# Analyzing Real Options and Flexibility in Engineering Systems Design using Decision Rules and Deep Reinforcement Learning

Cesare Caputo and Michel-Alexandre Cardin

*Dyson School of Design Engineering, Imperial College London*

**Abstract**. Engineering systems provide essential services to society e.g., power generation, transportation. Their performance, however, is directly affected by their ability to cope with uncertainty, especially given the realities of climate change and pandemics. Standard design methods often fail to recognize uncertainty in early conceptual activities, leading to rigid systems that are vulnerable to change. *Real Options* and *Flexibility in Design* are important paradigms to improve a system's ability to adapt and respond to unforeseen conditions. Existing approaches to analyze flexibility, however, do not leverage sufficiently recent developments in machine learning enabling deeper exploration of the computational design space. There is untapped potential for new solutions that are not readily accessible using existing methods. Here, a novel approach to analyze flexibility is proposed based on Deep Reinforcement Learning (DRL). It explores available datasets systematically and considers a wider range of adaptability strategies. The methodology is evaluated on an example waste-to-energy system. Low and high flexibility DRL models are compared against stochastically optimal inflexible and flexible solutions using decision rules. The results show highly dynamic solutions, with action space parametrized via artificial neural network. They show improved expected economic value up to 69% compared to previous solutions. Combining information from action space probability distributions along expert insights and risk tolerance helps make better decisions in real-world design and system operations. Out of sample testing shows that the policies are generalizable, but subject to tradeoffs between flexibility and inherent limitations of the learning process.

## 1 Introduction

Engineering systems fulfill highly important functions for society. They provide critical services to support human activity through power generation, transportation, supply chains for food production and delivery, water management, defense, healthcare, and telecommunications. Such systems operate for a long time, typically decades if not longer, and are exposed to important uncertainty and fluctuating conditions in their operational environment, but also in terms of markets, regulations, and technology. This uncertainty creates important risk and threats that may be amplified by high-level disruptions from ongoing climate change, pandemics, as well as cyber and physical terrorism. The 2021 blackouts in Texas, COVID crisis, and recurring attacks on oil and gas facilities (e.g., Pakistan, Yemen, Colombia, Iraq and Philippines recorded 989 attacks between 2010-2014) [1] show just how vulnerable these systems may be. Uncertainty also creates upside opportunities, for instance through the development of new markets, which open new economic prospects for the world. The development of GPS, smart phones, and CAFE emissions standards, to name a few, have pushed for cleaner and more sustainable mobility solutions and innovations over the last

 

decade. These in turn have contributed to the development of the new sharing economy [2], and created a whole new set of economic opportunities.

An important issue motivating this study is that current approaches to systems design and engineering may not lead to systems that are readily flexible and adaptable in the face of uncertainty – which is becoming an increasingly important consideration given current global conditions. For instance, the popular V-model in systems engineering [3] assumes that customer and requirements are known at the beginning of a project, and cascading design activities focus on enabling this particular view of the future. System designs are often optimized using deterministic assumptions for market, environmental conditions, regulations, and technologies, making them rapidly sub-optimal if operational conditions change. Industry often invests vast efforts to take advantage of economies of scale, thereby committing large resources upfront to reduce average production costs, while such resources may or may not be fully utilized in the future depending on market realizations. Such strategies lead to engineering systems that may be overly rigid, do not make best use of limited resources, and may be unable to quickly adapt and reconfigure following important disruptions. Such issues have given rise to a number of underperforming engineering systems – even important failures – in the face of uncertainty, despite well-functioning technologies e.g., Iridium satellite cell phone system [4], IUT Global waste-to-energy system in Singapore [5], Ghost Cities [6] in China, etc.

There is an exciting opportunity to take system design and engineering activities to the next level and contribute more systematically towards the development of more sustainable and resilient engineering systems in the future. In recent years, the field of *Flexibility in Design* has emerged as a concrete and systematic approach to enable better adaptability, reconfigurability, and evolvability in engineering systems [7]. Flexibility helps systems adapt and change in a cost-effective manner, considering arising uncertainty and risks. The field has evolved from *Real Options Analysis*, which quantifies the value of flexibility in irreversible investment projects [8]. Flexibility in Design helps developing and evaluating new computational tools to support early design activities, such as uncertainty modeling and quantification, creative system concept generation, design space exploration and optimization, and holistic process management [9]. The goal is to produce designs that reduce exposure to downside risks, while capitalizing on upside opportunities – essentially improving the tails of the performance distribution – thus resulting in higher expected performance and value overall. The design tools and processes help quantify expected performance to identify system configurations that provide better economic performance, sustainability, and resilience, in view of an uncertain future. Historical examples of flexible systems abound, such as the 25 the Abril bridge in Portugal, designed in the 1960s to accommodate additional transportation capacity through additional car lanes and railways, or the HCSC tower in Chicago, designed in the 1990s to accommodate vertical floor expansion in the future [10]. In both cases, flexibility was productively exercised to expand capacity when needed, such as when Portugal joined the European Union in the 1980s and saw booming economic activities requiring more transportation capacity. Flexibility was also beneficial when HCSC required more office space for its growing insurance activities and was able to nearly double the number of floors. What these two cases had in common is that flexibility had been carefully engineered in the design, and significantly reduced the costs of adaptation several years later.

A promising approach to analyze engineering systems for flexibility is to exploit a decision rule formulation. Decision rules are akin to *if-then-else* statements e.g., *if* demand exceeds a

certain threshold, *then* expand capacity, *else* do nothing. An important benefit from this formulation is that it combines both physical and managerial aspects into an elegant and succinct formulation, making it rather intuitive to use by system operators and decision-makers. It can be used with advanced techniques like stochastic programming or robust optimization to identify the best system design configurations. Decision rules act like signposts, or triggering mechanisms, that help operators identify the conditions when it is best to exercise flexibility in operations. It is important, however, to carefully enable the flexibility to be used in operations via decision rules in early design activities, from an engineering standpoint. Several studies have shown the benefits of such approach to flexibility analysis, as compared to standard real option methods based purely on dynamic programming [11-13]. One issue with the approach, however, is that decision rules are often generated by system designers based on a limited set of standard real option strategies e.g., expand or reduce capacity, stage capacity deployment, abandon a project doomed to fail, defer investment until favorable market conditions. They are elicited through designers' expertise with a system, using for instance creativity techniques like brainstorming or prompting [14], and may not completely explore the available design space. Thus, they may also leave potential value enhancements on the table.

While flexibility in engineering systems design leads demonstrably to highly valuable and performing flexible systems as compared to standard designs, there is still significantly untapped potential from exploiting recently developed techniques in machine learning and data science. For systems and mega-projects requiring large investments i.e., >$100 millions, the typical 10-30% improvement potential routinely seen in flexibility studies – often more – is non-negligible. Recent developments in data-driven fields can clearly help uncover new optimal flexible system design configurations, combinations of decision rules, and timings, that may complement strategies developed through human-led design activities and experience.

Deep Reinforcement Learning (DRL) presents several methodological similarities to prevailing state of the art methods in designing for flexibility. It is being increasingly investigated for potential applications in the field of engineering [15]. DRL involves formulating a sequential decision-making problem, with the objective of maximizing some reward, although with fewer limiting assumptions than with standard methods based purely on dynamic programming and Bellman's reward maximization equations [16]. A DRL formulation to flexibility analysis enables an agent (i.e., the system operator) to analyze different strategies through a heuristic trial-and-error learning process and *learn* the best actions at different points in time, based on statistical training on a range of potential scenarios.

Consequently, the proposed approach may help identify new dynamic adaptation strategies from the data, as opposed to relying solely on those elicited from standard real options and flexibility in design methods. It enables uncovering different combinations, timings, and resulting flexible designs that may not be considered explicitly by system designers, thereby expanding the design space exploration process.

Motivated by the above, this paper proposes a novel approach to analyze flexibility in system designs based on DRL. The approach is evaluated through a case study in sustainable waste-to-energy (WTE) system design. The study shows that a DRL approach helps generating new system configurations, combinations of decision rules and timings that may not be considered through current flexibility analysis tools. The approach produces highly dynamic solutions through learning valuable adaptation strategies from the data, while still exploiting the intuitive

nature of a decision rule formulation. The approach generates distributions of adaptation strategies that can in turn inform designers, decision-makers and operators on most widely used strategies, with the goal of designing systems that are more sustainable and resilient in the long term in the face of uncertainty.

# 2 Background and Related Work

## 2.1 From Real Options to Flexibility in Design

The development of Real Options – defined as the "right, but not the obligation, to change a system in the face of uncertainty" [8] – created the need to for new design methods and procedures to enable flexibility in engineering systems design. Flexibility in Design emerged in recent years as a field from such a need. Flexibility as a design concept, however, is not new, and has been studied extensively, for instance in manufacturing and product development [17, 18]. It has not been studied for as long in complex engineered systems. An important distinction between Real Options and Flexibility in Design is that the former focuses on quantifying the value of flexibility – effectively aiming to price real options – while the latter focuses on developing methods and procedures to embed flexibility in engineering systems design, as a value-enhancing mechanism. Both fields go hand in hand, since one enables quantification of the value of flexibility that is designed early on in each system. Flexibility in Design relies on value quantification in a similar fashion as in real options theory, but more as a mechanism to rank order possible design alternatives to support the decision-making process, and less to find the right "price" for the real options. It adapts the theory of real options to accommodate the needs of industry practice.

Over time several frameworks have been proposed to support the design process [19, 20], and reviews to organize the field of Flexibility in Design [21, 22]. More recently, Cardin [23] proposed a holistic five-phase framework to support such activities, also encompassing a number of tools and procedures to support design activities in each phase. The phases involve: starting from 1) an initial design establishing baseline performance, then 2) recognizing and modelling uncertainty affecting the system performance as to stimulate creativity, moving on to 3) generating performance-enhancing system design concepts leveraging flexibility, followed by 4) design space exploration and optimization. The framework is embedded within a holistic phase 5) to enable seamlessly the design process among relevant stakeholders. The proposed approach contributes mostly to phase 4 by providing a new approach to explore the design space.

## 2.2 Decision Rule Formulation

Standard approaches to quantify the value of flexibility rely on decision analysis and binomial lattice analysis, and simulations [24]. The expected value of flexibility is quantified as the difference between the expected payoffs from baseline and flexible designs. Decision analysis relies on decision trees and a backward induction process from dynamic programming [16]. The decisions available at each stage represent how the system can adapt. Binomial lattice analysis is similar to decision analysis, with the exception that in each stage the uncertainty can either go up or down relative to the previous state [25]. To reduce the number of possible outcomes, path independence is assumed, and lattice nodes can recombine.

One issue with the above approaches is that the decision to exercise flexibility is difficult to use in real-world operations. Among others, the folding back process in dynamic programming (i.e. starting from the end and folding back to initial time) is often criticized as not intuitive and

difficult to implement in practice [11]. The decision rule is essentially based on Bellman's expected reward maximization, and therefore does not provide much flexibility in terms of implementing other decision rules. It may also prove challenging in operations to identify the corresponding system state in a decision tree or lattice and determine the next best decision.

