APEM
journal

Advances in Production Engineering & Management

University of Maribor

Advances in Production Engineering & Management

Identification Statement

APEM Editorial

Advances in Production Engineering & Management

Contents

Scope and topics

Advances in Production Engineering & Management (*APEM journal*) is an interdisciplinary refereed international academic journal published quarterly by the *Chair of Production Engineering* at the *University of Maribor*. The main goal of the *APEM journal* is to present original, high quality, theoretical and application-oriented research developments in all areas of production engineering and production management to a broad audience of academics and practitioners. In order to bridge the gap between theory and practice, applications based on advanced theory and case studies are particularly welcome. For theoretical papers, their originality and research contributions are the main factors in the evaluation process. General approaches, formalisms, algorithms or techniques should be illustrated with significant applications that demonstrate their applicability to real-world problems. Although the *APEM journal* main goal is to publish original research papers, review articles and professional papers are occasionally published.

Fields of interest include, but are not limited to:

Additive Manufacturing Processes
Advanced Production Technologies
Artificial Intelligence in Production
Assembly Systems
Automation
Big Data in Production
Computer-Integrated Manufacturing
Cutting and Forming Processes
Decision Support Systems
Discrete Systems and Methodology
e-Manufacturing
Evolutionary Computation in Production
Fuzzy Systems
Human Factor Engineering, Ergonomics
Industrial Engineering
Industrial Processes
Industrial Robotics
Intelligent Manufacturing Systems
Joining Processes
Knowledge Management
Logistics in Production

Machine Learning in Production
Machine Tools
Machining Systems
Manufacturing Systems
Materials Science, Multidisciplinary
Mechanical Engineering
Mechatronics
Metrology in Production
Modelling and Simulation
Numerical Techniques
Operations Research
Operations Planning, Scheduling and Control
Optimisation Techniques
Project Management
Quality Management
Risk and Uncertainty
Self-Organizing Systems
Statistical Methods
Supply Chain Management
Virtual Reality in Production

Sustainable manufacturing – An overview and a conceptual framework for continuous transformation and competitiveness

Hussain, S.[a,*], Jahanzaib, M.[a]

[a]Industrial Engineering Department, University of Engineering and Technology, Taxila, Pakistan

A B S T R A C T

Sustainable manufacturing is advancing amidst the changing context and emerging paradigms. Business success and longterm sustainability today depends upon transforming continuously, and maintaining competitiveness through building context specific capabilities. To address these needs, this research presented a comprehensive overview of sustainable manufacturing, in first part, contributed by numerous aspects in the field. The second part proposes a conceptual framework comprising three interconnected elements: 'Ideal', 'Strategy', 'Architecture'. The ideal is best depicted as an exploration and choice context. Synthesis of stakeholders' desires and systematic discovery of opportunities, in context of larger containing systems, manifestes into desired attributes and characteristics of products, technology and enterprise system that are to be approached continuously. The strategy element is a match and transformation context. Strategic planning, focussed on continuous identification and building of capabilities, evolves into a broader concept of the business which enhances the firm's capacity to adapt to changing contexts and meet its proposed ends. The architecture is a function and execution context at the operational level. It combines capabilities, organization and operational structure, and value creation processes to perfom desired function in context of an agreed upon business concept. A systemic and potentially viable approach, embedded with specific capabilities to integrate sustainability into the core of a manufacturing business, is thus proposed which sets this research work apart. Contextual interrogation gives it a new way of constructing the big picture of the issue, i.e. sustainable manufacturing. This simplistic scheme is supplemented and guided by multi-aspect research in sustainable manufacturing, circular economy, capabilities, strategy and transformation, and systems thinking. The proposed framework is expected to fulfill the key needs of enterprises in sustainable manufacturing business, i.e. to transform and maintain competitiveness in a fast paced environment.

1. Introduction

Manufacturing is an indispensable feature of the modern society owing to its vital strength in terms of societal service, and pivotal role in contributing towards national and global economy. However, at the same time, these human activity systems collectively consume a considerable amount of resources, and generate wastes and emmissions thus affecting the environment and society [1]. It has been largely realized that natural resources and regenerative capacity of the ecosystem is not infinite. Current production and consumption patterns are being seen as largely responsible to assure a foreseeable future [2]. Consequently, manufacturing is evolving, within the framework of triple bottom line 3BL (environment, economy and society), for pursuing the big picture of sustainable development [3, 4].

ENVIRONMENT DOMAIN	ECONOMIC DOMAIN	SOCIAL DOMAIN
• Environmental issues (climate change, global warming, pollution, ozone layer depletion) • Eco-system concerns (waste and emissions, landfill, extinction of species, bio-diversity, restoration and conservation) • Policies, regulations and directives • Environmental performance • Impact assessment (frameworks, e.g. LCA) • Design philosophies (DfE, bio-mimicry, cradle to cradle etc.) • Resource conservation (minerals, energy, water) • Material and energy tracking (frameworks, e.g. materials flow analysis, energy flow analysis • Strategies (waste minimization, resource efficiency, eco-efficiency)	• Economic paradigm (e.g. Circular Economy) • Socioeconomic trends • Production and consumption patterns • Policies, business strategy, business model • Economic growth and development • Economic advantage, competitive advantage • Financial performance (value added, stakeholders' return, profit, potential financial benefits etc.) • Economic performance (products/operations/facilities) • Product lifecycle costs (frameworks, e.g. LCC) • Economic impact (products, manufacturing chain)	• Societal needs and values, issues and trends, lifestyle and culture • Social interaction and collaboration, feedback and inputs, perspective • Policies, regulations and directives • Organization and social behavior • Social performance, responsibility, reputation • Social value, social benefits (local, national, global) • Product benefits (customer, consumer, society) • Social equity, standard of living, quality of life • Health and safety , working conditions, employment opportunities, education and training, community well-being, • Social impact (frameworks, e.g. SLCA)

Fig. 1 Sustainability aspects relevant to manufacturing business

Sustainable Manufacturing (SM from here on) involves simultaneous consideration of multiple factors and their interaction across three dimensions. Interconnectivity and co-dependence of firms with others in the entire value network even inceases the breadth and overall complexity of the issue (SM), attributed to increased plurality and high degree of interdependence between elements [5-7]. Dealing with such a complex system necessitates a holistic understanding of the issue. Systems Thinking (ST from here on) is being recognized as one of the key competencies to address sustainability problems systemically across multiple domains and at different scales [8]. ST promotes holism which provides theoretical awareness of the system and avoids unintended consequences of sub-optimization [9, 10]. Sustainability, viewed as a system property, implies analysis and creation of sustainable value from the perspective of larger containing systems [11]. Yet, balancing the needs and benefits for 3BL requires robust methods incorporating an integratd analysis and evaluation of interaction among system elements. Nevertheless, exploring through vast literature, various aspects relevant to the issue (SM) are presented in three domains in Fig. 1.

Manufacturing enterprises, among others, are faced with an increasing complexity, uncertainty, change and diversity. There are a number of factors whose interplay is supposed to give emergence to this complex and diverse milieu. Naming few are; constantly shifting external environment, growing environmental and social concerns, inter-connectedness of industrial problems, faster learning and development needs, and transformation in societal needs etc. Additionally, there are concerns on rapidly depleting natural resources, and increasing price and volatility of raw materials and commodities [12]. Moreover, it has been realized that efficient manufacturing processes, technological innovation, optimization abilities, and other success factors in context of industrial production (e.g. economy of scale) are insufficient to ameliorate the overall situation [13]. Further, it has been learnt that gradual improvements at the level of product or process are inadequate to embrace long term business success and sustainability.

Sustainability frameworks and approaches, in relation to manufacturing, are generally aimed at development and implementation of sustainability strategy, sustainable business models, design and innovation capabilities, performance improvement and assessment, and decision making tools etc. Many of the proposed frameworks emphasize systemic management of sustainability concepts and requirements in context of broader systems, and a deep involvement of stakeholders. The proposed frameworks and approaches demonstrate some utility in different industrial aspects. But they lack, generally, in managing the issue (SM) holistically towards continuous

transformation and competitiveness in manufacturing enterprises. Moreover, despite widespread literature and enormous research on the issue, yet there is need to explore approaches which are potentially viable amidst the changing contexts to ensure long term sustainability. Nevertheless, researchers are learning about the capabilities of various frameworks that can bring about a change at the level of entire enterprise as a system. Numerous propositions can further increase the speed of learning and implementation.

