Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

Joint Optimization of Feedstock Procurement and Production Planning in AD: A Deep Learning-Integrated Stochastic Programming Framework

Ruosi Zhang[a], Michael Short[s]*

[a] School of Chemistry and Chemical Engineering, University of Surrey, Stag Hill, Guildford GU2 7XH, UK
* Corresponding Author: m.short@surrey.ac.uk.

ABSTRACT

Anaerobic digestion (AD) across Europe and the UK faces increasing economic and operational pressure from volatile feedstock supply under climate extremes. Existing stochastic programming (SP) approaches for feedstock planning often rely on limited historical observations and/or simplify yield uncertainty in ways that miss the joint, non-linear response of crops to weather variability, thereby understating downside supply risk. We develop an integrated decision-support framework that links climate uncertainty to AD procurement planning by coupling mechanistic crop simulation, generative surrogate modelling, and stochastic optimization. First, APSIM is used offline to generate a mechanistic yield knowledge base across weather trajectories and discrete planting-density choices. Then, a conditional GAN (CGAN) is trained to produce non-parametric joint yield samples for multi crops conditioned on scenario features and management, enabling fast Monte Carlo evaluation. At last, these samples are embedded in a two-stage SP that optimizes first-stage land allocation and planting densities, with second-stage recourse represented by spot-market purchases to cover shortfall. The architecture is designed for stage-based rolling updates as forecasts are progressively replaced by observations. We demonstrate the framework on a 15-ha unit under different contracts of biogas-equivalent output. Results reveal a target-induced regime shift in optimal procurement. Under moderate production targets, monocropping solutions minimize cost with negligible loss in reliability, reflecting broad operational indifference across several land allocation patterns. As contract levels approach the biophysical limits of a 15-ha system (~110×10^3 m^3), the optimizer transitions into a risk-reducing regime where maize–rye double cropping becomes increasingly necessary. At high targets, the feasible set collapses toward near-complete double cropping. At a representative contract of 140×10^3 m^3, the least-cost rye-only plan achieves 34.3% supply confidence, while full double cropping increases confidence to 85.7% but with higher cultivation cost. Even under full intensification, an irreducible tail risk remains (14.3% shortfall frequency), implying unavoidable reliance on spot-market procurement under extreme seasons. The APSIM–CGAN–SP integration translates climate-driven biophysical uncertainty into actionable procurement strategies. It reveals threshold behavior, tail-risk exposure, and the limits of intensification under fixed land constraints. The framework supports both pre-seasons planning and rolling in-season updates, providing a quantitative basis for contract feasibility assessment, hedging design, and resilient feedstock procurement.

Keywords: Biomass, Stochastic Optimization, Energy Systems, Surrogate Model, Planning, CGAN, Anaerobic Digestion

INTRODUCTION

Anaerobic digestion (AD) is a mature bioenergy technology that converts organic matter into biogas [1]. Unlike intermittent renewable sources such as solar or wind power, AD facilities often operate as baseload units, which means they rely on a steady and predictable feedstock supply. In practice, even short-term feedstock

shortages can trigger substantial economic losses or force unplanned shutdowns [2]. This dependence turns feedstock availability into a planning problem, operators must decide how much to grow, how much to contract, and how much risk to leave to the spot market.

Energy crops such as maize and rye are widely used as high-energy feedstock in AD systems [3]. Beyond providing high-energy biomass, these crops support greenhouse gas mitigation, energy security, and rural development [3]. However, climate-driven yield variability and extreme events increasingly translate agricultural uncertainty into operational risk for AD sites [4], [5]. Yield forecasting serves not as a standalone objective but as a foundational input for the generation of supply chain scenarios. These scenarios are subsequently embedded in procurement and production-planning frameworks to enable hedging strategies that account for operational variability and upstream supply uncertainty [6].

Process-based crop models (e.g., Agricultural Production Systems sIMulator (APSIM) [7]) can provide realistic yield responses but are computationally too expensive for real-time stochastic optimization. Purely data-driven machine learning (ML) models are faster, yet often lack physical consistency and can be unreliable under extremes such as heatwaves or floods [8]. This limits their usefulness for decision support, especially when yield uncertainty and needs to be propagated into procurement and production planning. This creates a need for surrogate-based modeling strategies, balancing physical rigor with mathematical tractability in process systems engineering [9].

Existing AD supply-chain studies have largely focused on logistics or facility siting under deterministic assumptions, overlooking the stochastic nature of biophysical production [10]. While integrated crop–operations frameworks have appeared in the ethanol sector [11], such crop-to-decision integration remains limited in AD. Mechanistic models like APSIM remain accurate but infeasible for real-time optimization, and no operational framework currently connects biophysical yield uncertainty to tactical procurement decisions [12].