To address these issues, a decision rules approach to flexibility analysis was proposed recently, and slowly taking precedence in the field [11, 26]. Flexibility Decision Rules (FDR) are typically employed in stochastic programming and robust optimization to alleviate intractability in the computational problems created. They include condition-go (akin to the *if-then-else* formulation), zero order, linear, and safety-first rules [27]. In the context of flexibility analysis, decision rules are not just mathematical tricks, they take on a full managerial meaning, giving intuitive policies as to when flexibility should be exercised, and how. An example FDR formulation may be that "*if* demand reaches threshold $\alpha$, *then* expand capacity by $\beta$, *else* do nothing". Given that the decision may be evaluated a regular time intervals, the problem can be modeled as a multi-stage stochastic programming or robust optimization problem, to identify the best values for decision rule variable $\alpha$ and design variable $\beta$ over the problem time horizon [26]. This formulation leads to intuitive policies readily applicable in operations and connects tightly what is done to the physical design to enable the flexibility early on to how it is managed in operations. The approach has been shown previously to quantify the value of flexibility in a similar manner as standard real option methods [11].

A decision rule formulation is embedded within the definition of a *flexible system design concept*, which comprises both a *strategy* and an *enabler* [23]. During early conceptual design activities, designers can choose a strategy to deal flexibly with certain uncertainty sources (e.g., price, demand) by considering standard real options (e.g., abandonment, expansion, etc.). They must also identify how to enable the flexibility in the early design phases, which includes considering both physical design aspects and a policy to determine when it is best to exercise the flexibility during operations (i.e., the decision rule). For example, engineers of the Lisbon bridge provided for additional strength to support a possible railway and car lane expansion, and the decision rule to expand (even though not documented) was certainly connected to new capacity needs emerging from increasing economic activities. The HCSC building was designed in the 1990s with stronger foundations, additional elevator shafts and a stronger structure overall to support additional floors, which was needed when expansion was implemented. These examples show that there is no "cookie-cutter" solution to design flexible systems. Different systems face different uncertainty sources and require different design configurations. Hence, a combination of creativity, guidance, and expertise is needed to identify the best strategies and enablers to embed valuable flexibility in a particular system.

### 2.3 Deep Reinforcement Learning

A decision-rule formulation is closely related to Markov-Decision Processes (MDP), which in turn is closely related to the mathematics of DRL. In a MDP formulation, an agent operates in a space characterized by current and future states, and actions connecting the different states. This formulation is very similar to a decision rule approach in flexibility analysis, since the *if-then-else* formulation needs first to determine the current state (i.e., observation of uncertain variable like demand or price) and take on certain actions (e.g., expand capacity, shut down operations) that will lead to a future state, with the goal of maximizing a certain discounted reward (e.g., cumulative profits) over time. MDPs can develop non-stationary, long term policies to operate in a complex environment, providing strong global convergence guarantees.

In their conventional form, however, they are only well suited for low dimensionality action and state spaces, as Markovian transition matrices can scale exponentially, with significant degradations in computational feasibility of policy evaluation at each decision step [28, 29].

DRL is a Markov-type machine learning approach with potential to alleviate some of these dimensionality limitations. It involves an agent interacting with an environment over time as part of a sequential decision-making problem, with the objective of maximizing a reward signal. Generally, at each time step ($t$), the agent ($A$) finds itself at state ($s_t$), and selects an action ($a_t$) following a reward maximization policy $\pi(a_t \mid s_t)$ – see Figure 1. As a result of taking the action, the agent transitions to a successive state $(s_{t+1})$, receiving a corresponding scalar reward ($r_t$), which are governed by the environment's dynamics or reward function, $R(s, a)$, and state transition probability, $P(s_{t+1}|s_t,a_t)$, respectively [28]. This repeated interaction is captured graphically in Figure 1, including an illustration of the role of neural networks within the process. The neural network approximates the reward value function learned during training, and may consist of several layers – hence the name "deep" [30]. In episodic environments, this continues until the agent reaches a user defined terminal state, at which point it restarts, although continuous environments are also present in several applications (i.e., operating a building demand response system). In training, the agent is fed a wide range of scenarios where they apply policy $\pi$ through a heuristic trial-and-error mechanism. Through several interactions with the environment, the agent seeks to develop an optimal policy $\pi^*$ to map and rank state-action pairs based on expected value, as captured in Eq. (1), where the objective is the maximization of discounted accumulated rewards per episode $\sum_t \gamma^t R_t$:

$$\pi^* = \arg max_\pi \, \mathbb{E}_{\tau\sim\pi(\tau)} \left[\sum_t R(s_t\,,\, a_t)\right] = \arg max_\pi \, \mathbb{E}_{\tau\sim\pi(\tau)} \sum_t \gamma^t R_t \tag{1}$$

The two primary approaches used to optimize Eq. (1) in DRL are value-based and policy gradient methods, originating from value and policy iteration in dynamic programming. In value-based methods, such as $Q$-learning, the objective is to improve iteratively the value estimate of each state action pair until it converges. The RL agent then follows the trajectory of highest $Q$ values to formulate the optimal policy, although it is not explicitly optimized. This is usually performed off-policy, such that each update can use data collected throughout training, even if collected using an old policy. In policy gradient methods, rather than updating the value function, the policy is parametrized directly. This involves on-policy updates and the use of a value function approximator to guide the policy update direction, typically done at the end of each episode [28]. In modern DRL, the distinction is not always clear as state of the art algorithms try to combine both approaches to yield performance improvements. Methods such as DQN, however, are closer to value-based approaches, A2C and TRPO to policy-gradient ones and others such as SAC, DDPG and TD3 combining some aspects of both [31]. Depending on the problem, the choice of algorithm can vary significantly.

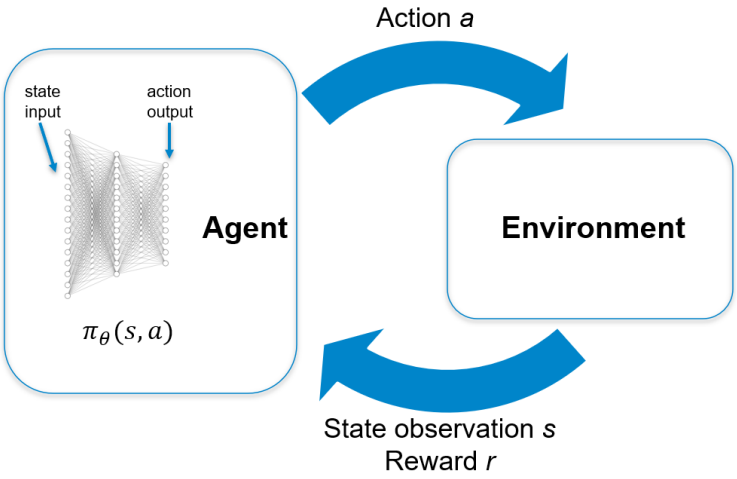


Figure 1: Graphical Overview of DRL

The majority of DRL modeling in engineering has thus far centered on more classical control problems which fit within a MDP, and stimulate agent learning [32]. A general framework for design optimization using DRL was proposed in [33] and was applied to the air foil angle of attack for a hypothetical stator with varying rotor designs. The work in [34] also shows the strong performance of DRL in the design task of fatigue resistant solutions for ship design compared to evolutionary methods. Both implementations, however, maintain limited scope within the overall problem, with overly simplistic reward functions or action spaces. There are several studies focusing on new algorithms or techniques to allow combination with more complex problems, but they tend to focus mostly on the mathematical or algorithmic side [28, 35-37]. More promising is the optimization performed in [38], where off the shelf DRL algorithms are implemented to design a microfluidic device for flow sculpting. Using a Double $Q$-Network architecture combined with Hindsight Experience Replay, the agent produces designs that perform consistently in the 99$^{th}$ percentile of the entire design space, compared to those obtained via much more computationally expensive approaches.

The potential for DRL to the objectives of this paper was perhaps most thoroughly investigated in [39], where it was first implemented to simultaneously solve the competing energy systems basic design and dispatch strategy. The authors, however, did not consider any decision-making flexibility through the project duration to adjust capacities over time based on realized uncertainties. In fact, to these authors' knowledge, there has been no previous study exploiting the potential of DRL in the context of enabling flexibility in engineering systems design.

# 3 Methodology

## 3.1 Generic Capacity Planning Model

This section introduces the generic capacity stochastic planning model used for the system case study in Section 4. Consider a planning horizon of $T$ time periods, and let $\theta$ denote the installed capacity, $i$ the discount rate and $d^t$ the demand in period $t$. Let $\xi = (\xi_1, \xi_2, \dots, \xi_t)$ be a scenario of uncertainty, where $\xi_t$ is a vector capturing the uncertainty observed in period $t$, thus able to account for multiple uncertainty sources. For example, $\xi^t$ may be used to model stochastic demand $(d^t)$ realizations in period $t$, as done later in the analysis. $S$ defines the set of all possible uncertainty scenarios $s$, assumed to be finite for simplicity, with corresponding

probabilities for each scenario $p_s > 0$ , $\sum_{s=1}^{S} p_s = 1$. The stochastic capacity planning model with no flexibility – referred as benchmark design – is formulated as in Cardin and Hu [40]:

$$Max\ ENPV = \sum_{s=1}^{S} p_s \left(-\mathcal{C}^0(\theta) + \sum_{t=1}^{T} \left(\frac{1}{1+\lambda}\right)^t [\mathcal{R}_s^t(\theta, \xi_s^t) - \mathcal{C}_s^t(\theta, \xi_s^t)]\right) \quad (2)$$

$$s.t.\ \ 0 \leq \theta \leq \theta_{max} \quad (3)$$

$$\mathcal{C}^0(\theta) = K(\theta)^{\alpha} \quad (4)$$

$$\xi_s^t \geq 0, \forall\, s, t \quad (5)$$

Variable $\theta_{max}$ captures the upper bound of $\theta$, $\mathcal{R}^t(\theta, \xi^t)$ is the revenue function for period *t*, $\mathcal{C}^t(\theta, \xi^t)$ is the cost function for period *t*, and $\lambda$ is the discount rate. Eq. (2) determines optimal rigid capacity under uncertainty by maximizing expected net present value (ENPV), considering initial capital costs (CAPEX) and annual net cash flows produced by the system. This metric is chosen because it captures both costs and benefits under the same performance attribute i.e., financial value. The constraints on maximum attainable capacity ($\theta_{max}$) are given by Eq. (3). Initial investment is estimated from Eq. (4), which is a power cost function accounting for economies of scale (EoS), with $K$ a constant coefficient parameter, and $\alpha$ the EoS factor. The magnitude of the EoS effect (as measured through $K$ and $\alpha$) can be estimated through statistical analysis of historical data on systems costs at different capacity levels (see Table A1 for assumptions). Alternatively, domain expertise combined with trends and cost data for similar engineering systems can be used.

### 3.2 Capacity Expansion Model with Decision Rules

Decision rules are embedded in the system model to act as triggering signals for exercising a particular flexibility in operations. Based on the information revealed in each time step, decision rules enable decision-makers to sequentially decide when to proceed with capacity expansion, the main flexibility strategy considered in this paper. The decision-maker must decide on capacity deployment ($x_t$) as part of a set of feasible decisions $X_t$ for period $t \geq 0$, including a zero-stage decision on initial capacity, before any uncertainty is revealed.