SM businesses have to seek ways to contribute to sustainable development objectives effectively, now and in the future, while advancing amidst the changing contexts. This needs continuous support and guidance that could systematically enable the enterprise's capability to meet its desired sustainability ends. Successful transformation requires more relevant and focused designs, and a mind-set change in relation to existential purpose of the firm, its stakeholders, society and environment. A framework, however, may be established that serves as a guideline for the enterprise to create and re-create itself, to fit the proposed ends. The conceptual framewok proposed in this research is an abstract concept structured by three interacting elements: 'Ideal', 'Strategy', and 'Architecture'. The 'Ideal' or ideal intent of the enterprise is chosen by stakeholders through exploration of opportunities within the contextual space. This proposition may be sussessfully realized provided the enterprise system is technologically and operationally capable enough and the one that can learn and adapt effectively. The 'Strategy' or strategic intent is formulated in context of ideal intent. It represents a mix of manufacturing and business level capabilities whose development and strategic deployment is potentially vital to gain competitive advantage and capture the proposed ends at large. Strategic planning and decision making is aimed towards evolving an innovative business concept that is operationally viable. The 'Architecture' at down stream converts this concept into reality. It formally describes the overall business architecture framework for SM including its functional aspects, structural issues, and value creation processes and activities. This simplistic scheme may herald long term business success and sustainability provided the concept is operationalized in a systematic and meaningful manner. The framework design and function is complemented by emerging concepts and approaches in ST, SM research and Circular Economy (CE from here on), capabilities, and various frameworks for strategy and architectrure. Contextual interogation gives it a new way of constructing the big picture of the issue, i.e SM.

Rest of the paper unfolds this scheme in quite a natural order. After the conceptual framework has been introduced in this section, the next section provides a comprehensive overview of the issue (SM) in various aspects and from multiple perspectives, to construct its contextual position. Section 3 describes proposed framework and its individual elements in detail. The relationship of proposed conceptual approach with social environment and level of economic development in a country along with its operationalization and main obstacles pertinent to a specific environment are discussed in Section 4. Finally, major findings from our discussion are drawn in Section 5, which concludes this research.

2. Sustainable manufacturing: An overview

SM has been received with even greater recognition in recent years as a comprehensive strategy for improving the economic, social and environmental performance of a firm. New management concepts (product lifecycle management and assessment, eco-effectiveness etc) , innovative technologies, and design and engineering approaches have been introduced to incorporate sustainability in manufacturing [14]. Research frameworks encompass numerous aspects at varying scale (enterprise, supply chain, manufacturing cell and operations etc.). Adopting a system design approach, Tonelli *et al.* [15] proposed a framework for sustainable industrial system involving supply chain design, sustainability performance measurement and management, and organisational change . Using DMAIC (Define, Measure, Analyze, Improve, Control) approach, Koho *et al.* [16] proposed a concept aimed at supporting Finnish manufacturing companies in realizing sustainable production while emphasizing measurement as basis of improvement. Subic *et al.* [17] developed a framework focused on development and assessment of relevant capabilities across the entire supply chain. Zhang *et al.* [7] yearned for a framework, in line with LCA and

SLCA, to integrate sustainability assessment with decision making methodology that assists in process planning at the production cell level. For sustainability in machining operations, Masood *et al.* [18] proposed a milling technique with the aim to enhance cutting tool life. Hassine *et al.* [19] presented a case study on sustainable turning operation utilizing a specifically developed multi objective optimization technique.

An emerging research interest is observed on creating value through sustainable industrial production systems (Sustainable Supply Chains SSCs and Closed Loop Supply Chains CLSCs). Govindan *et al.* [20] highlighted that reverse logistics (RL) and CLSC is now a revenue opportunity instead of a cost-minimization approach. Schenkel *et al.* [21] segregated value creation into four types of values (economic, environmental, information and customer) and presented an insight on maximizing the value by coordinating the forward and reverse supply chains. Ageron *et al.* [22] recognized that sustainability issues when integrated in SCM, as an intended strategy, contribute to business in terms of competitive advantage. Gopal Krishnan *et al.* [4] argued the necessity of devising sustainable operations along the value chain and approaching the 3BL in an integrated manner. Galal *et al.* [23] studied sustainability, mainly emission reduction, across an agri-food supply chain by developing a discrete-event simulation model.

From the system's perspective, SM needs to be analyzed in context of larger containing systems [11]. Understanding of external interactions becomes significant becacause of a broad range of factors across the product life cycle and their connectivity with broader systems (supply chain, market, and ecosystem) [24]. Fiksel [25] developed a triple value 3V model by dividing the physical world into three types of interacting systems (industrial, societal, and environmental) ,wherein the intricate dynamic linkages and flows of value among them may provide an architecture for system modelling. On the concepts of complexity theory, Moldavska [26] proposed a systemic model whereby a system contributes to not only its own sustainability, defined in terms of attractors, but also to that of larger containing system. From an operational perspective, Zhang *et al.* [7] proposed definitions and assessment framework for SM based on ST. Moldavska *et al.* [27] proposed a framework for sustainability assessment system for manufacturing organizations , guided by ST.

A sustainable product is just as effective and at least the same quality as the product it is replacing while produced using fewer resources and better materials. In addition to product's attributes, SM focuses on how the product is made, with the aim to minimize its impact. In a lifecycle view, it includes input materials, product design, production and distribution processes, and end of life (EoL) management. Within Lifecycle Thinking framework, manufacturers look beyond the supply chain to maximize the product performance throughout the life cycle. Lifecycle Engineering (LCE) is aimed at designing products that meet requirements for quality, cost, manufacturability, and consumer appeal, while at the same time minimising of environmental impacts. Product life cycle, if analysed systematically, may drive advancements in product and process technology, and other improvement strategies. Sharing and managing of product data, information and knowledge across this entire lifecycle forms the essence of Product Lifecycle Management (PLM) which seeks to integrate stakeholders, processes, resources and information in a product-centric knowledge environment, to make informed product decisions [28].

SM Principles include; using less energy and materials, using renewable energy, using fewer hazardous materials and toxic chemicals, closing the loop of resource flow, and designing products that can be easily and economically reused or re-manufactured and consume less energy during use [15]. Similar principles have been proposed by other researchers towards achieving sustainable ends. Alayon *et al.* [29] referred to 9 principles of sustainable production relevant to products, energy and material; economic performance; worker and community development; and natural environment. Esmaeilian *et al.* [30] reffered to seven main steps proposed by Ray Anderson, to become what he called a 'Prototypical Company of the 21st Century'. For long-term sustainability, economic growth needs to be decoupled from environmental constraints through technological changes and innovation in industrial systems [31]. CE and C2C are relevant frameworks here. Based on principles of Green Engineering, they approach sustainability from an eco-effective design perspective. Principles together with relatively new economic paradigm

of CE, and manufacturing paradigms (e.g. C2C, industry 4.0), may create sustainable competitive advantage alongside a positive impact on environment and society.

Value creation in CE, as popularized by EMF, is based principally on elimination of waste and use of renewable energy during the entire lifecycle. Resource yields are optimised by maximizing the number of technical cycles (repair, reuse, or remanufacturing) and the time spent in each of the multiple use cycles of product. Increased inflow of virgin materials into the economy may be substituted by cascading or diversifying the reuse of products across value chain. Moreover, the use of uncontaminated material streams (pure inputs) can increase collection and redistribution efficiency, and maintain material quality thus improving product longevity and material productivity. Furthermore, the principle of dematerialization(minimizing comparative materials use) is at play during processes of design, making and distribution of products to ensure reuseability with lesser changes, faster returns, and higher potential savings [12]. Contrary to the linear production pattern (take, make, waste), this regenerative mode is manifested in higher resource utility and optimization while minimising supply risks and enhancing of natural capital [32]. C2C design protocol, based on the strategy of designing ecologically intelligent products and processes, offers a good premise for CE. It redefines the problem to address its source, and shifts the end of pipe liabilities to product design. A design of industrial systems on this concept necessitates flow of materials in safe, regenerative, closed-loop cycles. To achieve an optimal effectiveness close to the natural system, inputs and outputs should be safe and beneficial [33, 34]. Industry 4.0 is an evolutionary industrial paradigm centered on achieving quality, efficiency and flexibility at the same time. Businesses need to converge their capabilities, to match with this era of fast paced and broad scale disruptions [35]. The manufacturing systems in 'Industry 4.0' utilize information and communication technology (ICT from here on) infrastructure in an internet of things to deliver smart products. Cross-linked value creation modules embedded with ICT and Cyber-Physical Systems (CPS) enable efficient allocation of resources , decentralized decision-making, and efficient work design and organization [36].

SM practices are aimed at reduction in each of energy, water, waste and emissions, improved product quality, enhanced corporate image and market competitiveness, better access to new markets, reduced costs, and increased profits. Practices range from very simple process improvements to innovative product design [3], and large investments in new technologies for cleaner production. Creation of wildlife habitats, promotion of community engagement and other external practices e.g. product stewardship can help integrate stakeholders' views into the business operations while contributing to a more sustainable world [37]. Firms primarily focus on conserving resource use and reducing physical waste which often widens to lifecycle sustainability through adoption of suitable policies and tools whereby an initial focus on internal operations grows to involve entire supply chain.

