A 6 pages Framework for Multi-criteria Self-Optimising Manufacturing Network

1st Frederic Menezes
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Alexandre Cesar Muniz de Oliveira
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Kenneth Brown
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Networked personalised production requires approaches for optimisation more robust to uncertainty, more reactive, and even proactive, assuming early action to deal with abrupt changes in connections with consumers and suppliers. The objective of this work is to propose a framework for Multi-Criteria Decision-Making in the context of Self-Optimising Manufacturing Network. Particularly, investigating the interaction between optimisation, simulation and reinforcement learning as a robust and flexible approach to tackle with uncertainty and multi-criteria optimisation, trying to balance economic, variability and sustainability indicators.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Manufacturing paradigms have shifted from mass production, mass customisation to mass personalisation model, enabling to proactively adapt to the high variety and low volume manufacturing but with mass efficiency [1].

A. Personalised

Personalisation lets businesses adopt a differentiation strategy to vie on additional worth for the client rather than competitors on value. Ultra-personalised products and Services (UPPS) are a market phase involving products tailored to the requirements of the individual in tiny series, up to ton size one. Though personalisation is not a brand new concept, in recent years, additive-producing and cooperative robots (cobots) have become accessible to a broad audience. Moreover, thanks to additive production, UPPS delivers a tier of personalisation at a reasonable value that was not doable in the past. In addition to decreasing acquisition prices, the marketplace for personalisation is predicted to grow due to customers becoming a lot stern [2].

There is some variance in semantical meaning and practical implementation between Mass Personalisation and Mass Customization. In scientific research, Mass Personalisation is evaluated as a pre-stage of Mass Customization, whereas in applications, it is understood as combined with Mass Customization. Once increasing medical care and the development of knowledge and Communication Technologies (ICT), new customer engagement in product or service individualisation procedures is disclosed as a collection of the latest necessities for co-creation and co-design activities. In this sense, personalisation is a further step in flexible manufacturing [3].

Flexible Manufacturing Systems (FMS) are considered an integral part of intelligent manufacturing systems and a foundation for digital manufacturing, The development of factories in an increasingly competitive industrial environment involves the study and analysis of some key elements of the FMS and managerial, technical and innovative efforts. A basic component of Industry 4.0 is the FMS, an advanced production system that interconnects clusters of machinery and logistics equipment, fully coordinated by computer systems. This system has characteristics that allow adaptation to the great diversity of operations and types of items produced, in order to guarantee delivery times and minimum manufacturing costs, in an uncertain economic environment [4].

Such systems are expected to be able to operate even in rapid market changes, without loss of efficiency, which favors the gain of competition in the international market, even subject to uncertainties arising from long periods of instability caused, for example, by pandemics [5] or wars.

B. Network

There is a worldwide tendency to change the globalised mass production mode into more local one but without giving up of high throughput and quality. Other requirements is about the capacity of responding to a production scale that dynamically changes with the uncertainty market conditions, and unstable or broken supply chains. Besides, regionalism and authenticity is desirable since customers have reacted even more sensible to sustainability and manpower health. Adapting both production and logistics demand optimisation of material flows and logistics functions so that mass-personalised products are delivered close to the market. [6].

In manufacturing network formation and operational planning, cost minimisation and profit maximisation have always been the fundamental drivers. A recent study proposes an innovative methodology to support manufacturing networks to take advantage of the big data stored in a virtual enterprise platform from where customer and manufacturer preferences are collected. Historical customer-related data is then used to compute a customer priority and insights to compute a manufacturer's reliability index. After that, an order-driven network feeds a mathematical model with the assumption that manufacturing processes of production orders can be divided into serial production stages with multiple candidate manufacturing units for each stage, with different cost structures and strategic partnership deals in a manufacturing network of complex products, with low production volumes and wide variety [7].

C. Self-organising manufacturing network

A dedicated Manufacturing Line (DML) adopts fixed automation to simultaneously control every production line machine to produce dedicated products at high volume. In contrast, Flexible Manufacturing System (FMS) consists of Computerised Numerical Control (CNC) machines that can produce various products, but its throughput is lower than that of DML. Modularity and system responsiveness provide Reconfigurable Manufacturing System (RMS) with the dual advantages of throughput in DML and product variety in FMS, but only allowing production within a product family. SOMS, however, is expected to produce personalised products at dynamic volume demands because of its changeable structure and adaptive manufacturing control [8].

The structural characteristics of SOMS are [8]:

- Distributed: The peer-to-peer connection between manufacturing things forms a network of multiple control nodes without a central controller;
- Bottom-up: Manufacturing control decisions are generated via local negotiations and coordination between manufacturing things instead of passively accepting commands from the top;
- Adaptability: Production volume and product variety can be responsively adjusted by adaptively changing system configuration and production;
- Individual autonomy: Manufacturing things can make decisions independently based on peer communications and adaptive control algorithms.

In [9], Self-organising manufacturing network is composed of autonomous manufacturing units, consisting of software tools, hardware equipment, and operators, connected in situation-dependent ways that can change their internal structure and functions with minimum external intervention to achieve optimal manufacturing operations and system performance in response to unforeseen conditions and evolution along time. The updated conditions relies on: Self-configuration (plug-and-produce paradigm), Self-optimisation (optimal autonomous behaviour), and Self-healing (external-free abnormality recovery).