Let $X \subseteq X_1 \times \cdots \times X_t$ denote the set of all feasible capacity decision sequences *x*, where $x = (\theta^0, \theta^1, .., \theta^t)$, and let $\xi^{[t]}$ represent the full history of the uncertain variable up to time $t$. A decision rule, also known as an implementable policy, $\mathcal{F}$, is a function that maps each scenario of uncertainty $\xi$ into a sequence of decisions *x* in *X* (i.e., $\mathcal{F}: S \rightarrow X$. $\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right)$ as shown in Eq. (*8*). The capacity decision made in each period $t$ for scenario $s$ (i.e., $\theta^t{}_s$) can be found from $\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right)$. The form of $\mathcal{F}$ varies for different problems, and a vector of parameters $\emptyset$ is used to characterize it, with the decision rule represented here as $\mathcal{F}_{\emptyset}$. Accounting for non-anticipativity constraints, the objective is to formulate a feasible decision rule to maximize total expected reward, based on a series of reward functions $(r_t \mathcal{F}_{\emptyset} \xi^{[t]}, \xi_t)$ with capacity decisions $(x_t)$ and uncertainty revealed $(\xi^{[t]})$ as inputs. The generic stochastic model with decision rules used to analyze flexibility in design is formulated as:

$$Max\ \ ENPV = \sum_{s=1}^{S} p_s( - \mathcal{C}^0(\theta_s^0) + \sum_{t=1}^{T} \left(\frac{1}{1+\lambda}\right)^t [\mathcal{R}_s^t \left(\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right), \xi_s^t\right) - \mathcal{C}_s^t \left(\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right), \xi_s^t\right) - \mathcal{H}_s^t(\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right), \theta_s^{t-1})]) \quad (6)$$

s.t. Eq. (3) – Eq. (5) and

$$\theta_s^t \in \Theta^t \quad \forall t, s \quad (7)$$

$$\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right) = \theta_s^t \quad \forall s, t = 1, 2, \cdots T \quad (8)$$

Here function $\mathcal{H}_s^t(\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right), \theta_s^{t-1})$ determines the costs associated with each capacity expansion decision and can vary depending on the problem. The general formulation is given by Eq. (*9*):

$$\mathcal{H}_s^t(\mathcal{F}_s^t \left(\xi_s^{[t]}, \theta_s^{t-1}\right), \theta_s^{t-1}) = \begin{cases} K(\theta_s^t - \theta_s^{t-1})^\alpha & if \quad \theta_s^t > \theta_s^{t-1} \\ 0 & otherwise \end{cases}, \forall s, t = 1, \dots T \quad (9)$$

The value of flexibility (VoF) is estimated as the difference in expected value between a flexible model and an inflexible benchmark, as captured in Eq. (*10*). It represents the maximum that a decision-maker should be willing to pay to embed flexibility in the system design.

$$\text{VoF} = ENPV_{Flexible} - ENPV_{Benchmark} \quad (10)$$

In Section 4, two flexible systems are considered: one using a decision rule formulation, the other using a deep reinforcement learning formulation. Both flexible models are developed based on generic assumptions captured in the capacity expansion model with decision rules described above.

## 4 Application Study

This section presents an example application of the proposed methodology to the analysis of a WTE system design in Singapore. The study analyzes tensions between two common approaches to systems design under uncertainty: 1) a centralized approach leveraging economies of scale, which is economically sound when EoS are strong under a set of predetermined scenarios, but requires significant capacity deployment early on, and; 2) a decentralized modular approach, which does not benefit as much from EoS, but helps managing uncertainty by deploying resources flexibly when needed, thereby extracting value from flexibility. The problem formulation is generic and applicable across a range of similar, distributed engineering systems, such as mini-grids, or other renewable energy systems. The study objectives are to show that the proposed approach to flexibility analysis based DRL may:

(a) Assist in the design of real-world engineering systems under uncertainty

(b) Improve expected value of projects compared to an inflexible baseline

(c) Enhance the value of flexibility as compared to alternative flexible solutions

(d) Yield dynamic and adaptive decision rules as compared to other flexible solutions, and

(e) Provide access to more flexible design solutions and decision rules as compared to existing approaches.

The benchmark system is a centralized large capacity WTE design located in the western area of Singapore and processing food waste produced by the city-state. Food waste must be transported to the centralized location from each of six sectors comprised under the Public Waste Collection (PWC) scheme. The flexible systems are decentralized WTE designs that build a small-scale plant in each of the waste collection sectors. They differ in how the capacity is deployed over time and space, governed by either Flexible Decision Rules (FDR), or DRL dynamics. PWC workers can choose to transport the collected waste to a sector's Anaerobic Digestion (AD) plant rather than the main centralized one. Both types of systems receive revenues from the production of electricity and refuse collection, while the primary costs are associated with capacity expansion, waste transport, disposal, maintenance, and land rent. The decentralized designs allow a significant reduction in transportation and collection costs but see increased average cost per unit processed due to limited EoS and modularity. The modularity, however, contributes to reducing exposure to downside conditions (e.g., lower amount of waste produced than expected in any given sector).

### 4.1 Model Development

The annual revenues associated with AD energy system operations are calculated as:

$$AR^t = R_D^t + R_E^t \tag{11}$$

$$R_D^t = d^t P_{to} \tag{12}$$

$$R_E^t = \min(d^t, \theta) E_g P_e \tag{13}$$

$$d^t = \sum_{i=1}^{6} d_i^t \tag{14}$$

where $d_i^t$ is the recycled sector $i$ food waste at year $t$ and $d^t$ the total recycled food waste in the city. The central site capacity is given by $\theta$. The food waste collection income is determined from tipping fee $P_{to}$ and additional revenues are found from the electricity generation rate $E_g$ and the unit selling price of electricity $P_e$. The annual system cost at year $t$ ($AC^t$) is instead determined from summation of the transportation cost $C_{Tran}^t$, disposal cost for waste residue $C_D^t$, the land cost $C_L^t$, and operational and maintenance (O&M) cost $C_{OM}^t$ as shown in Eq. (15). Eqs. (16) – (21) break down how each of these components is calculated.

$$AC^t = C_{Tran}^t + C_D^t + C_L^t + C_{OM}^t \tag{15}$$

$$C_{Tran_{centralised}}^t = C_{fuel} \sum_{i=1}^{6} \frac{d_i^t}{Cap_v} (D_{co\,i} + D_{Tr_i}) \tag{16}$$

$$C_{Tran_{DecentralizedFlex}}^t = C_{fuel} \sum_{i=1}^{6} \frac{d_i^t}{Cap_v} D_{co\,i} + C_{fuel} \sum_{i=1}^{6} \frac{d_i^t - \theta_i^t}{Cap_v} D_{tr\,i} \tag{17}$$

$$C_D^t = (min(d^t, \theta)\,(1 - \omega + \varepsilon) + max\,(d^t - \theta, 0))C_{dis} \quad (18)$$

$$C_L^t = \rho\theta \quad (19)$$

$$C_{OM}^t = \pi\mathcal{C}^0(\theta) = \pi[K(\theta)^\alpha] \quad (20)$$

$$\theta_i = \theta/6 \qquad i=1, 2,\cdots, 6 \quad (21)$$

For the centralized design, total transportation cost is computed from the sum of waste collection cost within each site, as well as transporting waste from each site to the central plant according to Eq. (*16*). $D_{co_i}$ is intra sector $i$ collection distance, and $D_{Tr_i}$ is the distance for collection from sector $i$ to the main site. $C_{fuel}$ is the unit cost for fuel consumption. In the decentralized system, transportation costs are found by first estimating excess untreated waste in each sector $i$ as $(d_i^t - \theta_i^t)$, and resulting number of trips required to transport the waste volume $\frac{(d_i^t - \theta_i^t)}{Cap_v}$ to yield total collection truck travel distance as shown by Eq. (17). The cost of disposal $C_D^t$ is determined as a function of the purity rate $\omega$, residue rate $\varepsilon$ and the unprocessed waste due to system wide capacity shortages for year $t$ from Eq. (*18*). All waste that must be disposed of is assumed to be incinerated at unit cost $C_{dis}$. Land rental cost $C_L^t$ is calculated based on installed capacity levels $\theta$ at year $t$, and the unit land rental fee $\rho$. The full list of assumptions and parameters can be found in Appendix.

Recycled food waste (FW) – also referred as the *demand* for food waste processing ( $d^t$)– is simulated using standard Geometric Brownian Motion (GBM), representing the overall increasing trend in food waste production, while accounting for random shocks and volatility:

$$d\xi_i^t = \mu\xi_i^t dt + \sigma\xi_i^t dW_t \qquad t = 1,2,\cdots T, i = 1,2,\cdots 6 \quad (22)$$

$$\xi^t = \sum_{i=1}^{6} \xi_i^t \qquad t = 1,2,\cdots T, i = 1,2,\cdots 6 \quad (23)$$

In Eqs. (22)-(23), variable $\xi_i^t$ captures the realized stochastic recycled food waste in sector $i$ ($S_i$) at year $t$ and $\xi^t$ is the total recycled food waste at year $t$. The mean growth rate $\mu$ and volatility $\sigma$ are estimated from historical food waste generation patterns in Singapore ($\mu$ = 12.3%, $\sigma$ = 16.3%) using standard statistical regression. Historical data is used for building the stochastic model, and enable direct comparison with results in Cardin and Hu [40]. The random variable $dW_t$ captures the Wiener process, modeling the stochastic error at time $t$, sampled from a standard normal distribution. $\xi^0$ (Or $d^0$) is the total recycled food waste as of 2013 in Singapore ($\xi^0 = d^0 = 274$ tonnes per day, or tpd), which is also the assumed starting point for the study. Using these assumptions, 2000 i.i.d scenarios are generated over the 15-year time horizon of the project, showing evolution of recycled food waste in each sector.

### 4.2 Flexible Decision Rule (FDR) Model

The capacity planning model for the flexible decentralized design under uncertainty can be formulated as in Eq. (24):

$$Max\ ENPV = \sum_{s=1}^{2000} p_s\Big(-\sum_{i=1}^{6}\mathcal{C}^0(\theta_{is}^0)(1+C_f) + \sum_{t=1}^{T}\sum_{i=1}^{6}\frac{AR_{is}^t(\theta_{is}^t,\xi_{is}^t) - AC_{is}^t(\theta_{is}^t,\xi_{is}^t) - C_{\tilde{e}xp_{is}}^t}{(1+\lambda)^t}\Big) \quad (24)$$

where $\xi_s^t$ is the total recycled food waste in scenario $s$ at year $t$, $\xi_{is}^t$ is the recycled waste in sector $i$ in scenario $s$ at year $t$ , $C_{exp}^t$ is the capacity expansion cost in year $t$ and $p_s$ is the probability of scenario $s$. The ENPV is maximized with respect to the initial capacity ($\theta_s^0$), the capacity expansion threshold ($\beta$), the decentralized threshold ($\tau$), and the number of expanded modular designs ($\gamma$). Capacity expansion decisions are guided by a set of *if-then-else* statement, with decision-making logic summarized in Figure 2. Procedure 1 determines first whether there is a need to expand capacity at a system level based on overall demand and installed capacity. Procedure 2 determines whether the expansion should proceed in a decentralized manner, or at the main site only (i.e., sector 1). If a decentralized expansion is chosen, Procedure 3 determines in which sector to add capacity, up to the maximum capacity $\theta_{max}$.