In mature electricity markets, stochastic programming (SP) is routinely used to hedge fuel contract risks under price volatility [13], [14]. AD procurement faces a similar make-or-buy challenge, but the dominant uncertainty arises not from market prices but from biophysical variation in crop yields. Yield variability translates into supply shortfalls and, consequently, greater pressure from volatile spot-market purchases. SP is therefore a natural framework for trading off upfront commitments (planting density and contracts) against scenario-dependent recourse actions and supply-gap penalties. If future yields can be predicted with adequate physical realism, managers can jointly optimize planting densities and purchase contracts to reduce exposure to costly spot-market feedstock.

To address the trade-off between biophysical realism in crop modeling and the computational demands of large-scale system decision-making, this paper proposes an integrated framework to support feedstock procurement decisions for ADs. This method utilizes the APSIM simulator to generate a high-fidelity dataset that captures the nonlinear response of crop yields to extreme weather events (such as heat waves). Based on this data, we train a physically-aware Conditional Generative Adversarial Network (CGAN) as a surrogate model, that simulates the complex mechanistic relationships between crops and conditions such as soil and weather. Finally, we embed this surrogate model generating probabilistic crop yield scenarios into a two-stage SP framework to evaluate land use and cultivation decisions under climate-driven uncertainty. By incorporating ensemble weather forecasts into the scenario generation process, this framework enables AD operators to optimize land allocation and planting density based on production objectives. The resulting decision-making tools reduce the risks associated with adverse yield outcomes and recourse costs related to volatile spot market feedstock procurement.

METHODOLOGY

The proposed framework (Figure 1) optimizes AD feedstock procurement by coupling biophysical process simulation, deep learning surrogates, and two-stage stochastic programming SP.

Problem Definition

A standard feedstock procurement agreement (FPA) for an AD site has been formalized, which specifies a contracted cultivation capacity A_{tot} (ha) and an annual baseload production target expressed as a feedstock requirement Q_{target} (m^3 biogas or t DM-equivalent). Internal supply is secured via fixed cultivation fees (covering seeds, management and land use). To ensure baseload operation, a minimum supply guarantee is effectively enforced by shortfall recourse.

The decision process is a two-stage stochastic problem. The first stage is here-and-now decision. Prior to the growing season, the manager determines the area allocation $\boldsymbol{A} = \{A_M, A_R, A_{DR}\}$ and planting densities $\boldsymbol{D} = \{D_m, D_r\}$ for maize (M), rye (R), and double-cropping (DR). It denotes sequential maize–rye double cropping within the same year, where rye and maize occupy less overlapping growing seasons on the same land base. DR is approximated in this prototype by combining monocrop yields with an annual reduction factor $\eta_{DR} \in (0, 1]$ that reflects a possible reduction in crop yield due to a shortened growing season or insufficient nutrients.

And the second stage is wait-and-see decision.

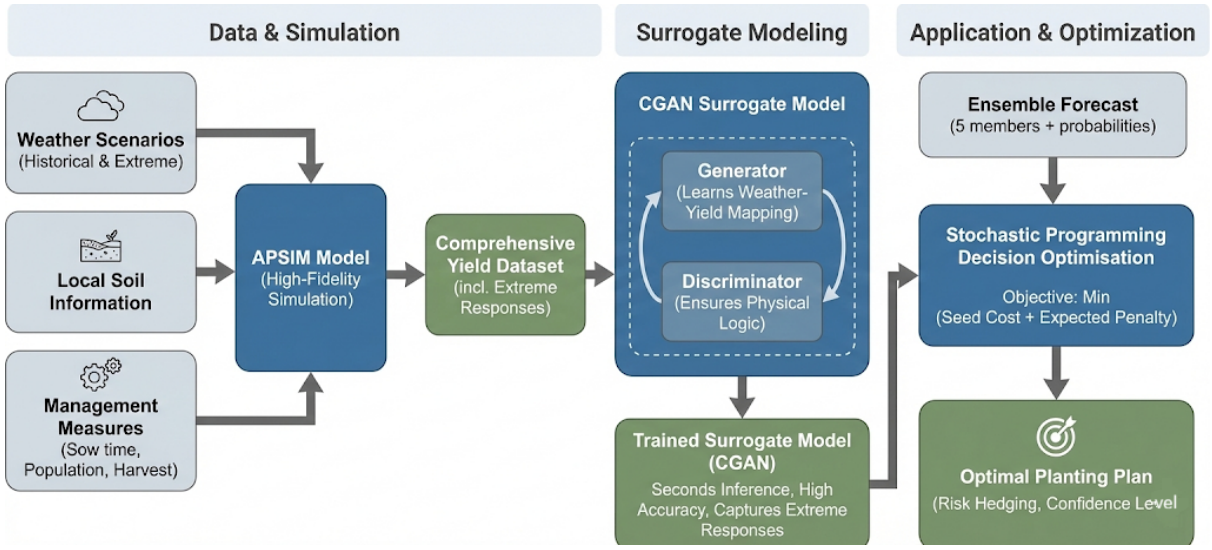

Figure 1. Integrated decision-support framework for AD feedstock procurement optimization under climate uncertainty, coupling biophysical simulation (APSIM), surrogate modelling (CGAN), and two-stage SP.

