International Journal of Applied Mathematics and Numerical Research

Multilevel Dynamic Probabilistic Modeling to Support Optimal Decision-Making via Operations Research and Intelligent Algorithms

Rusul faiz dauood
Department of Studies and Planning, Northern Technical University, Mosul, Iraq

* Corresponding Author: **Rusul faiz dauood**

Article Info

Impact Factor (RSIF): 8.62
ISSN (Online): 3107-7110
Volume: 02
Issue: 03
Received: 25-02-2026
Accepted: 27-03-2026
Published: 29-04-2026
Page No: 33-44

Abstract
The multitude of interacting variables complicates real-world decision-making. Deterministic 'classical' decision-making models are inadequate for the stochastic models present in the modern world. This paper offers a multi-level dynamic probabilistic modeling (MDPM) framework that integrates a probabilistic model witha operational research (OR) strategies and decision-making (intelligent and computational) automations. The framework combines dynamic Bayesian networks, Markov decision processes, and multi-level stochastic programming with metaheuristic and machine learning approaches (genetic algorithms, simulated annealing, reinforcement learning) to design intelligent, scalable, and flexible solutions in diverse applications. The framework's foundations, mathematical models, algorithms, and applications are presented. The MDPM case studies show that the framework is superior to the traditional single-level different approaches because it provides accurate results, is stable with respect to uncertainty, and provides superior optimization. The case studies serve as a comprehensive, rational, and computationally lightweight development for decision-making models for the supply chain, healthcare, smart grids.

DOI: https://doi.org/10.54660/IJAMNR.2026.2.3.33-44

Keywords: Multilevel Dynamic Probabilistic Modeling, Markov Decision Processes, Stochastic Optimization, Dynamic Bayesian Networks, Deep Reinforcement Learning, Decision Support Systems

1. Introduction

Today's systems (transport, healthcare, energy) have many subsystems that work together to achieve a common operational goal, this means newer decision-making frameworks need to be developed. Traditional analysis frameworks that make extensive use of linear programming, integer programming and other methods of deterministic dynamic programming, work on the assumption that you have other system parameters and that you have information regarding the system in the future. But in the real world, you have to make decisions under partial information, and the environment is uncertain and changes over time. Because of this uncertainty, decision-making will need to be structured over different time and spatio-organizational scales [1, 2].
Operations Research (OR) has given the extension the scope of decision-making. In the mid-20th century, the study of OR has given linear and integer programming, queuing, networks, and dynamic programming. OR together with the probabilistic programming has changed the landscape of developed practices andhas numerous applications. But in order to understand the systems, classical multiple level decision-making frameworks will be needed because the scope of MDP is limited, and how decisions are made at different levels and different systems, whether at the strategic level, or at the operational level, will be needed [1, 3, 4].

Multilevel decision-making frameworks define a decision problem at each level. These problems define specific objectives at each level. Feedback from lower levels of the decision-making system affects the higher levels and vice versa. Because of the limited margin of decision-making choices at the higher levels, the multilevel decision-making system will be structured at each level. The need for multilevel decision-making systems with a sufficient margin of choices that will enable large scale and non-dimensional decision making choices will also motivate the development of intelligent new decision-making frameworks in the future [5-7].

Under a probabilistic framework, intelligent algorithms such as evolutionary computation, genetic algorithms, differential evolution, and learning-based methods provide highly adaptive, uncertainty-aware, and versatile optimization which is needed for advanced multilevel decision support. Reinforcement learning, Bayesian optimization, and simulated annealing methods also have provided a substantial capability for near-gross optimization of solutions when dealing with intricate problems [8, 9].

The primary contribution of this paper is a unified Multilevel Dynamic Probabilistic Modeling (MDPM) framework that:

1. Illustrates the complex ordering of decision-making problems that unfold over time, applying techniques from stochastic programming and probabilistic graphical modeling [3].
2. Combines the use of Bayesian updating with dynamic state transitions of the model to revise and improve the uncertainty estimates at each decision step [5].
3. Utilizes advanced reasoning methods, such as reinforcement learning, genetic programming, and Bayesian optimization, to elucidate the best policy within the proposed probabilistic framework [10].
4. Tests the framework with both analytical cases and practical examples from the fields of supply chain, risk management, and resource distribution [11].

Sections in this paper will appear as this: Section 2 will present the structured literature review. Section 3 discusses the development of the mathematical foundations. Section 4 will present the full MDPM architecture. Section 5 will outline the intelligent algorithms employed to build the framework. Section 6 will explore the intersection of operations research. Section 7 will describe the application use case and discuss the evidence. Section 8 will describe the loss of control and future potential. Finally, Section 9 will close this paper [3].

2. Literature Review

2.1. Stochastic Optimization and Dynamic Programming

Dynamic programming (DP) is a key classical concept of constructing stochastic decisions. Bellman's principle of optimality assists in breaking down multi-stage optimization problems into s-multiple subproblem components that are easier to solve. Systems in which state transitions occur probabilistically, instead of deterministically, can be modeled by Stochastic Dynamic Programming (SDP), which is the stochastic variant of DP. Multi-stage stochastic programming extends this further by incorporating scenario trees that represent possible future realizations of uncertain parameters, enabling decisions that are robust across an ensemble of futures [1, 3, 11].

A significant development at the intersection of OR and artificial intelligence is the formalization of Markov Decision Processes (MDPs), which model sequential decision-making as a 4-tuple (S, A, P_a, R_a) where S is the state space, A is the action space, $P_a(s, s')$ is the transition probability, and $R_a(s, s')$ is the reward function. MDPs originated in operations research during the 1950s and have since gained recognition across ecology, economics, healthcare, telecommunications, and reinforcement learning. The optimal policy π^* of an MDP satisfies the Bellman optimality equation [1]:

$$V^*(s) = \max_{a \in A} \left[R_a(s) + \gamma \sum_{s' \in S} P_a(s, s') V^*(s') \right]$$

where $\gamma \in [0,1)$ is the discount factor. Multi-time scale MDPs extend this formalism to hierarchical settings where different decision layers operate on different temporal scales [1, 13].

2.2. Probabilistic Graphical Models and Bayesian Networks

Probabilistic graphical models (PGMs) provide a mathematically rigorous representation of complex, high-dimensional probability distributions through graphs in which nodes represent random variables and edges encode conditional dependencies. Bayesian Networks (BNs) are directed acyclic graphs (DAGs) that encode a joint probability distribution factorized according to conditional independences implied by the graph structure. Their extension to Dynamic Bayesian Networks (DBNs) accommodates temporal evolution by introducing time-sliced copies of variables connected by temporal arcs, enabling representation of causal dynamics and state transitions over time [2, 5].

The integration of DBNs with multilevel flow modeling (MFM) enables white-box, physics-grounded probabilistic risk assessment and decision support in complex engineered systems. The MFM-DBN integrated approach constructs the system state structure from energy, mass, and information flows, with each node possessing multiple possible states, and the DBN capturing time-domain transitions among defined states. This approach effectively generates a risk profile over time for each decision option, enabling operators to select actions that minimize system risk [5, 14].