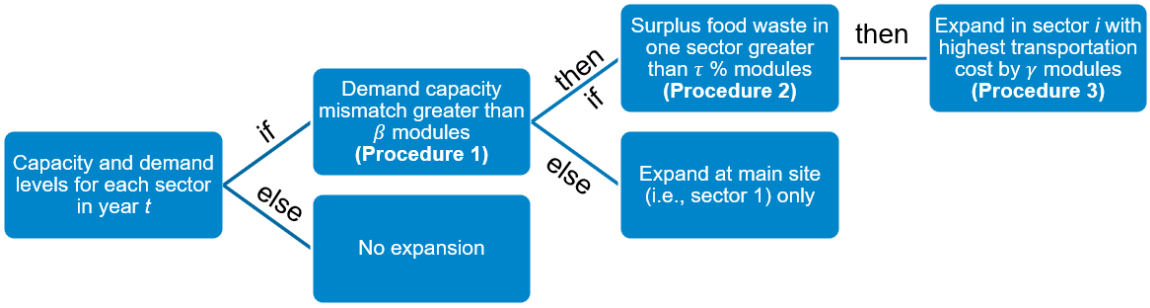


Figure 2: Decision tree for FDR approach on Decentralized Flexible WTE design

### 4.3 Deep Reinforcement Learning (DRL) Models

#### 4.3.1 Creating the representative environment

The Open AI gym format environment [31] is used to analyze two versions of the flexible WTE system using DRL. An initial model referred as DRL-LF (i.e., low flexibility) is developed to ensure performance is benchmarked against standard methods appropriately, where capacity expansion decisions are restricted to the same magnitude as for the FDR model (i.e., 200 tpd per expansion). The reasoning is to evaluate how dynamic the agent's decision-making process is when limited to the same actions as for the FDR model, to see what insights may be gained on optimal exercise time. The discrete action set for the DRL-LF agent then becomes:

- if action is = 0 → do not expand
- if action is = $a$ → expand capacity by 200 tpd in sector $i$

The states are defined through a box observation set as:

$$[\theta_1{}^t\ ,\ \xi_1{}^t\ ,\theta_2{}^t\ ,\ \xi_2{}^t,\theta_3{}^t\ ,\ \xi_3{}^t,\theta_4{}^t\ ,\ \xi_4{}^t,\theta_5{}^t\ ,\ \xi_5{}^t,\theta_6{}^t\ ,\ \xi_6{}^t]$$

where $\theta_i{}^t$ and $\xi_i{}^t$ represent the installed capacity and realized stochastic demand in sector $i$ for year $t$, respectively, as in the above sections. . The capacities within the observation are normalized by the minimum (0) and maximum (600 tpd) values for each site, while the upper bound on demand for all sectors is found from simulating 10,000 scenarios (independent of the environment) and finding the highest demand value in each sector to normalize those state variables. The starting capacity decision is left up to the agent in this initial implementation, although the magnitude is kept equal to that found with the FDR model, with the decision based on which sector to construct at $t$ = 0. Time awareness for the agent is implemented to help

address convergence stability and to produce better value approximations for near terminal states [41]. This approach does not violate non-anticipativity as project expected duration should be part of the description of the environment, and MDP dynamics. The reward function for each time step from Eq. 1 ($R_t$) is defined as the $NCF_t$ to reflect decision makers preferences more closely, also yielding a discounted reward policy of maximizing ENPV, according to Eq. (25):

$$R_t = NCF_t = \sum_{i=1}^{6} \frac{AR_{is}^t(\theta_{is}^t, \xi_{is}^t) - AC_{is}^t(\theta_{is}^t, \xi_{is}^t) - C_{exp}{}_{is}^t}{(1+\lambda)^t} \tag{25}$$

Building upon this initial implementation, a more complex DRL model is introduced allowing for greater flexibility (i.e., DRL-HF for high flexibility). The added complexity stems from the fact that the magnitude of capacity expansion decisions is not preset at 200 tpd. Rather, the agent can explore several different expansion levels. This adds much more dynamic decision making, subsequent systems design implications and can present a more holistic and complete overview of the potential added value for this methodology. In this environment, the observation space is kept the same, but the action space is changed to multi-discrete decisions such that:

- if action is = 0 → do not expand
- if action is = $a_j$ → expand capacity by $a_j$ in sector $i$

where $\boldsymbol{a} = [50, 100, 200]$ respectively for each sector and only one action allowed to be selected in each time step. The capacity increments are in discrete values of 50 tpd based on the AD technology considered (see Appendix). It should be noted that for DRL-HF, the original constraint where max capacity had to be limited to that of the benchmark design was relaxed to evaluate more decision-making possibilities. A constraint on max sector capacity of 600 tpd per sector is implemented to ensure benchmarking with a comparable system. This means, however, that the results should be taken more as a measure of a relative untapped potential for flexibility in the system given the original constraints were significantly relaxed.

#### 4.3.2 Algorithmic Approach

A policy gradient (PG) method is used based on the high dimensionality in action (7 and 19 possible actions for DRL-LF and DRL-HF models, respectively) and observation spaces (12 state variables). Value based methods typically struggle in this setting as the number of potential state-action pairs grows exponentially with action and observation space size, creating very large computational and memory requirements. While policy gradient methods may be less sample efficient, thus normally requiring more simulated episodes to learn a usable policy, this is not an issue here since design simulations are relatively inexpensive. In a highly stochastic environment such as here, PG methods can be advantageous as they output probability distributions over different actions to represent uncertain agent dynamics via a stochastic policy. Initially, this stochastic policy presents a high degree of randomness to allow agent exploration of different potential states and build a more complete representation of the environment dynamics. Over the course of training, this distribution converges to a deterministic policy and suggested action for each state, as the agent exploits rewards it has already experienced based on the policy update rules defined. This results in better convergence stability as well as a more generalizable policy than obtained via value methods. This class of

DRL methods also helps addressing perceptual aliasing, whereby it can better differentiate between very similar states to yield the optimal action probability distribution [32].

PG methods, however, do present a few technical issues. Excessive policy changes can hinder training, as it may be difficult to map changes between policy and parameter spaces directly. Even very small differences in parameters can cause significant variations in performance, and a single overly large "incorrect step" for the policy update can thus lead to great errors in approximated action values. Improper learning rate may anneal or overinflate the gradient, which may lead to policies being stuck in local optima. Trust Region Policy Optimization (TRPO) is an iterative method developed to address some of these limitations [42]. It uses a combination of the Minorize-Maximization (MM) algorithm and a "trust region" for policy updates. The MM algorithm approximates a lower bound for the received reward to guarantee monotonic improvements in policy, such that with each iteration it either improves or stays the same, theoretically converging to the optimal policy. This lower bound, also known as surrogate advantage, is calculated using Eq. (26):

$$\mathcal{L}\ (\varphi_k, \varphi\ ) = \frac{1}{1-\gamma}\ E_{s\ ,a\ \sim\ \pi_{\varphi_k}}\ \left[\frac{\pi_\varphi\ (a|s)}{\pi_{\varphi_k}(a|s)}\ A^{\pi_\varphi}(s,a)\right] \tag{26}$$

Here $\pi_\varphi$ denotes the updated policy and $\pi_{\varphi_k}$ the old policy used to estimate improvements. Generalized Advantage Estimation (GAE) is an improvement on these estimators that uses an exponentially weighted average of the *k*-step estimators [43]:

$$\widehat{A_t}^{GAE(\gamma,\lambda)}\ = (1-\lambda)\left(\widehat{A_t}^1 + \lambda\widehat{A_t^2} +\ \lambda^2\widehat{A_t^3} + \cdots\right) = \sum_{t=0}^{\infty}(\gamma\ ,\lambda)^l\delta_{t+1}^V \tag{27}$$

The trust region adjusts the step size for policy updates based on the objective function curvature and differences with the approximated one. This ensures the magnitude of policy updates is proportional to the confidence level within the trusted region. Anything outside this trusted region, where the approximation error is deemed too large, fails the improvement guarantee, and thus is not considered. To quantify the difference among the old and new policies for states already visited, restricting the size of the update step based on the trust region, the Kullback–Leibler (KL) divergence is calculated according to Eq. (28):

$$D_{KL}(\varphi||\varphi_k) =\ E_{s\sim\pi_{\varphi_k}}[D_{KL}(\pi_\varphi(\cdot\ |s)|\ \Big|\pi_{\varphi_k}(\cdot\ |s)\Big)] \tag{28}$$

Combining these equations yields the theoretical TRPO update according to Eq. (29):

$$\begin{gathered}\varphi_{k+1} = arg\ max_\varphi\ \mathcal{L}\ (\varphi_k, \varphi\ ) \\ \text{s.t. } D_{KL}(\varphi||\varphi_k) \le \delta\end{gathered} \tag{29}$$

where δ represents the KL-divergence limit. TRPO introduces a backtracking line search to the update rule, which addresses approximation errors introduced via Taylor series expansion required from Lagrangian duality theory [42]. The full pseudocode for the TRPO implementation is shown in Algorithm 1 below.

| | |
|---|---|
| **Algorithm 1:** Trust Region Policy Optimization (TRPO) with generalized advantage estimation (GAE) applied to WTE case study | |
| 1: | Initialize: Starting policy parameters $\varphi_0$, value function parameters $\emptyset_0$ , KL divergence limit $\delta$ , backtracking coefficient $\alpha$ , maximum number backtracking steps $K$ |
| 2: | **for** k = 0, 1, 2, … **do** |
| 3: | Generate stochastic demand scenarios ${\xi_k}^i$ according to Eqs. (22)-(23) |
| 4: | Collect set of trajectories $D_k$ by running policy $\pi_k$ $(\varphi_k)$ on the environment including decision for starting capacities in each sector at $t = 0$ |
| 5: | Estimate advantage ${A_t}^{\pi_k}$ using Eq. (*27*) based on current value function $V_{\emptyset_k}$ |
| 6: | Formulate estimates for policy gradient $\widehat{g_k}$ average KL-divergence Hessian-vector $\widehat{H_k}$ product function based on current sample |
| 7: | Use conjugate gradient algorithm to compute divergence gradient and trust region |
| 8: | Compute proposed policy change step $\Delta_k$ |
| 9: | Update the policy by backtracking line search as $\varphi_{k+1} = \varphi_k + \Delta_k$ |
| 10: | Refit value function based on observed trajectories and policy |
| 11: | **end for** |
| 12: | |

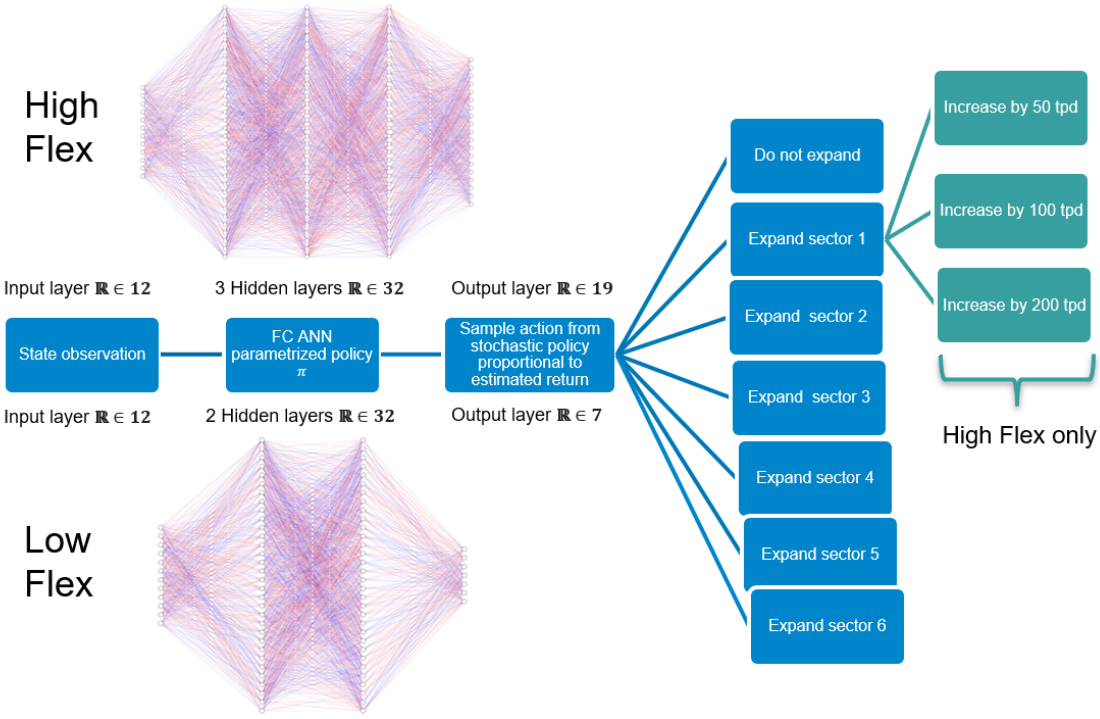


Figure 3: DRL decision process for both the low flex (DRL-LF) and high flex (DRL-HF) implementations. Magnitude of expansion limited to 200tpd in blue boxes for low flex implementation, no other expansion levels examined.