D. self-optimisation

Optimisation in a SOMN includes economic-related indicators, such as makespan, resource utilisation, workload, profit and costs, and lead time as well as non-economic ones, such as energy consumption, waste reduction, CO2 emissions, and other system-level performances related to sustainability [10], adaptability, scalability, reliability [8] and product diversification [**?**].

The self-optimisation function aims at adaptive allocating manufacturing resources and deciding the manufacturing tasks' schedules. Although heuristic algorithms have been widely applied for optimising manufacturing schedules, as found in [11], the such static algorithms do not work well for mass personalisation since the dynamic nature of the production schedules makes them continually re-optimised [8]

The highly personalised products require more adaptive and flexible production scheduling, while hard-coded heuristic algorithms with a predefined static model lack scalability and self-learning abilities. Additionally, the decentralised control architecture in a SOMN conflicts with heuristics algorithms since these algorithms need a central unit to make decisions. Therefore, SOMN needs a learning-based scheduling algorithm with decentralised characteristics to meet mass personalisation's dynamic scheduling requirements. The point is to learn to decide for different conditions and not to define decision-making for a scenario that rapidly changes [8].

In this sense, reinforcement learning has been seen as a possible answer [12], [13], presenting promising results in solving sequential decision-making problems in complex environments and providing a highly adaptable solution.

Future work on reinforcement learning relies on modelling more complex production environments and considering cooperative and competitive relationships between agents under limited production resources [8].

Recent work proposes an industrial knowledge graph (IKG)-based multi-agent reinforcement learning (MARL) approach to implement the so-called Self-X cognitive manufacturing network (configuration, optimisation, and healing), in which all the manufacturing things are organized and managed in a graph-based manner. Optimisation, in general, is a final process inside the intelligent behavior desired.) [1]

E. Sustainability

The literature models the sustainability issues according to the Triple Bottom Line (TBL) concept, marking the distinction among sustainability's environmental, economic and social dimensions [14]. Including sustainability issues is a crucial aspect of an efficient design and management of a Smart Manufacturing System with a social dimension. Higher reconfiguration capability of manufacturing systems leads to better environmental and economic performance and reduces energy consumption [15]. Environmental regulation requirements have been considered model constraints by stating formal rules for controlling pollutant emissions directly and indirectly [16].

Regarding supply chain issues, sustainability steps forward the green agenda by including social aspects beyond the more usual economic and environmental aspects [10]. Environmental and social concepts have been considered, maximizing job creation as a social objective and minimizing environmental pollution simultaneously [17]. Strategic interventions have been investigated that could help in greening the supply chain practices [18] and assessed by mathematical models, considering reliability for manufactured product distribution plans [19].

There have been investigated approaches to the smart manufacturing and economic sustainability problem by environmental-oriented, multi-objective reconfigurable manufacturing system design, including non-linear multi-objective integer program (NL-MOIP), and evolutionary approaches, showing the efficiency in metrics such as hypervolume, spacing metric, and cardinality of the mixed Pareto fronts [20].

Resuming, manufacturers are interested in providing custom-made and sustainable products, requiring flexible manufacturing systems and advanced production management to address efficient production, customization quality, and environmental performance. Reconfigurable Manufacturing Systems are required to produce cost-effective customization with high responsiveness and competitiveness. However, reconfiguration, in turn, requires multi-criteria decision-making regarding three integrated problems: process planning, scheduling, and layout optimization, in a multi-objective model method, minimizing the penalty for product tardiness, economic costs, venturous waste, and gas emissions.

F. Approaches

Optimisation and simulation-based approaches have been proposed to deal with several problems that arise around flexible manufacturing systems. Compared to optimisation methods, simulation can provide a more accurate estimate of the decision. Even taking a relatively long time to test all possible solutions, simulation is necessary for strategic and tactical decisions, where the number of choices is limited, but each decision must consider the operational phase to test its effect [21]

A combination of a production simulation system and an optimisation method, such as genetic algorithms, has been successfully used to find the appropriate order for a set of operations in machine tools available in FMS [22].

A simulation-based optimisation tool was used to design, plan and program under uncertainty scenarios. Simulators are useful for evaluating schedules, generating various likely scenarios and observing the impact on representative metrics of that schedule. An important step forward is closing the loop, that is, using the simulation to modify the constraints and goals of the optimisation solver [23].

The lack of established mathematical models to represent flexible manufacturing systems makes it difficult to better design such systems, with consequences for their performance. Mathematical modelling of complex systems also generally leads to large-scale models, making a model that includes all interactions between component subsystems intractable [21]–[23].

Thanks to ubiquitous sensors, on-site work-in-process information can be instantly perceived and automatically analysed, such as machine workload, task operation progress, inventory levels, and just-arrived demands and supply chain status. The data-rich production environment allows AI algorithms to perform data-driven self-adaptive and self-optimisation task scheduling [24].

Reinforcement learning (RL) is a popular unsupervised learning algorithm based on the trial-and-error method and is suitable for finding an optimal decision-making policy [25], consisting of environment, RL agent, and policy.

