Continent-wide planning of seed production: mathematical model and industrial application

Yanbin Zhu[a]*, Nilay Shah[a], Gabriel Carré[b], Simon Lemaire[c], Erick Gatignol[c], Patrick M. Piccione[d]

[a] Centre for Process System Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK

[b] Process and Production Technology, Technology and Engineering, Syngenta, 12 Chemin de l'Hobit, 31790 Saint-Sauveur, France

[c] Seed Operations EAME, Global Seed Operations, Syngenta, Route de Francescas, 47600 Nérac, France

[d] Process Studies Group, Technology and Engineering, Syngenta, Breitenloh 5, CH-4333 Münchwilen, Switzerland

Abstract

The seed supply chain is one of most sophisticated elements of the agricultural value chain with long lead times, fragmented structure and high levels of uncertainty. Since the seed industry has received less attention in research compared with other sectors in the agriculture industry, it has enormous potential for improvement due to the lack of comprehensive mathematical optimization applications, increasing competition within the industry and decreasing spare arable land worldwide. All of the existing optimization applications in the seed supply chain have concerned land allocation at the farm level as well as regional level processing and distribution after harvesting. This research closes the gap between farm level planning and regional level distribution through optimization of seed production planning at a regional level, taking account of a number of complex constraints and practical preferences. Compared to a "business as usual" approach, the proposed application can save up to 16% of the total cost as well as 9% land usage and effectively mitigate major risks in the planning phase. The method is evaluated using Syngenta's industrial case studies.

Keywords: Seed Supply Chain Planning, Stochastic Optimization, Cost Reduction, Land Usage Reduction

* Corresponding author. Tel.: +86 18610905825

Email address: yanbinzhu@outlook.com

Introduction

The global population is expected to rise by one-third to 9.7 billion in 2050. This increasing population around the world will lead to demands for additional crop production as well as effectively designed supply systems for these agricultural products. The majority of these agricultural products originate from seed, thus it is even more critical to meet the demands of seed supply. Previously, the cultivated seed was traditionally saved seed from the previous season. With the advancement of technology, commercial seed products have become the more reliable and effective option compared to stored seeds. The shift from using saved seed to purchasing commercial seed products has led to the growth of the seed industry in the recent decades. There is an emerging tendency of consolidation in global seed companies as well (Howard, 2009). This trend reveals the stress from competition and the need to design more efficient systems to produce and supply seed products. Seed supply chain management generally consists of the following components:

1. Seed production planning
2. Planting based on planned portfolio
3. Harvesting of crops
4. Processing of harvested crops to packaged products
5. Distribution and storage of final products
6. Sales

It is an extremely long and complicated process from the planning of production allocation to the final sales in the market in the agricultural industry. Additionally, the majority of seed product is hybrid, which normally includes two parent varieties, in turn requiring more effort in the planning stage, involving production and supply of parent varieties and consideration of genetic pollution risks in production. In general, the detailed production allocation plans must be finalized at least 6 months ahead of cultivation, and the general plans should be generated for the future three years. Ideally, all these plans should be able to be updated monthly or even weekly from the initial cultivation planning to harvesting and to final sales. In other words, many highly uncertain data sets, e.g. demand and yield, are predicted at least 6 months ahead.

Due to the complexity involved in the seed production process, the seed production planning process can be further refined into regional level planning and field level planning in each region. Based on the level of decisions, the scope of production planning can be divided into strategic, tactical and operational levels.

In summary, the seed supply chain is a long, complicated, fragmented and uncertain value chain. This indicates the necessity of integrating the entire process from seed growing to final sales. An effective technique to resolve and analyse all these issues is the design and optimization of effective seed supply chains. The objective of this study is to present an insight into the value that can be gained from implementing mathematical optimization in strategic seed production planning and build the initial foundation for a comprehensive framework of global supply chain optimization in the seed industry.

Problem Statement

The aim of this research is to develop a framework for the application of mathematical optimization in strategic seed production planning at a regional level. The problem statement can be summarized as follows.

Given:

i. Aggregated regional demand and opening stock information for different seed varieties per season
ii. Performance distribution of each growing area (yield distributions by variety and region)
iii. Regional land availability
iv. Additional business constraints
v. Additional technology and biological constraints

Determine:

i. Optimal allocation of land in each region for varieties considering total cost, production risk and plan robustness

Comhaire and Papier (2015) summarised the complexity of such tasks in Europe:

1. More than 12 months’ lead time between land allocation and seed harvesting
2. Diversified geographic locations for the entire supply process with more than hundreds of seed-processing sites and thousands of supply farms
3. Regulatory variations among the various sales regions

In our study, the additional complexities from the particularity of the industry as well as technology restrictions have been considered, for example, the deviations in sales strategies for seed products with various maturity levels, limitations due to pollination pollution and robustness to climate variability. These requirements are interpreted as linear equality and inequality constraints available to the model to be selected based on the specific planning strategies. As the final output is directly correlated with the individual performance of the field (i.e. yield), the potential risks from variations in yields need to be considered in the quest of formulating a reliable and realistic planning model. Additionally, the uncertainty in production derived from these potential risks is incorporated into the optimization process through stochastic optimization. Therefore, this problem can be precisely defined as the development and implementation of a decision-making optimization model for large-scale strategic seed production planning under uncertainty at a regional level.

Literature Review

The seed supply chain is considered as a subset of the general Agricultural Supply Chain (ASC). Due to the comparatively low attention on the ASC until the beginning of this century, there are only a few published comprehensive reviews in the ASC field.

Ahumada and Villalobos (2009) undertook a comprehensive review of planning models in the agri-food supply chain; A literature review was provided by Zhang and Wilhelm (2011) for the supply chain in the specialty crops industry with a particular focus on decision support models; Lowe and Preckel (2004) presented a brief review of key agribusiness problems as well as illustrations of the needs for further research including effective decision support tools for crop production and processing. To further summarize the previous studies in ASC, the majority of the research focuses on the strategic and operational optimization of supply chain with weak connections in between. Specifically in the seed industry, its fragmented structure, long lead times and complex biological restrictions lead to additional difficulties in comprehensively investigating its supply chain. Thus there are few studies which study the practical optimization of seed supply chains. The few existing research articles emphasise strategic optimization for land allocation and harvesting schedules at a farm level as well as operational optimization for processing and distribution after harvesting.

With respect to research on the land allocation problem, most of the studies concentrate on a land allocation plan for limited categories of agricultural products (unlike our case with more than a hundred varieties of seeds) for individual farmers and optimal land allocation at the farm level. Heady (1954) has been widely recognised as the first demonstration of Linear Programming (LP) in agricultural land allocation problems for farmers. Huh and Lall (2013) illustrated two stochastic programming models for the optimization problems of cropping land allocation and determination of irrigation water strategies under fluctuations in weather conditions and market prices. A multi-objective deterministic allocation model was proposed by Annetts and Audsley (2002) to simultaneously evaluate profitability and sustainability through manipulating the cropping and machinery scenarios for a single farm.

Optimization models for seed supply after harvesting were described by Zuo et al. (1991), who presented a deterministic mathematical model for a seed corn processing and distribution system with optimal solutions and sensitivity analysis. Junqueira & Morabito (2012) built a more integrated deterministic tactical planning model after harvesting with multiple stages, plants and products as well as circulation taxes for a Brazilian seed corn company.

The optimization of seed production planning at a regional level can be considered as the first step to link between strategic and operational decisions in a seed supply chain. The only directly related study regarding the seed production allocation optimization problem was proposed by Comhaire and Papier (2015), describing the difficulties in seed production planning and presenting some insights into the requirements for an ideal allocation tool. There were also several attempts to incorporate dynamic programming and stochastic optimization in seed production planning, e.g. the studies of Jones *et al.*(2001) and Jones *et al.* (2003). However, none of these have fully implemented mathematical optimization and incorporated detailed practical constraints as well as treated uncertainty comprehensively.

Therefore, there is a gap in the systematic development of a tool applying modern mathematical optimization techniques to seed production planning.

