben.

Var_{SVM}	Anzahl der Erntewochen	Wert	Var_{Stat}	Anzahl der Erntewochen	Wert
	44	1		46	556
	45	9		47	182
	46	116		48	108
	47	690		49	106
	48	181		50	42
$\text{Var}_{\text{SVM}}^{\text{Stabi}}$	49	3	$\text{Var}_{\text{Stat}}^{\text{Stabi}}$	51	6
	Anzahl der Erntewochen	Wert		Anzahl der Erntewochen	Wert
	48	206		45	1
	49	112		46	722
	50	18		47	175
	47	664		48	99
$\text{Var}_{\text{SVM}}^*$	Anzahl der Erntewochen	Wert	$\text{Var}_{\text{Stat}}^*$	49	2
	39	2		50	1
	40	12		Anzahl der Erntewochen	Wert
	41	582		38	507
	42	309		39	187
	43	76		40	182
$\text{Var}_{\text{SVM}}^*$	44	17	$\text{Var}_{\text{Stat}}^*$	41	085
	45	2		42	3
				43	9

Tabelle 5: Verteilung der Anzahl der Erntewochen für das Problem SHPNOC am Standort 0 für die sechs Modellvarianten.

Aus Abbildung 11 geht hervor, dass die Varianten $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{Stat}}^*$ eine wesentlich geringere Anzahl an Erntewochen anstreben (36 bzw. 38), als die anderen Varianten (44 bis 47). Var_{SVM} plant mit 44 Erntewochen. Dieser Wert wurde allerdings nur für ein Element s_i der Stichprobe angenommen. Am häufigsten führt die Planung mit Variante Var_{SVM} zu einer Ernte in 47 (96 %) Wochen. Die Variante Var_{Stat} plant mit 46 Wochen, und in der Tat wird auch in den meisten Fällen (55,6 %) in 46 Wochen geerntet. Die Varianten $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ planen mit 47 (66,4 %) und 46 (72,2 %) Erntewochen und erreichen, dass dieser Wert am häufigsten angenommen wird. Die Modellvariante $\text{Var}_{\text{Stat}}^*$ plant mit 38 Wochen und führt in den meisten Fällen zu einer Ernte in 38 (50,7 %) Wochen. Die Variante $\text{Var}_{\text{SVM}}^*$ plant mit 36 Wochen, erreicht dies aber nie. Am Häufigsten wird in der Modellvariante in 41 (58,2 %) Wochen geerntet.

Die Verteilung der Bewertungsfunktion f . Abbildung 12 zeigt die Verteilungen der Werte Δ_{med} , Δ_{max} und die Verteilung der Zielfunktion f .

Man erkennt sofort, dass die Verteilungen von f in den Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^*$ am schlechtesten sind und dort, im Vergleich zu den restlichen Varianten, hauptsächlich hohe Werte annehmen. Die Form der Kurven der Verteilungen in den Varianten Var_{Stat} , $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ sind sich sehr ähnlich. Alle drei Kurven haben zwei Hochpunkte. Einer liegt bei allen drei Varianten bei etwa 4000 bis 6000. Der andere Hochpunkt liegt bei Variante $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ etwa um die 2000 und bei den anderen beiden Modell-