Figure 3 gives an overview of the logical progression of the proposed approach following Algorithm 1, for both the DRL-LF and DRL-HF models. It compares and contrasts with the logic depicted for the FDR model in Figure 2. The state observation in any given year becomes an input into the feedforward Artificial Neural Network (ANN) parametrized policy, where an action is sampled from the underling probability distribution. One additional hidden layer is

implemented in DRL-HF compared to DRL-LF to account for added action-state complexity, with output layers matching action space size.

### 4.3.3 DRL Agent Training and Testing

Given the objective to yield a generalizable policy for agent-environment interactions, i.i.d. stochastic scenarios for FW processing demand are generated, as seen in line 3 of Algorithm 1. This leads to an infinite number of potential state action pairs, further justifying the use of a PG approach. DRL agent training is executed for 1.5 Million timesteps, equivalent to 100,000 simulated 15-year episodes for the agent, using $\boldsymbol{\delta}$ = 0.01, $\boldsymbol{\alpha}$ = 0.8 and maximum number of backtracking steps $\boldsymbol{K}$ = 10. Training process is fully non-anticipative, in that there is no look ahead bias for the agent, and operational decisions are purely based on state observations at the that time step. The policy built inside the ANN, therefore, while learning from past mistakes and interactions, is not based on any information on future states, as the i.i.d. scenario generation method ensures no trajectory inside the environment is the same. Furthermore, given the observation in the first few training episodes that the agent would try to build past the maximum capacity to capitalize on potential rewards, an imaginary reward penalty of –S$1 is given in time steps where the capacity is at $\theta_{max}$, and the selected action is not equal to 0. This is a common approach to deal with constraints in DRL, while balancing agent exploration tradeoffs, as further discussed in Section 6.

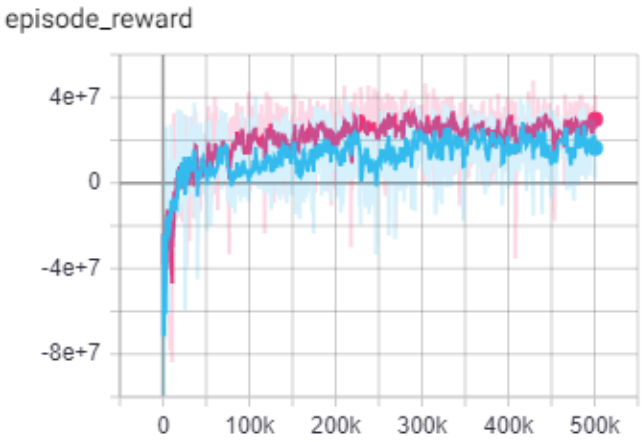


Figure 4: Training reward evolution for two different initializations of TRPO. Line smoothing factor of 0.8 applied to increase visual clarity.

The penalty is removed during the testing phase once the DRL agent has learned the form of the constraints implicitly through training. All benchmarking with results in Cardin and Hu [40] is conducted using the same sampling functions generating 2000 i.i.d. scenarios. To do this, the agent-environment interaction framework is slightly modified as successive state observations are explicitly defined for the ${\xi_i}^t$ parameters in the observation array defined in Section 4.3.1 , while the ${\theta_i}^t$ parameters are determined based on agent actions and expansion decisions. Testing is conducted with several subsets of 2000 scenarios to ensure consistency of results, and to assess the generalizability of the policy. The episode reward evolution of the first one third (500,000-time steps) of the training process for two different initializations of TRPO DRL-LF agent can be seen in Figure 4, showing the typical training reward evolution using this algorithmic approach. The different time-resolutions introduced in this section and

implemented for the remainder of this manuscript, as well as how they relate to each other, are summarized in Figure 5.


Figure 5: Summary of time resolutions used across training and testing process

# 5 Results and Discussion

## 5.1 Summary

Optimization results published in Cardin and Hu [40] provide readily usable guidelines to decision makers in terms of initial design, and triggering signals to exercise flexibility during operations. The results are used to compare with the results obtained in this study using DRL. For the Centralized Inflexible system used as benchmark, the authors showed via stochastic optimization that the system should deploy 600 tpd capacity at $t = 0$ in the Western area of Singapore and keep this capacity throughout the project lifetime. Considering a discrete modular capacity of 50 tpd for the flexible decentralized design (the Decentralized FDR model considered here), the authors showed that the optimal decision rule parameters are $\beta^* = 1$, $\tau^* = 0.5$, and $\gamma^* = 4$. In plain words, the optimal decision rule is that **if** *capacity mismatch at the system level in the year just ended is greater than 1×50 tpd = 50 tpd*, **and if** *capacity mismatch in sector i is greater than 0.5×50 tpd = 25 tpd*, **then** *expand capacity by 4×50 tpd = 200 tpd in that sector*. Note that the above rule contains a nested if statement, so that **if** *there is a capacity mismatch at the system level, but no sector satisfies the required capacity mismatch*, **then** *capacity is deployed at the main site (sector 1)*, **else** *no additional capacity is added to the system*. The logic of the FDR model is depicted in Figure 2. The DRL system solutions are more complex, so they are discussed further below.

The combined performance for the best Centralized Inflexible design, best performing Decentralized FDR design, as well DRL-LF and DRL-HF models proposed in this study are summarized in Figure 6. The simulation outputs show that embedding capacity expansion flexibility improves system performance significantly under uncertainty, with all flexible designs improving ENPV compared to the benchmark system. Implementing the DRL based

flexibility design process yields even greater performance improvements and extracts further value from flexibility than obtained with the FDR model.

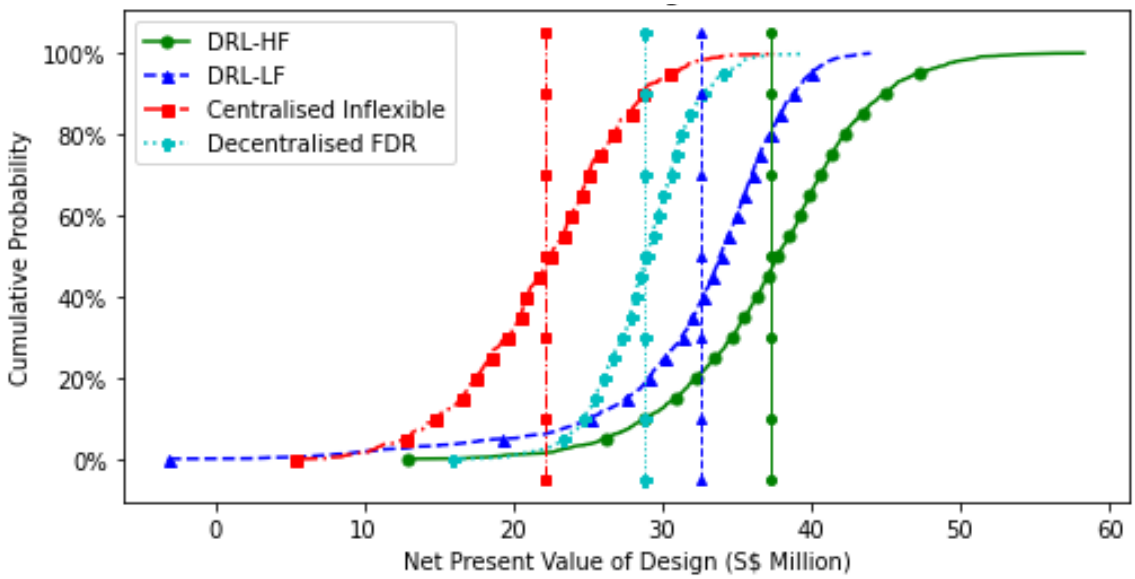


Figure 6: Cumulative distribution functions showing lifetime performance of different design alternatives. Vertical lines with corresponding markers represent ENPV of each model over 2000 simulated scenarios.

Table 1 provides further details on the stochastic performance results along different risk utility metrics. This is important to account for the possibility that decision-makers may show different risk preferences. For risk-neutral decision-makers, the analysis shows that both DRL models provide improved ENPV and extract additional VoF from uncertainty as compared to previous solutions. All flexible models improve exposure to downside risks (Value at Risk - VaR, 5%) as compared to the centralized system, which is important for risk-averse decision-makers. They also help risk-seeking system operators capture better upside potential (Value at Gain - VaG, 95%). Unsurprisingly, the DRL-HF model is able to extract the most value from uncertainty, along each metric. The results are discussed further below.

Table 1 : Design decision making table for WTE system (all values in Million S$)

| **Metric** | **(1) Centralized Inflexible** | **(2) Decentralized FDR** | **(3) DRL-LF** | **(4) DRL-HF** | **Best Solution?** |
|---|---|---|---|---|---|
| **ENPV** | 22.19 | 28.73 | 32.55 | 37.45 | DRL-HF |
| **VoF** | 0 | 6.54 | 10.36 | 15.26 | DRL-HF |
| **VaR, 5%** | 13.27 | 23.18 | 19.33 | 26.59 | DRL-HF |
| **VaG, 95%** | 30.41 | 34.25 | 40.06 | 47.11 | DRL-HF |

### 5.2 Relative Performance Analysis

The numerical results for the two DRL systems suggest that the DRL methodology has the potential to complement and improve existing approaches to design engineering systems for

flexibility. Looking back at the first objective from Section 4, the analysis shows that the proposed approach does help support design of real-world systems operating under uncertainty. The ENPV compared to the Centralized Inflexible benchmark is improved by 47% and 69% for the DRL-LF and DRL-HF implementations, respectively, and by 13% and 30% respectively as compared to the Decentralized FDR system. As an aside, no implementation relying only on dynamic programming is presented here, given the significant dimensionality of the problem – which is also an important limitation of such method. Also, it was shown previously that a decision rule approach can estimate to a very similar degree the value of flexibility obtained using dynamic programming [11], thus removing justifications for such comparisons.

The NPV improvement in any given scenario occurs because of better transportation cost management as compared to a Centralized Inflexible design. It is also due to exercising the expansion options at a more optimal time compared to a Decentralized FDR design. NPV instances greater than S$39 Million for the DRL-LF design were removed during ENPV calculations in Table 1 since they violated maximum installed capacity of 600 tpd, due to the stochastic nature of the policy. As this would allow them to capture more upside potential compared to alternative solutions, the choice of removing them was made to allow for more appropriate benchmarking against other methods. Nonetheless, this represents only the very tail end of the distribution (less than 1% of simulated scenarios), and thus Figure 6 is still representative of their relative performance.

It is observed that the DRL-LF design yields worse downside performance and overall variability per episode (equivalent to one 15-year simulation) compared to the FDR model. These indicators create valid concerns and suggest further fine tuning of the solutions may be needed to increase acceptability in a real-world setting, where VaR is an important metric for decision-making. The larger capacity expansion flexibility allowed by the DRL implementation may favor looking for upsides at the cost of neglecting downsides, although ultimately the effect of increased upside potential remains dominant. Including flexibility strategies more focused on addressing downside risks, such as abandonment or temporary shutdowns, could potentially aid in mitigating these effects, and yield a narrower CDF profile. This kind of trade-off is something that could be determined based on individual decision makers preference during design space exploration activities.

