A roadmap to solve two-stage stochastic problems implementing scenario reduction for agricultural production planning under uncertainty

3 authors:

Leonardo Talero
Autonomous University of Bucaramanga
64 PUBLICATIONS 71 CITATIONS

Juan David Marquez
Industrial University of Santander
3 PUBLICATIONS 1 CITATION

Henry Lamos
Autonomous University of Bucaramanga
40 PUBLICATIONS 66 CITATIONS

A ROADMAP TO SOLVE TWO-STAGE STOCHASTIC PROBLEMS IMPLEMENTING SCENARIO REDUCTION FOR AGRICULTURAL PRODUCTION PLANNING UNDER UNCERTAINTY

Talero-Sarmiento, Leonardo Hernan[1]*; Marquez, Juan[2]; Lamos, Henry[3]

Abstract

Decision-making in Agricultural Production Planning faces challenges due to the heterogeneous and uncertain sources, including but not limited to environmental conditions. (i) Temperature, rainfall, soil characteristics, and humidity. (ii) Non-constant factors in production such as hand-labor capacities, technology adoption barriers, finance availability, or low mechanization and automatization. (iii) Dynamic Sociopolitical context involving policies for or against rural development. Nevertheless, the farmer must decide over a defined decision horizon even if there is a lack of information. Consequently, this work's main objective is to provide a well-structured and detailed description of methods for solving optimization models under uncertainty scenarios; promptly, the Stochastic Optimization model and its two-stage strategy implementing forecasting scenarios. This work addresses a three-stage approach, starting with a section focusing on a (1) brief analysis of modeling trends for optimal decision-making during agricultural production planning under uncertainty, highlighting the (*) sources of uncertainty, and (*) comparing modeling strategies. The following section presents a detailed and illustrative (2) roadmap to develop a two-stage model from scratch, covering concepts of (*) large-scale optimization techniques such as decomposition and (*) processes to model the uncertainty of the parameters covering (*) scenario generation, including its probability estimation. (3) The final section shows an illustrative case to apply the Roadmap addressing a Two-stage stochastic problem. The case study uses the ARIMA modeling and backward reduction through the Kantorovich distance cluster strategy to represent the parameter trajectories, solving the deterministic equivalent formulation stochastic problem using the Benders decomposition technique. This work's critical contribution is that the Roadmap provides a valuable guide for practitioners, students, and researchers who need to model agricultural production planning under uncertain conditions providing a well-structured case study based on data reported (SP).

Keywords: Agricultural decision-making; Roadmap; Stochastic programming; Scenario reduction; Uncertainty; Two-Stage Optimization

1. Introduction

Mathematical optimization models to support decision-making in agricultural planning production represent a strategy widely used in the last decades (Bournaris et al., 2015; Osaki & Batalha, 2014; Pakawanich et al., 2020; Qian, 2021), considering several uncertain factors affecting the optimal resolution of the

[1] Universidad Autónoma de Bucaramanga, 0000-0002-4129-9163 [2] Universidad Industrial de Santander, 0000-0002-6969-4069 [3] Universidad Industrial de Santander, 0000-0003-1778-9768 *Correo de contacto: ltalero@unab.edu.co.

problem, leading to inappropriate decisions which impact the farmer's benefit. These aspects emerge from many uncertainty sources and connect to different modeling methodologies due to randomness or vagueness (Y. P. Li et al., 2009; Niu et al., 2016; Suo et al., 2011). Consequently, strategies such as deterministic optimization with fixed parameters represent an insufficient mechanism to support decision-making (M. Li & Guo, 2014; X. Li et al., 2017). Thus, this topic is under intense research using different approaches such as Stochastic Dynamic Programming (Lohano & King, 2009), Robust Optimization (Kazemzadeh et al., 2019; Randall et al., 2022), Markov Decision Processes (White, 2013), and Stochastic Programming (Birge & Louveaux, 2011; Sen, 2013; Shapiro et al., 2009) techniques. Two-stage Stochastic Programming supports proper agricultural decision-making in a two-stage framework (Zhang et al., 2017), and it is widely used to support various agricultural production activities and resources management (Kung et al., 2019; W. Li et al., 2016; Peña-Haro et al., 2011; Zhang et al., 2017). This work proposes a Roadmap for solving TSP models and their interaction with other stochastic programming methodologies using a fictional case study. The suggested Roadmap represents agricultural production planning issues under uncertain conditions.