The policy represents a function projection from state space to action space and provides appropriate action to the RL agent according to the perceived environment state. The environment can simulate a manufacturing system that transits in a predefined state machine from stochastic input data. The agent evolves its decision map from the expectation of reward from the trial-and-error method. The reward value fed back is used to evaluate the performance of the RL agent's action and optimise the policy until convergence is reached [24].

Mass-personalised production brings uncertainty to decision-making during process planning. Heuristic-based optimisation assumes in advance that manufacturing resources are static and makes a deterministic plan. Due to the uncertainty of the manufacturing environment, re-optimisation procedures are required whenever the input data is changed, including rolling the planning horizon. Deep reinforcement learning has been employed in such cases, aiming to promote fast response and accuracy by learning to decide and allowing advances with reusability and expandability of past decision-making experiences. Using RL requires masking algorithms to screen out infeasible machining operations at each decision step, and the Monte Carlo method to evaluate the policy [26].

Deep agents have outperformed traditional dispatch rules, such as first-in-first-out, profit maximisation, and penalty minimisation, in general, quickly converging to an above-average level. However, Dueling Double DQN agents show advanced stability and optimality regardless of the increasing training episodes [27].

In a multi-objective fashion, the reward function can consider weighted composition, as seen at [28], in which objectives regulate schedulers' performances and help them learn helpful knowledge for enriching their experience. In that work, the composite rewards regard the time-saving rate, energy-saving rate, machine utilisation rate, and workload distribution deviation, but a few works have looked for approaches to control adaptively the manufacturing system concerning all the criteria without composing a weighted reward.

G. GAP

Recent works indicate a growing trend towards reorganising manufacturing units in networks, with greater autonomy and engagement with a specialised consumer more attentive to working conditions, health, and sustainability [3]–[5]. At the same time, local norms and regulations must be respected as production restrictions. These units collaborate with local suppliers monitored by society and are responsible for providing specific raw materials to add regional value and authenticity to the final product [6].

Networked personalised production requires approaches for optimisation more robust to uncertainty, more reactive, and even proactive, assuming early action to deal with abrupt changes in connections with consumers and suppliers. More important than planning based on static data would be learning to make decisions based on recent experiences [8]. Online optimisation problems have this characteristic and are being solved by approaches based on reinforcement learning with satisfactory accuracy [29].

In [30], a smart scheduler to handle real-time jobs and unexpected events in smart manufacturing factories is proposed, including new composite reward functions in a multi-objective fashion. Based on deep reinforcement learning (DRL), the smart scheduler autonomously learns to schedule manufacturing resources in real time and improves its decision-making abilities dynamically. Our approach differs from the previous one by adding layers of actions from the decision to schedule an order that affects future decisions and the relationship with the other actors that participate in the manufacturing network. However, the core decision-making on the shop floor can be very well represented by the model proposed in [30].

H. Objective

The objective of this work is to propose a framework for Multi-Criteria Decision-Making in the context of Self-Optimising Manufacturing Network. Particularly, investigating the interaction between optimisation, simulation and reinforcement learning as a robust and flexible approach to tackle with uncertainty and multi-criteria optimisation, trying to balance economic, variability and sustainability indicators.

II. PROPOSAL

Considering the literature review about Personalised Mass Production and Self-Organised Manufacturing Networks, we are focusing on enabling technology based on reinforcement learning to provide a conclusive study of efficient policies for managing production in an uncertain and multi-criteria context.

A. Environment

The network of manufacturing units is a complex system with at least two elements with specific function: peripheral and backbone units. For simplicity, we are assuming a backbone unit as a central manufacture that applies structural features of interest of a large amount of customers scattered in the world (mass production stage). It is strategic to locate backbone units near big distribution hubs with facilities to reach other units in the network.

By other hand, a peripheral unit is responsible for printing specific and regional features (personalised stage). It is strategic to locate peripheral units near customer markets and regional supply chains. Indeed, we are dealing with flows of semi-finished products, set of local suppliers, set of distribution hubs, pursuing local objectives (rewards) and balancing overall multi-criteria objectives (control)

But the placement of unit is not the objective of this work, but only to decide flows of supply chain and product delivery. More specifically, we are focusing in a specific peripheral unit since such unit has to deal with regional suppliers and customer, i.e, the its ecosystem is more heterogeneous.

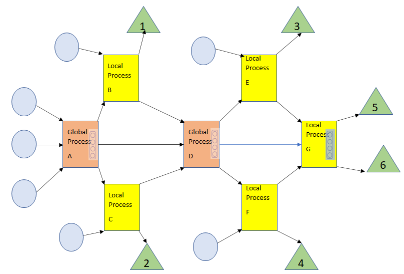

Fig. 1. Manufacturing network: 2 units dedicated only to structural features (backbone), 5 units dealing directly with customer hubs (triangles) and supplier (circles)

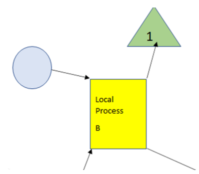

Fig. 2. LOCAL PROCESS

1) Specific assumptions: We are assuming an environment composed of autonomous manufacturing units negotiating raw material amounts and deadlines to face with foreseen demands of personalised products. Each manufacturing unit contributes with its production activity to the global multi-criteria objective in terms of costs, profit, and indicators related to sustainability issues.