After weather scenarios s are realized, the yield Q_s is observed. If realized yields are insufficient to meet Q_{target} due to adverse weather, shortfalls are mitigated via spot-market purchases at a premium price P_{buy}, while surpluses are stored or sold at a credit price P_{sell}.

To address data scarcity under extreme climates, the APSIM is used to generate a mechanistic knowledge base. We incorporate a pre-simulation spin-up using historical weather data to initialize the earlier soil environment (e.g., plant-available water capacity). This captures the biophysical memory, where residual moisture conditions from previous years interact with current climate anomalies to produce non-linear yield responses

DR yields are represented by $\hat{Y}_{DR,s} = \eta_{DR}(\hat{Y}_{M,s} + \hat{Y}_{R,s})$, where $\eta_{DR} \in (0,1]$ is an operational efficiency factor capturing inter-crop competition and harvesting windows.

Surrogate Modelling via CGAN

To bypass the computational cost of APSIM, a CGAN is employed as a surrogate. Unlike standard regressions, the CGAN maps the conditional inputs (weather W_s and density D) to a probabilistic yield manifold. The Generator G takes a latent noise vector $z^{(j)} \sim \mathcal{N}(0, I)$ as an additional input. By drawing N independent samples of z for each weather scenario s, the model generates a yield cloud:

$$\left[\hat{Y}_{M,s}^{(j)}, \hat{Y}_{R,s}^{(j)}\right] = G(W_s, D, z^{(j)}), j = 1, \dots, N \quad (1)$$

where $W_s \in R^{12}$ represents standardized weather anomalies (temperature and precipitation deviations) across key growth stages. Here $s$indexes the five climate regimes and $j$indexes CGAN samples within each regime. Only the weather space is reduced by clustering historical meteorology into five climate regimes (empirical probabilities $\{\pi_s\}$). No post-CGAN reduction is applied. For regime s, the CGAN generates N_s joint samples $\left[\hat{Y}_{M,s}^{(j)}, \hat{Y}_{R,s}^{(j)}\right]$, all retained; hence $S = \sum_s N_s$. Scenario weights are $p_{s,j} = \pi_s / N_s$. The CGAN internalises physiological overturn points where high density exacerbates water stress, ensuring the surrogate is physics aware.

This N-fold sampling is critical as it captures the inherent stochasticity of the biological system (e.g., non-linear interactions between soil moisture and micro-climates) that weather anomalies alone cannot fully explain. It transforms the surrogate from a point-estimator into a stochastic generator, providing the density information required for risk-aware optimization.

The optimization employs a sample average approximation approach to reconcile weather uncertainty. Climate uncertainty is represented by $s \in S$ representative ensemble members (e.g., $S = 5$), capturing large-scale climatic shifts. These encompass worst-case (dry/hot), best-case (wet/cool), and median trajectories. We apply a scenario reduction logic where the probability P_s of discarded ensemble members is redistributed to the closest representative member.

To bridge the gap between high-fidelity mechanistic simulation and stochastic optimization, we apply an impact-based scenario discretization approach. The initial APSIM simulations, which represent the full population of potential biophysical outcomes, are partitioned into five representative climate regimes based on their total annual biomass.

Specifically, we utilize a quantile-based binning method to ensure that the entire variance of climate-driven risks, from extreme failures to optimal harvests, is captured. Each bin represents 20% of the probability

space ($P_s = 0.20$), effectively discretizing the continuous yield distribution into five distinct states: Worst-case, Lower-Mid, Median, Upper-Mid, and Best-case.

The mean feature vector (weather anomalies) and mean biomass of each quintile serve as the representative scenarios for the CGAN surrogate and the subsequent two-stage stochastic model. This discretization ensures that the hedging decisions are grounded in the actual statistical density of the 30-year climatological baseline, rather than arbitrary weather samples.

Two-stage stochastic programming formulation

The objective is to minimize the expected total cost (ETC), reconciling deterministic investments with probabilistic recourse impacts:

$$minC_{cultivation}(x) + \sum_s \sum_{j=1}^{N_s} p_{s,j}\ C_{\text{recourse}}(x, s, j), \text{with} \sum_s \sum_j p_{s,j} = 1 \quad (2)$$