To the best of our knowledge, there is no literature which comprehensively describes the optimization of the seed supply chain through mathematical programming with a focus on seed production planning at a regional level.

Therefore, this research aims to develop an innovative application using mathematical optimization in seed production planning to link the strategic and operational management levels. It could act as the foundation for global optimization of the entire seed supply chain, and illustrate a successful industry – academic collaboration.

Mathematical Formulation

Due to the fragmented structure of the seed supply chain and its long lead times, some of the information used to formulate the mathematical model is uncertain, such as predicted yield and predicted demand with more than one year's lead time. The stochastic nature of various input information (e.g. demand forecast and yield) as well as the annual cycle in the production process make it practically meaningless to consider a dynamic process with more than one year's lead time, i.e. this problem is considered as a static planning problem. From the perspective of systematic development, it is more effective to assemble the fragments into a framework assuming all the uncertain information is deterministic, i.e. assume a deterministic optimization model. Afterwards, the uncertainty can be incorporated into the deterministic framework to increase the reliability of the model. In this study, a simplified two – stage stochastic optimization model is subsequently developed to demonstrate some of the insights associated with uncertainty.

Notation Table

For the sake of readability, the term hybrid indicates the final seed product to be sold to the market, male variety represents the parent male seed used to plant the hybrid.

Indices

c	Growing country
i	Growing region
j	Seed product (Hybrid)
k	Sales group
m	Male varieties
s	Scenario

Continuous non-negative parameter

D_j	Demand of hybrid j (bag)
$MnA_{i,j}$	Minimum amount of land to be allocated to hybrid j in growing region i (ha)
$MnGC_j$	Minimum number of growing countries of hybrid j
$MnGR_j$	Minimum number of growing regions of hybrid j
MnL_i	Minimum amount of land to be allocated in growing region i (ha)
MxI_i	Maximum number of isolations in growing region i
MxL_i	Maximum amount of land to be allocated in growing region i (ha)
$MxP_{i,j}$	Maximum production volume of hybrid crop j in growing region j (bag)
$MxPre_j$	Max weighted preference of hybrid j
$MxTVar_k$	Maximum average variability of sales group k
$MxVar_j$	Maximum variability of hybrid j
OP	Penalty cost for overproduction (($/bag)
OS_j	Opening stock of hybrid j (bag)
$Pre_{i,j}$	Preference index of hybrid j produced in growing region i
SP	Penalty cost for shortage ($/bag)
SPO_k	Penalty cost for overproduction for sales group k in stochastic model ($/bag)
$SPro_s$	Probability for the occurrence of scenario s
SPS_k	Penalty cost for shortage for sales group k in stochastic model ($/bag)
$SY_{s,i,j}$	Yield of hybrid j grown in growing region i in scenario s (bag/ha)

$UG_{i.j}$	Growing cost of hybrid j in growing region i per hectare ($/ha)
$UP_{i.j}$	Unit processing cost of hybrid j in growing region i ($/bag)
$UT_{i.j}$	Unit transport cost of hybrid j in growing region i ($/bag)
$V_{i,j}$	Variance of hybrid j harvested in growing region i
$Y_{i,j}$	Yield of hybrid j growing region i (bag/ha)
$\overline{Y}_j$	Average yield of hybrid j among all the growing regions (bag/ha)

Binary parameter

$CAM_{c,i}$	Binary parameter indicating the membership between growing region i and growing country c
$HPM_{j,m}$	Binary parameter indicating the membership between hybrid j and corresponding male variety m
$SG_{k,j}$	Binary parameter presenting the membership between sales group k and hybrid j

Continuous non – negative variables

CS_j	Deterministic closing stock of hybrid j (bag)
EOC	Expected value of overproduction penalty cost in stochastic model ($)
EPC	Expected value of processing cost in stochastic model ($)
ESC	Expected value of shortage penalty cost in stochastic model ($)
ETC	Expected value of transportation cost in stochastic model ($)

GC	Total growing cost (\$)
$LA_{i,j}$	Amount of land allocated to hybrid j in growing region i (ha)
OC	Total deterministic overproduction penalty cost (\$)
PC	Total deterministic processing cost (\$)
S_j	Deterministic shortage of hybrid j (unit)
SC	Total deterministic shortage penalty cost (\$)
$SO_{s,j}$	Total stochastic overproduction of hybrid j in scenario s (bag)
$SS_{s,j}$	Stochastic shortage of hybrid j in scenario s (bag)
STC	Total cost in stochastic model (\$)
$STP_{s,j}$	Total stochastic production of hybrid j in scenario s (bag)
TC	Total deterministic cost (\$)
TP_j	Total deterministic production of hybrid j (bag)
$TPre_j$	Total preference of hybrid j
$TraC$	Total deterministic transportation cost (\$)
$TVar_k$	Total average variability of sales group k
Var_j	Variability of hybrid j

Binary variables

$x_{i,j}$	1 if hybrid j is planned to be grown in growing region i; 0 otherwise
$xo_{s,j}$	1 if hybrid j is overproduced in scenario s; 0 otherwise
$xp_{i,m}$	1 if male variety m is grown in growing region i; 0 otherwise

$xu_{s,j}$	1 if hybrid j is under-produced in scenario s; 0 otherwise
$y_{c,j}$	1 if hybrid j is planned to be grown in growing country c; 0 otherwise

Deterministic Model

In this section, the framework of our mathematical model assuming deterministic information is introduced including the objective function as well as both the equality and inequality constraints. This serves to describe the deterministic planning model effectively. Later we shall introduce the stochastic elements.

Objective Function

The objective of this MILP is to minimize the total cost subject to a series of hard and soft constraints; the basic components for the seed production process are growing, processing and distribution, thus the total cost can be further classified into growing cost, processing cost and transportation cost. In this regional level model, the land allocated to each product j in each growing region i is regarded as a continuous decision variable and is the most important decision variable. The growing cost is estimated as the product of land usage and unit growing land cost, while both the processing cost and transportation cost are represented by production volume multiplied by the corresponding unit cost. An extra penalty cost for shortages (i.e. demand shortfalls) is generated to avoid infeasible solutions.

$$GC = \sum_i \sum_j LA_{i,j} \times UG_{i.j} \tag{1}$$

$$PC = \sum_i \sum_j LA_{i,j} \times Y_{i,j} \times UP_{i.j} \tag{2}$$

$$TraC = \sum_i \sum_j LA_{i,j} \times Y_{i,j} \times UT_{i.j} \tag{3}$$

$$SC = \sum_j SP \times S_j \tag{4}$$

$$OC = \sum_j OP \times CS_j \tag{5}$$

$$TC = GC + \ PC + \ TraC + SC + OC \tag{6}$$

Basic Constraints

The primary constraints for this problem can be categorized into six types; they are all essential to ensure the practical applicability of the seed production planning model.

Demand Balance

The demand constraint can be described as the mass balance accounting for total production, opening stock of product j and demand, with the shortage variable and closing stock variable being added to avoid any infeasibility due to unachievable demand or over-constrained production volume respectively.

$$TP_j = \ \sum_i Y_{i,j} \times LA_{i,j} \tag{7}$$

$$CS_j + D_j = OS_j + TP_j + S_j \tag{8}$$

Regional level land availability

The total available land in each growing region is considered as the maximum accessible and arable land belonging to the company and its potential third party contractors to be allocated in each region, meanwhile the minimum land allocation in each growing area is intended to represent the minimum contracted land to be used in each growing region, and this is an important consideration to ensure continuity of business in the region.

$$MnL_i \leq \sum_j LA_{i,j} \leq MxL_i \tag{9}$$

Land allocation

It is not profitable to allocate land in a particular region to a product below a minimum practical level. Thus, an additional binary decision variable ($x_{i,j}$) is implemented to indicate this discrete limitation.

$$x_{i,j} \times MnA_{i,j} \leq LA_{i,j} \leq x_{i,j} \times MxL_i \tag{10}$$

Biological constraint

Nowadays the majority of the seeds are grown as hybrids, i.e. the offspring of two parent varieties. Theoretically one hybrid is the offspring of one male variety and one female variety, and the nearby existence of any other male variety will affect the quality of pollination, i.e. pollution in pollination. Practically, the pollution in pollination can be effectively avoided by isolating hybrids with different male varieties. Based on the limitation from the previous constraint (Land allocation), the number of isolation blocks available to each grower is limited, thus the maximum number of isolation blocks within each growing region should be limited according to the limitations associated with growers. In order to properly interpret this constraint, an extra binary decision variable ($xp_{i,m}$) is used to indicate the presence of particular male varieties in each region. An extra binary parameter ($HPM_{j,m}$) is used to reveal the correlation between two types of decision variables ($x_{i,j}$ and $xp_{i,m}$) to ensure a correct mapping between varieties (j) and males (m).

$$\sum_m xp_{i,m} \leq MxI_i \tag{11}$$

$$HPM_{j,m} \times x_{i,j} \leq xp_{i,m} \leq \sum_j HPM_{j,m} \times x_{i,j} \tag{12}$$

Constraint (12) defines this variable and constraint (11) ensures that the maximum number of isolation blocks is not violated.