Integrating SM concepts and strategies into the core of business requires long-term decision making [29] and radically rethinking the existing business models. In a system's view, activities across all business functions need to be focussed and coordinated instead of isolated improvements in either of product design, manufacturing technology or product end-of-life management. SM strategies can be broadly categorized under the two themes or dimensions [14]; Eco-efficiency is aimed at increasing economic development through lowering of resource intensity throughout the life cycle [38] while Eco-effectivity is directly tied to the reduction of environmental or social burdens [14]. Strategies take on a different perspective at each level of enterprise. At the manufacturing process level, the focus is on improving efficiency and quality, optimization of process parameters, and reduction in processing cost. At the supply chain level, consideration may be given to the use of renewable energy across the chain, recirculation of products and materials into product lifecycles, and effective management of raw materials and services for an improvement in social and environmental impacts. At the enterprise level, the focus is on policy analysis, sustainable business development, and achievement of business objectives like productivity, competitiveness and profitability. Firms may struggle at the onset of their efforts towards SM. Yet many firms embrace an early success. Their efforts and the manner in which certain intiatives were taken could be mimiced and modelled for greater understanding.

Moreover, challenges encountered by others in the course of successful transformation may be helpful from the perspective of re-using the relevant experiences and knowledge.

Manufacturing businesses need support and guidance if they are to contribute to sustainable ends. Challenges and barriers faced need to be clarified. Future challenges include competition in terms of both inputs (resources) and outputs (products and services) , diversity of customer base, rapidly changing consumer lifestyles and preferences, increasing business risks and uncertainties, adherence to regulations, avoiding use of technologies that have adverse impacts on environment, and focussing on total economic, social, and environmental cost of a product instead of low cost production [39]. The financial aspect of sustainability initiatives can be an enormous obstacle to the overall progress. Capital investment decisions need to be based on cost benefit justification because technologies required in reducing plant emissions to zero might be both complex and costly. Other commonly encountered barriers include time and resource constraints, general business concerns, and corporate culture. Aligning internal company culture with sustainability objectives in day-to-day business operations is not an easy. Manufacturing businesses need fundamental changes in their missions, and significant improvements in ways of doing business [16]. Existing strategies, procedures and standards might need a radical transformation to adapt to this new role. Transformation might need a redesign of the organization which may pose difficulties in several aspects e.g. defining enterprise structure, boundaries, decision making processes, and resolving conflicting goals etc. [40].

Manufacturing businesses today are coped with challenges of increasing product variety and shortening product life cycles [5, 41]. While large and diversified product-mix can capture additional business opportunities, added economic benefits and enhanced consumer value, it also poses increased complexity and unpredictability to the manufacturing system. In case of SM, the concerns are even larger in terms of managing variety across the entire product life cycle. Manufacturing system's flexibility and adaptability is a critical requirement to accommodate changes brought about due to management of multiple generations of a variety of products. From the economic perspective, effective resource utilization, and integration of product design with manufacturing processes will have to be ensured to attain sustainable competitive advantage in these environments [5].

Sustainability assessment, performance measurement and reporting is a challenging task in manufacturing organizations [7, 27] due to the complexity and dynamics of manufacturing operations and diverse theoretical perspectives on sustainability requirements, often normative in nature [24]. WBCSD, GRI, OECD, ISO, UNEP etc. have developed standards and guidelines, yet there are difficulties in several aspects. GRI's general as well as sector-specific assessment and reporting guidelines need customization for their befitting to the needs of a firm and its stakeholders. Life Cycle Thinking (LCT), a more popular framework, integrates three elements-Lifecycle Assessment LCA, Lifecycle Costing LCC, and Social Lifecycle Assessment SLCA- which together provide a methodological framework for Life Cycle Sustainability Assessment (LCSA) [42, 43]. Although an LCA study has certain limitations and challenges, it is a robust and most widely used framework for evaluation of potential impacts of a product during entire life cycle [7]. However, SLCA and LCC methods are relatively less developed and lack scientific consensus - too few tools and fully developed databases are available - hence rarely conducted in process industries [43]. Despite huge amount of research in sustaiability assessment, there is an absence of comprehensive, universally accepted and widely applicable metrics. Organizations find limitations in terms of interpretation and implementation capabilities relevant to a variety of tools, indicators, and metrics on the list. Organizations are even recognizing the importance of developing sustainability tools and metrics specific to their unique business environment [27] . The real value of assessment lies in attaining of sustainability objectives and providing valuable input to decision-makers rather than external reporting only [27, 44]. To avoid sub-optimization, there is a need to develop frameworks and tools, from a unified and holistic perspective, capable to assess and monitor the entire organization as a system instead of focussing on individual areas and functions. Moreover, the value added through such an endeavour should be in balance with the resources required [27].

3. Proposed conceptual framework

Continuous renewal of enterprise intent and translating it into an implementation plan is a contested problem [45]. Understanding the nature and purpose of enterprise, as a starting point, may guide in establishing requirements from a functional perspective. An enterprise as a system consists of structure and processes which collectively perform economic function, characterized by products and services delivered in context of an environment [46, 47]. The principal objective of an enterprise is the 'development' that serves to contribute to the development of itself and the larger containing systems. Greater this capacitation, better the firms can attain their objectives [48].

The framework proposed in this research is based on enterprise's desire to contribute to the sustainability of society, environment and economy, and stay compettitive in relation to changing contextual position of SM. As presented in Fig. 2, the framework as a concept system contains three functional elements or subsystems interrelated with each other: 'Ideal', 'Strategy', 'Architecture', and is it self part of a larger whole, i.e the enterprise. The ideal is best depicted as an exploration and choice context. Systematic discovery of opportunities within all important business dimensions and a coevolved proposition of products and enterprise system are what underlye a holistic and clarified vision embedded with stakeholders' desires. The enhanced capacity to sense opportunities manifests into an improved understanging of what satisfies stakeholders, and which attributes (of product, system) can lead competitive advantage? The ideal intent thus evolved serves as a reference to be approached continuously throughout the transformation cycle. The strategy element is best represented as a match and transformation context. Strategic planning is aimed at identification of capabilities corresponding with opportunities; formulation of policies and guidelines to enble an effective and efficient use of enterprise's means; setting of business objectives and goals; and commitment of resources and assets. The desirable consequence of this planning process is a broader concept of the business. Continuous identification and development of capabilities enhances capacity to adapt to changing contexts. The effectiveness is reflected in the speed and degree to which specific capabilities can be created and matched with opportunities. The architecture, at the operational level, is a function and execution context. It combines capabilities, organization structure and business processes to perfom desired functions in context of an agreed upon business concept.

The context of the issue (SM) and hence the ideal intent keep changing. In response, the strategic planning reformulates the business concept and renews relevant capabilities. The architecture reconfigures capabilities and assets to adapt to changing business needs. This continuous process harnesses the capabilities and collective learning of the enterprise as a whole under dynamic business requirements. By operationalizing the individual elements and managing interaction among them, a manufacturing enterprise can be transformed continuously and systematically in relation to its desired sustainability ends. In subsequent part individual elements are explained in terms of their relationship with each other and with larger systems. Various aspects of SM are discussed alongside as deemed relevant to the nature and scope of each element.
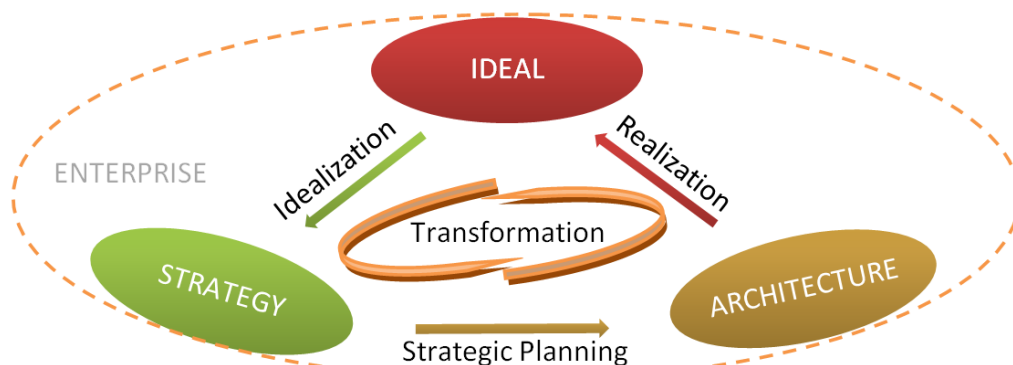

Fig. 2 Structural representation of framework

3.1 Ideal

Sustainable products should not only be safe for environment but also create opportunities for life-systems, in addition to satisfying the needs of business and society. Determining what the environment (life system) actually wants, to attain its own ends, is beyond scientific deliberation, and seems unattainable. Nevertheless, businesses can approach this limit by consistently turning social needs and desires to environmental advantage specifically, rather merely creating opportunities that could positively impact [49]. This could serve as an ideal to be approached endlessly. Apple has recently pronounced to move its entire supply chain to closed loop and renewable energy in the wake of undeniable climate change and depleting earth resources. Though it seems unattainable currently as the report says, "It sounds crazy, but we're working on it", Apple has embarked on its journey by launching projects related to capabilities enhancement and technology development [50].