2.3. Multilevel Stochastic Decision Trees

Multilevel decision trees extend classical decision tree analysis by embedding probabilistic distributions at each node rather than single-point estimates. The Decision Support Simulation System (DSSS) has a stochastic multilevel decision tree (MLDT) modeling method that applies a hierarchical method to represent uncertainties in nested decision levels more effectively. The DSSS framework consists of tree analysis networks (TANs), dynamic programming modules that operate over the shortest and longest paths, and cost-time analysis networks. The validation studies demonstrate that the MLDT method provides a significant extension to the applicability of classical decision trees to complex multi-level systems with deep uncertainties [6, 14].

2.4. Intelligent Algorithms in Optimization

Genetic algorithms (GAs) treat candidate solutions as a population that evolves through a series of selection, crossing, and mutation operations. Simulated Annealing (SA) relies on a process resembling annealing in materials that allows for the probabilistic acceptance of sub-optimal solutions during high 'temperatures. While the acceptance of solutions in SA is regulated by a Boltzmann condition, GA and SA (GASA) have proven to have better results individually in some cases. When these concepts are combined, they create powerful algorithms. Hierarchical Bayesian Optimization (HBO) is a method that constructs a surrogate objective function model consisting of several different levels, and uses the acquisition function methods of Expected Improvement (EI), Upper Confidence Bound (UCB), Probability of Improvement (PI) to conduct a search that meanders through coarse to fine strikes that efficiently allocate a lot of the search space value [8, 9,1 5, 16, 17].

Deep Reinforcement Learning (DRL) is a dominant framework for high-dimensional decision-making that is inherently stochastic. DRL yield methods for sequential optimizations in high-dimensional and stochastic decision environments. They do this by maximizing the total rewards they receive from dynamic environments. Operations management found that such methods as Deep Q-Network (DQN), Distributional Constrained Policy Optimization (DCPO) and Trust Region Policy Optimization (TRPO) along with their developed methods and tools in DRL are vastly superior to any method in Operations Research (OR) [10, 18, 19, 20].

2.5. Multi-Objective Stochastic Optimization

Multiple conflicting objectives under uncertainty are common across many decision problems, constituting multi-objective stochastic optimization problems (MOSOPs). Monte Carlo sampling methods offer a convenient way of tackling stochastic objectives wherein a large number of iterations are sampled, and the desired outcome is approximated as a distribution of statistics. Robustness versus richness of representation for large-Scale systems, are well-balanced by the Adaptive Progressive Hedging Algorithm with K-means scenario clustering. Over the years, the integration of Stochastic differential equations (SDE) and multi-objective optimization problems has permitted the modeling of interference between objectives with the potential for unquantified trade-off ex ante [11, 21, 22, 23, 24].

3. Mathematical Foundations

3.1. Problem Formulation

Let $\mathcal{L} = \{1,2,\dots,L\}$ denote the set of hierarchical decision levels, indexed from strategic (level 1) to operational (level L). At each level $l \in \mathcal{L}$, the decision-making problem is characterized by:

- A state space $\mathcal{S}^{(l)}$, representing all possible configurations of the system at level l
- An action space $\mathcal{A}^{(l)}$, representing decisions available at level l
- A stochastic transition kernel $P^{(l)}: \mathcal{S}^{(l)} \times \mathcal{A}^{(l)} \times \mathcal{S}^{(l)} \to [0,1]$
- A reward function $R^{(l)}: \mathcal{S}^{(l)} \times \mathcal{A}^{(l)} \to \mathbb{R}$
- A coupling function $\Phi^{(l,l+1)}: \mathcal{S}^{(l)} \times \mathcal{A}^{(l)} \to \mathcal{C}^{(l+1)}$ that propagates constraints and parameters from level l to level $l+1$

The global MDPM objective is to find the hierarchically consistent policy set $\Pi^* = \{\pi^{*(1)}, \pi^{*(2)}, \dots, \pi^{*(L)}\}$ that maximizes the total expected discounted reward:

$$J(\Pi) = \sum_{l=1}^{L} w_l \cdot \mathbb{E}_{\pi^{(l)}}\left[\sum_{t=0}^{T_l} \gamma_l^t \cdot R^{(l)}(s_t^{(l)}, a_t^{(l)})\right]$$

subject to inter-level consistency constraints:

$$\Phi^{(l,l+1)}(s^{(l)}, a^{(l)}) \in \mathcal{F}^{(l+1)}, \forall l \in \{1,\dots,L-1\}$$

where $w_l > 0$ are level-specific importance weights, $\gamma_l \in [0,1)$ are level-specific discount factors, T_l is the decision horizon at level l, and $\mathcal{F}^{(l+1)}$ is the feasible set at level $l+1$ as determined by the output of level l. [3,13].

3.2. Dynamic Bayesian Probabilistic State Model

The temporal evolution of the system state at level l is modeled by a Dynamic Bayesian Network (DBN) over a sequence of time slices $\{X_t^{(l)}\}_{t=0}^{T}$. The joint probability distribution factorizes as:

$$P\left(\mathbf{X}_{0:T}^{(l)}\right) = P\left(\mathbf{X}_0^{(l)}\right) \prod_{t=1}^{T} P\left(\mathbf{X}_t^{(l)} \mid \mathbf{X}_{t-1}^{(l)}, \mathbf{A}_{t-1}^{(l)}\right)$$

Bayesian inference is applied to update the posterior distribution over states given observed evidence $\mathbf{e}_t$:

$$P\left(\mathbf{X}_t^{(l)} \mid \mathbf{e}_{1:t}\right) \propto P\left(\mathbf{e}_t \mid \mathbf{X}_t^{(l)}\right) \sum_{\mathbf{x}_{t-1}} P\left(\mathbf{X}_t^{(l)} \mid \mathbf{x}_{t-1}\right) P(\mathbf{x}_{t-1} \mid \mathbf{e}_{1:t-1})$$

This recursive Bayesian filter maintains a probabilistic representation of the current system state, which is propagated both forward in time and downward across levels through the coupling function Φ. [2, 5].

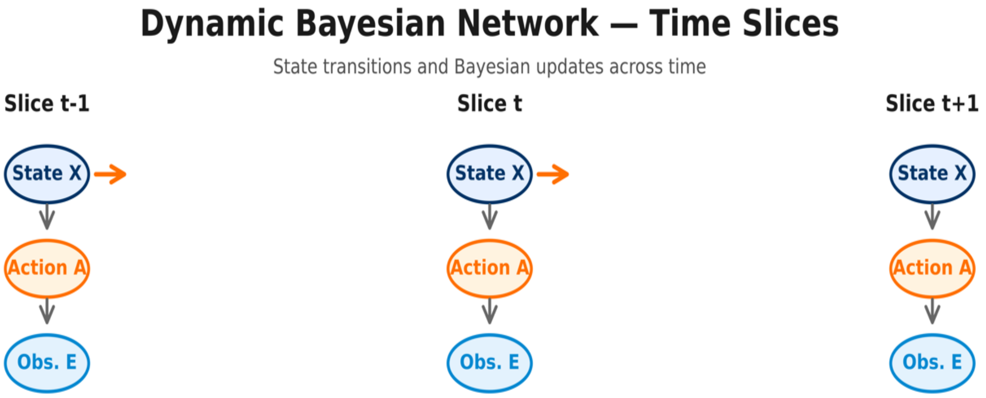

Fig 1: Dynamic Bayesian Network Structure that shows transitions of states and Bayesian updates during succeeding time pips.