2. Modeling trends for optimal decision-making under uncertainty

Nowadays, researchers use models with a system that considers uncertainty to make optimal and precise decisions. Most of these approaches essentially use three types of programming as the basis for modeling (or their respective combinations or hybrid models) mainly due to their quality in generating optimal combinations and resource allocations under conditions of uncertainty (Marquez et al., 2022; Wang et al., 2015): Stochastic Mathematical Programming (a.k.a., Stochastic Programming SP)(C. Li & Grossmann, 2021), Fuzzy Mathematical Programming (FMP)(Kumar, 2020), and Interval Mathematical Programming (IMP) (Ashayerinasab et al., 2018). **¡Error! La autoreferencia al marcador no es válida.** relates the main concerns of using every mathematical programming technique.

Table 1: Main uncertain mathematical programming strategies comparison

Comparison	FMP	IMP	SP
Advantages	It allows facing uncertainty through the expert's knowledge without excessive available data.	It allows the solution of problems under uncertainty considering insufficient or missing data through the definition of bounded intervals.	It allows fitting the probability distribution functions or random variables to set the parameter, reducing the uncertainty in the process.
Disadvantages	Based on the subjectivity derived from the expert knowledge. It can generate an inappropriate approach through membership functions.	It may present problems due to considering several parameters with uncertainty and establishing possible incorrect limits due to lack of information.	It requires vast amounts of available data to generate a suitable probability function that fits the parameters' uncertainty or models representing the historical behavior.