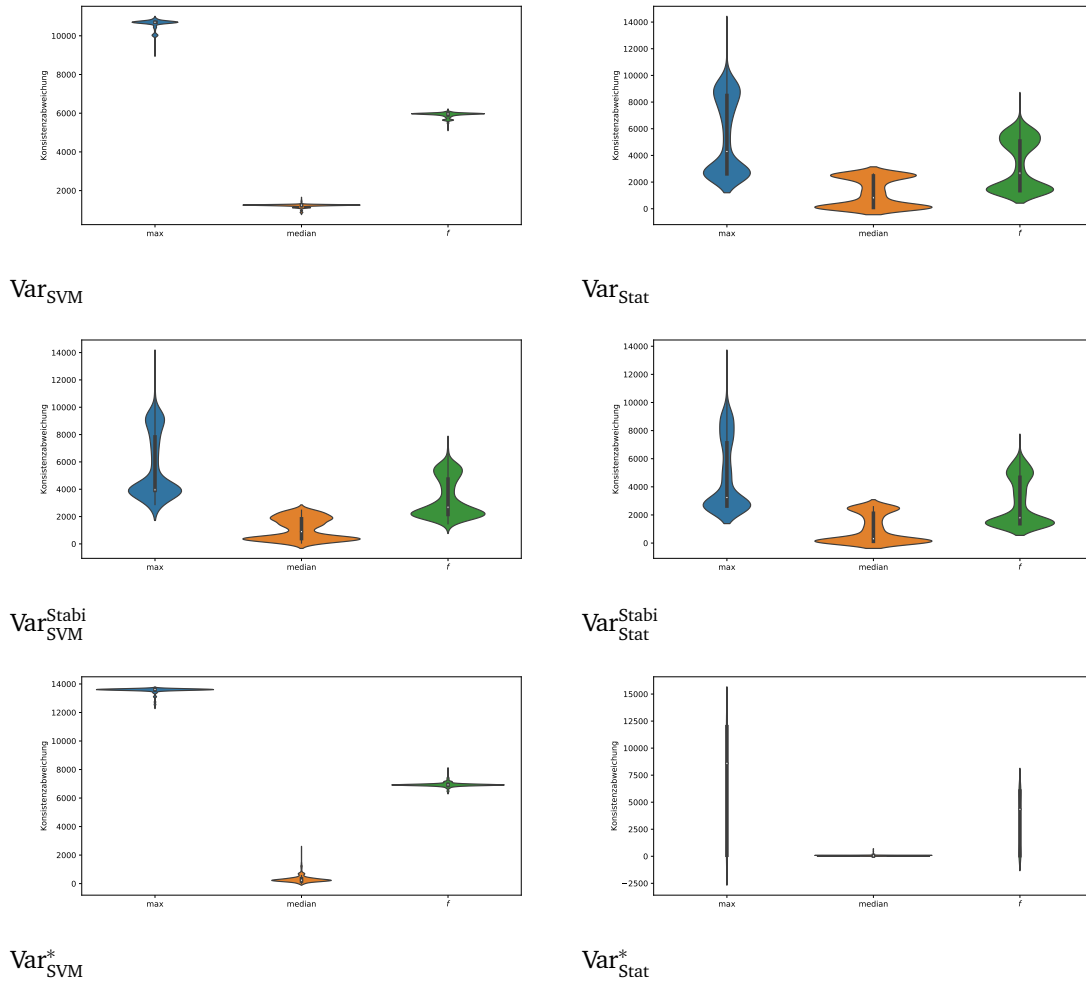


Abbildung 12: Verteilung der Zielfunktion f , sowie der Werte Δ_{\max} und Δ_{med} für das Problem SHPNOC am Standort 0 für die sechs Modellvarianten.

varianten etwas unter 2000. Das Verhältnis der beiden Hochpunkte zueinander muss noch betrachtet werden. In allen drei Varianten ist der Hochpunkt bei 2000 stärker ausgeprägt als derjenige bei 4000 bis 6000. Den stärksten Unterschied in der Ausprägung der Hochpunkte hat die Variante $\text{Var}_{\text{Stat}}^{\text{Stabi}}$, den schwächsten Unterschied die Variante Var_{Stat} . Auf Grundlage des Plots ist es schwierig, etwas Genaueres über die Variante Var_{Stat} auszusagen. Allerdings scheint die Verteilung, im Vergleich zu den Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^*$, recht gutartig zu sein und diese Variante liefert eindeutig die besten Ergebnisse in Bezug auf die Größe Δ_{med} . Dabei fällt auf, dass auch die Variante $\text{Var}_{\text{SVM}}^*$, obwohl die Verteilung der Bewertungsfunktion schlecht ist, was auf den die hohen Δ_{\max} Werte zurückzuführen ist, sehr gute Ergebnisse bezüglich Δ_{med} liefert. Es scheint so, als erziele die Variante $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ die besten Ergebnisse. Diese Annahme werden wir mithilfe eines Permutationstests zum Signifikanzniveau $\alpha = 5\%$ überprüfen. Wir werden die Werte für $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ mit denen für $\text{Var}_{\text{SVM}}^{\text{Stabi}}$, $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ und $\text{Var}_{\text{Stat}}^*$ vergleichen. Die Teststatistik wählen wir, wie schon in den Kapiteln zuvor.

Die Nullhypothese ist, dass die Mengen der Werte von f für die beiden Varianten Stichproben derselben Verteilung sind. Die Alternativhypothese ist jeweils, dass der Erwartungswert von f in Variante $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ kleiner ist als bei der anderen Variante. Abbildung 13 zeigt die Ergebnisse des Permutationstests. Man sieht die Verteilung der Teststatistik unter der Annahme, dass die Nullhypothese wahr ist. Die Linie t zeigt den Wert der Teststatistik für die nicht permutierten Daten. Bei den Vergleichen Var_{Stat} vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ist der Wert der Teststatistik für die gegebenen Zielfunktionswerte so

gegeben, dass er größer ist als durch Zufall in 100 % der zufälligen Werte zu erwarten. Im Vergleich $\text{Var}_{\text{Stat}}^*$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ist der Unterschied größer als in 99,5 % der Fälle durch Zufall zu erwarten wäre. Somit müssen wir in allen Permutationstests die Nullhypothese verwerfen und die Alternativhypothese annehmen. Wir müssen also davon ausgehen, dass die Variante $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ für das Problem SHPNOC und den Standort 1 die besten Ergebnisse liefert.