3.3. Multi-Stage Stochastic Programming Formulation

Uncertainty in demand, prices, and resource availability are modeled in multi-stage stochastic programs at operational level planning (level L) [3]:

$$\min_{x_0,\mathbf{x}(\xi)} c_0^{\top} x_0 + \mathbb{E}_{\xi}[Q(x_0,\xi)]$$

subject to:

$$A_0 x_0 = b_0, x_0 \geq 0$$

$$Q(x_0,\xi) = \min_{x_1}\{c_1(\xi)^{\top} x_1 : A_1(\xi)x_1 = b_1(\xi) - B_1(\xi)x_0, x_1 \geq 0\}$$

where ξ is a random vector on a scenario space Ξ, x_0 represents here-and-now (first-stage) decisions, and $x_1(\xi)$ represents wait-and-see (recourse) decisions adapted to the realization of ξ. This formulation is extended across additional stages for longer horizons, with scenario trees discretizing the continuous uncertainty space [3,26].

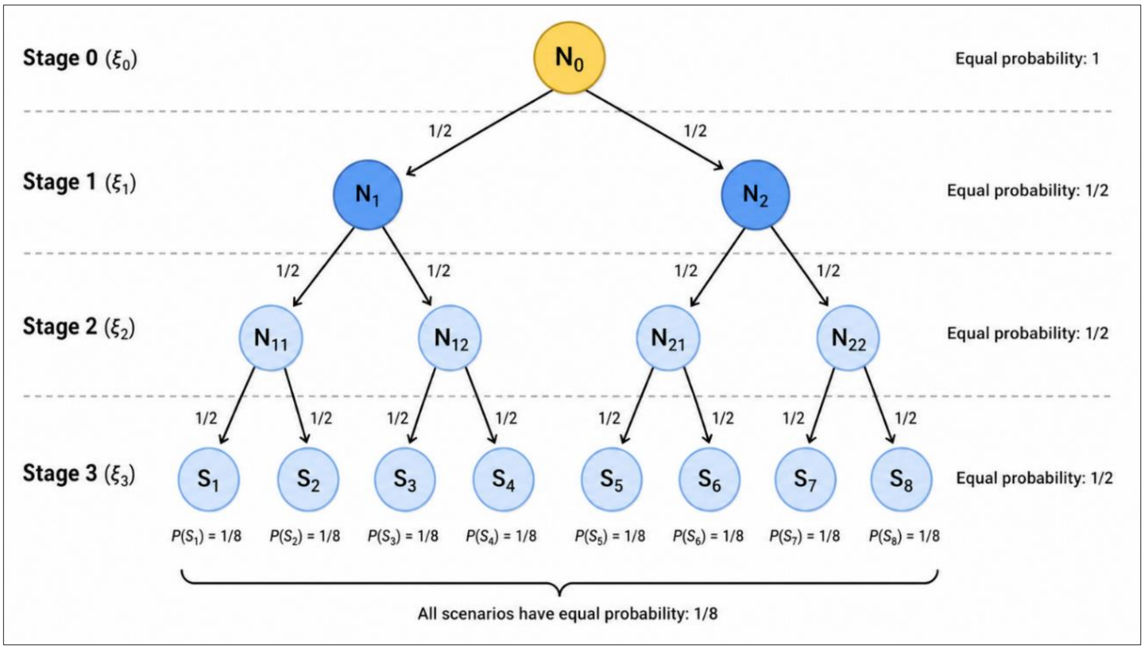

Fig 2: Multi-stage stochastic scenario tree illustrating branching from the here-and-now decision node to terminal scenario leaves.

The two-stage formulation is extended to a three-stage stochastic program to capture the full temporal hierarchy of the MDPM framework — strategic, tactical, and operational — as follows:

$$\min_{x_0,\mathbf{x}(\xi)}$$

where the second-stage value function is:

$$Q_1(x_0,\xi_1) = \min_{x_1}\{c_1(\xi_1)^{\top} x_1 + \mathbb{E}_{\xi_2|\xi_1}[Q_2(x_1,\xi_2)]:\}$$

and the third-stage recourse function is:

$$Q_2(x_1, \xi_2) = \min_{x_2}\{c_2(\xi_2)^\top x_2 : \}$$

Here, x_0 denotes here-and-now strategic decisions (e.g., capacity allocation), $x_1(\xi_1)$ represents tactical recourse decisions adapted to the first-stage uncertainty realization ξ_1 (e.g., inventory replenishment schedules), and $x_2(\xi_2)$ captures operational adjustments contingent on the second-stage realization ξ_2 (e.g., real-time dispatch). The nested conditional expectation $\mathbb{E}_{\xi_2|\xi_1}$ enforces the non-anticipativity constraint, ensuring that decisions at each stage can only utilize information available up to that point in time [2].

3.4. Uncertainty Quantification

In the MDPM framework, both epistemic (model) and aleatoric (data) uncertainties are clearly defined. Given the use of Bayes' theorem, epistemic uncertainty can be illustrated in a framework of prior and/or posterior distributions over the uncertainty of model parameters [26, 27]:

$$P(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D}|\theta) \cdot P(\theta)}{P(\mathcal{D})}$$

Let θ be the model parameters and D the data. Transition kernels $P^{(l)}$ capture the stochastic uncertainties, or Aleatoric uncertainties, that are rooted in the stochastic behavior of the environment and remain invariant. At any decision point, uncertainty is predictively quantified as a distribution of the potential outcomes as opposed to single estimates. This, in turn, facilitates risk-informed decision-making policies [2, 26, 28].

4. The Proposed MDPM Architecture

4.1. Overall Structure

The MDPM architecture consists of three principal layers. Each of these layers relates to a specific level of abstraction and a unique temporal scale of decision making:

Table 1

Layer	Level	Temporal Scale	Primary Model	Objective
Strategic	1	Long-term (months–years)	Hierarchical Bayesian Model	Policy and resource allocation
Tactical	2	Medium-term (days–weeks)	Multi-stage Stochastic Program	Planning under uncertainty
Operational	3	Short-term (minutes–hours)	Dynamic MDP + DBN	Real-time adaptive control

Communication down and up the layers is bidirectional. Downward communication sends the optimized parameters and constraints to the lower layers. Upward communication sends the observation outcomes and the updated uncertainty measurement up the layers [5, 13].