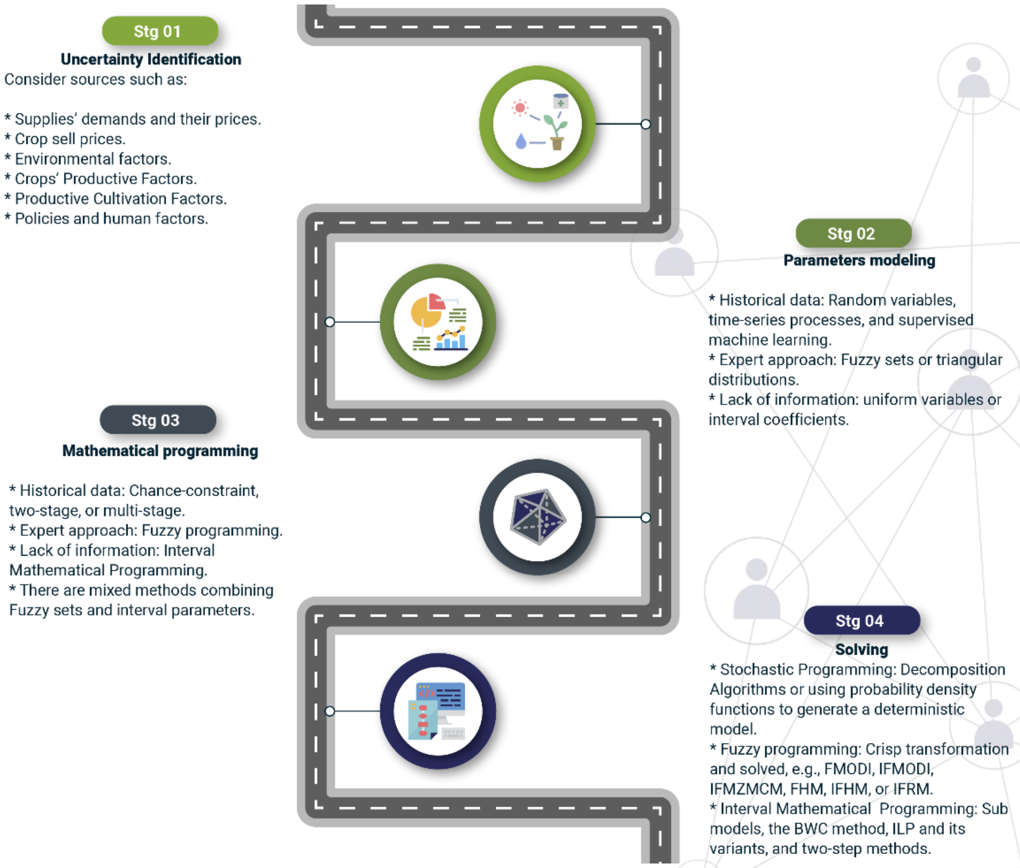
Source: Own elaboration

Therefore, these mathematical programming techniques relate to multiple ways to address data amount, quality, dependability, vagueness, and ambiguity (Y. P. Li et al., 2009; Suo et al., 2011). These types of uncertainty are present in multiple sources related to agricultural production and restrict the modeling of each uncertain factor. Specifically, there are three primary sources of uncertainty associated with agricultural production: climatic, socioeconomic, and productive sources of uncertainty. Climatic factors are primarily associated with elements that characterize climatic weather, such as rainfall, temperature, radiation, and humidity. Socioeconomic factors link the behavior of crop prices and input prices and the social environment of the region, such as population growth and feed demand. The productive factors are related to the productive capacities (e.g., available planting area), production schemes and planting patterns, and decision-makers' infrastructure.

3. Roadmap to solve two-stage stochastic optimization models

This study proposes a four-stage roadmap (Figure 1) for uncertain model building. On the other hand, this work exposes a theoretical case study approach based on Two-Stage Stochastic modeling.

Figure 1. Roadmap to address agricultural production planning under uncertainty



Source: Own elaboration

3.1 Uncertain sources identification

This stage identifies all sources of uncertainty. In this sense, the DM must examine the parameters associated with formulating the mathematical problem and determine which are subject to uncertain events. By identifying the sources of uncertainty (i.e., climatic, socioeconomic, and productive factors), the DM can establish the proper method to represent the parameter and the mathematical modeling approach suitable to address the problem.

3.2 Parameters modeling

If there is enough data and the kind of uncertainty is of random origin, modeling through stochastic strategies is the most appropriate approach. If data exists but expert participation is required, the fuzzy theory is the best modeling option. In conditions where data is scarce, defining intervals represents a conservative modeling strategy relevant to facing the problem. In each type of modeling, there are different approach strategies:

- Stochastic modeling: 1. Develop data trajectories to represent the parameter behavior (i.e., simulation of scenarios). 2. Using moment matching methods to fit a probability distribution function to historical data (Xu et al., 2012). 3. Internal sampling using the original distribution function (Høyland & Wallace, 2001). 4. Scenario reduction (Dupačová & Kozmík, 2017).

- Fuzzy modeling: Using fuzzy membership functions to represent the parameter.

- Modeling using intervals: 1. Establish crisp values associated with the maximum and minimum parameter bounds. 2. Establish variable functions to each bound relating a parameter variation regarding the period considered.

3.3 Mathematical programming

This stage addresses the model formulation considering the type of modeling strategies used to represent the uncertain parameter. Stochastic parameter modeling strategies are related to Chance-constraint, Two-stage, and multistage techniques. Methods based on a fuzzy theory related the Fuzzy Programming. On the other hand, the Interval Programming approach refers to case studies with a lack of information. Additionally, problems may have multiple types of uncertainty, which produce mixed strategies (e.g., Interval Two-stage Stochastic or Fuzzy Interval Programming) as the best problem modeling strategy.