Subject to:

$$A_M + A_R + A_{DR} \le A_{tot}, A_k \ge 0 \quad (3)$$

The recourse cost $\mathcal{L}_s^{(j)}$ captures the make-or-buy logic:

$$\mathcal{L}_s^{(j)} = P_{buy} \cdot max\left(0, Q_{target} - \hat{Q}_s^{(j)}\right) - P_{sell} \bullet max\left(0, \hat{Q}_s^{(j)} - Q_{target}\right) \quad (4)$$

where $\hat{Q}_s^{(j)}$ is the total simulated yield for scenario s and Monte Carlo draw j, $p_{s,j} = \pi_s/N$. The first term represents the shortfall penalty, and the second term is the credit from surplus sales or recovery. All results and comparisons refer to this total expected cost. Hence, plans with very low cultivation expenditures may still be suboptimal if they induce higher expected shortage or imbalance penalties in the recourse stage.

$$\hat{Q}_s^{(j)} = \sum_{c\in\{m,r\}} A_c \hat{Y}_{c,s}^{(j)} \eta_c\,, \quad (5)$$

where $Y_{c,s}$denotes the realised dry-matter yield (t/ha) under scenario s, and $\eta_c$is the crop-specific conversion to methane (e.g., $\eta_c = VS_c \times BMP_c, Nm^3\,CH_4/t$). This maps biomass outputs to the contractual unit consistently across monocrop and double-crop cases.

To ensure resilience against tail-risks, we evaluate the Cost Value-at-Risk ($VaR_{0.95}$). Let $TC_s^{(j)}$ be the realised total cost sample:

$$TC_s^{(j)} = Cost_{seed} + \mathcal{L}_s^{(j)} \quad (6)$$

where $VaR_{0.95}(TC)$ is defined as the 95th percentile of the scenario-weighted cost distribution. This metric allows the operator to quantify the financial exposure during synchronized biophysical failures, beyond simple expected values.

The framework supports rolling updates at crop growth-stage decision epochs τ (e.g., sowing, early growth, flowering, pre-harvest). At each τ, forecast components in $W_{s,\tau}$ are updated by replacing them with realised observations and refreshed ensemble information, yielding updated scenario inputs and (if required) updated spot price $P_{buy,\tau}$ from latest market quotes. The two-stage problem is resolved with updated inputs. After sowing, A and D remain fixed, while the rolling update refines the procurement hedge under increasingly accurate yield outlooks.

RESULTS AND DISCUSSION

Biophysical Yield Response and Sensitivity Analysis

Before optimising the procurement strategy, we first quantify the biophysical response of crop yield to planting density under climatic stress. Figure 2 shows APSIM-simulated dry matter yields of maize and rye as functions of planting density for five representative scenarios selected by ranking all ensemble members by APSIM-implied total annual biomass and taking predefined yield quantiles. In the upper-quantile (favourable) case, maize averages $17.5\ t\ DM \cdot ha^{-1}$and rye $11.0\ t\ DM \cdot ha^{-1}$. In contrast, in the lower-quantile (adverse) case the combined biomass availability declines markedly, with total dry matter reduced to approximately $20.6\ t\ DM \cdot ha^{-1}$. This quantile-based spread provides the empirical basis for the subsequent hedging optimisation, which trades off deterministic cultivation costs against the expected cost of shortfall procurement under low-yield realisations.

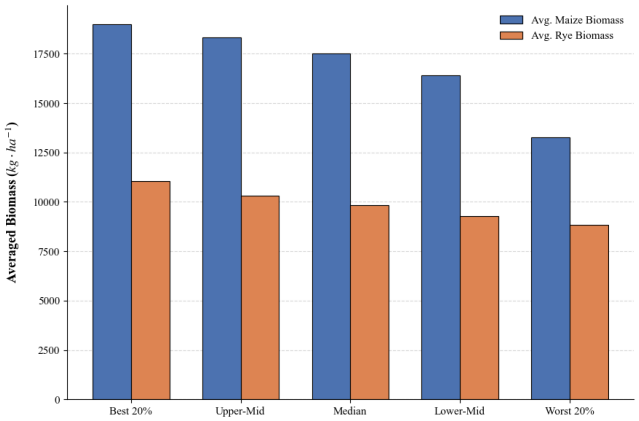

Figure 2. Mean crop biomass (dry matter) response for maize and rye under 5 climate ensemble scenarios, categorised by total yield performance quantiles ($kg \cdot ha^{-1}$).

The yield–density relationship is inherently non-linear and strongly regulated by inter-annual climate variability, where increased density does not monotonically