An SM enterprise delivers in context of a hierarchy of larger systems: industry (economic system); marketplace and society (social system); and eco-system or environment. Stakeholders' interests can significantly drive manufacturing towards sustainability and competitiveness [51]. This whole system's view helps understand the way a manufacturing enterprise interacts with its surroundings, throughout the entire life cycle of products. More pragmatically, however, the ideal can be chosen in relation to opportunities within present and future context of SM, continuously changing business environment, and SM principles, practices and industry callenges. Because external environment keeps changing, it necessitates proposing of products and technology in a larger context, i.e. emerging social and environmental needs, and trends and changes in business environment (e.g industry drivers, technological innovation, product lines and competitive priorities etc.). This concept is presented in Fig. 3. Through this systemic inquiry, stakeholders' desires can be translated into full range of potential business opportunities, and desired properties and boundaries of enterprise system. To validate the content of a proposed ideal, interactive planning (IP) suggests two means: 1) inter-subjectivity and consensus among stakeholders with regard to a proposal, and 2) experimental feasibility [52].

Creating a vision is a key requirement for successful organizational change, consistent with Ackoff's notion of an idealized design. Re-design and transformation are desired consequences of a thoughtfully synthesized vision. External vision (insights, beliefs, and assumptions) is based on interpretation of all the factors in the context that have an impact on firm's operations and performance. Context can be viewed as a field of related parts [53], constructed for an issue, say SM. Descriptions of contextual pattern within this socio-ecological field form the present context. In regards to future, scenarios are chosen that describe alternative, comparably plausible lines of changing overall field conditions. Although there is no way of designating the one future that is most likely to emerge [54], scenarios can facilitate cognition of multiple possible futures in concrete detail and guide the context specific design of new business concepts or innovating around existing ones. Business environment is a sort of design space, a context in which to conceive or adapt the business. In fast-paced competitive environments, assumptions about how market forces, industry forces and emerging trends unfold give us the design space to develop potential business options [55].

The boradened scope of activities in SM, and large number of actors along the value chain necessitate all concerned to have a voice in decisions pertaining to products, technology, market, and enterprise system at large. Guided by System's methods relevant to stakeholders' engagement, and Ackoff's interactive planning (IP) and his view of the organization as social systemic in nature, the process of co-creating the ideal intent can be made more valuable. Firm's stakeholders and their participation are centric to this process. We need to understand stakeholders' needs and interests, and pregressively combine them. By establishing a generative context for the engagement of diverse perspectives of multiple stakeholders, a synergistic proposition can be evolved that is meaningful in a context of use [56]. An important insight is that no participative process can include every possible perspective, i.e. comprehension is impossible. In an interconnected world, knowing about any situation has limits or boundaries [57]. In relation to system boundary, we need to determine what is in our control, who (people and issues) is to include, how to justify exclusions, and how to address marginalization.
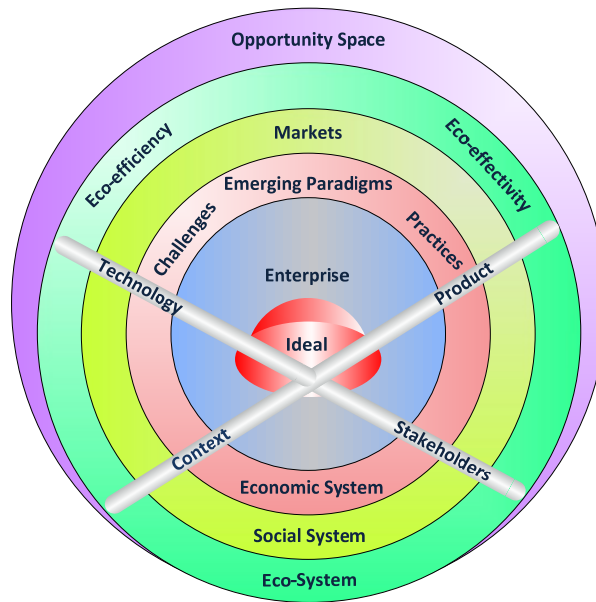
Fig. 3 Evolving system ideal

3.2 Strategy

An enterprise's quest towards its ultimate ends is corresponding with the capabilities possessed by its ends-seeking system. Strategic intent of the firm is sought mainly to convey the purpose of the enterprise and how it is expected to perform in relation to those ends. Strategic objectives set the direction for designing business solutions and shaping operational structure [58]. Policies govern the selection of means and instruments by which the objectives are to be pursued, and principles are formulations of values to be preserved in such selections [48]. Principles provide guidance in making strategically consistent choices from the perspective of sustainability. Capabilities, derived from strategic principles, are a core concept to communicate and realize strategic intent. Building of capabilities has emerged as a new approach under strategic planning [59] to maintain competitive advantage and attain long-term business sustainability in a turbulent environment. A firm needs specific capabilities to enhance its capacity to create sustainable value and adapt to changing circumstances faster than competitors. The strategic planning process, using ideal intent as as a set of inputs, yields strategic intent manifested in an agreed upon business concept. This process is presented in Fig. 4. Capabilities identified in this process are built alongside to enable meaningful execution of this business concept. Capabilities are superior business processes, represented as a set of knowledge, skills, and competences [60]. The capability of organization as a whole is a high-level routine (or collection of routines) which combines tasks from different business functions into a particular output or solution of significant importance to the firm [61]. Capabilities not only provide a platform for strategy but also impact growth and profitability which in turn provide means for further strengthing of them. For the manufacturing sector this view of strategy has important bearings on organization's skill base, and dynamic decision making especially on infrastructural aspects [59]. The principles, policies,

Fig. 4 Evolving strategic intent

and systems of actions, underlying the infrastructural choices (Organization, manufacturing planning and control, quality, new product introduction and human resources), govern and reinforce the capabilities and resources that affect a firm's operations.

Strategic importance of building capabilities is being realized beyond mere competitive reasons due to a shift in strategy focus from decreasing costs to increasing added value [62]. To embrace sustainability as a new business opportunity in manufacturing, a firm's key management function is to bridge the gap between strategic intent and required capabilities. Capabilities need to be developed in all of the important business dimensions (product, market, technology etc.) to capture entire range of opportunities. The future of manufacturing industry is characterized by capabilities in skills development, CE, technology and innovation, supply chain integration and digitization. Pre-production (R&D, product design) capabilities include those to develope innovative product concepts and its supporting technology. Within production segment, overall capabilities can be added to by undertaking fabrication of increasingly complex and innovative products, flexibile operations and optimized process planning, and improving resource efficiency through new fabrication methods. In post-production segment, capabilities are required in terms of efficiency and reliability to perform marketing, distribution, and EoL activities [35]. Both the individuals' learning capacity (competence) and capabilities of the organization are central to innovating the business in more sustainable ways and accommodate new needs of sustainability. Table 1 illustrates these capabilities along with salient requirements to attain them.

Table 1 Capabilities for sustainable manufacturing

Capability	Salient requirements
Integrating sustainability into the core of business	• Adopt a business perspective, and transform the traditional business mindset. • Innovation in business to turn environmental and social concerns to competitive advantage without incurring increased costs. • Articulate and maintain a holistic view of sustainable value aimed at proposing a balanced value and distribution of economic costs and benefits among all actors. • Judicious commitment of financial resources to investment opportunities. • Create differentiated and innovative business models and enhance capacity to create, adjust and replace business models.
Identification and evaluation of opportunities	• Gathering and filtering of information from both inside and out side the enterprise, and from core as well as the periphery of relevant business environment. • Comprehending the likely evolution of industry, technology, and marketplace. • Establish search and discovery methods and procedures, and allocate resources. • Enhance interpretive skills of individuals and learning capacity of the organization. • Aassessment of relevant opportunities through decision rules and informed conjectures.
Building of individual skills and competence	• Education and training focussed mainly on internalization of sustainability concepts, creative and innovation abilities, functional capabilities and life-cycle-thinking. • Building of cross-domain skills (technology, engineering, electronics, robotics, computational and computer sciences) to perform manufacturing functions in industries of future. • Cost-effective, innovative and exploratory teaching methods e.g. in-person training and coaching, experiential environments and digital methods. • Adopting a stuctured approach for capability alignment and assessment, i.e. tools, methods, procedures, quantitative targets, credible metrics etc.
Operationalization of CE	• Culture adaptation, transition processes, and strategic direction to appreciate difficulties. • Technological capabilities for an improved reverse logistic set-up aimed at high value recovery. • Strategic use of ICT to establish an efficient collaboration and knowledge sharing. • Capitalizing on information value to accelerate innovation in design, and identify more opportunities for additional value creation across the chain. • Product life cycle planning with a clear focus on product reuse/re-manufacturing. • Understanding materials (formulation, quality, innovation etc.), and a broader integration of material science into the business case and across value chain.
Product design & development	• Effective management of design process within a lifecycle thinking framework. • Capture and reuse the knowledge acquired throughout the lifecycle to modify existing designs. • Understanding the criticality of decisions at early stages of design in view of lifecycle costs. • Use fewer materials, renewable materials, non-hazardous materials, and substitute materials. • Evaluate materials against chosen criteria e.g. service life, recoverability, separateability etc. • Design products that consume lesser resources during production, packing, distribution, recovery, reuse, re-manufacture and recycling. • Evaluate competing product designs for price and ecological impact. • Design products that are re-usable, re-manufacturable, recyclable, and bio-degradable.