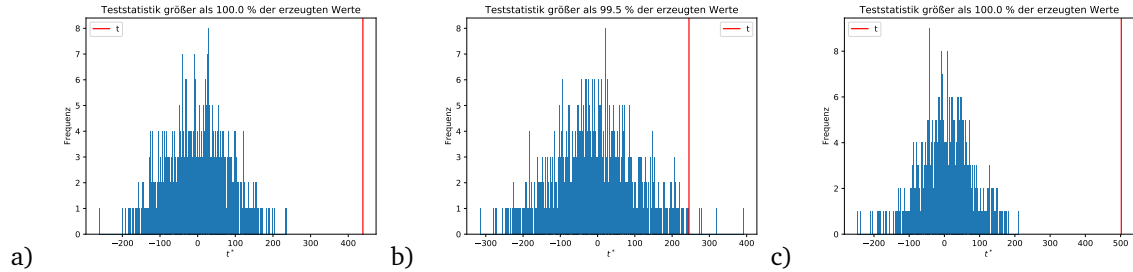


Abbildung 13: Geplante Erntemengen aller 70 Wochen für das Problem SHP für den Standort 0 und die vier verschiedenen Modellvarianten. a) zeigt den Vergleich von $\text{Var}_{\text{Stat}}^*$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$. b) zeigt den Vergleich von $\text{Var}_{\text{Stat}}^*$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$. c) zeigt den Vergleich von $\text{Var}_{\text{SVM}}^*$ vs. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$.

5.4 SHPNOC, Standort 1

Planung der Erntewochen. Abbildung 4 zeigt für alle 70 Erntewochen und alle Modellvarianten die Planungen aus Schritt 2, welche auf Grundlage der gegebenen Daten zu Standort 1 und dem Problem SHPNOC durchgeführt wurden. Man sieht sofort, dass die Planung in allen Fällen sehr gleichmäßig ist und nah bei der Kapazität liegt. Δ_{med} und Δ_{max} sind bei $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{Stat}}^*$ kleiner als bei den restlichen Varianten. Bei $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ liegt Δ_{med} bei 13 bzw. 12, bei den Varianten $\text{Var}_{\text{Stat}}^*$ und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ bei 6 bzw. 39. Bei $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{Stat}}^*$ ist Δ_{med} kleiner gleich 1. Der Δ_{max} -Wert von $\text{Var}_{\text{SVM}}^*$, $\text{Var}_{\text{Stat}}^*$ und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ liegt zwischen 218 bei $\text{Var}_{\text{SVM}}^*$ und 386 bei $\text{Var}_{\text{Stat}}^*$. $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ hat einen Δ_{max} -Wert von 62. In Variante $\text{Var}_{\text{Stat}}^*$ liegt Δ_{max} bei 5. Am niedrigsten ist der Wert in Variante $\text{Var}_{\text{SVM}}^*$ mit $\Delta_{\text{max}} = 1$. Dass die Varianten $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{Stat}}^*$ eine so viel bessere Planung für Schritt 2 ergeben, d. h. gleichmäßigere Planungen mit geringeren Fehlergrößen Δ_{max} und Δ_{med} , ist, wie schon an Standort 0, dadurch zu erklären, dass hier eine Kombination aus Aufteilung der Populationen auf die Erntewochen und Kapazität gesucht wird. Die gesuchte Kombination soll ein globales Minimum der Zielfunktion bilden, wo hingegen in den ersten vier Varianten ein lokales Optimum gesucht wird.