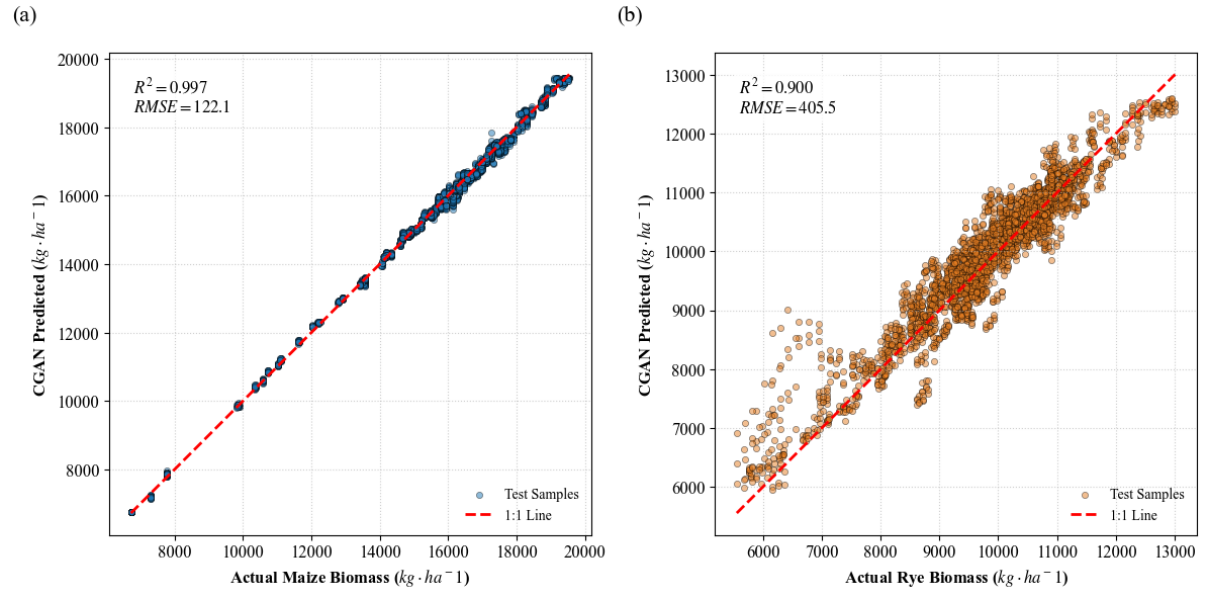

Figure 3. Validation of the CGAN surrogate model on the independent test dataset ($n = 3024$). (a) Maize biomass yield. (b) Rye biomass yield.

translate into higher productivity under water-limited conditions [15]. Increasing planting density generally accelerates canopy closure and increases leaf area index (LAI), thereby enhancing the potential for radiation interception. However, LAI is only a proxy for light capture and is not sufficient to determine final biomass or economic yield. Our mechanistic simulations confirm that while higher densities expand radiation interception capacity, they simultaneously exacerbate transpiration demand, potentially triggering yield collapse during synchronized heat-water stress. This non-linear coupling necessitates the use of a mechanistic solver like APSIM to capture the stochastic tipping points that purely statistical models often overlook. Empirically, this often results in a plateauing (or even declining) yield response beyond an intermediate density, with the optimal density shifting across weather scenarios, consistent with our motivation to treat density and area allocation as hedging decisions under uncertainty.

The non-linear yield response and the pronounced quantile gap indicate that a deterministic, average-year calibration is insufficient for procurement planning. Instead, the decision must account explicitly for the distributional tail where shortfall costs concentrate. This motivates the stochastic optimisation model, which hedges the plant's operational costs against low-probability, high-impact adverse seasons. Implementing this framework with APSIM directly is infeasible because the optimisation requires repeated evaluation of yield distributions under many scenarios and decision candidates. To enable fast probabilistic sampling while retaining biophysical realism, we train a CGAN as a surrogate of APSIM.

The predictive performance of the CGAN surrogate model was evaluated using an independent test dataset ($n = 3024$ scenarios). The parity plots in Figure 3 illustrate the model's ability to replicate the complex biophysical outputs of the APSIM platform across diverse climatic conditions and management intensities.

As shown in Figure 3a, the CGAN achieved superior predictive fidelity for maize biomass ($R^2 = 0.997$). The data points are tightly clustered along the parity line across the entire productivity spectrum. This indicates that the generative network successfully captures the primary physiological drivers of maize growth under summer conditions. In contrast, the prediction for rye biomass (Figure 3b) exhibited higher variance ($R^2 = 0.90$). While the general trend remains robust, the increased dispersion suggests that the secondary crop in the system is subject to more complex stochastic influences that are harder for the model to generalize.

Beyond parity and RMSE/R^2, we compared the marginal density of maize and rye yields (empirical vs CGAN) and observed close alignment overall; modest overestimation of rye in the lowest decile may slightly favor double-crop hedging under severe-drought scenarios, a point considered in the sensitivity discussion.

A critical observation in the rye validation is the systematic over-estimation in low-productivity regimes. This phenomenon is driven by two main factors. First, unlike maize, which grows during the relatively stable summer window, rye's phenology spans the winter-spring transition. It is highly sensitive to events such as late-spring frosts or erratic rainfall distributions. While APSIM's

mechanistic equations may trigger near-zero yields once stress thresholds are exceeded. As a function approximator, the CGAN tends to soften such hard threshold behavior, producing a less discontinuous lower tail (regression-to-the-mean). The smaller absolute magnitude of rye yields further reduces the signal-to-noise ratio in the joint-output model, potentially biasing learning toward the dominant maize gradients. Importantly, the surrogate shows no drift and preserves the directional response to planting density, supporting its use as a computationally efficient generator of yield samples for Monte Carlo-based stochastic optimization.