Distribution limitations

Due to the differences in the policy of each country, it is impossible to transport certain biological products between certain countries. The maximum production volume for each product in each growing region is thus applied to simplify this distribution limit at a regional level.

$$Y_{i,j} \times LA_{i,j} \leq MxP_{i,j} \tag{13}$$

Commercial importance of the products

From the perspective of the business, the commercial importance of each product depends on its maturity as a product and potential within the market, i.e. the sales targets for the products with various relative significance levels are different. The most important products with highest potential are the newly developed seeds with lowest product maturity as well as highest future sales potential. The mature seeds are less important and the old products with high product maturity as well as lower potential in the market are grown with the lowest priority. The level of importance represents the desired "stability" of production; more important products require more stable production, i.e. lower risk of variation in production. Even though it is difficult to directly interpret this constraint in a deterministic model, it can be indirectly estimated through two types of extra constraints: minimum number of growing regions/countries and tuneable risk constraints.

In practical field production, a very unfavourable event could lead to the complete loss of all the crops within entire field. It has been known to happen, but it is impossible to precisely quantify or forecast all the unpredictable factors/events, e.g. extreme weather conditions and sudden changes of temperature. The first factor can be considered as the "foundation" to manually enforce the minimal number of growing regions and growing countries, so that the overall production can be made more robust by diversifying the growing regions and countries, i.e. limiting the maximum amount of production of one hybrid in one particular location. In case of an extreme event occurring, the diversification will reduce the risk of total production loss. The number of growing regions for a hybrid can be evaluated by the summation of binary variable $x_{i,j}$ for all growing regions. The new binary decision variable ($y_{c,j}$) is added to indicate the existence of seed production in each country. Similar to the biological constraint, an extra binary parameter is implemented to reveal the correlation between $x_{i,j}$ and $y_{c,j}$.

$$MnGR_j \leq \sum_i x_{i,j} \tag{14}$$

$$MnGC_j \leq \sum_c y_{c,j} \tag{15}$$

$$x_{i,j} \times CAM_{c,i} \leq y_{c,j} \leq \sum_i x_{i,j} \times CAM_{c,i} \tag{16}$$

Risk Constraints – Yield Uncertainty

Even though the uncertainty in production cannot be comprehensively captured in a deterministic model, the risk involved in the optimization model can be constrained based on the maximum level of uncertainty in production. Unlike the basic constraints described previously, the strength of the risk constraints is determined by the risk appetite in a specific problem setting. Two approaches are introduced by the company to limit the level of stability.

1. “Preference” of Production
2. “Variability” of Production

The preference of production is decided using a user defined preference index determined by the industry specialists based on experience with different varieties and different growing regions to manipulate the risk involved in production. The preference index is defined for each valid variety and growing region pair. A lower preference index represents a lower historical risk associated with that hybrid – region allocation as well as higher stability in harvesting. The preference index is subjective and implies a qualitative but accessible view of risk. Meanwhile, it reveals less dependence on the accuracy of the risk information due to its high level of abstraction in the risk of production. In other words, the constraint for the preference of production can only be qualitatively considered as an indicator of risk.

The total preference index of product j is normalized by average land allocation of hybrid j.

$$TPre_j = \frac{\bar{Y}_j \times \sum_i Pre_{i,j} \times LA_{i,j}}{\mathrm{D}_j} \quad (17)$$

Referring to the commercial importance of the products described previously, the industrial specialists have reference values for the importance of each product, and based on these references, they will estimate the levels of preference of each product through their experience. A simplified approach to achieve such a limitation is to define the upper bound of the total preference of each product, noting that this is a risk measure and hence constrained from above (the effect of this measure can be explained through sensitivity analysis):

$$TPre_j \leq MxPre_j \quad (18)$$

As introduced previously, preference can only be qualitatively considered as an indicator of risk. In case where variances of the yields (assuming the yield is normally distributed) are available with limited accuracy, the variability constraints presented in (19) – (22) can be applied instead. The variability is defined as the ratio of deviation caused by variance in harvesting normalized by demand. The variability of each product is constrained by a manually determined upper bound.

$$Var_j = \frac{\sum_i Y_{i,j} \times LA_{i,j} \times V_{i,j}}{D_j} \quad (19)$$

$$Var_j \leq MxVar_j \quad (20)$$

The advantage of the variability constraint is its quantitative representation of the risk in production, thus it can be incorporated together with a minimum number of growing regions/countries to capture the importance of different seeds. In terms of sales targets, vital products require a stable planting strategy and cost is less critical while the declining products are driven by cost regardless of stability or potential shortage. From the perspective of harvesting, the newly developed seeds with less mature technology are inherently unstable. It is not practical only to manage the variability through constraint (19) and constraint (20) for each product. Additionally, the average variability within each sales group is constrained to simultaneously achieve different objectives.

$$TVar_k = \frac{\sum_j SG_{k,j} \times Var_j}{\sum_j SG_{k,j}} \quad (21)$$

$$TVar_k \leq MxTVar_k \tag{22}$$

A similar approach can be applied for the preference constraints, however it is harder to “tune” these as they have a higher level of subjectivity and approximation compared with variability, which is a statistical quantity.

To summarize the risk constraints, either preference or variability should be selected to control and quantify the stability of production in the case of insufficient predicted future information. The preference constraint is more qualitative and has less dependence on the risk information but less controllability, while the variability is more flexible and practical to manipulate but it requires quantitative uncertainty information (e.g. variance).

The combination of minimum numbers of growing regions/countries, which qualitatively reduce the effect from unexpected extreme events, and tuneable risk constraints, which constrain the level of stability in production, is introduced to reduce the risk in production for the deterministic model.

In the next section, we extend the deterministic analysis to include stochastic elements with a focus on yield uncertainty.

Stochastic Model

The risk associated with production in this problem can be specifically defined as the uncertainty in the yield which is only observed during harvesting. The effect of uncertainty can only be partially reduced in the previous deterministic model. In order to further optimize the seed production planning process together with risk, both stochastic information and the modified model incorporating uncertainty should be implemented instead. In this study, the insights and benefits of applying stochastic optimization are demonstrated by assuming a description of uncertainty (e.g. predicted yield distribution) is available with practical reliability. Hence the practicality of the model can be further improved and the impacts from uncertainty can be further eliminated through a simplified two – stage stochastic optimization.

For stochastic optimization of seed production planning, the scenario – based approach is applied to construct the deterministic equivalent for the stochastic linear programme. Normally, the algorithm used in a deterministic model is also applicable for this type of stochastic model.

In this section, the reformulations required to convert the deterministic model to a stochastic model are described first followed by the specifications of objective function and constraints for the reformulated model.

Reformulation of the model

The deterministic model introduced previously can be regarded as the framework for modelling seed production planning. The remaining issue to be considered is the effect from uncertain information. In this study, a simplified two – stage stochastic optimization method is applied to balance the influences from uncertainty.