Table 1 (Continuation)

	• Evaluate reusability, and remanufacturability of design alternatives via pre-determined criteria e.g. technological and economic feasibility, and environmental benefits. • LCA, cost-benefit analysis and financial modelling to reveal potential designs for incorporation. • Incorporate design for sustainability (DfX) and other strategies e.g. design for upgrade, design for disassembly, design for modularity, and design for reliability etc.
Production and delivery	• Innovation and flexibility in production and process technology. • Choose and develop production processes that are efficient, resource conserving, processing of minimum material, and less polluting. • Understand technology, energy and material as three interrelated components, vital to improving resource efficiency. • Effective deployment and use of innovative technologies (e.g. ICT) for process control, energy saving, production efficiency, and safety etc. • Efficient transportation and logistics to reduce environmental impact and cost of distribution.
End of life (EoL) management	• Establishment of reverse logistics activities, and their alignment with after market and remarketting businesses. • User friendly and cost efficient collection and segregation depending on product quality, and effective segmentation to support dynamic recovery and economics of value capture. • Capitalize on knowledge acquired during lifecycle to facilitate EoL processes. • Devise processes that enable product re-usability with minimum possible resources to gain best environmental and economic advantages. • Leverage external resources to increase the throughput and performance of EoL processes e.g. outsourcing the core acquisition, integrating the supply chain with aftermarket divisions, and accomplishing economies of scale in re-manufacturing by sharing resources with others.
Standards	• Establish norms and rules that lead to a significant reduction of activities and transaction costs. • Strategic use of standards to add to the firm's capabilities and manufacturing base, re-engineering of the technology and processes, and redesign of products. • Adoption of international standards towards a faster knowledge acquisition and technological transition. • Establishing uniform standards along the value chain for standardization of product-related information, complete life cycle assessment of products, sustainability performance evaluation of all upstream and downstream partners, and optimization of end product sustainability. • Efficient collaboration and data exchange, distribution of right information in a fast and reliable way, and holistic management of complex tasks through the use of communication standards.
Allignment and collaboration	• Extending capabilities in terms of exploiting additional resources through broad-based external search and close interaction and collaboration with other enterprises, entities and institutions in the business environment and society. • Strategic involvement of customers in product design and development with the aim to include their perspective in value proposition, discover new opportunities and business models, develop more successful and radical value creation solutions, reduce product costs, anticipate the potential for applying new technologies, and retain customers longer. • Extending capabilities in terms of exploiting additional resources or shifting certain activities (e.g. procurement, product design etc.). • Establishing strategic partnering with supplier to capture benefits of economies of scale by outsourcing to contract manufacturers and large suppliers, assimiliating the capability of suppliers in terms of those capabilities that are closely linked and mutually reinforcing, capitalizing on rapid innovations by component suppliers ahead of the competition, improving supplier operations, and optimizing supplier-manufacturer interface to reduce material costs and gain a better insight on environmental impact reduction opportunities along the supply chain. • Leveraging benefits of co-specialization and complementary innovations towards differentiated product offerings and unique cost savings, improved performance, sustainable competitive advantage, and enhanced entrepreneurial capacities.

3.3 Architecture

The architecture serves the purpose of realizing an agreed upon business concept, formulated as part of firm's strategic intent. It links business strategy with business roles and processes, and structures the responsibility over business activities along the value chain [46]. It determines effectiveness in attaining firm's objectives keeping in view their strategic impact. From a functional perspective, it manages interaction among structure, processes and activities, and resources and capabilities to deliver sustainable value to all stakeholders. This concept is presented in Fig. 5. Within its functional framework, it utilizes a common and preferably customized knowledge base that allows manipulation of information about different business aspects from a variety of perspectives.
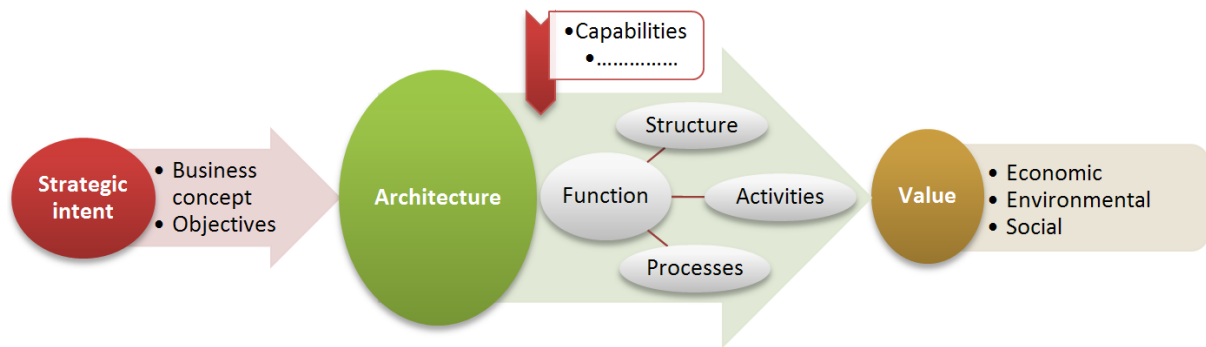
Fig. 5 Architecture framework for sustainable value creation

In system's view, architecture design involves a holistic understanding of how function, possible structures and processes should interact and communicate with each other to produce desired outputs, within a larger context. The planning, learning, and control function is thus an integral element here, responsible to evaluate the effectiveness of organization as a whole [10]. In functional design process, a comprehensive review of the strategic intent is conducted first to understand business priorities, relationship among objectives and transformation requirements. BA role is then determined in relation to business needs including required capabilities and core activities. The detailed business analysis next establishes a clear view of the business in terms of its organization and value chain structure. The final solution dictates initiatives (e.g. investment analysis, new product rollout, capability outsourcing etc.) and measurement criteria relevant to set objectives. This process, however, is ongoing and continues to redeem the once finalzed structure, processes, activities and even the business concept. Analogous to Checkland's [63] abstract concept of 'system as an adaptive whole', an architecture responds to the changing needs and requirements, interactively manages the value creation factors, and establishes a fit between problem space and proposed solution. The desirable emergent properties characterizing such a system could be e.g. a unique and differentiated business value or improved sustainability performance etc.

Environmental and social concerns today, changing market needs, technological innovations, and capabilities based competition are driving manufacturing businesses towards sustainability. Structural flexibility on the other hand must complement the prevalent dimension of the competition; based on either of new product offerings, market dominance or technological leadership, towards an efficient and conflict free internal environment. A multidimensional mode of structure – capable of producing different outcomes in the same or different environments – can interactively manage the interdependent business dimensions thus eliminating the need for a change of emphasis from one business orientation to another [10, 48]. [45] Strategic intent is interdependent with firm's operational structure which establishes relations between people, process, technology, managerial aspects, and information. The operational structure must ensure that wealth creation, cost reduction, and environmental and social perfromance are improved simultaneously to support the sustainability strategy.

Processes are primarily concerned with how the actual output(s) is to be produced through an integrated chain of activities performed by different groups in an organization, with a strong interface among them. Processes are technologically driven and subject to continuous change and improvement. A firm must comprehend the relevance of technologies to emerging competitive challenges, understand the flow and interface between system elements, and acquire operational knowledge of processes. In addition, these technological processes should maintain compatibility with already in place organizational processes that are concerned with creating integration, alignment and synergy among the parts of an organization [10]. A firm's core competence (unique resources and strengths) drives its core processes. In case of SM, the core processes in the chain include design and development of new products, production and delivery of products, and EOL management. Salient requirements in terms of capability of these processes have been illustrated in Table 1. Each process must be designed to achieve its priorities. Operational effectiveness of these processes may be assessed with key metrics corresponding to those

key activities in the value chain that influence the critical dimensions of sustainability performance. In addition to key activities, the firm needs to perform other essential activities e.g. to train its workforce, to develop new capabilities, and to link new activities to the existing activity system [64]. The nature of products and services, and the activities performed to deliver them collectively determine the structure of value chain [46]. The choices on content, structure and governance of activities shape the architecture in terms of where these are to be performed and what kind of resources are required to perform them? The entire collection of activities can be managed conveniently by aggregating them at different levels (major functions at top-level and specific activities at lowest level) [64], or segregating them according to their core, supporting or peripheral nature.