Resource constraints and uncertainty conditions (modelled by probability distribution) are high impact to the decision making. Some variables that are unforeseen in a real scenario: lead time expected, amount and time for raw material deliverance, etc.

For simplicity, no predatory competition among the manufacturing units are permitted. We understand that this side effect can be prevented by not allowing negotiation of prices, only amounts in order to avoid units can choose which demand to supply. Thus, the only factor to be considered is timely. Since, deals can be agreed or not, decision chains may connect or disconnect flows by producing only cost-effective products.

Concerns about variability are introduced by specific and regional features that are aggregated to structural ones. So we can design a final product as a sequence of manufacturing units where it passes for. In Figure 3, each coloured line means a sequence of units responsible for aggregating a specific feature and introducing variety. All lines start from a backbone manufacturing unit (orange square) and finish at a customer hub (green triangle). For example, the products P41 and P42 can be assembled (or represented) by the following sequences, $\{A2, B, D1, E\}$ and $\{A3, C, D2, F\}$ respectively.

2) Sequential Decision Process: In sequential decision problems, the utility function is obtained from the state reached as the result of the whole sequence of agent's action. The action is made from a set of available decisions, and is based on a decision rule under certain circumstances,

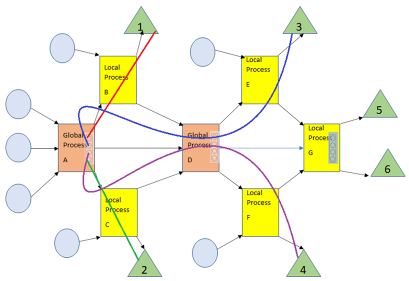

Fig. 3. Production flows through the manufacturing network showing sequences 2-assembled products $\{A1, B\}$, $\{A4, C\}$ and 4-assembled ones $\{A2, B, D1, E\}$, $\{A3, C, D2, F\}$

considering a complete information of the system state. For each decision, an immediate utility value is gotten and the system transits to a new state. The objective of a sequential decision problem is to find the optimal policy, i.e., a sequence of decision rules for each future time step [31].

a) Data and variables: At following, we present a set of variables and functions to model the sequential decision process in a very simplified way.

```
balance := BA[1..M]
income := IN[1..M]
outcome := OU[1..M]
yard := YA[1..*]
```

In this modelling, we are dealing with a simplified inventory considering the balance for M *glocal* features (global and local) [6]. The *balance* is used in the decision process since we can prioritise product demands for which the balance is sufficient to cover the production with minimum need of ordering raw material.

The whole inventory regards not only in-site raw material, but such ones that are incoming (*income*: already ordered, but not delivered) as well as those already expected to be employed in the lineup of production, i.e., *outcome*, both in the same dimension of features available in the system. The availability of the whole inventory $\overrightarrow{AV}$ can be calculated by:

$$\overrightarrow{AV} = \overrightarrow{BA} + \overrightarrow{IN} - \overrightarrow{OU} \tag{1}$$

The inventory considers not only raw material available but finished products that are ready to be delivered but, for some reason, they lost the due-time window and remain in the yard. Non-delivered products are represented by demands with status value of stored.

Since, each product is represented by a very specific array of features, the matching between a new product request and a previously stored one is very unlikely. So such comparison is a modelled by a *matching function*, $\mathcal{M}(FT_d, \overrightarrow{YA})$ that considers similarity degree at matching between a given demand d, featured by FT, and all product stored in the yard, YA. The decision of ordering a missing raw material, $\mathcal{O}(\overrightarrow{AV})$, is taken over the availability, in case of the matching function fails in finding a similar product already in the yard.

Demands are received at specific time windows in batches from each customer hub identification CU. A time stamping is used to sort the demands in the system, considering the input DI and output DO times. Recalling output time as the due time and must be sufficient to be covered by expected lead time of the product in question.

```
demands := DE <CU,DI,DO,TP,PR,CO,AM,SP,PE,LT,VA
requests := RQ[1..M]
manifest := MN[1..M]
material := MT[1..M]
usage := US[1..M]
price := EU[1..M]
```

Each demand has its expected cost CO, SP consumption if kept in inventory, and net profit PR for each unit requested in the amount AM as well as some penalty PE for accepting and not delivering the batch. ST and $FT[1..M]$ refer to the demand status and the Boolean feature array, respectively. The feature array works as a personalised signature defining the variability degree of the demand. The remaining demand attributes LT,VA,SU,BU are computed by specific functions shown in equations xyz.

$$TP_d = DO_d - DI_d \tag{2}$$

$$CO_d = AM_d \cdot \overrightarrow{FT}_d \cdot \overrightarrow{EU} \tag{3}$$

$$SP_d = \gamma(AM_d \cdot \overrightarrow{FT}_d) \tag{4}$$

$$LT_d = \tau(AM_d \cdot \overrightarrow{FT}_d) \tag{5}$$

where τ is a function regarding the amount, features and current time expectations for production.

$$VA_d = \upsilon(\overrightarrow{FT}_d) \tag{6}$$

Sustainability is calculated in function of the carbon print, energy consumption, environment impact or anything associated to. A general rule is the more features are aggregated to the finishes product the more environment impactful and consequently less sustainability is achieved in the end.

$$SU_d = \sigma(\overrightarrow{FT}_d) \tag{7}$$

$$ST_d = \{0, 1, 2, 3, 4, 5\} \tag{8}$$

in which different status values are assigned for the production process, i.e., received, ready, rejected, produced, stored and delivered.

$$BU_d = \beta(\bigcap\{\overrightarrow{FT}_d^{t-w}\}) \tag{9}$$

in which β returns the expectation for business considering the popularity of the features of the last demands arising since time w.

b) Procedures and decision rules: The local process under investigation in this work can be isolated from the remaining manufacturing network considering a few interactions along with local customers and suppliers. Therefore, the decision-making process can be significantly simplified to a decision sequence that advances or cancels the production stages of manufactured goods depending on the risk of profit or loss in the face of circumstances and uncertainties.