Since this model is built to demonstrate the value of a stochastic model, only the yields of crops are considered as uncertain parameters and all the practical decisions are assumed to be made in the first stage, i.e. land allocation of hybrid by growing region. The actions in the second stage are harvesting, processing and transportation, as a simplified stochastic model in this study, all these practical decision variables from the second stage are neglected due to limited availability of practical data. Only the decision variables regarding closing stock and shortage in production of hybrid by scenario are considered from the second stage to avoid infeasibility due to shortage or overproduction. The dependent variable related to harvesting is the total volume of production and volume of underproduction (shortage). An extra variable for overproduction is also generated due to the variations in production volumes in the second stage. The demand balance is reformulated with the additional dimension of scenarios.

$$STP_{s,j} = \sum_i SY_{s,i,j} \times LA_{i,j} \tag{23}$$

$$SO_{s,j} + D_j = OS_j + STP_{s,j} + SS_{s,j} \tag{24}$$

From a practical point of view, both shortage which leads to unfulfilled demand and reduced customer satisfaction and overproduction which leads to additional inventory cost) are not desired. Shortage and overproduction cannot occur simultaneously for a particular hybrid, thus two additional binary variables are generated to avoid the possible coexistence of overproduction and underproduction.

$$xo_{s,j} \leq SO_{s,j} \leq xo_{s,j} \times D_j \tag{25}$$

$$xu_{s,j} \leq SS_{s,j} \leq xo_{s,j} \times D_j \tag{26}$$

$$xo_{s,j} + xu_{s,j} \leq 1 \tag{27}$$

Considering perishability and sensitivity of seed products, overproduction may even be more critical than shortage to ensure operational efficiency of the company. A penalty cost for overproduction is added to quantify the effect of overproduction and minimize the impact from it. The expected values of overproduction and shortage cost are calculated in the objective function. The penalty cost for the overproduction and shortage can be formulated with a penalty cost dependent on each sales group.

$$EOC = \sum_k \sum_s \sum_j SG_{k,j} \times SPro_s \times SPO_k \times SO_{s,j} \tag{28}$$

$$ESC = \sum_k \sum_s \sum_j SG_{k,j} \times SPro_s \times SPS_k \times SS_{s,j} \tag{29}$$

In terms of actions related to processing and distribution, the deterministic processing and transportation cost are reformulated to the corresponding mathematical expectation of all possible scenarios.

$$EPC = \sum_s \sum_i \sum_j SPro_s \times LA_{i,j} \times SY_{s,i,j} \times UP_{i.j} \tag{30}$$

$$ETC = \sum_s \sum_i \sum_j SPro_s \times LA_{i,j} \times SY_{s,i,j} \times UT_{i.j} \tag{31}$$

Objective Function

The deterministic equivalent objective function is the summation of the deterministic growing cost as well as expected values of processing, transportation, overproduction penalty and shortage penalty cost.

$$STC = GC + EPC + ETC + EOC + ESC \tag{32}$$

Due to the incorporation of the yield probability distribution, the objective function can be specified based on the relative importance of the seed product. Both penalty costs for overproduction and shortage can be manipulated independently according to the specific targets for different varieties.

Constraints

Other than the reformulated demand balance introduced previously, the land availability, existence of land allocation, biological constraint and distribution limit described in the basic constraints of the deterministic model can be applied without any modification since there are no scenario – dependent second stage variables involved. Regarding the constraint concerning the relative importance of each product, it is interpreted directly through sales group dependent tuneable penalty costs in the objective function. However, the minimal number of growing regions can be incorporated with relaxed bounds as insurance to avoid any significant impact from inaccurate uncertainty descriptions. For example, the minimum number of growing regions is set to 4 for one certain seed product in the deterministic model, and that value can be reduced, e.g. 3 or 2, in the stochastic model to reduce the effects from inaccurate information while preventing unnecessary subjective restrictions from this type of constraint. The additional risk constraints implemented in the deterministic model should not be utilized in the stochastic model, otherwise the subjective restrictions from risk constraints will affect the optimality and feasibility of stochastic optimization.

Demand uncertainty

Compared with risk in production, demand will be more challenging to accurately predict as it will be dependent on various stochastic process, futures prices, exchange rate and farmers’ choices. Combined with internal uncertainty, such as inventory level for next year (at the time of production planning for next season, sales for the current season are not finished yet), accurate demand forecasting will not be a simple problem being solved in the short term. From the perspective of production planning optimization, theoretically, it will be an option to extend the model to incorporate demand uncertainty through stochastic optimization or robust optimization.

However, practically it is not a currently feasible option due to the unavailability of accurate data for demand forecasts.

Case Study and Results

Real world data from Syngenta have been incorporated into the models introduced previously to undertake case studies and evaluate the benefits of optimization, and form the basis of a decision support tool. The industrial application of deterministic models is described initially followed by the application of the reformulated stochastic model. The purpose of this study is to present insights into the value of implementing strategic optimization in seed production planning. Before the implementation of mathematical optimization in the strategic seed production planning process, planning experts from the company used an iterative approach to manually allocate the required production amounts into available production regions within a spreadsheet tool.

All the mathematical models described in this section are built in GAMS 24.5.6 and solved by the commercial solver CPLEX 12.6.1.

Initial Case Study

As an initial motivational case study to explore the expected benefits of this approach, only one key performance indicator (KPI) is applied to evaluate the quality of allocation: total cost. This initial deterministic optimization framework (only objective function and demand balance, i.e. LP) is applied to a real industrial case; Syngenta AG provided real data on two crops representing 200+ products to be allocated to 20+ growing regions and 10+ growing areas on one continent. The optimal allocation plans are then compared with the original seed production plans generated by Syngenta. The results of the optimization contrasted with the original allocation plans of Syngenta can be revealed and divided into two aspects: reductions in cost and improvements in planning process efficiency.

Cost Reductions

The initial model can be used as either the case to quantify the value of optimization or a starting point for the implementation of a more detailed deterministic model. Taking the original allocation plans for the two crops selected in Syngenta as examples, the following reductions in cost can be observed for total cost:

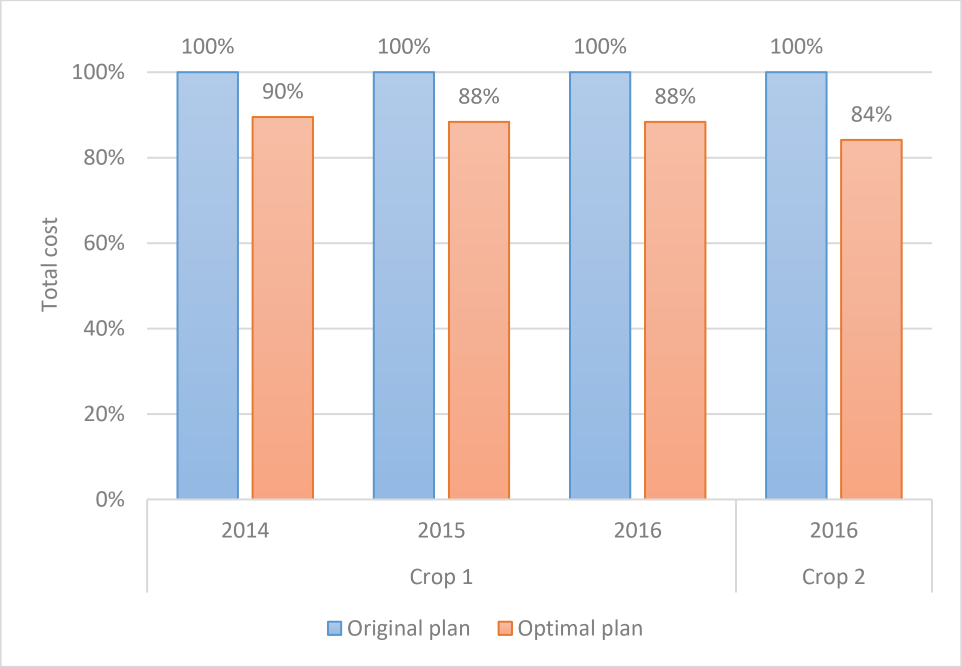

Fig. 1 Comparison between original and optimal plan

For confidentiality reasons, the total cost calculated by the original plan (developed manually by experts) value has been used as a normalization factor, but it can be seen that in terms of percentage, 10% - 16% savings have been achieved in total cost. These savings are equivalent to 5 to 10 million US dollars per year for the regional level case study. The major sources of this cost savings are the better use of land, balancing cost, yield and availability for over 100 varieties in 20 regions.