Nonetheless, there are certainly scenarios where the flexible solution is not guaranteed to be stochastically dominant relative to the non-flexible solution, given the associated costs of flexibility, and form of the design problem itself, highlighted by the long downside tail for DRL-LF in Figure 6. Visual inspection of DRL-LF/DRL-HF policy and scenarios suggests that this may be due to simulations with relatively higher initial demand and relatively much lower growth rates over time in specific sectors. This can lead the DRL-LF agent to overbuild in certain sectors in early years at a significantly higher cost than for the inflexible design, possibly based on overestimations of future demand evolution within the neural network policy, while also incurring higher transportation costs. Interestingly, the more modular capacity expansion allowed in DRL-HF is much less sensitive to these cases, as smaller capacities are normally deployed in early years in these scenarios. This helps to spread out the investment risk over time, and allows the episode policy to be progressively updated according to the uncertainty realized, limiting the instances of uncorrectable overbuilding in certain sectors and associated downside tail magnitude compared to DRL-LF.

### 5.3 Capacity Evolution and Action Probability Scenario Example

Figure 7 compares capacity evolution for an example distributed demand scenario for the optimal Decentralized FDR, DRL-LF and DRL-HF designs. The scenario is characterized by unusually high waste amount in sector 3 (S3), and in sector 4 (S4). The FDR model yields expansions in sectors S6 in year 2 and in S1 in year 6. The DRL-LF design instead can recognize the excess demand in S3 and expands there in year 3, while expanding in S6 later in year 7. This is a better expansion strategy than for the FDR model, as the difference between sectors S3 and S6 demand outweighs the extra transportation cost in early years. The DRL-HF model matches demand and capacity even more closely, with incremental expansion decisions in sectors S1, S3, S4 and S6 at various stages of the project when the action is estimated to return the highest expected value. Given more dynamic policies and this specific food waste scenario across S1-S6, the project NPV increases from S$31.5 Million for the Decentralized FDR design to S$32.6 Million and S$34.8 Million for the DRL-LF/HF designs, respectively.

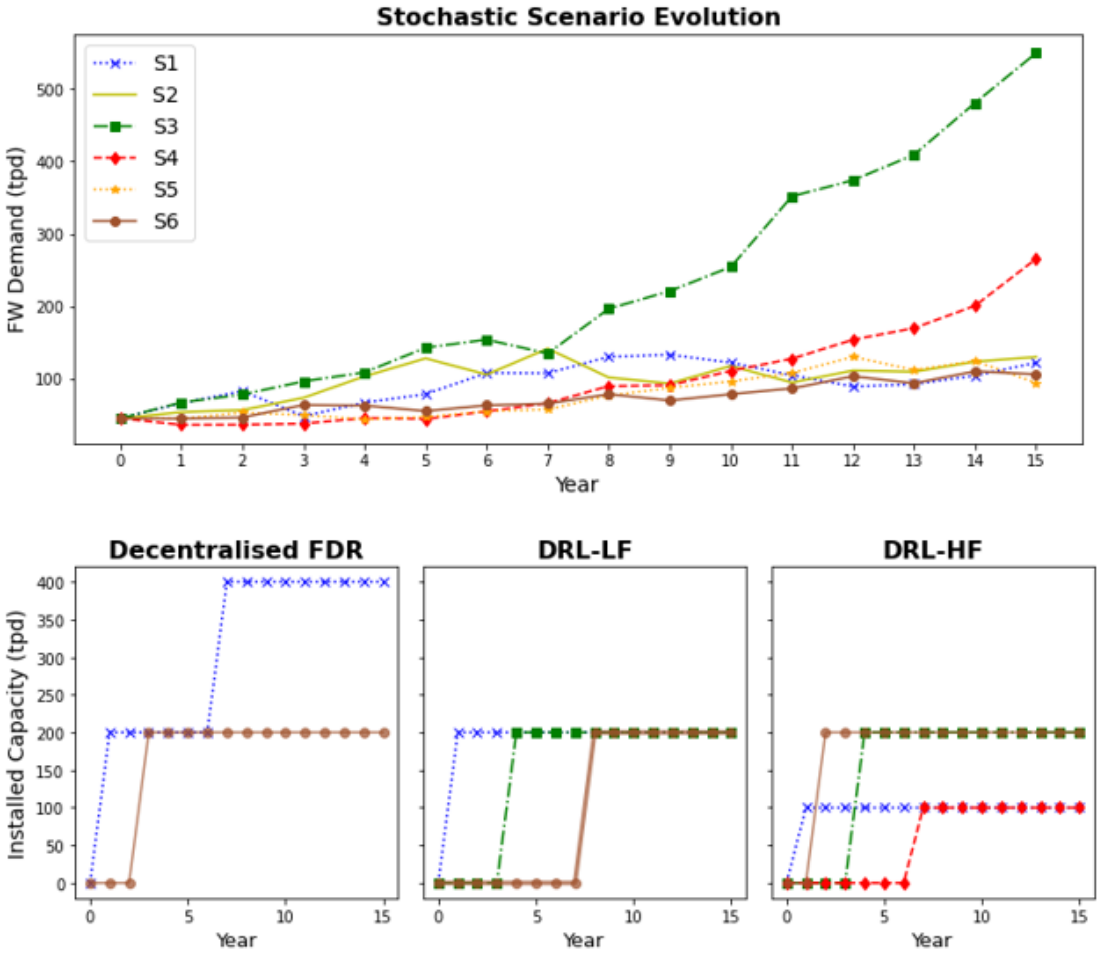


Figure 7: Example distributed food waste scenario leading to demand scenarios in sectors S1-S6 (top), corresponding design capacity expansion for FDR (bottom left), DRL-LF (bottom center) and DRL-HF designs (bottom right). Legend in top plot applies to all items.

Figure 8 gives a visual representation of the decision-making process for the DRL-LF agent over time, as a function of sampling probability for different actions, using the same distributed demand scenario in Figure 7. Following Algorithm 1, the sampling probabilities for a specific action reduces significantly once that action has been selected, or if another capacity expansion

decision is executed, as it is unlikely that two expansion decisions in consecutive time periods are made. Gradual changes are normally attributed to additional uncertainty realizations made in each year as the project is underway. Using year 2 observation as input, the FDR model follows the top nodes from Figure 2 as Procedure 1 and 2 are *true*, leading to an expansion decision in sector S6 (see Figure 7). When the same state observation is given as an input to the DRL agent, following the logical progression from Figure 3, the most likely action returned is *no expansion* with a 65% probability. Instead, the expansion probability rises significantly for sector S3 (41%) in year 3 based on observed increasing demand, leading to that action being selected. This suggests that the agent can consider the problem more holistically, as while the per sector transport costs are highest for sector S6 that year, it may estimate that total costs in the long run will be lower if expanding in sector S3, while continuing transportation from S6. In most years, transport costs averaged 3-5% of total system cost, thus basing the expansion decision solely on that may also miss out on some potential value adding flexibilities.

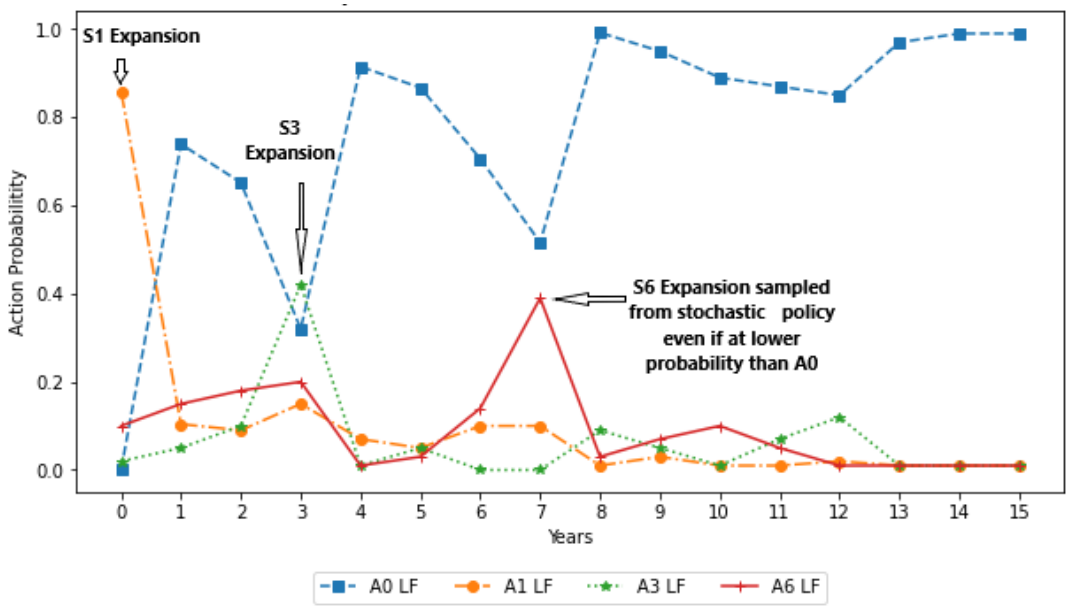


Figure 8: Action sampling probabilities for the example distributed demand scenario in Figure 7, for DRL LF system. Actions maintaining a probability of being selected below 5 % through the project duration are not included.

In real-world system operations, a decision maker would be able to refer to the optimal policy and access the associated action space probabilistic distribution based on state observations (see Figure 8). Looking at year 7, about halfway through the project, can better highlight how this distribution of actions could be used to fit the risk criteria of individual engineering design projects and influence the choice to deviate, or continue with the deterministic policy. For the FDR model, the system is not able to consider expansion, as the max installed capacity of 600 tpd has already been met. The DRL-LF system also suggests no expansion (i.e., action A0) as most likely (51%), however also presents expansion in sector 6 as feasible (39% probability) – and it ultimately chooses it for this episode due to the stochastic sampling of the policy. Risk-tolerance information, however, could help guide this sampling process (or deviation from deterministic policy) in a more systematic and value enhancing manner, as long as the difference in probability between two or more actions is relatively small. In this case, a more risk averse decision maker may decide to follow the deterministic policy (i.e., action A0), minimizing investment risk while also integrating one additional year of uncertainty realization

before possibly making an expansion decision. A more risk prone decision maker, however, may recognize that proceeding with sector 6 expansion in year 7 could allow an additional year/capacity for increasing waste processing revenues compared to the deterministic policy, which does carry significant investment risk in case demand does not materialize. In this particular demand scenario, the deviation from the deterministic policy to proceed with early sector 6 expansion would lead to an increase in NPV for the project, although waiting an additional year is likely to provide better protection against possible downsides, hence the risk tolerance tradeoffs. On the other hand, in year 3, it may be more likely for a risk-averse decision maker to choose to diverge from the deterministic policy (A3) and select no expansion (A0) in light of their very similar sampling probability (note that this was not the case in this episode), for the same reasons discussed above. Nonetheless, there is one obvious DRL action recommendation for the majority of system operation, thus if some risk tolerance information is available for decision making, possible deviation from the deterministic policy should only be considered for years 3 or 7 in this case, as shown in Figure 8.