We conducted a sensitivity analysis by sweeping the contracted annual production target from 80 to $200 \times 10^3\ m^3$, spanning regimes from readily achievable to practically infeasible for a 15-ha farm system. For each target level, we solved the land-allocation problem over three cropping options: maize monocropping, rye monocropping, and maize–rye double cropping. Visualised the resulting optimal area shares in Figure 4. Each point on the curve corresponds to the minimum-cost solution (under the defined reliability metric) obtained by enumerating feasible density pairs and searching over land allocations with 1 ha resolution.

Figure 4 reveals a clear regime shift in the optimal allocation as the production target increases. At low targets $(\approx 80 - 110 \times 10^3\ m^3)$, the optimiser predominantly allocates land to maize and rye monocropping, indicating that under moderate contractual requirements, a single-crop strategy provides sufficient biomass supply without relying on intensification. An additional feature in Figure. 4 is the intermittent appearance of a small DR share within the low-target regime, embedded in an otherwise single crop dominated allocation. This should be interpreted as a near-indifference band rather than a qualitative regime change. Under the objective of minimizing expected total cost (seed expenditure plus expected shortfall purchasing), several allocation patterns can become quasi-equivalent in this range. Introducing a marginal area of the DR system may slightly reduce tail shortfall exposure under adverse realizations, which can offset its yield-penalty factor and additional seed inputs. Because the objective differences are small and land is discretized at 1-ha resolution, the optimizer may switch between near-tied allocations across adjacent targets, producing the visually inserted sliver. Importantly, this indicates operational robustness in the low-target regime, that small changes in allocation do not materially affect expected cost or reliability.

Around $110 \times 10^3\ m^3$, the allocation begins to transition. This marks a threshold for operational sustainability, as further increases in target yields bring mono-crop strategies close to their biophysical limits, posing significant risks under unfavourable yield expectations. Beyond this threshold, the optimiser progressively reallocates area away from monocropping and toward intensification via double cropping, which becomes the dominant strategy as targets approach $135 - 140 \times 10^3\ m^3$.

For higher targets ($> 140 \times 10^3\ m^3$), the solution space collapses toward an all-out intensification regime in which almost all land is assigned to double cropping. This indicates that, for a fixed land base, meeting high contractual requirements relies primarily on stacking production within the year rather than marginal tuning of a single crop. Overall, Figure 4 summarises how increasing contractual pressure drives a qualitative change in the optimal procurement strategy, from simple monocropping to increasingly intensive multi-crop deployment. This marks a clear shift in land-allocation behaviour and reveals distinct planning regimes.

To illustrate the decision complexity at a risk-sensitive contract level, we fix the production target at $Q_{tar} = 140 \times 10^3 m^3$ for the 15-ha system and evaluate the cost–reliability trade-off over the feasible land-allocation and density options. We define supply confidence as the probability that internal production meets the contractual target:

$$Confidence = Pr(\hat{\mathrm{Q}} \geq Q_{\mathrm{tar}}) \tag{7}$$

It is estimated via CGAN-based Monte Carlo sampling (scenario-weighted) as the fraction of simulated realizations in which annual supply exceeds the target.

Table 1 summarises three representative strategies spanning the economic-to-reliability spectrum. Land is reported as D/M/R ha, where D denotes maize–rye double cropping, M maize monocropping, and R rye monocropping. Pop denotes planting densities for maize/rye. At the low-input extreme (0/0/15), allocating all land to rye minimises cultivation cost (£5, 536) but achieves only 34.3% confidence, implying frequent reliance on external procurement under adverse yield realizations. A balanced strategy (7/1/7) increases confidence to 59.4% at a moderate cost (£7, 009), reflecting that reallocating hectares toward intensification and diversification can materially reduce downside exposure. The high-reliability extreme (15/0/0) assigns all land to double cropping, achieving the highest confidence observed in this set (85.7%) at the highest cost (£8, 364). Notably, even the most intensive plan does not reach 100% confidence, indicating residual exposure to extreme climate realizations under a fixed 15-ha land constraint. In those cases, recourse actions (spot-market purchases) remain necessary. Practically, these strategies correspond to different managerial priorities. If the operator is cost-constrained and can tolerate frequent recourse purchases, the low-input rye plan is preferred. If a balanced hedge is required, the mixed allocation offers a mid-cost improvement in reliability. If the minimum supply guarantee compliance is the dominant objective, the solution shifts toward full intensification via double cropping.