The rules can be non-deterministic ones since they are tackling with unforeseen events subject to high degree of uncertainty. For example, a supply cannot be delivery in time or, even so, the stochastic production lead time may make it impossible to deliver on time.

The procedure at following represents the rules concerning the decision in different stages of production, including supply chains negotiation, matching, planning, and the final stages inherent to reject, accept-store and accept-dispatch demands:

```
Receive demand, raw material
Match demand with inventory
while inventory NOT covers the demand requirements
    Order raw material for demand
    Receive demand, raw material
    Match demand with inventory
end
if demand is OutTime
    Reject demand
else
    Plan production of demand
    if product is OutTime
        Store product
    else
        Dispatch product
    end
end
```

In Figure 4, one can see the sequence of states for each demand received. Circles represent procedures and state values assumed while arrows connect states depending on certain conditions. Traced lines are those connecting final states in which a reward or penalty can be applied demands reaching.

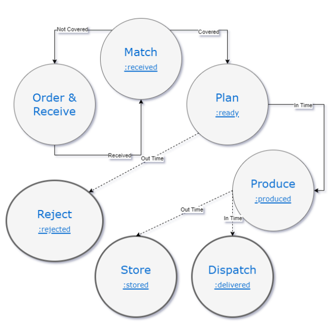

Fig. 4. State space

Procedures *Order&Receive* and *Match* compose the negotiation procedure, in which suppliers are contacted to provide raw material needed to cover the demanded product. All demands arriving are managed to be prioritised those more promising, considering profit and resources available, in order to meet production. Occasionally, demands may lapse even without receiving confirmation of raw materials' delivery in time. Naturally, such demands are ruled out in a real scenario, but this connection to the *Rejected state* is neglected in the Figure 4.

Since a given demand reach the *ready state* (*Plan state*) and the risk of out time production (*Job-shop makespam*) is acceptable the demand is put in the production lineup. *Dispatch* and *Store* are final states for a finalised production, but with well-succeeded delivery or not, respectively.

The decision is taken on different operational stages of production, dealing with contracting supplies and amounts, starting production, and dispatch batches. Furthermore, and fundamentally, it is necessary to decide which demands should be prioritised considering the system's current state.

c) Priority queue: Decision policy has been represented by hyper-parameters associated with different decision-making ranking criteria that can be used depending on operational circumstances. The different rankings induce the evaluation of the system and can be used to pursue different goals.

In this work, queues are proposed to manage different priorities for meeting demands (customer orders). Each priority queue is ordered by an index that considers different information observed from the production environment, such as cost/profit ratio, the penalty for non-production, product variability, etc.

d) Indexes (hyper-parameters): At following, indexes are defined to offer different views over the environment's observed variables. The lesser is the value computed the greater is the priority for a given demand.

The index h_1 is computed for all demand d considering the rate between cost and profit. Recalling that profit is already discounted from all production costs, i.e., it is a net profit.

$$h_{1,d} = \frac{CO_d}{PR_d} \tag{10}$$

The index h_2 is an inverse proportion for the lead time of a given demand.

$$h_{2,d} = \frac{1}{LT_d} \tag{11}$$

The index h_3 is an inverse proportion for the penalty for not producing a given demand.

$$h_{3,d} = \frac{1}{PE_d} \tag{12}$$

The index h_4 is an inverse proportion for the feature variety of a given demand. Recalling that variety is a subjective concept calculated by summing up the included feature vector (the more features, the more variety included).

$$h_{4,d} = \frac{1}{VA_d} \tag{13}$$

The index h_5 makes the same for the presumed sustainability of a given demand. Recalling that variety is a subjective concept calculated by a sustainability function, enough flexible to incorporate issues related with green-agenda suppliers or carbon footprint, as examples.

$$h_{5,d} = \frac{1}{SU_d} \tag{14}$$

The index h_6 is an direct proportion for the remaining time to produce a given demand.

$$h_{6,d} = (DO_d - t) \tag{15}$$

The index h_7 is an inverse proportion for the expectation for business of a given demand.

$$h_{7,d} = \frac{1}{BU_d} \tag{16}$$

e) Multi-criteria evaluation: There are many possibilities to design the objective function in a multi-criteria way. In the global context, it is desirable trying to balance the behaviour of the entire manufacturing network, leaning each decision step to favour the objectives of profit or sustainability. In the local context, the objective can go further to deal with regional requirements associated to, for example, law regulations, local customer featuring attendance, favouring local green suppliers, etc. Of course, these are antagonistic goals as the manufacturing unit is likely to pay more for each action aimed at harmonising with local communities.