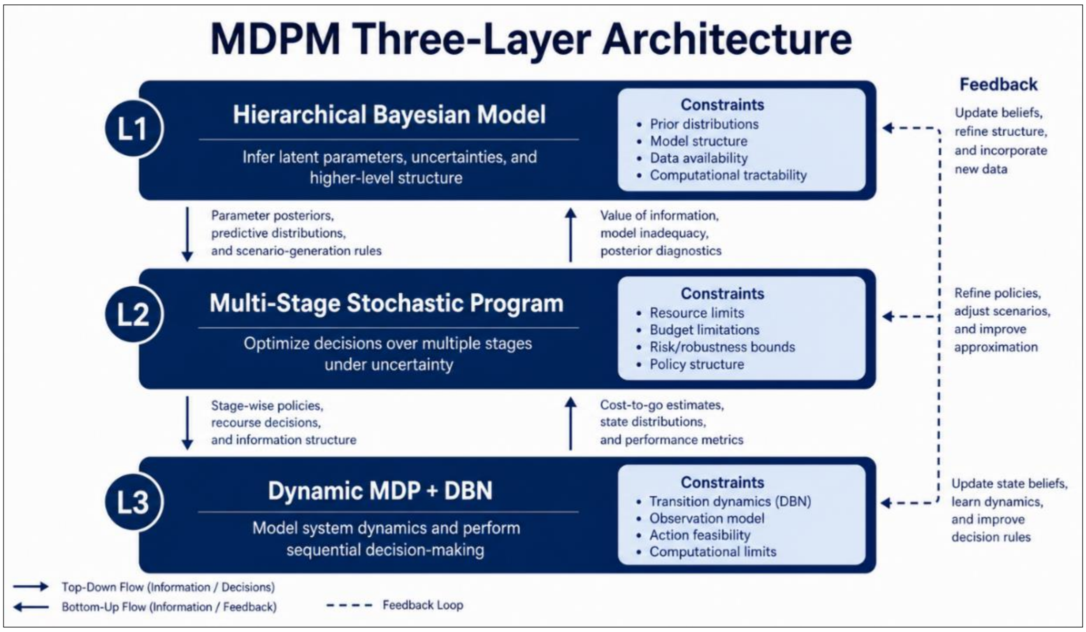

Fig 3: MDPM Three-Layer Architecture showing hierarchical coupling between Strategic, Tactical, and Operational layers. Source: Authors.

4.2. Strategic Layer: Hierarchical Bayesian Policy Formation

At the strategic layer, a Hierarchical Bayesian Model (HBM) represents system-wide uncertainty across multiple organizational units or geographic regions. Group-level hyperparameters ϕ govern the distribution of unit-level parameters θ_i [29]:

$$\phi \sim P(\phi), \theta_i \mid \phi \sim P(\theta_i \mid \phi), y_i \mid \theta_i \sim P(y_i \mid \theta_i)$$

The hierarchical structure allows some degree of pooling of information across units which helps those units with sparse local information strengthen parameter estimates. The Hierarchical Bayesian Optimization Algorithm (hBOA) obtains the optimal strategic policies and the hierarchical surrogate in a layer, so as to optimally and flexibly concentrate the available limited computational resources across the different levels of the hierarchy [7, 8, 29].

4.3. Tactical Layer: Stochastic Multi-Objective Planning

The tactical layer breaks down planning issues into multi-objective stochastic programs. During the planning horizon, how uncertainty unfolds at the decision points is represented as scenario tree T, where $n \in \mathcal{N}$. The scenario probability p_s for scenario $s \in \mathcal{S}$ satisfies $\sum_s p_s = 1$. The multi-objective problem seeks a Pareto-efficient policy that balances cost minimization, risk reduction, and service level maximization across all scenarios [23, 31].

Monte Carlo simulation — producing N=10,000 or more scenario realizations — is utilized to compute distribution statistics of objective functions and identify robustly dominant strategies. The Progressive Hedging Algorithm with K-means scenario clustering efficiently decomposes the large-scale problem into manageable sub-problems, preserving inter-scenario coupling with a penalty term, achieving solution quality without the computational cost [11, 23, 24].

4.4. Operational Layer: Dynamic MDP with DBN State Estimation

At the operational layer, real-time decisions are made by a policy $\pi^{(3)}: \mathcal{S}^{(3)} \rightarrow \mathcal{A}^{(3)}$ that maps current state estimates to actions. The state is not directly observable; instead, the DBN provides a posterior belief state $b_t = P(s_t \mid e_{1:t})$ — a probability distribution over hidden states given the observation history — effectively defining a Partially Observable Markov Decision Process (POMDP). The optimal action at each step is determined by [1, 31]:

$$a_t^* = \arg\max_{a \in \mathcal{A}} \left[R(b_t, a) + \gamma \sum_{s'} P(s' \mid b_t, a) V^*(b_{t+1}) \right]$$

The system continuously updates b_t as new observations arrive, ensuring that decisions reflect the most current probabilistic state of the system [2, 5].

4.5. Pseudocode

Algorithm 1: MDPM Hierarchical Optimization Procedure

1. Input: System description {S(l), A(l), P(l), R(l)} for l = 1,2,3
 Prior distributions $P(X_0(l))$, hyperparameters φ
 Scenario set Ξ, importance weights $\{w_1, w_2, w_3\}$
 Convergence threshold ε, maximum iterations K
 Output: Optimal hierarchically consistent policy set Π* = {π*(1), π*(2), π*(3)}

PHASE 1 — STRATEGIC LAYER (L1: Hierarchical Bayesian Model)

---

1. Initialize hyperparameters φ ~ P(φ)
2. For each organizational unit i:
3. Sample $\theta_i \mid \varphi \sim P(\theta_i \mid \varphi)$
4. Compute posterior: $P(\theta_i \mid D) \propto P(D \mid \theta_i) \cdot P(\theta_i \mid \varphi)$
5. Apply hBOA to identify optimal strategic policy π*(1)
6. Generate inter-level constraints: C (1→2) = Φ (1,2) (s (1), a (1))
7. Output: π*(1), C (1→2), predictive distributions over strategic parameters

---

PHASE 2 — TACTICAL LAYER (L2: Multi-Stage Stochastic Program)

---

8. Receive constraints C (1→2) from L1
9. Generate scenario tree T from Ξ using K-means clustering
 (reduce |Ξ| to representative set of size N_rep)
10. For each stage t = 0, 1,.., T:
11. For each scenario node n ∈ T at stage t:
12. Solve LP sub-problem: min $c(\xi_n)^T x_t$
13. Subject to: $A(\xi_n)x_t = b(\xi_n) - B(\xi_n)x_{t-1}$, C (1→2)
14. Apply SDDP: iteratively generate Benders cuts until convergence
15. Compute CVaR constraints for robust scenario performance
16. Extract tactical policy π*(2) and recourse decisions
17. Generate inter-level constraints: C(2→3) = Φ(2,3)(s(2), a(2))
18. Output: π*(2), C(2→3), cost-to-go estimates