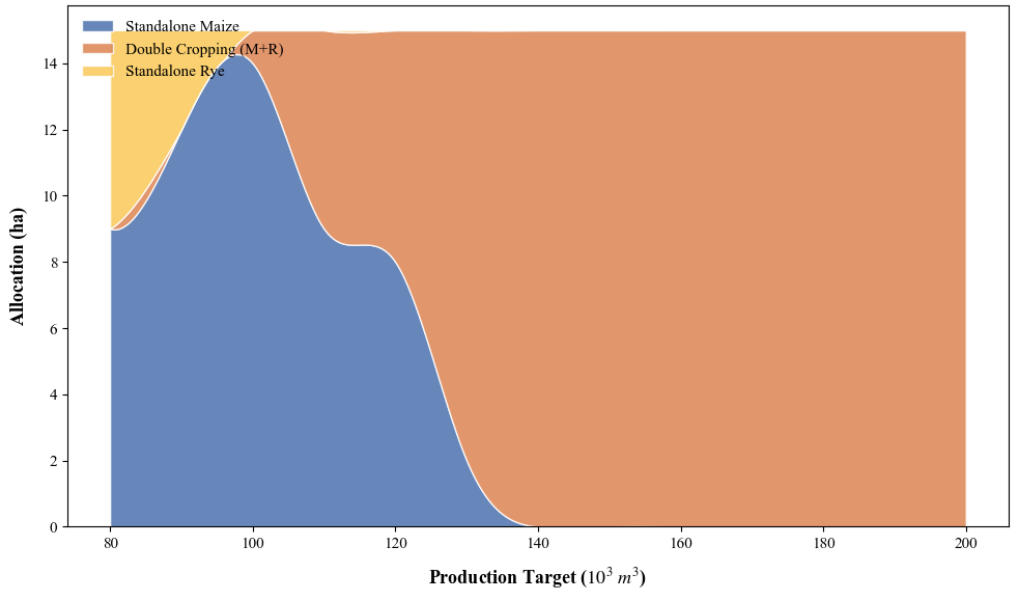

Figure 4. Optimal land allocation across production targets for a 15-ha system. For each annual production target ($80 - 200 \times 10^3$ m^3), the optimization selects the area (ha) allocated to maize monocropping (blue), rye monocropping (yellow), and maize–rye double cropping (orange). The stacked areas visualise the evolution of the optimal strategy as contractual requirements increase, revealing a threshold-like transition from monocropping-dominated solutions to intensification via double cropping.

Overall, the $140 \times 10^3 m^3$ target represents a regime where the system transitions from cost-first planning to explicit risk mitigation. Hectare-level allocation choices produce large changes in confidence, while marginal cost increases eventually yield decreasing reliability gains. This provides a quantitative basis for contract feasibility discussions, either accepting a target with residual procurement exposure, or complementing agronomic optimisation with target flexibility and/or external buffering arrangements.

Table 1: Strategic decision matrix for optimal land allocation under a high production target (140×10^3 m^3). The table compares the performance of three Pareto-optimal regimes: Economic, Balanced, and Reliable, extracted from the climate-resilience simulations. Allocations are presented as hectares of double-cropping (D), standalone maize (M), and standalone rye (R).

Land (D/M/R ha)	Pop (M/R)	Total Cost (£)	Confidence (%)
0.0/0.0/15.0	8/250	5, 535.71	34.30
7.0/1.0/7.0	8/250	7, 008.57	59.40
15.0/0.0/0.0	8/250	8, 364.29	85.70

Although the numerical experiments in this paper focus on pre-season planning, the proposed architecture is explicitly designed for stage-based rolling updates. As the season progresses, forecast components in the scenario feature vector $W_{s,\tau}$ can be progressively replaced by realized observations, and procurement price inputs ($P_{spot,\tau}$) can be refreshed from market quotes. This update decreases predictive uncertainty in the CGAN-generated yield distribution and enables rapid re-evaluation of shortfall exposure under the same committed planting plan. Each solve is a two-stage SP with first-stage cropping/procurement decisions and second-stage recourse. Designing explicit mid-season controls is left for the journal version. The surrogate's low evaluation cost is precisely what makes such sequential re-optimisation operationally feasible.

CONCLUSION

This study addresses the robustness of AD feedstock procurement under extreme climate uncertainty by developing a closed-loop decision support system encompassing weather scenario synthesis, mechanistic yield generation and stochastic planting decision optimization with rolling updates. Through typical case study, the study concludes that:

- The CGAN surrogate trained on APSIM simulations reproduces non-linear yield responses to climate variability and management intensity with strong out-of-sample agreement. This provides realistic, distributional yield inputs for downstream

stochastic optimization, enabling planners to hedge against downside under-yield tail events rather than relying on mean-year assumptions.