4. Discussion

The proposed conceptual approach rests on some key tenets, elaborated in section 3, which collectively could enable long-term sustainability (ability to transform and maintain competitiveness) for manufacturing businesses; among them greater capacitation of businesses to contribute to sustainable development, understanding stakeholders' needs and intersts, a coevolved vision and business proposition, enhanced capacity to discover opportunities systematically, learning capacity, competence and adptability of individuals and organization, building of context specific capabilities in all core segments of the business and their strategic use towards creating added value, innovative solutions and faster responce to changing market needs and technological challenges, and holistic understanding of problems and solutions in a larger context to improve economic, environmental and social perfromance simultaneously . The existing social environment and level of economic development of a country could be relevant/ significant in view of implementation of proposed concept across its entire manufacturing sector. Specific strategies and policies that consider both economic growth and social development are vital to nurture an environment for sustainable business development. Leading a competitive advantage demands educated and competent manpower. Quality of education has a direct impact on skill levels of individuals. Equal opportunities for education, training, employment and health increase productivity and growth. Exploration and capturing of new business opportunities is a knowledge intensive and enterpreneurial endeavour. National policies for matching skills with business needs could add to the ability of businesses to transform and create new value. Strong financial means in a country and better access to those, along with other supporting conditions, may encourage entrepreneurship and pursuits of competitiveness and innovativeness. National spending on research and development could pave the way for strong collaborative networks in a society which foster co-creation.

The proposed conceptual framework serves as a guideline in enabling continuous transformation and competitiveness of a manufacturing enterprise. However, it is not an implementation methodology or a procedure in itself that explicilty defines any preference as to how to get there. It will be instantiated using a chosen or developed methodology; conceptual frameworks, generally, do not include or imply any process model for implementation. Nevertheless, useful insights have been drawn from relevant research in reference to enterprise. First, formalization or standardization of proposed framework's elements, in context of specific manufacturing environment, is critical for its consistency and common understanding across management levels and common people. An explicit management procedure may be established to enable systematic implementation and execution through relevant tools and techniques. Potential challenges and any unforeseen conditions or obstacles, posed by the dynamics of the external or internal environment, may be identified. Building a robust culture of sustainability is vital to successful implementation of proposed concept. Sustainability education and skills in the domain of change management and organizational behaviour are critical to deployment process. Top level management must lead the implementation process to elevate its status. Employees who understand day-to-day business routines may be encouraged to get them involved in generating innovative ideas. Implementation objectives should be defined. Moreover, it needs to be assured that specified initiatives and activities contribute to transformation goals throughout the course of im-

plementation. Communicating to the employees and other stakeholders that why the company has chosen this continuous transformation path, i.e. "ideal—strategy—architecture" is essential. Pertinent here is to frame and relate stakeholders' apprehensions in an understandable form. Company wide communiqué, at the bare minimum, may include business needs, implementation issues and requirements, and expected benefits of such a transformative change or redesign.

Despite above recommendations, an average manufacturing environment may yet face multiple challenges on the way to become a sustainable competitive firm in accordance with transformation path proposed in this research. Capturing opportunities for sustainable value creation and building and evaluation of capabilities (Table 1) are no less than a paradigm shift for such environments. The proposed concept is centred on sustainability being a core component of the business (strategy) along with its continual renewal to gain long-term success while average manufacturers most often target short-term economic gains. Nevertheless, a deliberate support from government and society, in various aspects, could play a significant part to bring them at par with proposed requirements through a systematic incremental approach. Identification and evaluation of opportunities (idealization stage) could be a knowledge intensive and costly affair posing resource constraints. This process may be facilitated through a network of manufacturers, researchers and domain experts thereby helping traditional manufacturers seek alternate products and new markets to create competitive advantage. Experts in cultural assessment and change management can assist in developing and promoting a company culture needed for integrating sustainability into the everyday business. Government can sponsor participation mechanisms during idealization and strategic planning stage that could enable businesses to create new value innovatively with existing resources instead of huge investments in new technology. Evaluation of resource efficiency and lifecycle impacts needs specialized skills wherein provisions can be made for free of cost availability of such services. All such initiatives, at significantly low cost to businesses, could motivate them towards implementation of proposed concept.

5. Conclusion

Despite enormous research on SM, rarely the studies have focussed on continuous business transformation according to stakeholders' desires while maintaining competitiveness in a changing context. This research tried to fill this gap by first recognizing the need for a holistic and systemic representation of the issue along with a potentially viable approach embedded with capability to integrate sustainability concepts in manufacturing. A comprehensive overview of the issue, contributed by numerous aspects, provided an overall design context. A holistic and integrated process of planning and realization was aspired in relation to sustainability goals and a systematic transformation. Ultimately, framework was designed to provide a simplistic and progressive approach that could produce consistently the sustainable outcomes, and meet proposed ends. The three elements that jointly constitute this approach, i.e. 'Ideal', 'Strategy', and 'Architecture', have been presented from their functional perspective and contextual relation with each other and larger context. A systemic platform is thus established by constructing and operationalizing the individual elements. The dependency and relationship of elements with each other drives the conclusion that when executed in isolation, each element could still create value but limited to its context of use and initiating choices.The holistic view can maximize the cumulative sustainable value by managing the interaction among its elements, and maintaining the traceability from proposed ideal to strategic intent, and finally to architecture which has implications for functional effectiveness of the enterprise. Transformation can be more effectively achieved and repeated by accomplishing what is necessary and in concert with stakeholders'desires within each of the three elements. The enhanced capacity of enterprise to satisfy these requirements is the key to attain proposed sustainability ends. Continuous identification and re-newal of capabilities build this capacity to exploit opportunities for sustainable value creation.

Embracing this conceptual framework and its systematic operationalization will drive continuous transformation of a SM business in a fast paced environment. It also seems fair to conclude that proposed framework can play a critical role in maintaining competitiveness since it

suggests several capabilities whereby challenges can be addressed through meeting salient requirements underlying each capability. However, the framework is not yet embedded with methodology, procedures or tools for implementation, and there is room for further comprehension in terms of implementation strategy. Further research is needed in actual settings to confirm whether it can achieve the claimed ojectives in diverse manufacturing environments. The experience so gained will demonstrate its conceptualized role of integrating sustainability requirements and addressing challenges, pertinent to manufacturing businesses.

References

[1] Duflou, J.R., Sutherland, J.W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M., Kellens, K. (2012). Towards energy and resource efficient manufacturing: A processes and systems approach, *CIRP Annals,* Vol. 61, No. 2, 587-609, doi: 10.1016/j.cirp.2012.05.002.

[2] Bi, Z. (2011). Revisiting system paradigms from the viewpoint of manufacturing sustainability, *Sustainability*, Vol. 3, No. 9, 1323-1340, doi: 10.3390/su3091323.

[3] Garetti, M., Taisch, M. (2012). Sustainable manufacturing: Trends and research challenges, *Production Planning & Control*, Vol. 23, No. 2-3, 83-104, doi: 10.1080/09537287.2011.591619.

[4] Gopalakrishnan, K., Yusuf, Y.Y., Musa, A., Abubakar, T., Ambursa, H.M. (2012). Sustainable supply chain management: A case study of British aerospace (BAE) systems, *International Journal of Production Economics*, Vol. 140, No. 1, 193-203, doi: 10.1016/j.ijpe.2012.01.003.

[5] ElMaraghy, H., Schuh, G., ElMaraghy, W., Piller, F., Schönsleben, P., Tseng, M., Bernard, A. (2013). Product variety management, *CIRP Annals*, Vol. 62, No. 2, 629-652, doi: 10.1016/j.cirp.2013.05.007.

[6] Jawahir, I.S., Bradley, R. (2016). Technological elements of circular economy and the principles of 6R-based closed-loop material flow in sustainable manufacturing, *Procedia CIRP*, Vol. 40, 103-108, doi: 10.1016/j.procir.2016.01.067.

[7] Zhang, H., Calvo-Amodio, J., Haapala, K.R. (2015). Establishing foundational concepts for sustainable manufacturing systems assessment through systems thinking, *International Journal of Strategic Engineering Asset Management*, Vol. 2, No. 3, 249-269, doi: 10.1504/IJSEAM.2015.072124.

[8] Lönngren, J., Svanström, M. (2016), Systems thinking for dealing with wicked sustainability problems: Beyond functionalist approaches, In: Leal Filho, W., Nesbit, S. (eds.), *New developments in engineering education for sustainable development*, Springer, Cham, Switzerland, 151-160, doi: 10.1007/978-3-319-32933-8_14.

[9] Jackson, M.C. (2006). Creative holism: A critical systems approach to complex problem situations, *Systems Research and Behavioral Science*, Vol. 23, No. 5, 647-657, doi: 10.1002/sres.799.

[10] Gharajedaghi, J. (2011), *Systems thinking: Managing chaos and complexity: A platform for designing business architecture, Third edition,* Elsevier, Burlington, USA.

[11] Gaziulusoy, A.I. (2015). A critical review of approaches available for design and innovation teams through the perspective of sustainability science and system innovation theories, *Journal of Cleaner Production*, Vol. 107, 366-377, doi: 10.1016/j.jclepro.2015.01.012.