Efficiency of planning

According to the experience of several industry specialists in Syngenta, it previously took one month to generate one allocation plan. By using mathematical programming, around 10 hours is used to set up the model and less than a second is needed to generate the plan presented in Fig. 1 using a standard laptop with guarantee of optimum. In other words, the implementation of mathematical programming in seed production planning can dramatically boost the efficiency of the entire process.

Case Study I: Deterministic Model

The data for two of major product families of Syngenta have been tested for the pilot industrial implementation of the complete deterministic model. Since similar benefits can be obtained from both products, only the results for one type of crop in 2014 with 100+ hybrids, 100+ male varieties, 20+ growing regions and 10+ growing countries are introduced in this paper. In order to present the business value of the model, KPIs are defined before the execution of optimization. Only the total cost and total land usage are presented as KPIs in this section for the purpose of simplification.

In the previous subsection, the initial optimization model with only the demand balance constraint was tested using real world data from Syngenta. In this section, all the additional constraints described previously have been integrated into the model and the corresponding production plans have been generated and analysed in detail. Similarly as for the preceding industrial case study, all the solutions are compared with the original seed production plan from Syngenta. Moreover, the preference and variability constraints have been analysed separately. With respect to product groups defined in this study, only three groups are considered: New, Mature and Declining. The new products are regarded as the most important products with the objective of maximizing stability regardless of any additional cost, while the production cost of the declining products should be minimized with lowest production priority regardless of stability in production. The mature products lie between new and declining products with an objective of minimizing total cost with reasonable levels of stability.

All the additional constraints are applied to improve the practicality and applicability of the seed production planning; however the most appropriate combinations of constraints applicable to each product as well as each year will be highly dissimilar and should be specified by the company at the time of planning. In this case study, all the basic constraints introduced previously are incorporated while the risk constraints have been tested separately through the following combinations.

Table 1 - Combinations of constraints

Scenario	Constraints

S1	Basic constraints
S2	Basic constraints + Preference constraint (Average 2.50)
S3	Basic constraints + Preference constraint (Average 2.25)
S4	Basic constraints + Variability constraint (New Group Average 0.25)
S5	Basic constraints + Variability constraint (New Group 0.25, Mature Group 0.35)

The constraints included in the basic constraints set are demand balance, regional level land availability, existence of land allocation, biological constraint, distribution limit and commercial importance of the products (only minimum number of production regions/countries). In **S2** and **S3**, the maximum average preferences for each seed product family are set at 2.50 and 2.25 respectively. In **S4**, only the overall average variability ($MaxTotalVariability_k$) of the new group is constrained at 0.25, while both the overall average variability of the new group and the mature group have been constrained at 0.25 and 0.35 respectively in **S5**.

In order to maintain consistency with the original planning process from Syngenta, the minimum and maximum land availabilities are defined by the actual amount of contracts assigned by Syngenta. Since these contracts are assigned based on preceding experience instead of any optimization, there are some extra limitations from these contracts. In spite of such additional restrictions, the seed production planning process can still be improved significantly as shown in Fig. 2 and Fig. 3. Note that all the results presented are normalised using data from the original plan from Syngenta to preserve confidentiality.

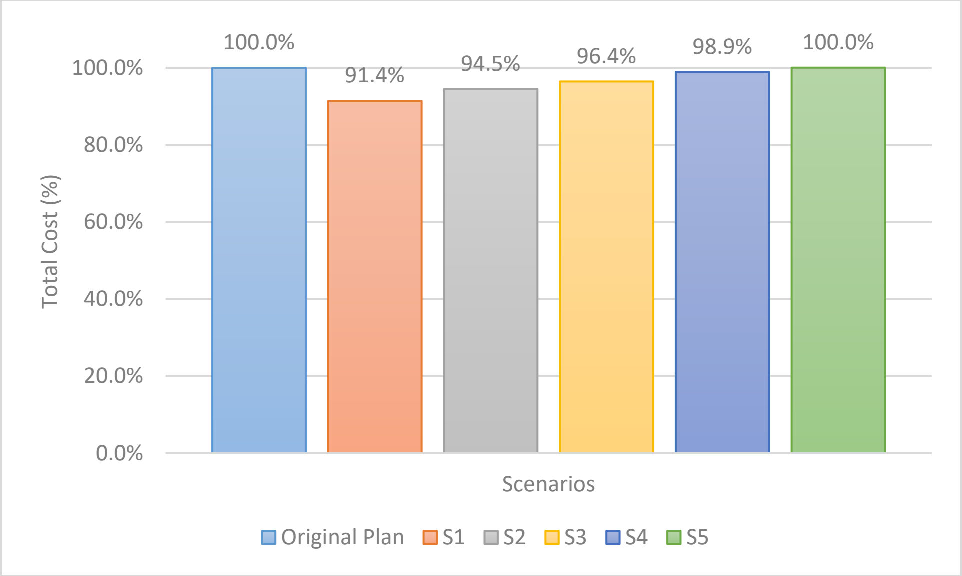

Fig. 2 - Comparisons of total cost

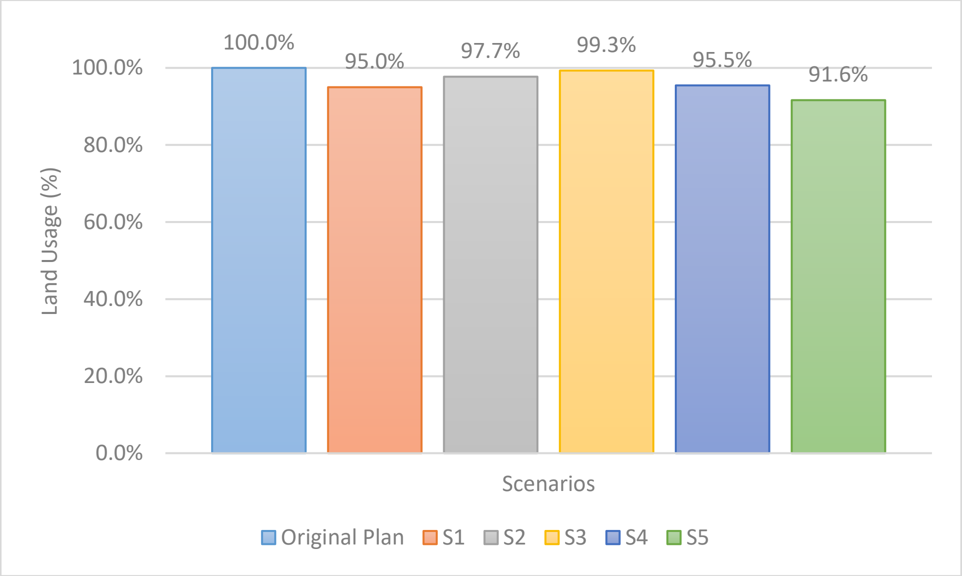

Fig. 3 - Comparisons of total land usage

Without any explicit limitation on risk, the use of deterministic optimization can save at least 8.6% of cost and 5% land usage even with additional subjective restrictions (through manually defined minimum land availability). All these savings represent a significant increase in profitability and sustainability of the company since the implementation of optimization can not only reduce the cost in production and land usage but also provide the possibility to easily transfer any business information or requirement to the production plan instantaneously.

In terms of the risk constraints, the impacts and drawbacks of preference and variability constraints are already discussed previously thus they will not be specifically discussed in detail in this case study.

The addition of preference constraints further restricts the feasible range for selection and therefore reduces the savings compared with the original seed production plan. Meanwhile, a slight change in constrained average preference index can lead to a large deviation in both total cost and land usage. Even though the preference index is unable to quantify precisely the risk associated with seed production, it provides an approach to evaluate the effect of the risk in case of limited access to accurate statistical information, e.g. the predicted yield distribution.