### 5.4 Incorporating the Computational Insights in Decision-Making

In the DRL-LF implementation, the primary insight is the optimal triggering signal to exercise the flexibility, as capacity expansion decisions are limited to the level obtained from the FDR approach. In terms of starting capacity, the same conclusion is reached as in the FDR model, to install the initial capacity of 200 tpd in Sector 1 (main site). This is a result of this decision being the most likely to lead to the lowest transportation cost over system lifetime based on the assumptions presented in Appendix. In terms of optimal capacity expansion strategy, similar conclusions are also reached. The FDR approach yields expansion in sectors S1 and S6 for most simulated scenarios, resulting from those sites presenting the highest collection distance and thus transportation cost for undersized capacity. Exceptionally high demand scenarios in sectors S4 and S5 do occasionally lead to expansions within their area, but this is often constrained by the form and resulting timing of the decision rules. The DRL-LF approach presents a broader distribution of possible actions. Other than no expansion (A0), the highest probability is found for expansion in sector S1, followed by sectors S6, S5, S4, S3 and S2 in that order, which seems nearly optimal given the assumptions presented in Appendix.

Evaluating the DRL-HF design yields further insights and an increased value of flexibility compared to the DRL-LF implementation. In terms of starting capacity, the optimal suggested rule is between 100-200 tpd of initial capacity installed at the main site. The findings on expansion magnitude primarily confirm the results from the decision rule approach as the highest probabilities are generally found for 200 tpd action for each sector, to take best advantage of economies of scale. The magnitude of the EoS factor is an important determinant, and it is expected that for reduced EoS the probability of selecting smaller, more modular expansions would increase. In terms of expansion decision distribution, a visual inspection of past actions distribution shows that unlike in the DRL-LF implementation, the solution can likely be further optimized. The primary indicator is a relatively high probability of selecting action A5 (expand by 100 tpd in Sector S2) compared to the DRL-LF implementation. Aside from cases of extremely high demand in S2, this is unlikely to be the optimal decision as more significant reductions in transportation costs can be achieved by expanding in other sectors. This suggests that the DRL-HF agent is somewhat stuck in a local optimum, although more training and different initialization conditions offer potential solutions to this issue.

In a practical setting, in any given year, the system state (capacity and demand in each sector) would become an input to the optimal policy, which would return an action probability distribution as introduced in Figure 3 and presented in Section 5.3. This distribution can be used as a proxy for evaluating the estimated expected value of different actions. Having access to this kind of distribution rather than a single recommended action obtained via FDRs, could allow decision makers to integrate their own expertise and insights into policy outputs for choosing system operation strategy, increasing the applicability to real world situations. The maximum relative difference in probability of selecting different actions to justify a potential deviation from the deterministic policy is something that will vary among engineering systems, based on associated uncertainty and action space size, with potential adjustments through decision makers risk profile and project specific insights. Future work could focus on determining the appropriate thresholds to warrant considering deviation from the deterministic policy, and the resulting implications for system operation. In most decision time-steps, however, it is still expected that the deterministic policy would be followed, with decision makers progressively updating system state description year to year and implementing the resulting recommended action.

Furthermore, say an important but possibly uncertain parameter like fuel cost changes significantly by year 2, the DRL agent can observe the change, and dynamically adjust the policy to match the new environment based on a few computationally inexpensive interactions. This is often not the case in standard methods to analyze flexibility, where initial assumptions are usually maintained throughout implementation to compute stochastically optimal decision rules – although approaches such as post-optimality sensitivity analysis can help determine the thresholds for which these solutions break down. These examples and the scenario presented in Section 5.3 give some intuition on how the action distribution changes based on underlying state observations, and thus how the current state of the system could be used as an input into the DRL policy at various stages, helping to determine preferred operation strategy.

### 5.5 Reproducibility and Other Limitations

Comparing the original results in the study by Cardin and Hu [40] to the DRL solutions, the performance improvements are visually striking, but their interpretation should be conducted carefully. Among others, the stochastic nature of the TRPO policy means that there are cases where the DRL agent violates the constraints imposed on maximum capacity. This is a result of the distribution for action selection, where the probability of "illegal" actions shrinks very significantly throughout the training process, but never reaches zero. While an attempt was made to manually remove these instances during the testing process, it is possible there may have been an inflation of upside performance. More advanced approaches such as chance constrained DRL offers potential solution for safety critical engineering systems design [44]. There may be logistical constraints which make implementation of DRL solutions in real life more complex or more costly than it appears. Furthermore, the dynamic and adaptable nature of DRL decision making is not as readily understood in terms of an actual project's implementation. The policies produced may not be fully generalizable, and the question of how best to incorporate computational insights into the design process and system operations remains an open question for future research.

To evaluate results limitation, an important issue is whether the DRL agent overfits the environment's characterization. If this occurs, it may exhibit a significant drop in performance as soon as stochastic parameters (e.g., volatility $\sigma$, drift $\mu$) are changed, as has been recorded

in other studies [45]. Out of sample testing is done to assess the potential magnitude of this effect, where the agent trained on the parameters and assumptions presented thus far is tested using scenarios sampled from different distributions i.e., using $\sigma_{low}$ = 8% and $\sigma_{high}$ = 25% in Eqs. (22)-(23).

Table 2 captures the percentage deviation from the results obtained using $\sigma_{central}$ = 16.7%, shown in Table 1. For the case with $\sigma_{low}$ = 8%, ENPV is improved for all designs, and the greatest increments come from the DRL models. VoF worsens for the FDR design – a typical observation when a flexible system faces less volatility, since there is less uncertainty to deal with – but it improves for both DRL models. This observation is striking, but also seems counter to the literature on real options analysis. The case with $\sigma_{high}$ = 25% suggests there is an interesting interaction between the ability to adapt dynamically, and reinforcement training. VoF improves for the FDR model due to increased volatility – the expected flip side of the observation above – but it worsens for both DRL models. This may be because with higher volatility, the scenarios are too different from those encountered during training, and the optimal adaptation strategy is less efficient.

The significant improvements in VoF for both DRL models in the low volatility case, however, suggest that the policy learned during training can generalize well for *some* level of unseen conditions (i.e., lower volatility scenarios), and that the agent is able to extract additional VoF from such conditions. The results suggest, however, that the generalizability of the policy could be related to the degree of uncertainty experienced in the testing conditions, compared to that found during training. Nevertheless, this result is interesting as it illustrates the tradeoff between potential value gained from having the ability to adapt, and the limitations of training based on a particular set of scenarios. Further studies may help understand when a particular model no longer generalizes well to unseen conditions or level of uncertainty.

Table 2: Out of sample testing of performance of design alternatives considered in this study

| Case | Metric | (1) Centralized Inflexible | (2) Decentralized FDR | (3) DRL-LF | (4) DRL-HF |
|---|---|---|---|---|---|
| $\sigma_{low}$ = 8% | ENPV | +5.6% | +3.8% | +8.6% | +6.4% |
| | VoF | 0% | -2.3% | +15.0% | +7.6% |
| | VaR | +37.9% | +41.4% | +54.6% | +14.3% |
| | VaG | -11.8% | -5.2% | -3.1% | -5.9% |
| $\sigma_{high}$ = 25% | ENPV | -11.6% | -6.3% | -11.4% | -8.8% |
| | VoF | 0% | +11.7% | -10.97% | -4.73% |
| | VaR | -107.5% | - 26.6% | -59.6% | -34.6% |
| | VaG | +8.5% | +4.1% | +2.3% | +3.4 % |

It should also be noted that the limitations and sensitivity of results discussed in the above section are, to a certain degree, inherent to the fact that historical data was used to build the stochastic model in order to allow benchmarking with previous results. Future work should explore the performance of this approach when completely unforeseen events and disruptions are considered, for instance using jump diffusion models, or disruptions that are manually generated. While recent studies suggest that flexibility may play an important role in dealing with unforeseen events [26, 46, 47], more research is needed to evaluate how the proposed approach fares under such conditions.

## 6 Conclusions

This study proposes a novel methodology based on deep reinforcement learning (DRL) to complement existing approaches to analyze flexibility – also referred as *real options* – in engineering systems design. The proposed approach helps explore alternative solutions or configurations through thorough analysis of the data more systematically and uncovers solutions that may not be considered using standard design methods. Building upon an approach based on decision rules and existing design frameworks, an example implementation is shown on the design and analysis of a waste-to-energy (WTE) system in Singapore, with capacity expansion flexibility across spatial and temporal dimensions. The results confirm that embedding flexibility in engineering systems under uncertainty improves expected performance significantly as compared to stochastically optimal, but inflexible designs. Furthermore, they show that the DRL approach produces highly adaptable and dynamic design and decision rules to navigate an uncertain environments, thereby further improving the solutions identified in a previous study [40]. Integrating the approach as part of the framework developed in [9] also shows potential for supporting the phases focusing on concept generation, as well as design space exploration.

There are several limitations that can be addressed through future research. Future efforts could investigate integrating additional flexibility (i.e., all generic real options) and uncertainty sources (i.e., fuel costs, waste purity rate, etc.) into the analysis, to help determine the problem dimensionality for which this approach is no longer suitable. Increased integration of year 0 design decisions, accounting for technical enablers, could help uncover further value for the proposed approach. Multi-objective reward functions, looking at sustainability of the whole project through ESG metrics, for example, could also be implemented with some very important operational insights. The introduction of random “shocks” or perturbations to environment dynamics could help investigate the potential for DRL to design more resilient engineering systems. More advanced statistical analysis of agent actions under different conditions could also help to clarify operational insights and reformulation of the modeling outputs as usable guidelines. Finally, it would be interesting to understand what kind of environments and design problems are better suited for a stochastic rather than deterministic sampling policy on action space (i.e., select the highest probability), and the resulting implications on system performance.

## 7 Acknowledgments

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/R513052/1 for the Dyson School of Design Engineering, Imperial College London, and the World Bank Group.

# 8 References

[1] Pate, A., 2015, "Terrorism Trends with a Focus on Energy and Mining," START - National Consortium for the study of terrorism and responses to terrorism, University of Maryland.

[2] Jemielniak, D., and Przegalinska, A., 2020, Collaborative society, MIT Press.

[3] Haskins, C., Forsberg, K., Krueger, M., Walden, D., and Hamelin, D., "Systems engineering handbook," Proc. INCOSE, pp. 13-16.

[4] MacCormack, A., Herman, K., 2001, "The Rise and Fall of Iridium," Harvard Business School, Cambridge, MA, United States.

[5] Lim, R., and Ng, K., 2011, "Recycling Firm IUT Global Being Wound Up," The Business Times, Marshall Cavendish Business Information, Singapore.

[6] Shepard, W., 2016, "An Update On China's Largest Ghost City - What Ordos Kangbashi Is Like Today," Forbes.

[7] de Neufville, R., and Scholtes, S., 2011, Flexibility In Engineering Design, MIT Press, Cambridge, MA, United States.

[8] Trigeorgis, L., 1996, Real Options: Managerial Flexibility and Strategy in resource Allocation, MIT Press, Cambridge, MA, United States.

[9] Cardin, M.-A., 2014, "Enabling Flexibility in Engineering Systems: A Taxonomy of Procedures and a Design Framework," ASME Journal of Mechanical Design, 136(1), pp. 1-14.

[10] Guma, A., Pearson, J., Wittels, K., de Neufville, R., and Geltner, D., 2009, "Vertical phasing as a corporate real estate strategy and development option," Journal of Corporate Real Estate.

[11] Cardin, M.-A., Xie, Q., Ng, T. S., Wang, S., and Hu, J., 2017, "An approach for analyzing and managing flexibility in engineering systems design based on decision rules and multistage stochastic programming," IISE Transactions, 49(1), pp. 1-12.