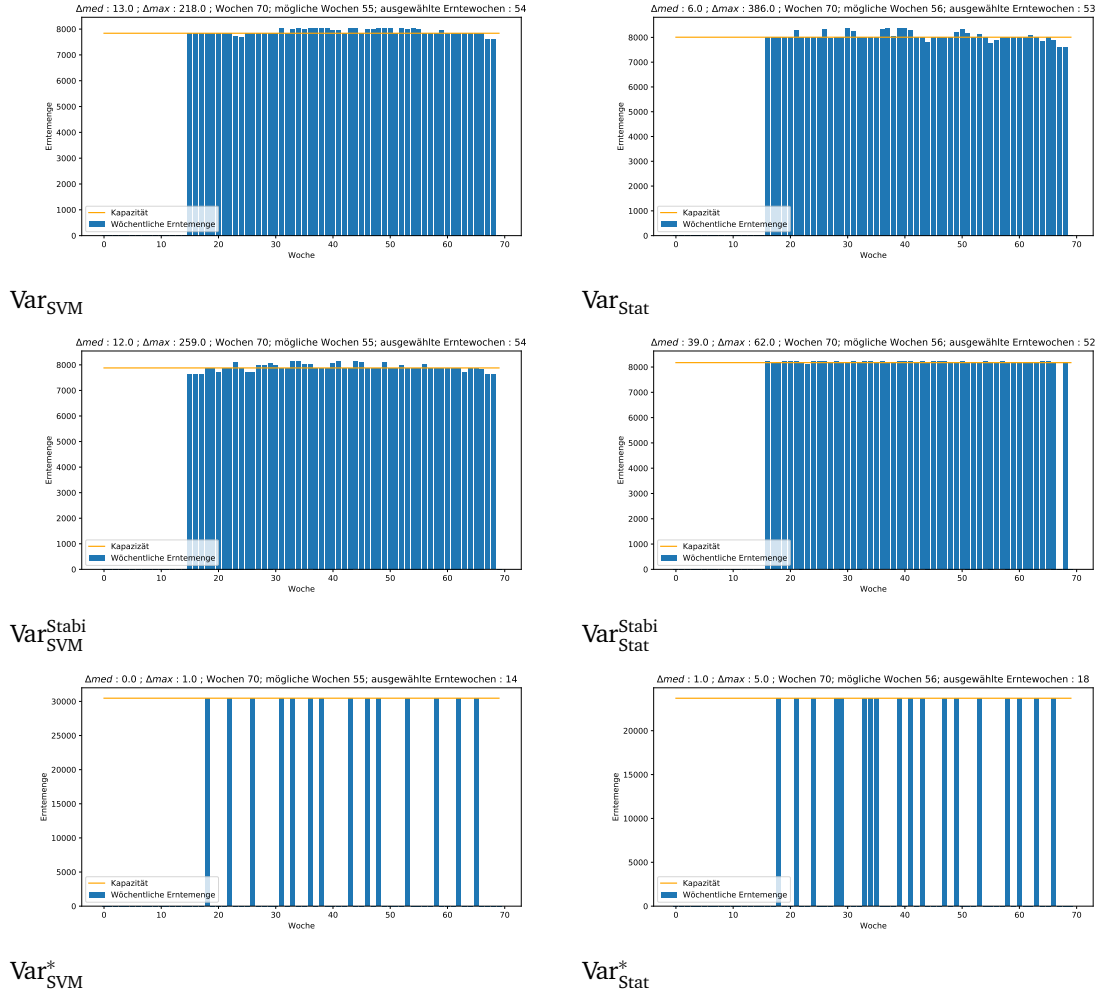


Abbildung 14: Geplante Erntemengen aller 70 Wochen für das Problem SHP für den Standort 0 und die vier verschiedenen Modellvarianten.

Anzahl der Erntewochen. Tabelle 6 zeigt die Verteilung der Anzahl der Erntewochen, welche sich aus den sechs Varianten aus den Elementen der Stichprobe $s \in S = (s_1, \dots, s_{1000})$ ergeben. Am instabilsten in Bezug auf die Anzahl der Erntewochen ist die Variante $\text{Var}_{\text{SVM}}^*$. In dieser Variante wird mit 14 Wochen geplant, was jedoch nie erreicht wird. Am häufigsten wird in 23 Wochen (27,4 %) geerntet. Die restlichen Modellvarianten sind bezüglich der Erntewochen deutlich stabiler. Die Varianten Var_{SVM} , Var_{Stat} , $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ernten in nahezu allen Fällen in der geplanten Wochenanzahl. Variante $\text{Var}_{\text{Stat}}^*$ liefert leicht schlechtere Ergebnisse. Hier wird mit 18 Wochen geplant und in 70,4 % der Fälle auch in 18 Wochen geerntet. Die Instabilitäten der Varianten $\text{Var}_{\text{SVM}}^*$ und $\text{Var}_{\text{Stat}}^*$ bezüglich der Wochenanzahl können dadurch erklärt werden, dass sie mit wesentlich weniger Wochen planen als die anderen Varianten. Die Tatsache, dass Variante $\text{Var}_{\text{Stat}}^*$ dabei dennoch so viel stabiler ist als Variante $\text{Var}_{\text{SVM}}^*$ könnte daran liegen, dass in Schritt 1 für Variante $\text{Var}_{\text{Stat}}^*$ günstigere Erntewochen für jede Population bestimmt wurden.

Var_{SVM}

Anzahl der Erntewochen	Wert
53	3
54	994
55	3

 Var_{Stat}

Anzahl der Erntewochen	Wert
52	1
53	992
54	7

 $\text{Var}_{\text{SVM}}^{\text{Stabi}}$

Anzahl der Erntewochen	Wert
53	3
54	989
55	8

 $\text{Var}_{\text{Stat}}^{\text{Stabi}}$

Anzahl der Erntewochen	Wert
51	3
52	974
53	20
54	3

 $\text{Var}_{\text{SVM}}^*$

Anzahl der Erntewochen	Wert
18	8
19	27
20	67
21	213
22	271
23	274
24	116
25	24

 $\text{Var}_{\text{Stat}}^*$

Anzahl der Erntewochen	Wert
18	704
19	173
20	75
21	32
22	12
23	3
24	1

Tabelle 6: Verteilung der Anzahl der Erntewochen für das Problem SHPNOC am Standort 1 für die sechs Modellvarianten.