Similarly as for the preference constraint, the total cost increases with increased limitations on variability, however the user may also prioritise the stability due to its ability to quantify the risk with a practical interpretation.

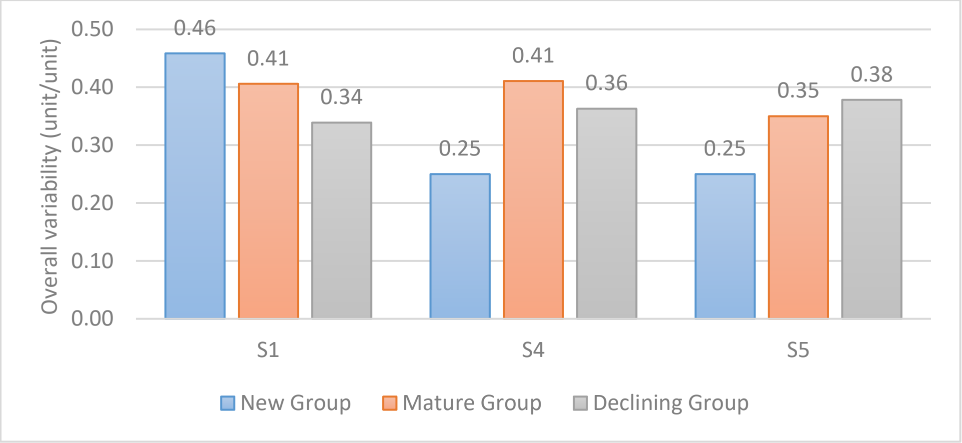

Fig. 4 - Overall variability for different sales groups

Although the savings in cost gradually diminish as the restrictions on stability increase, the robustness of the allocation plans also increases. Without any limitation on variability, the new product group would experience the highest variability due to the relatively low maturity in technology and the declining product group would experience much lower levels of variation due to its higher maturity and predictable yields. However the primary target is to obtain lower variability in the new group. The implementation of this variability constraint enables satisfaction of this difficult requirement. However, it is not straightforward to quantify the commercial benefits in the improved stability through the variability which represents the maximum ratio of deviation compared with demand. Hence, a stochastic model is introduced to quantify better the direct benefits brought from the improved stability in the next section.

Case II: Stochastic Optimisation Model

The stochastic model is implemented with statistically simulated yield distribution data of one type of crop from Syngenta AG. The yield distribution itself is generated based on historical data and a statistical approach, but the details of distribution generation are not discussed as they are beyond the scope of this study. Even though the results stated here are only for insights into the benefits of stochastic optimization, the implications are already extremely valuable and the method is particularly attractive to implement at full scale.

Since the purpose of this case study is to evaluate the value of stochastic optimization and the total production volume is uncertain due to uncertain yield, total cost and total production volume are used as the main KPIs. Additionally, the cost per bag is illustrated as an effective indicator of the effectiveness of seed production planning.

All the yield distributions (for all the hybrid and growing region combinations) have been approximated by three scenarios: 20th percentile, 50th percentile and 80th percentile. These scenarios are represented as lower bound, mean and upper bound respectively in the following figures.

The size of this test case is similar to the previous one, which considered 100+ hybrids, 100+ male varieties, 20+ growing regions and 10+ growing countries.

The penalty costs of both overproduction and shortage are approximated in this case study, and they should be replaced by practical commercial penalties in an actual implementation of the stochastic model in the decision support tool by the industrial collaborator.

Improvement in robustness

Since the land allocation plan is determined much earlier than the actual harvesting, all the scenarios are actually based on the same land allocation plan (i.e. the first stage variables) and the analysis is undertaken to illustrate the differences in performance within each scenario. Although the land allocation remains the same for all scenarios, the processing cost, transportation cost and the production volume are different due to the variations in uncertain yield, i.e. the total cost, total production and cost per bag will be dissimilar in various scenarios.

In order to compare the deterministic and stochastic solutions, various values of yield within each scenario are applied to simulate total cost, production units and cost per unit separately under each scenario.

The comparisons of total cost savings, total production savings and cost per unit savings compared with the original plan from Syngenta between the deterministic solution and stochastic solution are presented in Fig. 5, Fig.6 and Fig.7 respectively.

In the cases where “Mean Yield” or “Upper Bound Yield” are realised in practice, the cost savings generated by the deterministic solutions will be more significant. However, if the actual yield is closer to the “Lower Bound Yield”, the stochastic solution could potentially lead to 13.2% savings in cost while only 0.5% savings against the default plan can be achieved by the deterministic solution, demonstrating the benefits of implementing stochastic optimisation model under yield uncertainty.

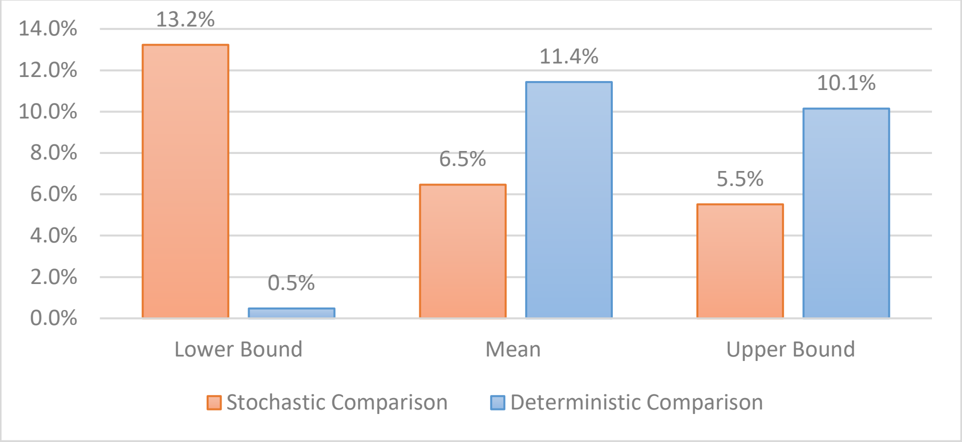

Fig. 5 – Total cost comparison between deterministic and stochastic solutions

Considering the purpose of production, it should be as close as possible to the demand in all possible scenarios. Although the stochastic solution indicates slight over-production in the case of “Mean Yield”, less shortage and overproduction are expected in the cases of “Lower Bound Yield” and “Upper Bound Yield” respectively.

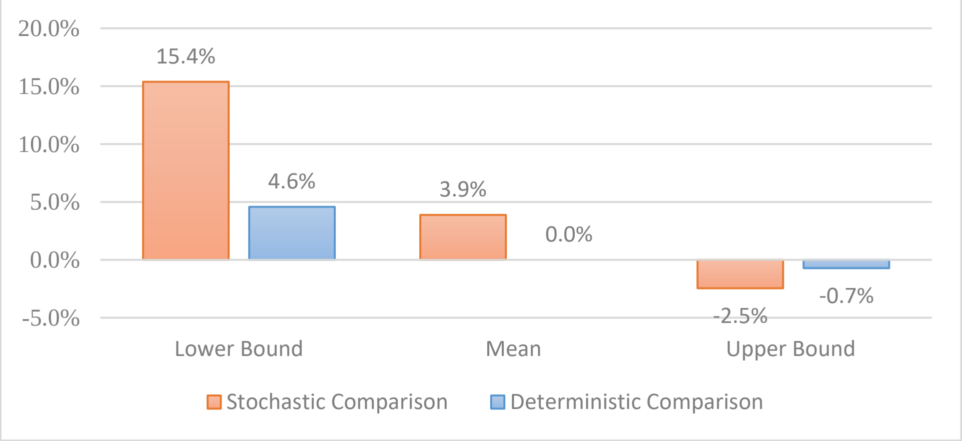

Fig. 6 – Total production comparison between deterministic and stochastic solutions

To integrate the influences from total cost and total production, the cost per bag is calculated. It presents a similar trend as shown for total cost while the differences in the lower bound scenario have increased to 24.8% in the stochastic solution compared with -4.3% in the deterministic solution.