[12] Cardin, M.-A., Zhang, S., and Nuttall, W. J., 2017, "Strategic Real Option and Flexibility Analysis for Nuclear Power Plants Considering Uncertainty in Electricity Demand and Public Acceptance," Energy Economics, 64, pp. 226-237.

[13] Zhang, S., and Cardin, M.-A., 2017, "Flexibility and real options analysis in emergency medical services systems using decision rules and multi-stage stochastic programming," Transportation Research Part E: Logistics and Transportation Review, 107, pp. 120--140.

[14] Cardin, M.-A., Kolfschoten, G. L., Frey, D. D., de Neufville, R., De Weck, O. L., and Geltner, D. M., 2013, "Empirical evaluation of procedures to generate flexibility in engineering systems and improve lifecycle performance," Research in Engineering Design, 24(3), pp. 277-295.

[15] Andriotis, C., and Papakonstantinou, K., 2019, "Managing engineering systems with large state and action spaces through deep reinforcement learning," Reliability Engineering & System Safety, 191, p. 106483.

[16] Bellman, R., 1952, "On the Theory of Dynamic Programming," Proc. Natl. Acad. Sci. U. S. A., Santa Monica, CA, United States, pp. 716-719.

[17] Sethi, A. K., and Sethi, S. P., 1990, "Flexibility in Manufacturing: A Survey," The International Journal of Flexible Manufacturing Systems, 2, pp. 289-328.

[18] Linsey, J. S., Green, M. G., van Wie, M., Wood, K. L., and Stone, R., "Functional Representations in Conceptual Design: A First Study in Experimental Design and Evaluation," Proc. Proceedings of the American Society for Engineering Education Annual Conference and Exposition.

[19] Nilchiani, R., and Hastings, D. E., 2007, "Measuring the Value of Flexibility in Space Systems: A Six-Element Framework," Systems Engineering, 10(1), pp. 26-44.

[20] Mikaelian, T., Nightingale, D. J., Rhodes, D. H., and Hastings, D. E., 2011, "Real Options in Enterprise Architecture: A Holistic Mapping of Mechanisms and Types for Uncertainty Management," IEEE Transactions on Engineering Management, 54(3), pp. 457-470.

[21] Ferguson, S., Siddiqi, A., Lewis, K., and de Weck, O. L., 2007, "Flexible and Reconfigurable Systems: Nomenclature and Review," ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE, Las Vegas, NV, United States.

[22] Saleh, J. H., Mark, G., and Jordan, N. C., 2008, "Flexibility: a Multi-Disciplinary Literature Review and a Research Agenda for Designing Flexible Engineering Systems," Journal of Engineering Design, 1, pp. 1-17.

[23] Cardin, M.-A., 2014, "Enabling flexibility in engineering systems: a taxonomy of procedures and a design framework," Journal of Mechanical Design, 136(1).

[24] Copeland, T., and Antikarov, V., 2003, Real Options: A Practitioner's Guide, Thomson Texere, New York,

NY, United States.

[25] Cox, J. C., Ross, S. A., and Rubinstein, M., 1979, "Options Pricing: A Simplified Approach," Journal of Financial Economics, 7(3), pp. 229-263.

[26] Caunhye, A. M., and Cardin, M.-A., 2017, "An approach based on robust optimization and decision rules for analyzing real options in engineering systems design," IISE Transactions, 49(8), pp. 753-767.

[27] Garstka, S. J., and Wets, R. J.-B., 1974, "On decision rules in stochastic programming," Mathematical Programming, 7(1), pp. 117-143.

[28] Sutton, R. S., and Barto, A. G., 2018, Reinforcement learning: An introduction, MIT press.

[29] Markov, A. A., 1954, "The theory of algorithms," Trudy Matematicheskogo Instituta Imeni VA Steklova, 42, pp. 3-375.

[30] Li, Y., 2017, "Deep reinforcement learning: An overview," arXiv preprint arXiv:1701.07274.

 

[31] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W., 2016, "Openai gym," arXiv preprint arXiv:1606.01540.

[32] Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A., 2017, "Deep reinforcement learning: A brief survey," IEEE Signal Processing Magazine, 34(6), pp. 26-38.

[33] Yonekura, K., and Hattori, H., 2019, "Framework for design optimization using deep reinforcement learning," Structural and Multidisciplinary Optimization, 60(4), pp. 1709-1713.

[34] Cui, H., Turan, O., and Sayer, P., 2012, "Learning-based ship design optimization approach," Computer-Aided Design, 44(3), pp. 186-195.

[35] Zhang, J., Wang, Z., and Zhang, H., 2018, "Data-based optimal control of multiagent systems: A reinforcement learning design approach," IEEE transactions on cybernetics, 49(12), pp. 4441-4449.

[36] Baker, B., Gupta, O., Naik, N., and Raskar, R., 2016, "Designing neural network architectures using reinforcement learning," arXiv preprint arXiv:1611.02167.

[37] Van Moffaert, K., Drugan, M. M., and Nowé, A., "Scalarized multi-objective reinforcement learning: Novel design techniques," Proc. 2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL), IEEE, pp. 191-199.

[38] Lee, X. Y., Balu, A., Stoecklein, D., Ganapathysubramanian, B., and Sarkar, S., 2019, "A case study of deep reinforcement learning for engineering design: Application to microfluidic devices for flow sculpting," Journal of Mechanical Design, Transactions of the ASME, 141(11), p. <xocs:firstpage xmlns:xocs=""/>.

[39] Perera, A. T. D., Wickramasinghe, P., Nik, V. M., and Scartezzini, J.-L., 2020, "Introducing reinforcement learning to the energy system design process," Applied Energy, 262, p. 114580.

[40] Cardin, M.-A., and Hu, J., 2016, "Analyzing the tradeoffs between economies of scale, time-value of money, and flexibility in design under uncertainty: Study of centralized versus decentralized waste-to-energy systems," Journal of Mechanical Design, 138(1).

[41] Pardo, F., Tavakoli, A., Levdik, V., and Kormushev, P., "Time limits in reinforcement learning," Proc. International Conference on Machine Learning, pp. 4045-4054.

[42] Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P., "Trust region policy optimization," Proc. International conference on machine learning, PMLR, pp. 1889-1897.

[43] Schulman, J., Moritz, P., Levine, S., Jordan, M., and Abbeel, P., 2015, "High-dimensional continuous control using generalized advantage estimation," arXiv preprint arXiv:1506.02438.

[44] Petsagkourakis, P., Sandoval, I. O., Bradford, E., Zhang, D., and Chanona, E. A. d. R., 2020, "Constrained reinforcement learning for dynamic optimization under uncertainty," arXiv preprint arXiv:2006.02750.

[45] Zhang, C., Vinyals, O., Munos, R., and Bengio, S., 2018, "A study on overfitting in deep reinforcement learning," arXiv preprint arXiv:1804.06893.

[46] Caunhye, A. M., and Cardin, M.-A., 2018, "Towards More Resilient Integrated Power Grid Capacity Expansion: A Robust Optimization Approach with Operational Flexibility," 6th International Symposium on Reliability Engineering and Risk Management, Singapore.

[47] Blume, S. O. P., Sansavini, G., and Cardin, M.-A., 2021, "Fuzzy control-enabled flexible short-turning for real-time disruption management in urban transit systems," Manuscript in Preparation.

[48] National Environmental Agency, 2013, "Solid waste management," http://app2.nea.gov.sg/energy-waste/waste-management/refuse-disposal-facility.

[49] Shell, 2014, "Shell Station Price Board," http://www.shell.com.sg/products-services/on-the-road/fuels/price-board.html.

[50] National Environment Agency, 2011, "Environmental Protection Division Report ".

[51] IUT Global Pte Ltd, 2006, "9.5 MW food waste based grid connected power project implemented by IUT Singpaore Pte Ltd," Clean Development Mechanism.

[52] RIS international Ltd, 2005, "Feasibility of Generating Green Power through Anaerobic Digestion of Garden Refuse from the Sacramento Area," MacViro Consultants.

[53] Singapore Power, 2013, "Electricity fariff rate," http://www.singaporepower.com.sg/irj/portal?NavigationTarget=navurl://41c8e6a3faf48bb168af2c222faa8ee4&windowId=WID1366188757420.

[54] Hu, J., and Cardin, M.-A., 2015, "Generating Flexibility in the Design of Engineering Systems to Enable Better Sustainability and Lifecycle Performance," Research in Engineering Design, 26(2), pp. 121-143.

[55] Bai, R., and Sutanto, M., 2002, "The practice and challenges of solid waste management in Singapore," Waste Management, 22(5), pp. 557-567.

[56] Evangelisti, S., Lettieri, P., Borello, D., and Clift, R., 2014, "Life cycle assessment of energy from waste via anaerobic digestion: A UK case study," Waste Management, 34(1), pp. 226-237.

# 9 Appendix

Table A1: Parameters and assumptions for the waste-to-energy system design problem. All values and assumptions are the same as in the study by Cardin and Hu [40].

| Parameters | Definition | Assumptions | Comments and Source |
|---|---|---|---|
| $C_{dis}$ | unit cost for disposing residues | S$77 /ton | This is the dispose cost which should be paid by AD plants to dispose of residues in incineration plants [48]. |
| $Cap_v$ | vehicle capacity for collecting wastes per trip | 25 tonnes | |
| $C_{fuel}$ | unit cost for fuel consumption | S$0.4/km | The price of diesel fuel is S$ 1.625 in Singapore [49]. It is assumed one liter of diesel fuel can last for 4km of travel distance. |
| $D_{co_i}$ | distance for collecting wastes within sector $i$ | 54km | It is assumed that the distances for collecting waste within the 6 sectors are the same. This distance is assumed based on Google map |
| $D_{Tr_i}$ | distance for transporting wastes from sector $i$ to the main sector | 0km, 20km, 25km, 29km, 36km, 40km | 0 km represents the transporting distance of the main site. It is assumed that no additional effort is required to transporting the wastes to the main site if the wastes are collected within the main site. The rest numbers are represented as the distances from other sectors to the main site and are assumed based on Google map. |
| $d^t(d_i^t)$ | amount of recycled food wastes at year $t$ (in sector $i$) | 274 tpd<br>(ton per day) | The total recycled food wastes in Singapore in 2013 is 100,000 tonnes [50]. |
| $E_g$ | electricity generation rate | 230 kwh/ton | It is assumed that the biogas generation rate is 150 $m^3$/t [51]. The electricity conversion rate is 35% and only 80% of the generated electricity can be sold to power grid [52]. |
| $P_{to}$ | unit tipping fee for food wastes | S$65 /ton | It is assumed to be slightly lower than $C_{dis}$ to encourage organic waste separation |
| $P_e$ | unit selling price for electricity | S$0.27/kwh | It is estimated based on the Singapore electricity tariff in 2013 [53]. |
| $K$ | Coefficient parameter for cost function | 305,288 | It is estimated based on the real data from [52]. Detailed analysis can be found in [54]. A reasonable range of economies of scale factor is 0.6 to 1.0. |
| $\alpha$ | economies of scale factor | 0.8 | |
| $\rho$ | unit land rental fee for the installed capacity | S$816/tpd | It is generated based on [11]. |

| | | | |
|---|---|---|---|
| σ | residues rate for food wastes | 5% | The residues rate for incineration technology is 10% in Singapore [55]. The resides rate for AD is assumed to be less than 10% since it has high efficiency [56]. |
| ω | purity rate for food wastes | 70% | In Singapore, the food wastes from industrial and commercial areas have about 30%-40% impurities [5] |
| $T$ | lifecycle period | 15 | Long-term lifecycle period |
| $\lambda$ | discount rate | 8% | A general discounted rate |