Hence, we design a prototype for the multi-criteria objective function considering initially: profit, sustainability and local customer featuring attendance, all of them defined in the interval $[0, 1]$. Profit can be calculated by maximising the profit PR, without considering CO once, for simplicity, we define PR as the net profit for each batch with delivered status:

$$\max \sum_{d \in ST = :delivered} PR_d \tag{17}$$

Sustainability also can be computed in a similar way, but considering every batch already scheduled to the lineup (included delivered and stored):

$$\max \sum_{d \in ST = :produced} SU_d \tag{18}$$

Local customer featuring attendance can be associated to the variability degree since the featuring array is design to attend local customer requirements and hence the more features aggregated the more specific is the customer reached with:

$$\max \sum_{d \in ST = :produced} VA_d \tag{19}$$

As observed, the evaluation functions depend on qualifying demand batches produced and delivered. Decision-making depends on factors inherent to the demands and the system's state. We can represent the decision variable as:

$$x_d = \begin{cases} :produced & \text{if} \quad \text{demand is to be produced} \\ :delivered & \text{if} \quad \text{demand is to be produced and delivered} \\ 0 & \quad\; otherwise \end{cases} \tag{20}$$

f) Knapsack Job-shop connection problem: It is worth mentioning there are two very popular built-in optimisation problems running inside the modelled manufacturing unit: *knapsack* and *job-shop* problems. The matching process is something about to decide what to produce, maximising the profit, PR, or other criterion, but subject to the resources (raw materials and time) needed. This is a typical knapsack dilemma. Consider also a special case in which the resources, either raw material and time are dynamically changed at each time step.

Besides, the planning process have to decide the sequence of operations on each machine, in a typical job-shop problem, which can be described as a industrial plant composed of a set of machines that should process a set of different tasks (demands), in which each task is composed of a group of operations (stages) to be processed in a given order. So each product may have different routes with different lengths [32].

In a typical job-shop modelling, constraints guarantee that each task has a single successor and predecessor on each machine, ensuring each task's first operation is completed after the respective processing and setup times. The $\mathcal{K}$ notation indicates the k^{th} machine visited by the task. The partial lead time $LT(\mathcal{K})_d$ is the expected processing time (part of LT_d) of demand d in the machine $\mathcal{K}$, considering previous demands $\mathbf{d}$ processed in the same machine. The expected completion time $C(\mathcal{K})_d$ measured at a $\mathcal{K}$ machine is given by:

$$C(\mathcal{K})_d = C(\mathcal{K})_{\mathbf{d} \leq d} + LT(\mathcal{K})_d \tag{21}$$

where $C(\mathcal{K})_d$ converges do the global makespan, $C(max)$, usually associated to the final job-shop problem performance, in an offline approach.

By the knapsack hand, once priorities are assigned to the demands ready to be produced, the schedule starts from the highest priority one but respecting the available capacity. The available production capacity, $\mathfrak{C}(t)$, in given instant t, is associated to the elapses between the due time, DO_d, and the current time t:

$$\mathfrak{C}(t) = \max(0, DO_d - t) \tag{22}$$

The schedule strategy takes place before the production starts considering a Knapsack and Job-shop joint approach, in which a demand is only put inline when there is sufficient time to finish the product, prioritising the profitable ones, but once initialised, the demand must be finished. The schedule rule can be defined as:

$$x_d = \begin{cases} :produced & \text{if} \quad C(max)_d \leq \mathfrak{C}(t) \\ :rejected & \quad\; otherwise \end{cases} \tag{23}$$

in which $C(max)_d$ is the expected completion time regarding the lineup workload.

We recall the mathematical model as

$$\max \sum_{d \in ST = :delivered} (PR_d) + \sum_{d \in ST = :produced} (SU_d + VA_d) \tag{24}$$

subject to

$$C(\mathcal{K}, t)_{\mathbf{d} \leq d} + LT(\mathcal{K})_d \leq \max(0, DO_d - t) \tag{25}$$

for all d and $\mathcal{K}$ in a given instant t.

A joint knapsack and job-shop approach has been proposed to deal with multiple tasks completed at the same time, when dynamic priorities in changeover operations must be considered [33]. The proposal is formulated by a flexible job-shop scheduling with changeover priorities based on the knapsack model [33].

g) Reward: The reward function is associated with a given demand d reaching one of the successful states: stored or delivered. Recalling stored is a status related to well-succeeded demand production but that lost the deadline to be dispatched. The product is so updated in the inventory possibly to be delivered in a future order.