Die Verteilung der Bewertungsfunktion f . Abbildung 12 zeigt die Verteilungen der Werte Δ_{med} , Δ_{max} und die Verteilung der Zielfunktion f . Die Zielfunktion bei Variante $\text{Var}_{\text{SVM}}^*$ nimmt, anders als die Werte der Zielfunktion in den restlichen Varianten, ausschließlich sehr hohe Werte an. Die Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ nehmen im Maximum Werte um die 4000 an. Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ liegen bei 5000. Die Variante $\text{Var}_{\text{Stat}}^*$ nimmt maximal Werte bei ca. 12000 an. Die Variante $\text{Var}_{\text{SVM}}^*$ erzielt im Maximum Werte bei in etwa 22500. Die maximalen Werte von Δ_{max} liegen in allen Varianten in der Größenordnung der jeweiligen Kapazität und sind darauf zurückzuführen, dass bei manchen $s_i \in S$ einige wenige Populationen in einer Woche geerntet werden, in der keine Ernte geplant war. Die Zielfunktion nimmt in allen Varianten, außer in $\text{Var}_{\text{SVM}}^*$, am häufigsten Werte nahe null an. Die Verteilung bei Variante Var_{SVM} wirkt etwas schlechter als in Var_{Stat} , $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$.

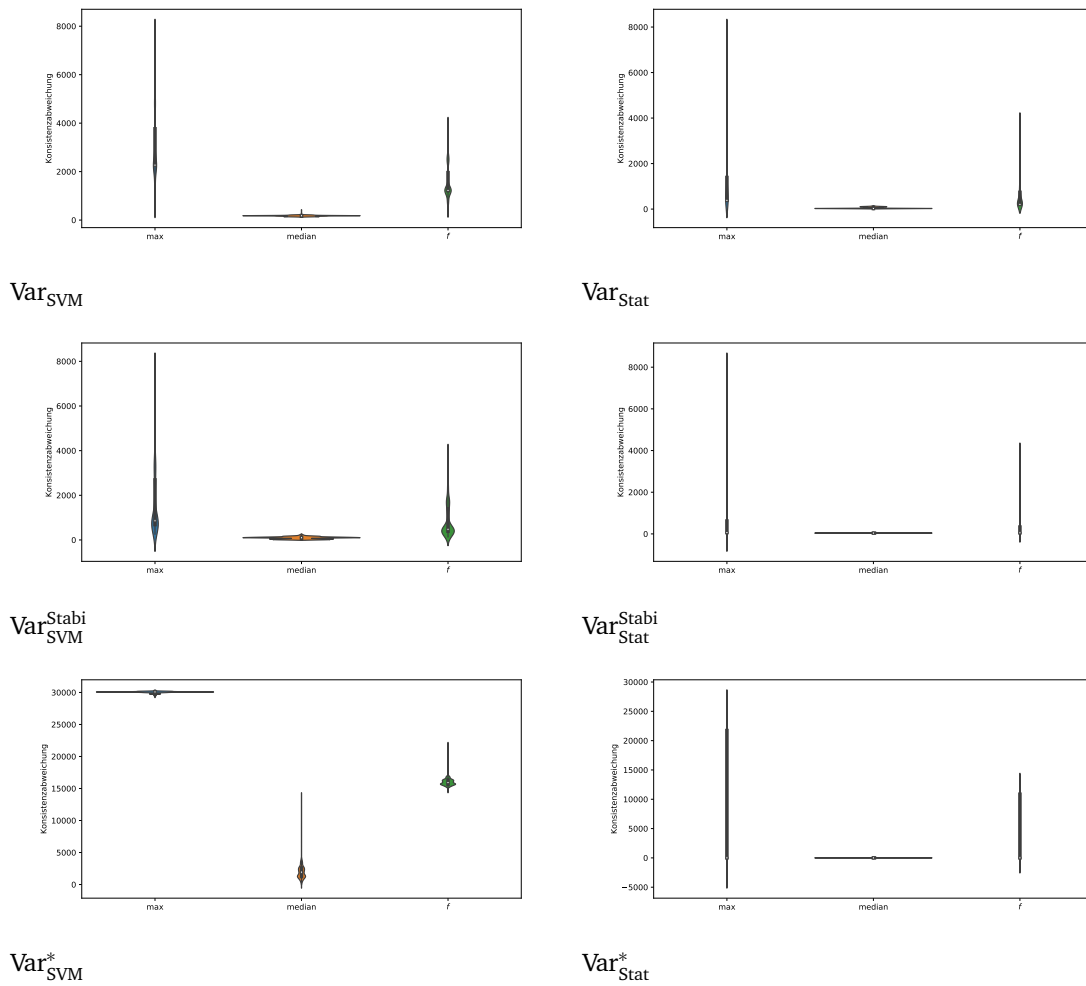


Abbildung 15: Verteilung der Zielfunktion f , sowie der Werte Δ_{\max} und Δ_{med} für das Problem SHPNOC am Standort 1 für die sechs Modellvarianten

6 Diskussion

Das folgende Kapitel hinterfragt die in der Arbeit vorgestellten Methoden und Resultate.