3.4 Mathematical problem solving

Various associated solution strategies regarding the type of mathematical programming used for modeling exist. Due to the computational cost considering the number of scenarios to represent the parameter, Stochastic programming is solved using decomposition strategies (e.g., Dantzig-Wolfe, Column Generation, Benders, among others) for large-scale problems or deterministic approaches using probability distribution functions. Fuzzy programming has various techniques to transform the fuzzy model into a deterministic approximation, such as Intuitionistic Fuzzy Modified Distribution Method (FMODI), Intuitionistic Fuzzy Min-Zero Min-Cost Method (IFMODI), Intuitionistic Fuzzy Min-Zero Min-Cost Method (IFMZMCM), Intuitionistic Fuzzy Hungarian Method (IFHM), and Intuitionistic Fuzzy Reduction Method (IFRM). In the case of using intervals, strategies such as Best-Worst Case, model transformation into deterministic version using the bounds of the parameters, and ILP variants represent proper solution alternatives.

4. Hypothetical Case-Study: A two-stage model for Agricultural Production Planning implementing back-ward scenario reduction and Benders Decomposition

Consider that optimizing Agricultural Production Planning (APP) over weekly time horizons associated with crop sales prices is a stochastic decision problem. Since the DM must decide the optimal crop scheduling despite the lack of information about crops' sell prices. The DM uses historical data to forecast each crop's expected price using ARIMA models. This work discretizes the time-stochastic process into weekly intervals; for this purpose, using the crop sale-price data from the Colombian Bureau of Statistics Departamento Administrativo Nacional de Estadistica – DANE Agronet (2020). In the city of Bogota, between January 2015 and December 2020. The simulation covers sixteen weeks ahead of four time-series Lettuce (*Lactuca sativa*) ARIMA(9,0,6), Pumpkin (*Cucurbita moschata*) ARIMA(8,1,8), Coriander (*Coriandrum sativum*) ARIMA(8,0,8), and Pea (*Pisum sativum*) ARIMA(8,0,5).

Scheduling decisions for APP focus on determining the area's portion to cultivate the four crops during two seasons, assuming the agricultural harvesting and sale at the end of each period (each one is equal to eight weeks), subject to grown maturity periods constraints and crop demands. Therefore, the arable areas dedicated to crops two and four are the first-stage variables. Thus, their stochastic prices, $q(s)$, in the objective function could be substituted for their expected values. Besides, the DM must satisfy demands for the four crops in both seasons. Due to the cultivation period and harvest time of crops two and four, it would not be possible to meet the demand for them in season 1. For simplicity, consider this decision-making process as a two-stage problem. For that, it is necessary to assume that the expected sell prices at the end of the first stage are known and equal to the whole simulation's average price at week eight. From there, the simulated scenario approximates the stochastic load process with a scenario tree, generating a well-defined and deterministic optimization model with a particular sparsity structure due to the variables related to the scenario's modeling strategy. The following equations show the deterministic equivalent formulation.

$$\max z = p_{i,s}x_{i,s} + \varepsilon\{Q(s)\} \tag{1}$$

Subject to:

$$\sum_{s=1}^{S} x_{i,s} \leq 1, \forall i, Arable\ area\ in\ season\ 1\ for\ each\ scenario\ s \tag{2}$$

$$x_{i,s} \geq 0 \tag{3}$$

Where:

$$\varepsilon\{Q(s)\}\ is\ equivalent\ to\ its\ expected\ value: \sum_{s=1}^{S} \pi_s\, q_{i,s} y_{i,s} \tag{4}$$