Something important to be considered in customised production systems concerns the difficulty of completely matching specific incoming demand with a product already in inventory. Probably, the decision to stock the lot may not bring benefits to the company unless there is storage space available. Even so, it is not the type of action that should be encouraged as a policy, but just a consequence of a bad decision to accept production within a very short period of time (high risk of stranding). For this case, the stored state must provide a reward only if there is space available and, eventually, if there is not, it is changed into a penalty proportional to the violation of the space constraint. On the other hand, the reward for the delivered state must be leveraged by the nominal profit defined for the lot.

$$RW_d = \begin{cases} \frac{|YARD| - SP_d}{|YARD|} & \text{if demand d is } _{:stored} \\ \frac{PR_d \cdot AM_d}{\sum_{\delta \leq d} PR_\delta} & \text{if demand d is } _{:delivered} \\ 0 & \text{if demand d is } _{:rejected} \end{cases} \tag{26}$$

B. Implementation details

1) Define environment parameters:

2) Define environment procedures:

```
DE, MT = readData()
while True:
    AV = BA + MT + IN - OU
    BA, OU, YA = match_demand_with_inventory(DE, YA, AV)
    if not inventory_covers_demand(AV, DE):
        IN = order_raw_material(AV, IN)
        DE, MT = readData()
    else:
        break
if is_demand_out_of_time(DE):
    reject_demand(DE)
else:
    DE = proceed_production_of_demand(DE)
    if is_product_out_of_time(DE):
        store_product(DE)
    else:
        dispatch_product(DE)
```

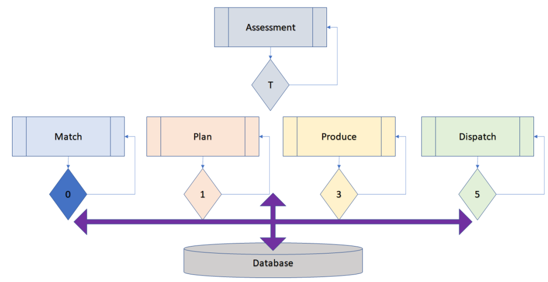

Fig. 5. SEQUENTIAL DECISION PROCESS

A woman and a girl in white dresses sit in an open car.

3) Modules:

C. Figures and Tables

a) Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation "Fig. 6", even at the beginning of a sentence.

TABLE I
TABLE TYPE STYLES

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 6. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted expression "one of us (R. B. G.) thanks …". Instead, try "R. B. G. thanks…". Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [?]. The sentence punctuation follows the bracket [?]. Refer simply to the reference number, as in [?]—do not use "Ref. [?]" or "reference [?]" except at the beginning of a sentence: "Reference [?] was the first …"

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [?]. Papers that have been accepted for publication should be cited as "in press" [?]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [?].

REFERENCES

[1] P. Zheng, L. Xia, C. Li, X. Li, and B. Liu, "Towards self-x cognitive manufacturing network: An industrial knowledge graph-based multi-agent reinforcement learning approach," *Journal of Manufacturing Systems*, vol. 61, pp. 16–26, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612521001643

[2] I. Torn and T. Vaneker, "Mass personalization with industry 4.0 by smes: a concept for collaborative networks," *Procedia Manufacturing*, vol. 28, pp. 135–141, 2019, 7th International conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV2018). [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978918313647

[3] G. Baranauskas, "Mass personalization vs. mass customization: Finding variance in semantical meaning and practical implementation between sectors," 2019.

[4] A. Florescu and S. A. Barabas, "Modeling and simulation of a flexible manufacturing system—a basic component of industry 4.0," *Applied sciences*, vol. 10, no. 22, p. 8300, 2020.

[5] A. Owida, N. M. Galal, and A. Elrafie, "Decision-making framework for a resilient sustainable production system during covid-19: An evidence-based research," *Computers & Industrial Engineering*, vol. 164, p. 107905, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360835221008093

[6] E. Rauch, P. Dallasega, and D. T. Matt, "Distributed manufacturing network models of smart and agile mini-factories," *International Journal of Agile Systems and Management*, vol. 10, no. 3-4, pp. 185–205, 2017.

[7] S. Sadic, J. P. de Sousa, and J. A. Crispim, "A two-phase milp approach to integrate order, customer and manufacturer characteristics into dynamic manufacturing network formation and operational planning," *Expert Systems with Applications*, vol. 96, pp. 462–478, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417307418

[8] Z. Qin and Y. Lu, "Self-organizing manufacturing network: A paradigm towards smart manufacturing in mass personalization," *Journal of Manufacturing Systems*, vol. 60, pp. 35–47, 2021.

[9] Y. Lu, X. Xu, and L. Wang, "Smart manufacturing process and system automation – a critical review of the standards and envisioned scenarios," *Journal of Manufacturing Systems*, vol. 56, pp. 312–325, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S027861252030100X

[10] S. Azarakhsh, H. Sahebi, and S. M. Seyed Hosseini, "Design of a sustainable integrated crude oil manufacturing network with risk cover and uncertainty considerations: a case study," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2021.

[11] R. Ogunsakin, N. Mehandjiev, and C. A. Marín, "Bee-inspired self-organizing flexible manufacturing system for mass personalization," in *International conference on simulation of adaptive behavior*. Springer, 2018, pp. 250–264.

[12] B. Cunha, A. M. Madureira, B. Fonseca, and D. Coelho, "Deep reinforcement learning as a job shop scheduling solver: A literature review," in *International Conference on Hybrid Intelligent Systems*. Springer, 2018, pp. 350–359.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[14] J. Elkington, "Cannibals with forks: The triple bottom line of 21st century business," *Journal of Business Ethics*, vol. 23, no. 2, pp. 229–231, 2000.

[15] M. Bortolini, F. G. Galizia, and C. Mora, "Reconfigurable manufacturing systems: Literature review and research trend," *Journal of manufacturing systems*, vol. 49, pp. 93–106, 2018.