[12] World Economic Forum (prepared in collaboration with Ellen MacArthur Foundation and McKinsey & Company). Towards the circular economy: Accelerating the scale-up across global supply chains, from *http://www3.weforum.org/docs/WEF_ENV_TowardsCircularEconomy_Report_2014.pdf*, accessed August 23, 2016.

[13] Teece, D.J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance, *Strategic Management Journal*, Vol. 28, No. 13, 1319-1350, doi: 10.1002/smj.64.

[14] Oertwig, N., Galeitzke, M., Schmieg, H.-G., Kohl, H., Jochem, R., Orth, R., Knothe, T. (2017). Integration of sustainability into the corporate strategy, In: Stark, R., Seliger, G., Bonvoisin, J. (eds.), *Sustainable manufacturing: Challenges, solutions and implementation perspectives*, Springer, Cham, Switzerland, doi: 10.1007/978-3-319-48514-0.

[15] Tonelli, F., Evans, S., Taticchi, P. (2013). Industrial sustainability: Challenges, perspectives, actions, *International Journal of Business Innovation and Research*, Vol. 7, No. 2, 143-163, doi: 10.1504/IJBIR.2013.052576.

[16] Koho, M., Tapaninaho, M., Heilala, J., Torvinen, S. (2015). Towards a concept for realizing sustainability in the manufacturing industry, *Journal of Industrial and Production Engineering*, Vol. 32, No. 1, 12-22, doi: 10.1080/21681015.2014.1000402.

[17] Subic, A., Shabani, B., Hedayati, M., Crossin, E. (2012). Capability framework for sustainable manufacturing of sports apparel and footwear, *Sustainability*, Vol. 4, No. 9, 2127-2145, doi: 10.3390/su4092127.

[18] Masood, I., Jahanzaib, M., Haider, A. (2016). Tool wear and cost evaluation of face milling grade 5 titanium alloy for sustainable machining, *Advances in Production Engineering & Management*, Vol. 11, No. 3, doi: 10.14743/apem2016.3.224.

[19] Hassine, H., Barkallah, M., Bellacicco, A., Louati, J., Riviere, A., Haddar, M. (2015). Multi objective optimization for sustainable manufacturing, application in turning, *International Journal of Simulation Modelling*, Vol. 14, No. 1, 98-109, doi: 10.2507/IJSIMM14(1)9.292.

[20] Govindan, K., Soleimani, H., Kannan, D. (2015). Reverse logistics and closed-loop supply chain: A comprehensive review to explore the future, *European Journal of Operational Research*, Vol. 240, No. 3, 603-626, doi: 10.1016/j.ejor.2014.07.012.

[21] Schenkel, M., Caniëls, M.C.J., Krikke, H., van der Laan, E. (2015). Understanding value creation in closed loop supply chains – Past findings and future directions, *Journal of Manufacturing Systems*, Vol. 37, Part 3, 729-745, doi: 10.1016/j.jmsy.2015.04.009.

[22] Ageron, B., Gunasekaran, A., Spalanzani, A. (2012). Sustainable supply management: An empirical study, *International Journal of Production Economics*, Vol. 140, No. 1, 168-182, doi: 10.1016/j.ijpe.2011.04.007.

[23] Galal, N.M., El-Kilany, K.S. (2016). Sustainable agri-food supply chain with uncertain demand and lead time, *International Journal of Simulation Modelling*, Vol. 15, No. 3, 485-496, doi: 10.2507/IJSIMM15(3)8.350.

[24] Fiksel, J. (2003). Designing resilient, sustainable systems, *Environmental Science & Technology*, Vol. 37, No. 23, 5330-5339, doi: 10.1021/es0344819.

[25] Fiksel, J. (2012). A systems view of sustainability: The triple value model, *Environmental Development*, Vol. 2, 138-141, doi: 10.1016/j.envdev.2012.03.015.

[26] Moldavska, A. (2016). Model-based sustainability assessment – An enabler for transition to sustainable manufacturing, *Procedia CIRP*, Vol. 48, 413-418, doi: 10.1016/j.procir.2016.04.059.

[27] Moldavska, A., Welo, T. (2016). Development of manufacturing sustainability assessment using systems thinking, *Sustainability*, Vol. 8, No. 1, doi: 10.3390/su8010005.

[28] Terzi, S., Bouras, A., Dutta, D., Garetti, M., Kiritsis, D. (2010). Product lifecycle management – From its history to its new role, *International Journal of Product Lifecycle Management*, Vol. 4, No. 4, 360-389, doi: 10.1504/IJPLM.2010.036489.

[29] Alayón, C., Säfsten, K., Johansson, G. (2016). Conceptual sustainable production principles in practice: Do they reflect what companies do?, *Journal of Cleaner Production*, Vol. 141, 693-701, doi: 10.1016/j.jclepro.2016.09.079.

[30] Esmaeilian, B., Behdad, S., Wang, B. (2016). The evolution and future of manufacturing: A review, *Journal of Manufacturing Systems*, Vol. 39, 79-100, doi: 10.1016/j.jmsy.2016.03.001.

[31] Haas, W., Krausmann, F., Wiedenhofer, D., Heinz, M. (2015). How circular is the global economy?: An assessment of material flows, waste production, and recycling in the European union and the world in 2005, *Journal of Industrial Ecology*, Vol. 19, No. 5, 765-777, doi: 10.1111/jiec.12244.

[32] Ellen MacArthur Foundation. Towards a circular economy: Business rationale for an accelerated transition, from *https://www.ellenmacarthurfoundation.org/assets/downloads/TCE_Ellen-MacArthur-Foundation_9-Dec-2015.pdf*, accessed October 7, 2016.

[33] McDonough, W., Braungart, M., Anastas, P.T., Zimmerman, J.B. (2003). Peer reviewed: Applying the principles of green engineering to cradle-to-cradle design, *Environmental Science & Technology*, Vol. 37, No. 23, 434-441, doi: 10.1021/es0326322.

[34] McDonough, W., Braungart, M. (2002). Design for the triple top line: New tools for sustainable commerce, *Corporate Environmental Strategy*, Vol. 9, No. 3, 251-258, doi: 10.1016/S1066-7938(02)00069-6.

[35] World Economic Forum. Manufacturing our future: Cases on the future of manufacturing, from *http://www3.weforum.org/docs/GAC16_The_Future_of_Manufacturing_report.pdf*, accessed May 2, 2017.

[36] Stock, T., Seliger, G. (2016). Opportunities of sustainable manufacturing in industry 4.0, *Procedia CIRP*, Vol. 40, 536-541, doi: 10.1016/j.procir.2016.01.129.

[37] Miroshnychenko, I., Barontini, R., Testa, F. (2017). Green practices and financial performance: A global outlook, *Journal of Cleaner Production*, Vol. 147, 340-351, doi: 10.1016/j.jclepro.2017.01.058.

[38] Abdul Rashid, S.H., Evans, S., Longhurst, P. (2008). A comparison of four sustainable manufacturing strategies, *International Journal of Sustainable Engineering*, Vol. 1, No. 3, 214-229, doi: 10.1080/19397030802513836.

[39] Chatha, K., Butt, I. (2015). Themes of study in manufacturing strategy literature, *International Journal of Operations & Production Management*, Vol. 35, No. 4, 604-698, doi: 10.1108/IJOPM-07-2013-0328.

[40] Kappelman, L.A., Zachman, J.A. (2013). The enterprise and its architecture: Ontology & challenges, *Journal of Computer Information Systems*, Vol. 53, No. 4, 87-95, doi: 10.1080/08874417.2013.11645654.

[41] van Iwaarden, J., van der Wiele, T. (2012). The effects of increasing product variety and shortening product life cycles on the use of quality management systems, *International Journal of Quality & Reliability Management*, Vol. 29, No. 5, 470-500, doi: 10.1108/02656711211230481.

[42] Halog, A., Manik, Y. (2011). Advancing integrated systems modelling framework for life cycle sustainability assessment, *Sustainability*, Vol. 3, No. 2, 469-499, doi: 10.3390/su3020469.

[43] Friedrich-Schiller-University Jena. Roadmap for sustainability assessment in european process industries, Technical report, from *https://www.researchgate.net/publication/301823430*, accessed 13 May 2016.

[44] Zhang, H., Haapala, K.R. (2015). Integrating sustainable manufacturing assessment into decision making for a production work cell, *Journal of Cleaner Production*, Vol. 105, 52-63, doi: 10.1016/j.jclepro.2014.01.038.

[45] Hendrickx, H.H.M. (2015). Business architect: A critical role in enterprise transformation, *Journal of Enterprise Transformation*, Vol. 5, No. 1, 1-29, doi: 10.1080/19488289.2014.893933.

[46] University of Cambridge. State-of-practice in business modelling and value-networks, emphasising potential future models that could deliver sustainable value, from *http://www.sustainvalue.eu/publications/D2_1_Final_Rev1_0_web.pdf*, accessed November 10, 2016.