Subject to:

$$\sum_{s=1}^{S} x_{i,s} + y_{i',s} \leq 1, \forall i \neq \{1,3\}, \forall\ i' \neq \{2,4\}, Arable\ area\ in\ season\ 2 \tag{5}$$

$$\sum_{s=1}^{S} yield_i\big(x_{i,s} + y_{i,s}\big) \geq 2, \forall i \neq \{2,4\}, Lettuce\ and\ Coriander\ joint\ demand \tag{6}$$

$$\sum_{s=1}^{S} yield_i * y_{i,s} \geq 1, \forall i \neq \{1,3\}, Pumpkin\ and\ Pea\ demand \tag{7}$$

$$x_{i,s} = y_{i,s} \forall i \neq \{1,3\}, \forall s, Area\ in\ both\ seasons\ are\ the\ same\ for\ long - term\ crops \tag{8}$$

$$x_{i,s} = x_{i,s'} \forall i, \forall s, s', nonanticipativity\ conditions \tag{9}$$

$$y_{i,s} \geq 0\ , yield = \{2.14, 2.153, 2.11, 0.948\} \tag{10}$$

Wherein variables x_i and $y_{i,s}$ are the area's portion to cultivate the i crop in seasons one and two, respectively, for each scenario s. The Objective Function maximizes the income of the crops over the planning period with sell prices $p_{i,s}$ in season one and $q_{i,s}$ in season two. π_s is the probability of each scenario s. The two first constraints limit the total arable area in each season. Remarkably, crops two and four can be cultivated only in the first season and harvested in season two. The following two constraints satisfy the crop demands assuming crop yield is constant in both periods. Finally, the last constraint forces the nonanticipativity conditions. Notice that, even if the model only considers four crops and two stages, it contains 1,600 continuous decision variables since $Q_v = 2 * \mathrm{i} * S, i = 4$ 3,200,080,000 inequalities constraints $(C_i = 2 * S^{\mathrm{i}} + 2 * S^{\mathrm{i}'}, i = 4, i^{'} = 2)$, and 1,196 equalities constraints (C_e), with $C_e = 2 * S + (S - 1) * (i), i = 4$, therefore, the sparsity constrain matrix has 3,200,081,196 rows and 1,600 columns. Considering that solving this large-scale problem is cumbersome and exhausting, it is necessary to simplify it. Notice that the number of constraints depends strongly on the number of scenarios; hence, selecting the most representative paths in the tree scenario could decrease the optimization problem's dimension with minimum lost accuracy in uncertainty parameters modeling.

Bearing that the final decision tree needs to keep the original phenomena' stochastic information, it is necessary to generate a subset of the most representative paths and reassign its probabilities. Due to the scenario size (200 paths), this work applied a backward reduction algorithm getting 20 finals scenarios in 0.2 seconds with a relative error equal to $\varepsilon_{Rel} = 0.44$, using a desktop computer with 3.40 GHz and 12 GB memory in MATLAB R2019b 64-bit language. The final model contains 160 decision variables, 320,800 inequalities constraints, and 116 equalities constraints, reducing the problem size considerably. This work uses the linprog function on MATLAB R2019b 64-bit in a computer with the same characteristics described early to solve the optimization model using the simplex-dual algorithm and after developing a Benders Decomposition (BD) script (Rahmaniani et al., 2017; Taşkin, 2011). The first strategy obtains the optimal solution with 8.678060 seconds elapsed, while BD takes 14.369723 seconds.

5. Conclusions

This work explains the theoretical background to develop an SP using the two-stage optimization model. Its structure allows researchers, students, instructors, or practitioners to identify optimal decision-making elements regarding agricultural production planning under an uncertain context, highlighting uncertain sources and leading methods for mathematical modeling under uncertainty reported in the literature (i.e., Fuzzy Programming, Interval Mathematical Programming, and Stochastic Mathematical Programming). This work explains didactically the strategy based on data reported (SP). This work summarizes these approaches into a Roadmap, exhibiting a briefcase implementing the two-stage optimization and its deterministic equivalent formulation, which includes scenario simulation and reduction to represent scenarios and solve the large-scale problem by implementing a decomposition algorithm.

6. Acknowledgments

MinCiencias' Becas Bicentenario Grant fully funds this research.