- The optimal planting structure (land allocation across monocropping vs double cropping, and discrete planting densities) is highly target dependent. As the contracted gas target increases, the decision space compresses and the optimal policy exhibits regime shifts. Low targets admit multiple low-cost feasible allocations, whereas high targets push the system toward intensification and diversification, which reveals threshold-like behavior consistent with biophysical and land constraints.
- A cost–reliability trade-off evaluation at the intensified target level ($140 \times 10^3 m^3$) shows residual exposure even under maximal intensification. In our case study, the most reliable strategy still requires recourse procurement in approximately 14.3% of realizations. It quantifies the remaining vulnerability to extreme climate outcomes under a fixed land base and highlights a practical carrying-capacity ceiling for contract design.
- In practice, the framework yields three actionable strategy classes: economical, balanced, and reliability oriented. Multiple strategies allow decision-makers to select policies aligned with budget and risk preferences. The rolling (stage-based) update capability allows the procurement hedge to be revised as forecast uncertainty collapses through the season, improving decision timeliness without changing already-committed sowing decisions.

AUTHOR IDENTIFIERS

Author ORCIDs:
Ruosi Zhang: 0000-0002-4931-9771

REFERENCES

1. Alengebawy A, Ran Y, Osman AI, Jin K, Samer M, Ai P. Anaerobic digestion of agricultural waste for biogas production and sustainable bioenergy recovery: a review. Environ Chem Lett 22:2641-2668 (2024). https://doi.org/10.1007/s10311-024-01789-1
2. Espinoza Pérez AT, Camargo M, Narváez Rincón PC, Alfaro Marchant M. Key challenges and requirements for sustainable and industrialized biorefinery supply chain design and management: a bibliographic analysis. Renewable and Sustainable Energy Reviews 69:350-359 (2017). https://doi.org/10.1016/j.rser.2016.11.084
3. Souza GM, Ballester MVR, de Brito Cruz CH, Chum H, Dale B, Dale VH, Fernandes ECM, Foust T, Karp A, Lynd L, Maciel Filho R, Milanez A, Nigro F, Osseweijer P, Verdade LM, Victoria RL, Van der Wielen L. The role of bioenergy in a climate-changing world. Environmental Development 23:57-64 (2017). https://doi.org/10.1016/j.envdev.2017.02.008
4. Nyéki A, Neményi M. Crop yield prediction in precision agriculture. Agronomy 12:2460 (2022). https://doi.org/10.3390/agronomy12102460
5. Mall RK, Gupta A, Sonkar G. Effect of climate change on agricultural crops. Current Developments in Biotechnology and Bioengineering :23-46 (2017). https://doi.org/10.1016/b978-0-444-63661-4.00002-5
6. Lepnaan Dayil J, Akande O, Mahmoud AED, Kimera R, Omole O. Challenges and opportunities in machine learning for bioenergy crop yield prediction: a review. Sustainable Energy Technologies and Assessments 73:104057 (2025). https://doi.org/10.1016/j.seta.2024.104057
7. D. P. Holzworth *et al.*, "APSIM – Evolution towards a new generation of agricultural systems simulation," *Environ. Model. Softw.*, vol. 62, pp. 327–350, Dec. 2014, doi: 10.1016/j.envsoft.2014.07.009.
8. Tamayo-Vera D, Wang X, Mesbah M. A review of machine learning techniques in agroclimatic studies. Agriculture 14:481 (2024). https://doi.org/10.3390/agriculture14030481
9. Bhosekar A, Ierapetritou M. Advances in surrogate based modeling, feasibility analysis, and optimization: a review. Computers & Chemical Engineering 108:250-267 (2018). https://doi.org/10.1016/j.compchemeng.2017.09.017
10. Mayerle SF, Neiva de Figueiredo J. Designing optimal supply chains for anaerobic bio-digestion/energy generation complexes with distributed small farm feedstock sourcing. Renewable Energy 90:46-54 (2016). https://doi.org/10.1016/j.renene.2015.12.022
11. Kim S, Kim S. Hybrid simulation framework for the production management of an ethanol biorefinery. Renewable and Sustainable Energy Reviews 155:111911 (2022). https://doi.org/10.1016/j.rser.2021.111911
12. Y. Shi *et al.*, "Deep Learning Meets Process-Based Models: A Hybrid Approach to Agricultural Challenges," *ArXiv Prepr. ArXiv250416141*, 2025.
13. Gal N, Milstein I, Tishler A, Woo CK. Fuel cost uncertainty, capacity investment and price in a competitive electricity market. Energy Economics 61:233-240 (2017). https://doi.org/10.1016/j.eneco.2016.11.014

14. Chabar R, Pereira M, Granville S, Barroso LA, Iliadis N. Optimization of fuel contracts management and maintenance scheduling for thermal plants under price uncertainty. 2006 IEEE PES Power Systems Conference and Exposition :923-930 (2006). https://doi.org/10.1109/psce.2006.296437
15. Lobell DB, Roberts MJ, Schlenker W, Braun N, Little BB, Rejesus RM, Hammer GL. Greater sensitivity to drought accompanies maize yield increase in the U.S. midwest. Science 344:516-519 (2014). https://doi.org/10.1126/science.1251423