[16] M. C. Carneiro, G. P. Ribas, and S. Hamacher, "Risk management in the oil supply chain: a cvar approach," *Industrial & Engineering Chemistry Research*, vol. 49, no. 7, pp. 3286–3294, 2010.

[17] N. Moradinasab, M. Amin-Naseri, T. J. Behbahani, and H. Jafarzadeh, "Competition and cooperation between supply chains in multi-objective petroleum green supply chain: A game theoretic approach," *Journal of Cleaner Production*, vol. 170, pp. 818–841, 2018.

[18] A. Adam, N. Zakuan, S. A. U. A. A. Bichi, U. Shettima, M. Ali, R. B. Almasradi *et al.*, "Supply chain sustainability practices of oil servicing firms in the downstream sector of nigeria's oil and gas industry," *Journal of Economic Info*, vol. 6, no. 4, pp. 11–14, 2019.

[19] B. Wang, H. Zhang, M. Yuan, Z. Guo, and Y. Liang, "Sustainable refined products supply chain: a reliability assessment for demand-side management in primary distribution processes," *Energy Science & Engineering*, vol. 8, no. 4, pp. 1029–1049, 2020.

[20] S. Gao, J. Daaboul, and J. Le Duigou, "Process planning, scheduling, and layout optimization for multi-unit mass-customized products in sustainable reconfigurable manufacturing system," *Sustainability*, vol. 13, no. 23, p. 13323, 2021.

[21] M. Sahnoun, Y. Xu, F. B. Abdelaziz, and D. Baudry, "Optimization of transportation collaborative robots fleet size in flexible manufacturing systems," in *2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO)*. IEEE, 2019, pp. 1–5.

[22] J. E. Abu Qudeiri, "Production simulator system for flexible routing optimization in flexible manufacturing systems," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 231, no. 7, pp. 1237–1247, 2017.

[23] G. G. Castañé, H. Simonis, K. N. Brown, Y. Lin, C. Ozturk, M. Garraffa, and M. Antunes, "Simulation-based optimization tool for field service planning," in *2019 Winter Simulation Conference (WSC)*. IEEE, 2019, pp. 1684–1695.

[24] Y. Zhang, H. Zhu, D. Tang, T. Zhou, and Y. Gui, "Dynamic job shop scheduling based on deep reinforcement learning for multi-agent manufacturing systems," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102412, 2022.

[25] Y. G. Kim, S. Lee, J. Son, H. Bae, and B. Do Chung, "Multi-agent system and reinforcement learning approach for distributed intelligence

in a flexible smart manufacturing system," *Journal of Manufacturing Systems*, vol. 57, pp. 440–450, 2020.
[26] W. Wu, Z. Huang, J. Zeng, and K. Fan, "A fast decision-making method for process planning with dynamic machining resources via deep reinforcement learning," *Journal of manufacturing systems*, vol. 58, pp. 392–411, 2021.
[27] J. Tang and K. Salonitis, "A deep reinforcement learning based scheduling policy for reconfigurable manufacturing systems," *Procedia CIRP*, vol. 103, pp. 1–7, 2021, 9th CIRP Global Web Conference – Sustainable, resilient, and agile manufacturing and service operations : Lessons from COVID-19. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212827121008398
[28] T. Zhou, H. Zhu, D. Tang, C. Liu, Q. Cai, W. Shi, and Y. Gui, "Reinforcement learning for online optimization of job-shop scheduling in a smart manufacturing factory," *Advances in Mechanical Engineering*, vol. 14, no. 3, p. 16878132221086120, 2022.
[29] C. D. Hubbs, H. D. Perez, O. Sarwar, N. V. Sahinidis, I. E. Grossmann, and J. M. Wassick, "Or-gym: A reinforcement learning library for operations research problem," *CoRR*, vol. abs/2008.06319, 2020. [Online]. Available: https://arxiv.org/abs/2008.06319
[30] T. Zhou, H. Zhu, D. Tang, C. Liu, Q. Cai, W. Shi, and Y. Gui, "Reinforcement learning for online optimization of job-shop scheduling in a smart manufacturing factory," *Advances in Mechanical Engineering*, vol. 14, no. 3, p. 16878132221086120, 2022. [Online]. Available: https://doi.org/10.1177/16878132221086120
[31] M. A. Strøm, C. F. Rehn, S. S. Pettersen, S. O. Erikstad, B. E. Asbjørnslett, and P. O. Brett, "Combining design and strategy in offshore shipping," in *Marine Design XIII*. CRC Press, 2018, pp. 147–162.
[32] Y. Yuan, H. Xu, and J. Yang, "A hybrid harmony search algorithm for the flexible job shop scheduling problem," *Applied Soft Computing*, vol. 13, no. 7, pp. 3259–3272, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494613000677
[33] H. Milne, O. Adesina, R. Campbell, B. Friesen, and M. Khawaja, "Flexible job-shop scheduling with changeover priorities," in *Machine Learning, Optimization, and Data Science*, G. Nicosia, V. Ojha, E. La Malfa, G. La Malfa, G. Jansen, P. M. Pardalos, G. Giuffrida, and R. Umeton, Eds. Cham: Springer International Publishing, 2022, pp. 611–625.