7. Conflict of interest

The authors declare that they have no conflict of interest.

8. References

Ashayerinasab, H. A., Nehi, H. M., & Allahdadi, M. (2018). Solving the interval linear programming problem: A new algorithm for a general case. *Expert Systems with Applications*, *93*, 39–49. https://doi.org/10.1016/j.eswa.2017.10.020

Birge, J. R., & Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer New York. https://doi.org/10.1007/978-1-4614-0237-4

Bournaris, Th., Papathanasiou, J., Manos, B., Kazakis, N., & Voudouris, K. (2015). Support of irrigation water use and eco-friendly decision process in agricultural production planning. *Operational Research*, *15*(2), 289–306. https://doi.org/10.1007/s12351-015-0178-9

Dupačová, J., & Kozmík, V. (2017). SDDP for multistage stochastic programs: preprocessing via scenario reduction. *Computational Management Science*, *14*(1), 67–80. https://doi.org/10.1007/s10287-016-0261-6

Høyland, K., & Wallace, S. W. (2001). Generating Scenario Trees for Multistage Decision Problems. *Management Science*, *47*(2), 295–307. https://doi.org/10.1287/mnsc.47.2.295.9834

Kazemzadeh, N., Ryan, S. M., & Hamzeei, M. (2019). Robust optimization vs. stochastic programming incorporating risk measures for unit commitment with uncertain variable renewable generation. *Energy Systems*, *10*(3), 517–541. https://doi.org/10.1007/s12667-017-0265-5

Kumar, P. S. (2020). Algorithms for solving the optimization problems using fuzzy and intuitionistic fuzzy set. *International Journal of System Assurance Engineering and Management*, *11*(1), 189–222. https://doi.org/10.1007/s13198-019-00941-3

Kung, C. C., Cao, X., Choi, Y., & Kung, S. S. (2019). A stochastic analysis of cropland utilization and resource allocation under climate change. *Technological Forecasting and Social Change*, *148*(December 2017), 119711. https://doi.org/10.1016/j.techfore.2019.119711

Li, C., & Grossmann, I. E. (2021). A Review of Stochastic Programming Methods for Optimization of Process Systems Under Uncertainty. *Frontiers in Chemical Engineering*, *2*. https://doi.org/10.3389/fceng.2020.622241

Li, M., & Guo, P. (2014). A multi-objective optimal allocation model for irrigation water resources under multiple uncertainties. *Applied Mathematical Modelling*, *38*(19), 4897–4911. https://doi.org/https://doi.org/10.1016/j.apm.2014.03.043

Li, W., Feng, C., Dai, C., Li, Y., Li, C., & Liu, M. (2016). An inexact risk management model for agricultural land-use planning under water shortage. *Frontiers of Earth Science*, *10*(3), 419–431. https://doi.org/10.1007/s11707-015-0544-1

Li, X., Huo, Z., & Xu, B. (2017). Optimal allocation method of irrigation water from river and lake by considering the fieldwater cycle process. *Water (Switzerland)*, *9*(12). https://doi.org/10.3390/w9120911

Li, Y. P., Huang, G. H., Huang, Y. F., & Zhou, H. D. (2009). A multistage fuzzy-stochastic programming model for supporting sustainable water-resources allocation and management. *Environmental Modelling and Software*, *24*(7), 786–797. https://doi.org/10.1016/j.envsoft.2008.11.008

Lohano, H. D., & King, R. P. (2009). A Stochastic Dynamic Programming Analysis of Farmland Investment and Financial Management. *Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie*, *57*(4), 575–600. https://doi.org/10.1111/j.1744-7976.2009.01171.x

Marquez, J., Talero-Sarmiento, L. H., & Lamos, H. (2022). Multistage Stochastic Programming to Support Water Allocation Decision-Making Process in Agriculture: A Literature Review. *IOCAG 2022*, 26. https://doi.org/10.3390/IOCAG2022-12307