6.1 Kritik an der Problemstellung

Zunächst muss angemerkt werden, dass die Problemstellung Syngentas nicht präzise formuliert wurde. Es werden mehrere Zielvorgaben festgelegt, ohne genauer auf die Gewichtung dieser Ziele einzugehen, geschweige denn eine Zielfunktion anzugeben. Aus diesem Grund haben wir uns für eine parametrisierte Zielfunktion entschieden, deren Parameter bei Bedarf angepasst werden können.

Des Weiteren ist die Vorgabe zu kritisieren, dass insbesondere im Problem SHP alle Populationen angepflanzt werden müssen, auch wenn dadurch der Fehler größer wird und man häufiger mehr erntet, als die Kapazität hergibt. Dies könnte in der Realität zu Überproduktion oder Verschwendung von Gütern führen.

6.2 Minimierung des Erwartungswertes

An unseren Modellen muss man kritisieren, dass sie SHP bzw. SHPnoC nicht exakt in angemessener Zeit lösen. Allerdings sollte aufgrund dessen, dass man nicht weiß, wie die Zufallsvariable der GDUs verteilt ist, grundsätzlich kein Modell dazu in der Lage sein, die Probleme exakt zu lösen. Als Evaluationskriterium haben wir den Erwartungswert festgelegt. Allerdings wird dieser an keiner Stelle aktiv minimiert, sondern wir hoffen, einen niedrigen Erwartungswert durch Minimierung der Zielfunktion und Berechnung stabiler Erntewochen und Aussaatzeitpunkte zu erhalten. Genauer gesagt wählen wir stabile Erntewochen als mögliche Erntewochen und diejenigen Aussaattage, die mit der größten Wahrscheinlichkeit zu einer Ernte in einer bestimmten Woche führen. Dies machen wir in der Hoffnung, dass es zur Stabilität der in Schritt 2 berechneten Lösung beiträgt.

Ein weiterer Versuch, den Erwartungswert zu minimieren, war es, den Stabilitätswert einer Woche mit im MILP zu berücksichtigen. Man kann festhalten, dass die Varianten, die in Schritt 2 das MILP mit den Stabilitätswerten verwendeten, in beiden Szenarien und allen Standorten die besten Ergebnisse lieferten. Bei dieser Variante ist allerdings zu kritisieren, dass wir den Parameter γ für die Gewichtung des Stabilitätswertes auf den Wert 1 setzen, ohne dies genauer begründen zu können.

Die besten Ergebnisse lieferte meistens die Verwendung der Variante, die in Schritt 1 unter den möglichen Erntewochen einer Population die statistisch häufigsten wählt und in Schritt 2 das MILP mit Stabilitätswerten löst. Es fällt auf, dass die Varianten, die in Schritt 1 nicht die SVM zur Berechnung verwenden, bessere, d. h. stabilere Ergebnisse, erzielen.

6.3 Kritik an der Evaluation und an der Zufallsvariable \mathbb{G}

Um die von uns vorgestellten statistischen Methoden zu nutzen, war davon auszugehen, dass die vorliegende Stichprobe der Zufallsvariable \mathbb{G} zu klein ist. Um diese zu vergrößern, treffen wir die Annahme, dass die GDU-Werte verschiedener Tage unabhängig voneinander sind, doch diese Annahme spiegelt vermutlich nicht die Realität wider. Man müsste sich darüber Gedanken machen, ob man die Stichprobe auf eine andere Art und Weise vergrößern kann, die versucht, dies zu berücksichtigen. Wir evaluieren unsere Ergebnisse, indem wir die Verteilung der Zielfunktion f betrachten, die von einer zufälligen Stichprobe der Größe 1000 erzeugt wird. Es ist fraglich, ob eine Stichprobe der Größe 1000 wirklich ausreicht, um diese Verteilung gut zu repräsentieren, da wir alleine 33^{487} verschiedene GDU-Verteilungen für den Planungszeitraum kennen. Des Weiteren müssen wir anmerken, dass die historischen GDU-Daten in unserem Kriterium keine gesonderte Rolle bei der Evaluation spielen. An dieser Stelle wollen wir deshalb untersuchen, wie unsere Planungen für das Problem SHP am Standort 0 unter den historischen GDU-Daten abschneiden. Abbildung 16 zeigt die Verteilung von Δ_{\max} , Δ_{med} und f für die verschiedenen Modellvarianten auf den historischen GDU-Werten der Jahre 2009 bis 2019.