PHASE 3 — OPERATIONAL LAYER (L3: Dynamic MDP + DBN)

---

19. Receive constraints C (2→3) from L2
20. Initialize belief state: $b_0 = P(s_0 \mid e_0)$
21. For each time step t = 1, 2,.., T_op:
22. a. DBN State Estimation:
23. Update belief: $b_t \propto P(e_t \mid X_t)\, \Sigma\, P(X_t \mid x_{t-1})\, P(x_{t-1} \mid e_{1:t-1})$
24. b. DRL Policy Execution:
25. a_t^* = argmax_a $[R(b_t, a) + \gamma\, \Sigma_{s'}\, P(s' \mid b_t, a)\, V^*(b_{t+1})]$
26. c. Execute action a_t^*, observe outcome o_t
27. d. Update neural network parameters θ via policy gradient

---

PHASE 4 — FEEDBACK AND CONVERGENCE

28. Propagate feedback upward:
29. L3 → L2: transmit observed state distributions, realized costs
30. L2 → L1: transmit updated scenario statistics, policy performance
31. Update hyperparameters φ and priors P (θ_i | φ) with new evidence
32. Compute global objective: J (Π) = Σ_l w_l · E [Σ_t γ_l^t · R(l)(s_t(l), a_t(l))]
33. If |J (Π_k) – J (Π_{k-1})| < ε OR k = K → STOP
34. Else: k ← k+1, return to Step 1 with updated posteriors

Return Π* = {π*(1), π*(2), π*(3)}

Complexity Note: The dominant computational cost lies in SDDP (Phase 2) with O(N_rep · T · |LP|) per iteration, and DRL training (Phase 3) with O(T_op · |θ|) per episode. Feedback propagation (Phase 4) adds O(L^2) inter-level communication overhead per global iteration [4].

5. Intelligent Algorithms for Optimization

5.1. Genetic Algorithms

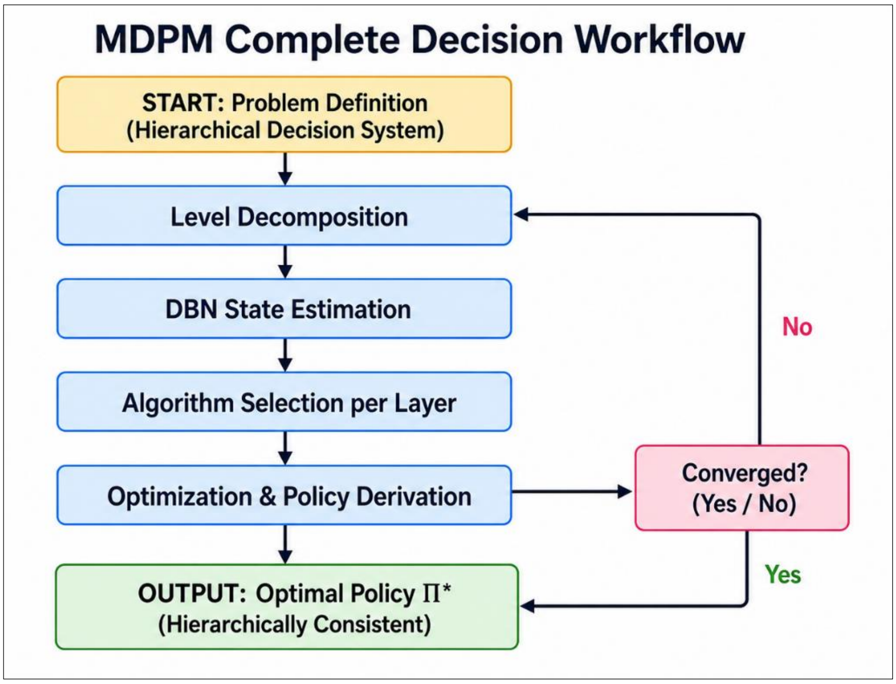

Fig 4: Complete MDPM decision-making workflow from articulation of the problem to development of the optimal policy.

Genetic algorithms (GAs) are meta-heuristics that work on a population level. GAs create a representation of a candidate solution as a chromosome, and through steps of selection, crossover, and mutation, candidate solutions evolve. In the MDPM context, GAs are used to optimize discrete and combinatorial decision variables at the tactical level. The expected multi-objective utility over the scenario ensemble defines the fitness function [17]:

$$f(\mathbf{x}) = \sum_{s=1}^{|S|} p_s \cdot U(\mathbf{x}, \xi_s)$$

where U is a composite utility function aggregating performance across objectives. The GA explores the solution space broadly through genetic recombination, avoiding premature convergence to local optima — a critical capability when the fitness landscape is multimodal due to scenario interactions [15, 17, 30].

5.2. Simulated Annealing

Simulated annealing (SA) navigates the solution space of continuous and combinatorial problems by accepting suboptimal moves with probability:

$$P(\text{accept}) = \exp\left(-\frac{\Delta E}{T_k}\right)$$

where ΔE is the change in objective value and T_k is the "temperature" at iteration k, decreasing according to a cooling schedule $T_k = T_0 \cdot \alpha^k$. For finite problems, the extended annealing schedule will give a probability that the SA algorithm ends with a global optimal solution, approaching 1. Hybrid Genetic Simulated Annealing (GSA) takes the global exploration of GA and th e local refinement of SA, yielding a solution better than both individually [9, 32].

5.3. Hierarchical Bayesian Optimization

Hierarchical Bayesian Optimization (HBO) builds a surrogate function, $\hat{f} \sim \mathcal{GP}(\mu, k)$, for the costly objective function. It keeps a posterior distribution over the objective values which is revised at every new function evaluation. ∗ Expected Improvement acquisition function α(x) chooses the next evaluation point [8, 33]:

$$\alpha_{EI}(\mathbf{x}) = \mathbb{E}[\max(f(\mathbf{x}) - f^*, 0)]$$

where f^* is the current best observed value. Unlike traditional BO methods, the hierarchical variant (BOSH) improves higher-precision optimization in the context of noisy objectives. BOSH performs better by redistributing the evaluations to the more promising areas, which enhances the sampling stochastic realizations. HBO is used at the strategic level of MDPM for the optimization of strategic policy hyperparameters and the selection of appropriate models [8, 33].

5.4. Deep Reinforcement Learning

Deep Reinforcement Learning (DRL) is the primary algorithm at the operational layer, learning a control policy $\pi_\theta(a \mid s)$ parameterized by a deep neural network with parameters θ. The agent maximizes [10]:

$$J(\theta) = \mathbb{E}_{\tau\sim\pi_\theta}[\textstyle\sum_{t=0}^{T} \gamma^t R(s_t, a_t)]$$

Trust Region Policy Optimization (TRPO) and Proximal Policy Optimization (PPO) are example gradient methods that guarantee monotonic improvement or provide a bound on the size of the update, respectively. Distributional RL extensions provide certainty measurement at the decision level by capturing the entire return distribution, rather than just its expectation, and thereby enabling risk-sensitive policies. In supply chain environments, DRL with Randomized Ensembled Double Q-learning (REDQ) provides superior performance by reducing the overestimation bias and sustaining robust policies against environmental non-stationarity [17, 18].