Ministerio de Agricultura. (2020). *Agronet*. Agricultural Historical Sale Prices. https://www.agronet.gov.co/estadistica/Paginas/home.aspx?cod=11

Niu, G., Li, Y. P., Huang, G. H., Liu, J., & Fan, Y. R. (2016). Crop planning and water resource allocation for sustainable development of an irrigation region in China under multiple

uncertainties. *Agricultural Water Management*, *166*, 53–69. https://doi.org/10.1016/j.agwat.2015.12.011
Osaki, M., & Batalha, M. O. (2014). Optimization model of agricultural production system in grain farms under risk, in Sorriso, Brazil. *Agricultural Systems*, *127*, 178–188. https://doi.org/https://doi.org/10.1016/j.agsy.2014.02.002
Pakawanich, P., Udomsakdigool, A., & Khompatraporn, C. (2020). Robust production allocation model for an agricultural cooperative with yield uncertainty and similar revenue constraints. *Computers and Electronics in Agriculture*, *168*, 105090. https://doi.org/https://doi.org/10.1016/j.compag.2019.105090
Peña-Haro, S., Pulido-Velazquez, M., & Llopis-Albert, C. (2011). Stochastic hydro-economic modeling for optimal management of agricultural groundwater nitrate pollution under hydraulic conductivity uncertainty. *Environmental Modelling and Software*, *26*(8), 999–1008. https://doi.org/10.1016/j.envsoft.2011.02.010
Qian, X. (2021). Production planning and equity investment decisions in agriculture with closed membership cooperatives. *European Journal of Operational Research*, *294*(2), 684–699. https://doi.org/https://doi.org/10.1016/j.ejor.2021.02.007
Rahmaniani, R., Crainic, T. G., Gendreau, M., & Rei, W. (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, *259*(3), 801–817. https://doi.org/10.1016/j.ejor.2016.12.005
Randall, M., Montgomery, J., & Lewis, A. (2022). Robust temporal optimisation for a crop planning problem under climate change uncertainty. *Operations Research Perspectives*, *9*, 100219. https://doi.org/10.1016/j.orp.2021.100219
Sen, S. (2013). Stochastic Programming. In *Encyclopedia of Operations Research and Management Science* (pp. 1486–1497). Springer US. https://doi.org/10.1007/978-1-4419-1153-7_1005
Shapiro, A., Dentcheva, D., & Ruszczyński, A. (2009). *Lectures on Stochastic Programming* (1st ed.). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718751
Suo, M. Q., Li, Y. P., & Huang, G. H. (2011). An inventory-theory-based interval-parameter two-stage stochastic programming model for water resources management. *Engineering Optimization*, *43*(9), 999–1018. https://doi.org/10.1080/0305215X.2010.528412
Taşkin, Z. C. (2011). Benders Decomposition. In *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470400531.eorms0104
Wang, R., Li, Y., & Tan, Q. (2015). A review of inexact optimization modeling and its application to integrated water resources management. *Frontiers of Earth Science*, *9*(1), 51–64. https://doi.org/10.1007/s11707-014-0449-4
White, C. C. (2013). Markov Decision Processes. In *Encyclopedia of Operations Research and Management Science* (pp. 934–937). Springer US. https://doi.org/10.1007/978-1-4419-1153-7_580
Xu, D., Chen, Z., & Yang, L. (2012). Scenario tree generation approaches using K-means and LP moment matching methods. *Journal of Computational and Applied Mathematics*, *236*(17), 4561–4579. https://doi.org/10.1016/j.cam.2012.05.020
Zhang, C., Li, M., & Guo, P. (2017). Two-Stage Stochastic Chance-Constrained Fractional Programming Model for Optimal Agricultural Cultivation Scale in an Arid Area. *Journal of Irrigation and Drainage Engineering*, *143*(9), 05017006. https://doi.org/10.1061/(asce)ir.1943-4774.0001216