Es ist bemerkenswert, dass die Modellvarianten Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ scheinbar keine schlechteren Ergebnisse liefern, als die Varianten Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$. In allen Varianten besitzt die Verteilung der Bewertungsfunktion einen Hochpunkt. Der Wert an dem dieser Hochpunkt angenommen wird, liegt dabei um die 4000, aber die Verteilung der Zielfunktion bei Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ ist stärker gestreut, als bei den anderen beiden Varianten. Man erkennt insbesondere, dass auch die Verteilung von Δ_{\max} bei Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$ nicht so stark gestreut ist wie bei den Varianten Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$. Die Verteilung von Δ_{\max} nimmt bei Var_{Stat} und $\text{Var}_{\text{Stat}}^{\text{Stabi}}$ deutlich kleinere, aber auch deutlich größere Werte an als bei den Varianten Var_{SVM} und $\text{Var}_{\text{SVM}}^{\text{Stabi}}$. Darauf wird auch der Unterschied in der Verteilung der Zielfunktion zurückzuführen sein.

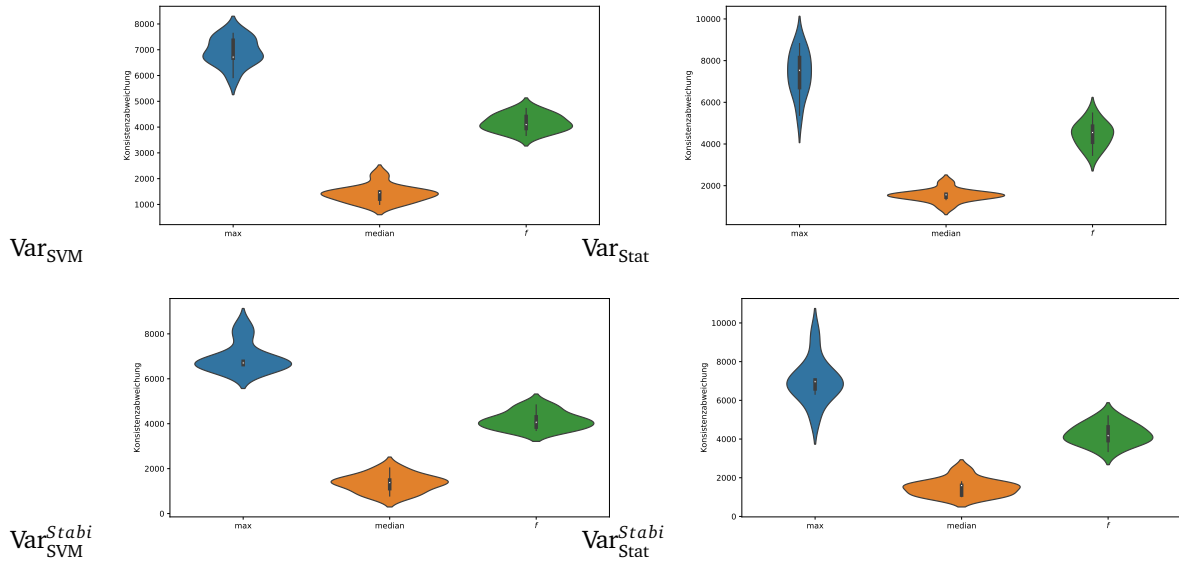


Abbildung 16: Verteilung der Zielfunktion f , sowie der Werte Δ_{\max} und Δ_{med} für das Problem SHP am Standort 0 für die vier Modellvarianten. Die Planungen wurden auf die originalen GDU-Daten aus den Jahren 2009 bis 2019 angewendet.

7 Resümee

Wir präsentierten einen stochastischen Ansatz zur Optimierung der erwarteten Ernteertragskonsistenz von Aussaatplanungen für eine Menge an Saatspopulationen.

Um die GDU-Werte zu repräsentieren, beschreiben wir diese als Zufallsvariable. Wir vergrößern die Stichprobe unter der Annahme, dass die GDU-Werte an zwei Tagen unabhängig voneinander sind.

Zunächst berechnen wir im ersten Schritt für jede Population eine Menge an Wochen, in denen man hofft, diese Population ernten zu können. Dazu verwenden wir zwei verschiedene Verfahren, die sowohl für das Problem mit gegebener Kapazität (SHP) als auch für das Problem, in dem man die Kapazität zusätzlich berechnen muss (SHPNOC), verwendet werden. Das eine Verfahren wählt diejenigen Wochen aus, die auf der Stichprobe der GDU-Werte häufig Erntewochen waren. Ein anderes Verfahren, welches wir für diesen Schritt verwenden, beruht auf einer Vorhersage der GDU-Werte des Planungszeitraums mithilfe einer SVM. Auf Grundlage dieser Vorhersage werden dann mögliche Erntewochen für jede Population bestimmt.

Als Nächstes berechnen wir für das Problem eine Aufteilung der Populationen auf die möglichen Erntewochen. Eine von uns vorgestellte Variante für das Problem SHP verwendet ein exaktes MILP, welches die maximale, absolute Differenz und den Median dieser absoluten Differenzen aus Erntekapazität und wöchentlichem Ernteertrag minimiert. Das sekundäre Optimierungsziel wird sein, die Anzahl der Erntewochen zu minimieren. Eine weitere Variante versucht gleichzeitig den minimalen Stabilitätswert aller Erntewochen zu maximieren. Der Stabilitätswert repräsentiert dabei die Wahrscheinlichkeit, dass alle Populationen, die in der Woche geerntet werden sollen, auch in dieser Woche geerntet werden.