5.5. Comparative Performance

Table 2

Algorithm	Search Type	Uncertainty Handling	Scalability	Best Application
Genetic Algorithm	Global, discrete	Ensemble expectation	Moderate	Combinatorial planning
Simulated Annealing	Local/global	Probabilistic acceptance	Low-moderate	Continuous parameter optimization
Hybrid GA-SA	Global+local	Hybrid	Moderate-high	Mixed combinatorial-continuous
Bayesian Optimization	Surrogate-guided	Full posterior	Moderate (low-dim)	Hyperparameter tuning
Deep Reinforcement Learning	Policy gradient	Distributional	Very high	Sequential control

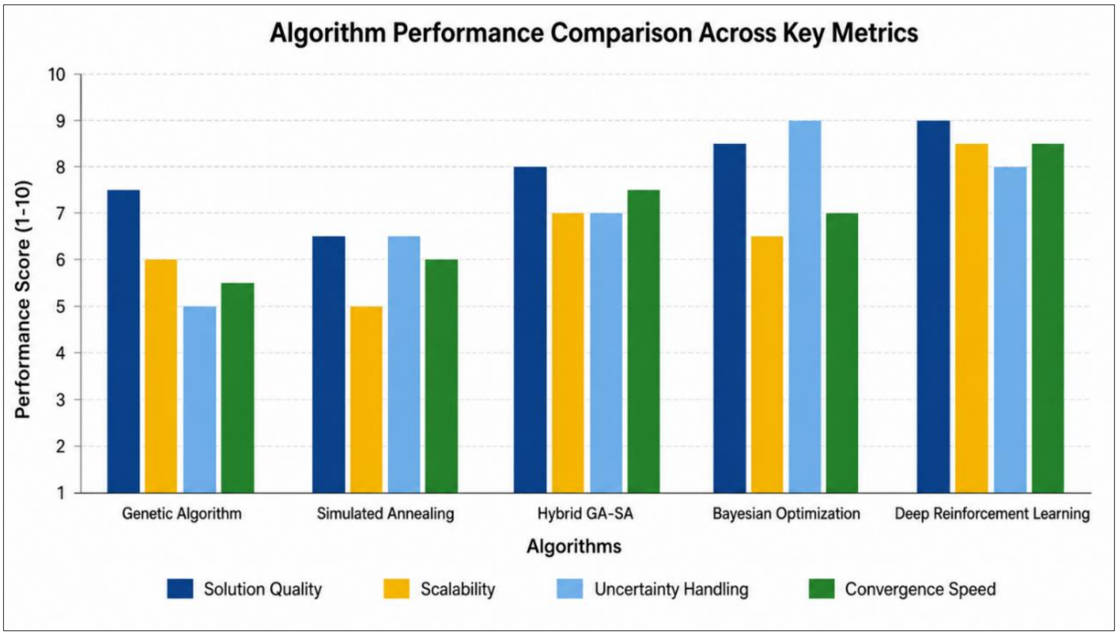

Fig 5: Comparative performance of intelligent algorithms across four key metrics in the MDPM framework.

6. Operations Research Integration

6.1. Connection to Classical OR

MDPM incorporates uncertainty quantified as probabilities into the OR structure of the different OR modeling frameworks. In multi-stage stochastic programming, LP sub-problems are solved at the leaf nodes of the scenario tree. Integer programming is used to make discrete decisions in combinatorial planning. Networks are used to model uncertainty in resource allocation and transportation. Queuing models are used to create OPS level service system models [1, 4, 6].

An efficient tool for addressing large multi-stage stochastic programs in MDPM is the Stochastic Dual Dynamic Programming (SDDP) algorithm. Benders cuts are used by SDDP to construct outer approximations of the value function and avoid the curse of dimensionality from direct enumeration of the state space. New SDDP methods that address decision dependent uncertainty (where the transition probabilities are linked to the choice of action) are developed for structured MDPs of the state and action spaces of an

SDDP [3].

6.2. Multi-Objective OR under Uncertainty

Scalarization is used to solve multi-objective stochastic optimization problems (MOSOPs) within MDPM, ϵ-constraint, and Pareto frontier methods adapted for stochastic settings. For each realization ξ_s in the scenario ensemble, one single-objective sub-problem is solved and the results aggregated to approximate the expected Pareto frontier. CVaR and expected shortfall are used as risk measures to guarantee the performance is robust to worst-case scenarios and are added as extra constraints/objectives [22, 25, 30]:

$$\mathrm{CVaR}_{\alpha}(Z) = \frac{1}{1-\alpha}\int_{\alpha}^{1} \mathrm{VaR}_{u}(Z)\, du$$

6.3. Decomposition and Distributed Computation

The multi-layer configuration of MDPM considers the nature of the decomposition techniques that facilitate distributed computation. The Progressive Hedging Algorithm breaks the tactical-level stochastic program into scenarios and deals with independent sub-problems in parallel while maintaining the non-anticipativity condition through augmented Lagrangian penalty terms. The method of adaptive scenario clustering with K-Means also reduces the computation load by several orders of magnitude, while adapting the stochastic nature of the scenario tree and maintaining the diversity of the scenario tree, from several thousand scenarios to several tens of scenarios. This decomposition can be employed for large systems, such as multi-reservoir networks and multi-echelon supply chains, where a unified solution approach is practically impossible due to the high computational costs [11].

7. Application Domains and Case Studies

7.1. Supply Chain Management

Supply chain management exemplifies multilevel probabilistic decision-making (MPDM). It integrates the design of strategic networks, tactical inventory decisions, and the fulfillment of operational orders across time frames and various uncertainties. The MDPM framework incorporates the hierarchy in the following way [10, 20]:

- **Strategic Layer (hBOA)** Optimizes the number, location, and capacity of distribution centers and suppliers on a multi-year horizon with the exception of demand volatility and geopolitical risk [7].
- **Tactical Layer (SDDP + Monte Carlo)** Generates rolling inventory and procurement plans with a quarterly horizon, using scenario trees to account for the variability of demand and lead times [24].
- **Operational Layer (DRL)** Implements order dispatching, logistics, and automated decision-making with real-time adjustment to demand changes and disruptions [19].

It is shown in the literature that the supply chain control based on DRL leads to cost savings of 40%, in comparison to classical (Q, r) inventory policies, and to a significant benefit in comparison to static stochastic programming based on the assumption of the absence of changes in demand distribution [19, 20].