From the company's perspective, the robustness of production is much more important than possible benefits from any single scenario as reliabilities and predictability of risk are always more critical for them as part of managing uncertainty in operations.

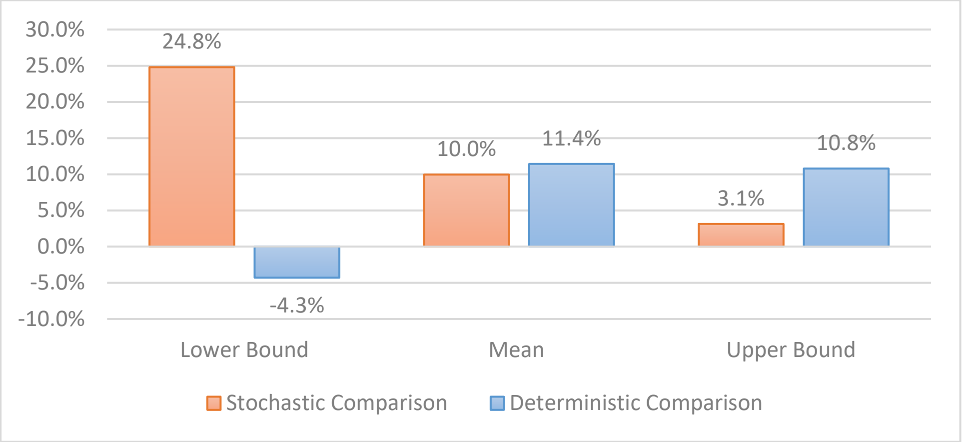

Fig. 7 – Cost per bag comparison between deterministic and stochastic solutions

In summary, the risk in production can be effectively balanced by the implementation of a stochastic model. Although the performance of a deterministic model can be better in certain possible future scenarios, the improved robustness of the solution presented by the stochastic model is more important for practical application due to unpredictability of the future yields.

Adjustable multiple objectives

In terms of the multiple types of sales targets, products with various degrees of importance are simultaneously optimized with differentiated objectives. The sales group concept defined in the previous case study is applied to meet the specified business requirements for different sales groups. Since the statistically simulated yield distribution mentioned previously can only predict actual yield distribution with a limited level of accuracy, 11 scenarios are applied to increase the accuracy of the distribution in the stochastic model without overfitting of input data (due to limited level of yield prediction accuracy). In other words, 11 scenarios are used to estimate the continuous yield probability distribution: 5th percentile, 15th percentile, 25th percentile, 35th percentile, 45th percentile, 50th percentile, 55th percentile, 65th percentile, 75th percentile, 85th percentile and 95th percentile.

For the "new" group, the aim can be summarized as minimizing the probability of shortage regardless of any extra cost or overproduction. Similarly, the objective of the "declining" group is to minimize the overall cost and probability of overproduction, where a certain level of shortage is acceptable. Regarding the “mature” group, the target can be identified as the minimal expected total cost with a limited level of shortage and overproduction. Based on the model introduced previously, the probability of shortage and overproduction can be manipulated through modifying the tuneable penalty cost for overproduction and shortage.

The different objectives from various sales groups can be met simultaneously through differentiating various penalty costs. The following figures illustrate an example of the production distributions of different groups within one optimization model.