Für das Problem SHPNOC verwenden wir drei verschiedene Ansätze, um die Aufteilung zu berechnen. Für die ersten beiden Ansätze wird eine initiale Kapazität berechnet. Anschließend wird mithilfe der berechneten Kapazität eine Aufteilung der Populationen mit den MILPs, für das Problem SHP, bestimmt. Danach wird für diese Aufteilung dann eine neue Kapazität berechnet. Das Berechnen der Aufteilung und der neuen Kapazität wird dann so lange wiederholt, bis sich die Kapazität nicht mehr wesentlich ändert. Der dritte Ansatz verwendet eine Umformulierung des MILPs von SHP, die eine simultane Berechnung der Aufteilung und der Kapazität ermöglicht. Die ersten beiden Varianten entwickeln wir in der Hoffnung, dass die Kapazität bei ihnen nicht zu hoch gesetzt wird, denn wenn die

Kapazität hoch ist kann auch schnell ein großer Fehler resultieren.

Der letzte Schritt ist für beide Probleme identisch. Hier wird für jede Population ein Aussaatzeitpunkt bestimmt. Dieser soll dazu führen, dass die Population auch tatsächlich in der zuvor geplanten Woche geerntet wird. Dazu wird das Datum, als Aussaatzeitpunkt ausgewählt, was für die GDU-Verteilungen in unserer Stichprobe am häufigsten zu einer Ernte in der entsprechenden Woche führt.

Besonders die Varianten, die sowohl das MILP mit Stabilitätswerten, als auch die Berechnung der möglichen Erntewochen mithilfe der Statistik verwenden, erzeugen in unseren Untersuchungen die besten Resultate.

In zukünftigen Arbeiten, sollte die Variante mit dem angepassten MILP und die Variante die, die Stabilitätswerte verwendet, zusammengefasst und untersucht werden. Weiter sollte man versuchen die MILPs in Schritt 2 so anzupassen, dass sie nicht die Zielfunktion f minimieren, sondern den Erwartungswert der Zielfunktion f . Dies könnte man erreichen, indem man mit dem zu erwartenden Ernteertrag einer Saatspopulation für die geplante Woche arbeitet, anstelle des absoluten Ernteertrags. Für die Evaluation könnte man die historischen GDU-Daten mehr berücksichtigen und das Modell ohne die vergrößerte Stichprobe und nur auf Grundlage der historischen Daten trainieren. Auch sollte noch versucht werden das Problem umzuformulieren und nicht mehr die Forderung zu haben, dass jede Population angebaut werden muss.

Literatur

- [1] *Syngenta crop challenge 2021*. 2020. URL: <https://www.ideaconnection.com/syngenta-crop-challenge/%20challenge.php>.
- [2] Michael R. Garey und David S.O Johnson. “Sets and Partitions”. In: *Computers and Intractability; A Guide to the Theory of NP-Completeness*. USA: W. H. Freeman & Co., 1990, S. 223. ISBN: 0716710455.
- [3] Florian Jarre und Josef Stoer. “Lineare Programme, Beispiele und Definitionen”. In: *Optimierung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, S. 9–21. ISBN: 978-3-642-18785-8. DOI: 10.1007/978-3-642-18785-8_2. URL: https://doi.org/10.1007/978-3-642-18785-8_2.
- [4] Florian Jarre und Josef Stoer. “Das Simplexverfahren”. In: *Optimierung*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, S. 23–66. ISBN: 978-3-642-18785-8. DOI: 10.1007/978-3-642-18785-8_3. URL: https://doi.org/10.1007/978-3-642-18785-8_3.
- [5] Werner Dinkelbach. “Ganzzahlige lineare Programmierung”. In: *Operations Research: Ein Kurzlehr- und Übungsbuch*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1992, S. 109–158. ISBN: 978-3-642-77162-0. DOI: 10.1007/978-3-642-77162-0_4. URL: https://doi.org/10.1007/978-3-642-77162-0_4.
- [6] Mariette Awad und Rahul Khanna. “Support Vector Regression”. In: *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Berkeley, CA: Apress, 2015, S. 67–80. ISBN: 978-1-4302-5990-9. DOI: 10.1007/978-1-4302-5990-9_4. URL: https://doi.org/10.1007/978-1-4302-5990-9_4.
- [7] Gianluca Bontempi. “Permutation test”. In: *Statistical Foundations of Machine Learning*. Universite Libre de Bruxelles, 2017.
- [8] L. Gurobi Optimization. *Gurobi Optimizer Reference Manual*. 2021. URL: <http://www.gurobi.com..>

Erklärung

Hiermit versichere ich, dass ich die Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Düsseldorf, im März 2021

(Jan van Grimbergen)