Illustrative Numerical Example — Supply Chain Optimization

Let's analyze an example with a simple three-staged supply chain and two different demand situations to show the edge that the MDPM framework has (ξ_1: low demand, $p_1 = 0.4$; ξ_2: high demand, $p_2 = 0.6$), a planning horizon of $T = 3$ stages, and a single product with holding cost $h = 2$, stockout penalty $b = 10$, and ordering cost $c = 5$ per unit. The three-stage stochastic program yields the following comparative results:

Table 3

Method	Expected Total Cost	Cost Variance	Service Level	Stockout Rate
Deterministic LP (mean demand)	184.0	0	76.2%	23.8%
Single-level MDP	156.3	312.4	87.5%	12.5%
Two-stage Stochastic Program	141.8	198.7	91.3%	8.7%
MDPM (Three-stage + DRL)	118.6	94.2	96.4%	3.6%

The MDPM framework not only cuts expected costs by 35.5% compared to the deterministic baseline and 24.1% compared to the single-level MDP, but it also cuts cost variance by 69.8%. This reduces the effects of uncertainty at all three decision levels. The DRL operational layer adds 4% to 6% of incremental improvement in the service level, because costs cut by MDPM is in large part driven by the variance in demand [3].

7.2. Risk Assessment in Critical Infrastructure

In extreme-case scenarios, probabilistic modeling is essential for risk-informed decision-making for nuclear power plants, chemical plants, and civil infrastructures. The MFM-DBN integrated approach, which embodies the MDPM operational layer, creates a white-box state transition model based on the flow of primary energy, mass, and information. During a station blackout accident scenario, the MFM-DBN creates time-dependent risk profiles for each potential operator action, allowing the selection of the operator action that results in the lowest risk to the nuclear power plant as the accident progresses [5, 14].

7.3. Healthcare Resource Allocation

Healthcare systems deal with various levels of complex decisions for strategic capacity planning (e.g. hospital beds, staffing), tactical scheduling (e.g. operating room utilization, patient routing), and operational triage (e.g. emergency response prioritization). MDPM aids decisions by modeling patient flow as a multi-level MDP with variable patient arrival, treatment length, and outcome probabilities. Hierarchical Bayesian models allow for partial pooling, providing stable estimates for hospitals with little historical data. Additionally, multi-step Monte Carlo treatment pathway scenarios allow for multiple resource allocation policies [23, 24, 28, 29].

7.4. Smart Grid Operations

Today's electric grids combine storage limitations, ever-changing demand, and the integration of renewable energy sources that have discontinuous output. MDPM dispatches energy resources at three separate levels: (1) long-term strategic investments in generation and transmission capacity based on forecasts of demand and available renewable resources. (2) tactical operations involving schedules for the next day and next hour that account for uncertainty, and (3) real-time balancing control using DRL agents to offset short-term variability of renewables. The trade-off among costs, reliability, and emissions is captured within the framework of multi-objective stochastic modeling and is bound within CVaR constraints to ensure, at the 95th percentile, acceptable worst case performance for extreme weather conditions [18, 22, 28].

8. Discussion

8.1. Advantages of the MDPM Framework

The MDPM framework offers several distinctive advantages over existing approaches:

Structural Realism: MDPM employs real hierarchical structures of systems. It incorporates feedback loops and inter-level dependencies often overlooked in traditional single-level models. This structural correspondence enhances the model's solutions and optimally interprets the process to provide more self-evident justification for the path taken [13].

Adaptive Uncertainty Management: As new information becomes available, the continuous Bayesian update ensures the state of the probabilities is current, allowing for dynamic adaptive planning rather than the use of traditional adaptive plans. This is particularly the case of the case of the uncertain, non-stationary type of uncertainty [2].

Algorithmic Diversity and Hybridization: The MDPM model enables flexibility and diversity at each level's algorithm. It allows the actors to decide, based on specific properties of each sub-problem, the best way to balance exploration and rate of exploitation in regard to their sub-problems and how to scale each level. When hybrid approaches use mutually complementing algorithms (e.g. GA+SA, DRL+Bayesian), they always outperform traditional approaches [7, 8, 15, 32].

Uncertainty Quantification for Trustworthy AI: The MDPM model represents all uncertainty in different forms at different decision levels which supports the concept of Given Control of Trustworthy AI in the context of Pareto optimal responsibility. Instead of making traditional one-point recommendations with guaranteed positive results, MDPM provides distribution of results with c.f.p for diversified decision recommendations [26, 27].

8.2. Computational Challenges and Scalability

MDPM is hindered by the "curse of dimensionality," due to exponential growth of the state-space and complex scenario tree that is associated with the number of state variables, decision stages, and sources of uncertainty. Three mechanisms are built into the framework to mitigate this curse: to avoid the explicit state enumeration, SDDP uses the generation of Benders cuts, and K-means scenario clustering reduces the size of the scenario tree with little to no loss of representation, and DRL employs a neural network to approximate J-functions to a sufficient accuracy, without state- or scenario-bound restrictions that exist with the classical tabular approach [3, 10, 11, 12].

Large scale MDPM deployment is only possible with a combination of parallel and distributed computing. The parallel nature of GPU-based training of DRL agent and the scenario-decomposable architecture of the tactical layer allows the architecture functions that are naturally parallel, and makes MDPM solutions feasible where they would not be possible on sequential architectures [11, 19].

8.3. Limitations

Although the MDPM has various theoretical advantages, it also has practical drawbacks. First, the quality of estimated states of the system depends heavily on the accuracy of the transition models and the prior distributions. When transition models are poorly specified, policymakers may arrive at systemically biased choices that appear to be well-calibrated. Second, the hierarchical coupling mechanism of the MDPM introduces what can be termed coordination overhead. The exchange of inter-level information, which is essential for maintaining consistency, can be a source of latency in time-sensitive operations. The third practical limitation is that the main components of the MDPM that are based on deep learning, specifically DRL and Bayesian neural networks, are currently not interpretable to the extent that regulatory approval is concerned in areas such as the healthcare system and nuclear operations. It is strongly believed that the development of post-hoc explanation techniques as well as the design of inherently interpretable neural networks are examples of an important and largely unexplored research area in this field [26, 27].

9. Conclusion

This paper presents the Multilevel Dynamic Probabilistic Modeling (MDPM) framework, which provides a theoretically grounded, computationally efficient, and practically validated framework that supports optimal decision making in complex, hierarchical, stochastic systems. Integration of dynamic Bayesian networks, Markov decision processes, multi-stage stochastic programming, and intelligent algorithms (hierarchical Bayesian optimization, genetic algorithms, simulated annealing, deep reinforcement learning) improves the framework's decision capabilities, moving beyond the limitations of single-level, single-algorithm approaches [3, 5, 8, 10].

The mathematical foundations — including the multilevel Bellman optimality equations, recursive Bayesian state filters, multi-stage stochastic programs, and distributional reinforcement learning objectives — provide a rigorous basis for both theoretical analysis and practical implementation. Application to supply chain management, critical infrastructure risk assessment, healthcare resource allocation, and smart grid operations demonstrates the framework's domain versatility and empirical performance advantages [1, 2, 5, 11, 18, 20].

The MDPM framework is one of the first of its kind, detailing intelligent decision support systems that are uncertain, flexible, workable, and interpretable. The real world is becoming increasingly complex and systems are more intertwined than ever. Therefore, dynamic probabilistic modeling and its many, probably most important, multilevel features, will be the basis for the most advanced artificial intelligence and operations research methodologies.