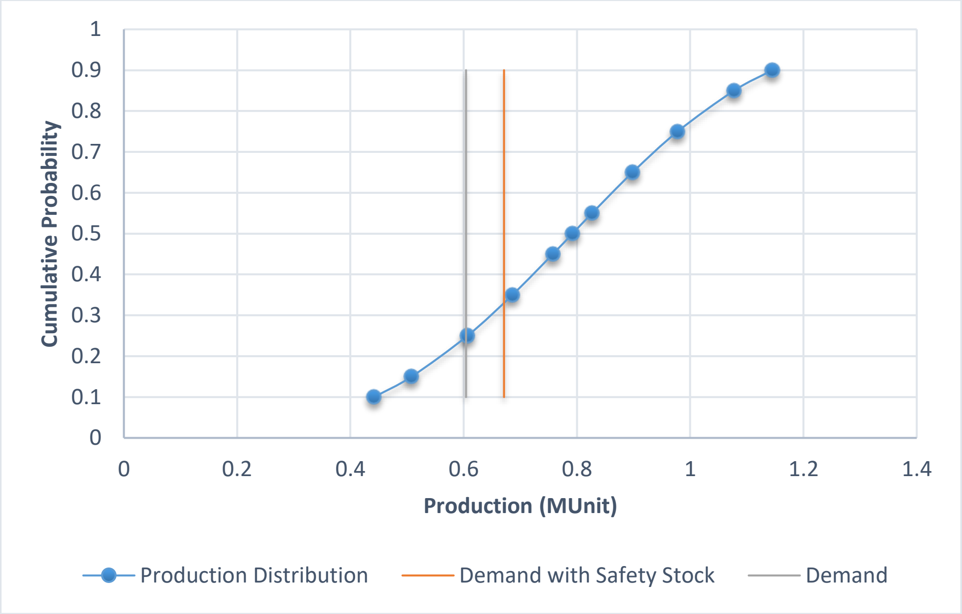

Fig. 8 - New group cumulative production probability distribution

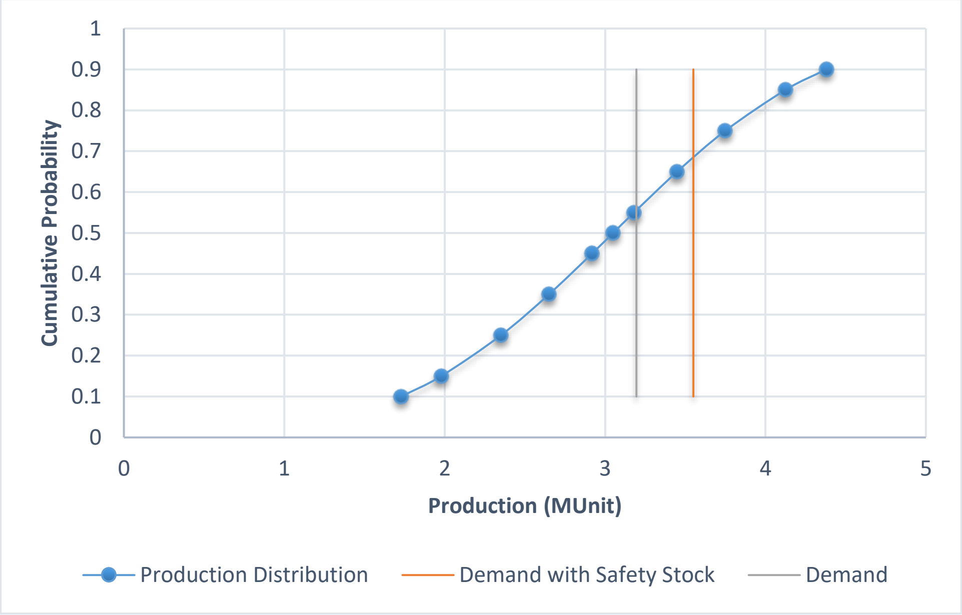

Fig. 9 - Mature group cumulative production probability distribution

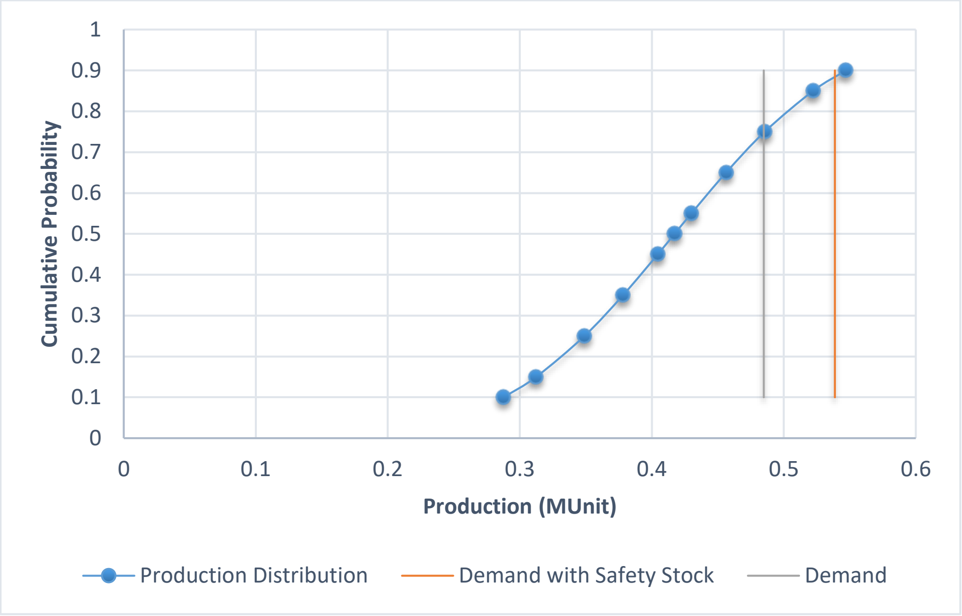

Fig. 10 - Declining group cumulative production probability distribution

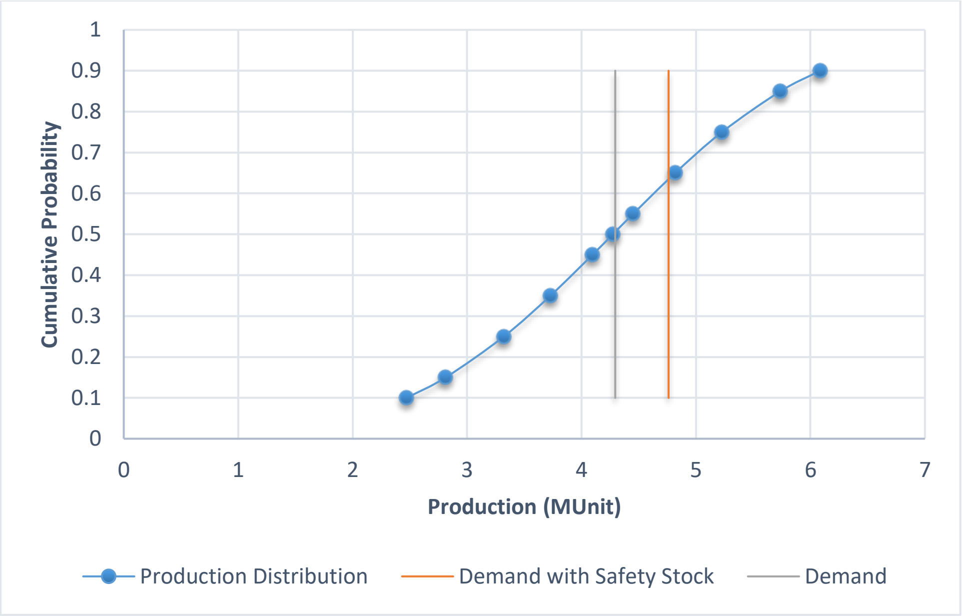

Fig. 11 - Total cumulative production probability distribution

The production volumes indicated in Fig.8 – Fig.11 are all based on simulated data without any indication of the actual production volumes in Syngenta AG.

Compared with the risk constraint in the deterministic model, the probabilistic distributions of production can directly quantify the value of improved stability through probability, i.e. a lower probability of shortage indicates a lower risk or higher robustness. When it comes to the trade – off problem between stability and cost, the model itself provides the flexibility to modify the balance between these two criteria while the final decision can be made by decision – makers in the company through incorporating the conditions from all the other aspects of the business at the time of actual planning.

Conclusions

In this paper, models for deterministic MILP and two – stage stochastic optimization are proposed for a real – world regional level seed production planning problem. The objective function is to minimize the overall cost including growing, processing and transportation cost. The demand balance, land availability, land allocation, biological constraints, distribution limitations and the commercial importance of different product groups have been taken into account as constraints. The deterministic model covers all the basic constraints together with the risk constraints to evaluate the effect of the risks in production and forms the basis for a toolkit used by the industrial partner. The stochastic model incorporated uncertainty in yield to optimize the effects from uncertain information.

To conclude the impacts from this research, this study addresses the research gap in the application of mathematical optimization in the seed industry with a focus on seed production planning at a regional level. It enables a transition from an experience-based approach in seed supply chain planning to a mathematical programming-based approach in industry. Quantitatively through industrial continent – wide case studies of the seed products in Syngenta, the model can save up to 16% in overall cost as well as 9% in total land usage, while up to 24.8% in cost per bag can be saved in the future by considering uncertainty. Moreover, the efficiency of planning can be boosted significantly by reducing the time used to generate a land allocation plan from one month to a few minutes.

Future research directions include global scale seed production allocation optimization at the field level with additional implementation of multi-objective optimization and links to harvest and production process scheduling.

Acknowledgement

The authors would like to thank the Centre for Process System Engineering and Syngenta AG for their support for this research.

Reference

Ahumada, O. & Villalobos, J.R. (2009) Application of planning models in the agri-food supply chain: A review. *European Journal of Operational Research*. [Online] 196 (1), 1–20. Available from: doi:10.1016/j.ejor.2008.02.014.

Annetts, J.E. & Audsley, E. (2002) Multiple objective linear programming for environmental farm planning. *The Journal of the Operational Research Society*. [Online] 53 (9), 933–943. Available from: doi:10.1057/palgrave.jors.2601404.

Comhaire, P. & Papier, F. (2015) Syngenta uses a cover optimizer to determine production volumes for its European seed supply chain. *Interfaces*. [Online] 45 (6), 501–513. Available from: doi:10.1287/inte.2015.0812.

Heady, E.O. (1954) Simplified Presentation and Logical Aspects of Linear Programming Technique. *Journal of Farm Economics*. [Online] 36 (5), 1035–1048. Available from: http://www.jstor.org/stable/1234313.

Howard, P.H. (2009) Visualizing consolidation in the global seed industry: 1996-2008. *Sustainability*. [Online] 1 (4), 1266–1287. Available from: doi:10.3390/su1041266.

Huh, W.T. & Lall, U. (2013) Optimal crop choice, irrigation allocation, and the impact of contract farming. *Production and Operations Management*. [Online] 22 (5), 1126–1143. Available from: doi:10.1111/poms.12007.

Jones, P.C., Kegler, G., Lowe, T.J. & Traub, R.D. (2003) Managing the seed-corn supply chain at Syngenta. *Interfaces*. [Online] 33 (October 2015), 80–90+104. Available from: doi:10.1287/inte.33.1.80.12718.

Jones, P.C., Lowe, T.J., Traub, R.D. & Kegler, G. (2001) Matching Supply and Demand: The Value of a Second Chance in Producing Hybrid Seed Corn. *Manufacturing & Service Operations Management*. [Online] 3 (2), 122–137. Available from: doi:10.1287/msom.3.2.122.9992.

Junqueira, R. a R. & Morabito, R. (2012) Production and logistics planning considering circulation taxes in a multi-plant seed corn company. *Computers and Electronics in Agriculture*. [Online] 84100–110. Available from: doi:10.1016/j.compag.2012.02.019.

Lowe, T.J. & Preckel, P. V. (2004) Decision Technologies for Agribusiness Problems: A Brief Review of Selected Literature and a Call for Research. *Manufacturing & Service Operations Management*. [Online] 6 (3), 201–208. Available from: doi:10.1287/msom.1040.0051.

Zhang, W. & Wilhelm, W.E. (2011) OR/MS decision support models for the specialty crops industry: a literature review. *Annals of Operations Research*. [Online] 190 (1), 131–148. Available from: doi:10.1007/s10479-009-0626-0.

Zuo, M., Kuo, W. & McRoberts, K.L. (1991) Application of Mathematical Programming to a Large-scale Agricultural Production and Distribution System. *Journal of the Operational Research Society*. [Online] 42 (8), 639–648. Available from: doi:10.1057/jors.1991.130.