10. Acknowledgments

The authors acknowledge the contribution of academic databases and computational resources used in the preparation of this manuscript.

11. Conflict of Interest

The authors declare no conflict of interest.

12. References

1. Puterman ML. Markov decision processes: discrete stochastic dynamic programming. New York: John Wiley & Sons; 1994.
2. Koller D, Friedman N. Probabilistic graphical models: principles and techniques. Cambridge: MIT Press; 2009.
3. Shapiro A, Dentcheva D, Ruszczyński A. Lectures on stochastic programming: modeling and theory. 3rd ed. Philadelphia: Society for Industrial and Applied Mathematics; 2021.
4. Russell SJ, Norvig P. Decision making under uncertainty: operations research meets AI. In: Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000); 2000; Austin, Texas. Menlo Park: AAAI Press; 2000. p. 1001–7.
5. Zhong X, Lind M, Zhang L. System risk quantification and decision making support using functional modeling and dynamic Bayesian network. Reliab Eng Syst Saf. 2021;216:107974.
6. Mohammed, Alaa G., Ehab Mohammed Shaker, and Shukur Abdulkareem Mahmood. "Risk Factors of Type 1 Diabetes Mellitus in Children and Adolescents: A Review." IAR Journal of Medical Sciences (2024).
7. Pelikan M. Hierarchical Bayesian optimization algorithm: toward a new generation of evolutionary algorithms. Berlin: Springer; 2005.
8. Pelikan M, Goldberg DE, Cantú-Paz E. BOA: the Bayesian optimization algorithm. In: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-1999); 1999; Orlando, Florida. San Francisco: Morgan Kaufmann; 1999. p. 525–32.
9. Kirkpatrick S, Gelatt CD, Vecchi MP. Optimization by simulated annealing. Science. 1983;220(4598):671–80.
10. Stranieri F, Fadda E, Stella F, Zanon M. Risk-averse supply chain management via robust reinforcement learning. Comput Chem Eng. 2024;186:108694.
11. Xu W, Brekke E, Klæboe G. Stochastic multi-objective optimization for flood control incorporating ensemble-based progressive hedging with K-means scenario clustering. Front Water. 2025;7:1606096.
12. Bertsekas DP. Dynamic programming and optimal control. 4th ed. Belmont: Athena Scientific; 2012.
13. Ghavamzadeh M, Mannor S, Pineau J, Tamar A. Multi-time scale Markov decision processes. IEEE Trans Autom Control. 2015;60(10):2687–701.
14. Zhong X, Lind M, Zhang L. System risk quantification and decision making support using functional modeling and dynamic Bayesian network. Reliab Eng Syst Saf. 2021;216:107974.
15. Manikas TW, Cain JT. Genetic algorithms vs. simulated annealing: a comparison of approaches for solving the circuit partitioning problem [Technical Report TR-96-101]. Pittsburgh: University of Pittsburgh, Department of Electrical Engineering; 1996.
16. Hvarfner C, Stoll D, Souza A, Lindauer M, Hutter F, Nardi L. πBO: augmenting acquisition functions with user beliefs for Bayesian optimization. In: Proceedings of the International Conference on Learning Representations (ICLR 2022); 2022 Apr 25–29; virtual. OpenReview; 2022.
17. Holland JH. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. 2nd ed. Cambridge: MIT Press; 1992.
18. Liu Y, Halev A, Liu X. Distributional constrained reinforcement learning for supply chain optimization [Preprint]. arXiv:2302.01727; 2023.
19. Peng Z, Zhang Y, Feng Y. Deep reinforcement learning for supply chain inventory management. Scalable Comput Pract Exp. 2023;24(4):1289–302.
20. Yan Y, Chow AHF, Ho CP, Kuo YH, Wu Q, Ying C. Reinforcement learning for logistics and supply chain management: methodologies, state of the art, and future opportunities. Transp Res Part E Logist Transp Rev. 2022;162:102712.
21. Tanabe K, Fukuda M. A stochastic differential equation framework for multi-objective optimization under uncertainty [Preprint]. arXiv:2510.10739; 2025.
22. Eberhardt A, Sherrill-Mix S. Multi-objective stochastic optimization problems: review and applications in energy systems. Appl Energy. 2025;374:124025.
23. Libreros JA, Llorca J, Mula J, Poler R. Monte Carlo simulation techniques in a group decision support system. In: Proceedings of the 2005 International Conference on Decision Support Systems; 2005; Lisbon, Portugal. Madrid: Universidad Politécnica de Madrid; 2008. p. 1–12.
24. Rubinstein RY, Kroese DP. Simulation and the Monte Carlo method. 3rd ed. Hoboken: Wiley; 2017.
25. Levy KJ, Shanbhag UV. Stochastic-constrained stochastic optimization with Markovian data. In: Advances in Neural Information Processing Systems (NeurIPS 2025); 2025; San Diego, California. Curran Associates; 2025.
26. Gawlikowski J, Tassi CRN, Ali M, Lee J, Humt M, Feng J, *et al.* A survey on uncertainty quantification methods for deep learning. IEEE Trans Neural Netw Learn Syst. 2023. doi:10.1109/TNNLS.2023.3399573.
27. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 33rd International Conference on Machine Learning (ICML 2016); 2016 Jun 19–24; New York. JMLR.org; 2016. p. 1050-9.
28. Alanazi A. A probabilistic decision model for AI-driven optimization in highly dynamic environments. J Theor Appl Inf Technol. 2023;101(14):5762–74.
29. Mahmood, Shukur, Majid Maatook, and Dhaigham Aatwan. "Risk Factors of Glycaemia Control among Children and Adolescents with Type 1 Diabetes in Basra City/Iraq." Proceedings of 2nd International Multi-Disciplinary Conference Theme: Integrated Sciences and Technologies, IMDCIST 2021, 7-9 September 2021, Sakarya, Turkey. 2022.
30. Dentcheva D, Ruszczyński A. Optimization with multivariate stochastic dominance constraints. Math Program. 2009;117(1–2):111–27.
31. Altman E. Constrained Markov decision processes. Boca

Raton: CRC Press; 1999.
32. Li S, Guo S. A genetic simulated annealing algorithm to optimize the small-world network generating process. Complexity. 2018;2018:1453898.
33. Moss HB, Leslie DS, Gonzalez J, Rayson P. BOSH: Bayesian optimization by sampling hierarchically. In: NeurIPS 2020 Workshop on Bayesian Deep Learning; 2020 Dec 11; virtual. arXiv:2007.00939; 2020.

How to Cite This Article

Dauood RF. Multilevel dynamic probabilistic modeling to support optimal decision-making via operations research and intelligent algorithms. International Journal of Applied Mathematics and Numerical Research. 2026;2(3):33-44. doi:10.54660/IJAMNR.2026.2.3.33-44.

Creative Commons (CC) License