[47] Grassl, W. (2012). Business models of social enterprise: A design approach to hybridity, *ACRN Journal of Entrepreneurship Perspectives*, Vol. 1, No. 1, 37-60.

[48] Ackoff, R.L. (1999), *Re-creating the corporation: A design of organizations for the 21st century*, Oxford University Press, USA.

[49] Mathews, F. (2011). Towards a deeper philosophy of biomimicry, *Organization & Environment*, Vol. 24, No. 4, 364-387, doi: 10.1177/1086026611425689.

[50] Apple Inc. Environmental responsibility report, from *https://images.apple.com/environment/pdf/Apple_Environ mental_Responsibility_Report_2017.pdf*, accessed May 15, 2017.

[51] Ocampo, L.A., Clark, E.E., Tanudtanud, K.V.G., Ocampo, C.O.V., Impas Sr, C.G., Vergara, V.G., Pastoril, J., Tordillo, J.A.S. (2015). An integrated sustainable manufacturing strategy framework using fuzzy analytic network process, *Advances in Production Engineering & Management*, Vol. 10, No. 3, 125-139, doi: 10.14743/apem2015.3.197.

[52] Haftor, D.M. (2011). An evaluation of R.L. Ackoff's interactive planning: A case-based approach, *Systemic Practice and Action Research*, Vol. 24, No. 4, 355-377, doi: 10.1007/s11213-010-9188-y.

[53] Rhyne, R.F. (1994). Contextual discipline: Its essentiality within social-systems analysis, *Technological Forecasting and Social Change*, Vol. 47, No. 3, 277-292, doi: 10.1016/0040-1625(94)90069-8.

[54] Amer, M., Daim, T.U., Jetter, A. (2013). A review of scenario planning, *Futures*, Vol. 46, 23-40, doi: 10.1016/j.futures.2012.10.003.

[55] Osterwalder, A., Pigneur, Y. (2010), *Business model generation: A handbook for visionaries, game changers, and challengers*, John Wiley & Sons, Hoboken, New Jersey, USA.

[56] Midgley, G. How can co-creation produce better outcomes for people and communities?, from *https://www.researchgate.net/publication/303752603*, accessed July 24, 2016, doi: 10.13140/RG.2.1.2373.3360.

[57] Midgley, G. (2000), *Systemic intervention: Philosophy,methodology and practice*, Springer Science, New York, USA, doi: 10.1007/978-1-4615-4201-8.

[58] Ackoff, R.L. (1990). Strategy, *Systems Practice,* Vol. 3, No. 6, 521-524, doi: 10.1007/BF01059636.

[59] Hayes, R.H., Pisano, G.P. (1996). Manufacturing strategy: At the intersection of two paradigm shifts, *Production and Operations Management*, Vol. 5, No. 1, 25-41, doi: 10.1111/j.1937-5956.1996.tb00383.x.

[60] Bilge, P., Seliger, G., Badurdeen, F., Jawahir, I.S. (2016). A novel framework for achieving sustainable value creation through industrial engineering principles, *Procedia CIRP*, Vol. 40, 516-523, doi: 10.1016/j.procir.2016.01.126.

[61] Mills, J., Platts, K., Bourne, M. (2003). Competence and resource architectures, *International Journal of Operations & Production Management*, Vol. 23, No. 9, 977-994, doi: 10.1108/01443570310491738.

[62] World Economic Forum. The future of manufacturing: Driving capabilities, enabling investments, from *http://www3.weforum.org/docs/Media/GAC14/Future_of_Manufacturing_Driving_Capabilities.pdf*, accessed May 2, 2017.

[63] Checkland, P. (2012). Four conditions for serious systems thinking and action, *Systems Research and Behavioral Science*, Vol. 29, No. 5, 465-469, doi: 10.1002/sres.2158.

[64] Zott, C., Amit, R. (2010). Business model design: An activity system perspective, *Long Range Planning*, Vol. 43, No. 2-3, 216-226, doi: 10.1016/j.lrp.2009.07.004.

A hybrid grey cuckoo search algorithm for job-shop scheduling problems under fuzzy conditions

Yang, F.[a,b,*], **Ye, C.M.**[a], **Shi, M.H.**[a]

[a]Business School, University of Shanghai for Science and Technology, Shanghai, P.R. China
[b]College of Management, Henan University of Traditional Chinese Medicine, Zhengzhou, Henan, P.R. China

A B S T R A C T

This paper aims to acquire the precise makespan or delivery period in job-shop scheduling (JSP) under fuzzy conditions. To this end, the author designed a grey scheduling model and a hybrid grey cuckoo search (HGCS) algorithm in the following steps. Firstly, three- and four-parameter interval grey numbers were introduced to depict the fuzzy makespan and delivery period, respectively; then, the possibility measure and necessity measure were defined, and the tardiness credibility index was proposed to estimate the probability of job tardiness. After that, a grey mixed integer programming model was developed to minimize the mean tardiness credibility, and the HGCS was proposed to solve the model. Finally, simulations were conducted on the classical example of $6(3) \times 6$. The results show that the proposed algorithm outperformed the basic cuckoo search. The research findings shed new light on the JSP under fuzzy conditions.

1. Introduction

Job-shop scheduling (JSP) is a common NP-hard scheduling problem in modern industry. The problem is about the processing of a number of jobs on several machines. Each of these jobs needs to go through multiple processes and can run on several machines. The constraints of the JSP are as follows: the jobs must be treated in a feasible sequence of processes and follow the same order of priority; the same job should be processed only once on the same machine; the same machine can only process one job at a time. Under the above constraints, the JSP aims to optimize performance indices like the minimum makespan or the minimum tardiness through rational arrangement of the process sequence and start time. Many algorithms have been developed to solve classical JSPs. For instance, Zhang *et al.* [1] proposed a hybrid meta-heuristics algorithm based on shuffled frog leaping algorithm (SFLA), intelligent water drops algorithm (IWD) and path relinking (PR) algorithm. Modrák *et al.* [2] converted an *m*-machine problem into a 2-machine problem, and created a Johnson algorithm to solve it. Chaudhry *et al.* [3] presented a genetic algorithm with process planning and scheduling. Considering sequence flexibility, this algorithm solves classical JSPs using the spreadsheet of domain independence. Huang *et al.* [4] suggested solving classical JSPs by sequence flexibility and minimized the makespan with an improved genetic algorithm.

In classical JSPs, the production is scheduled under the ideal condition, that is, all factors are determined. In real-world production scheduling, however, both makespan and delivery period are highly indeterminate owing to uncertainties like machine breakdown, unfavourable environment or wrong decisions. To reflect the exact production situation, it is necessary to treat the makespan and delivery period as fuzzy variables.

Recent years has seen much attention being paid to the JSP under fuzzy conditions. The most common solution is fuzzy scheduling, that is, describing the fuzzy makespan and delivery period with triangular and trapezoidal fuzzy numbers. For instance, Ishii *et al.* [5] proposed the concept of fuzzy delivery period after exploring various jobs, and then investigated single- and double-machine open-shop scheduling under fuzzy delivery period. Noori-Darvish *et al.* [6] created a two-objective fuzzy programming model for open-shop scheduling, whose sequence relies on the job preparation time, fuzzy makespan and fuzzy delivery period. Focusing on two-machine scheduling, Gharehgozli *et al.* [7] put forward a fuzzy mixed integer programming model for scheduling sequence dependent on the job preparation time and fuzzy delivery period. Simeunović *et al.* [8] built up a workforce scheduling model based on artificial neural network.

For triangular and trapezoidal fuzzy numbers, both sides are monotonous linear functions. Therefore, linear, uniform variations can be expected for the probability function of the fuzzy number value on both sides of the most likely value (i.e. the window of the most likely makespan and the most likely delivery period). Nevertheless, the probability functions of makespan and delivery period are not necessarily linear in actual production. The possible linear and nonlinear functions of the two indices are shown in Fig. 1.

Considering the fuzziness of makespan and delivery period, a scheduling constraint should be expressed as an interval rather than a fixed value. Moreover, the ideal makespan and delivery period in the classical JSPs are often the most likely values in fuzzy situations; the probability functions for the deviation from the most likely values are not necessarily linear ones. These features are consistent with those of interval grey numbers.

In grey system theory, an interval grey number has different probabilities in taking different values. This unique feature has been widely discussed in the academia. For example, reference [9] introduces the interval grey number into multi-criterion optimization, with the aim to solve various fuzzy decision-making problems in the real world. Considering the impact of fuzzy grey number on prediction results, reference [10] converts the interval grey number sequence into a real number sequence and establishes a prediction model of interval grey number. Based on the definition of the basic interval number, references [11-18] put forward the three-parameter interval number and a multi-attribute decision-making method. Among them, reference [11] discusses the grey target decision-making with known maximum probability of attribute values: the index weight was determined through subjective or objective weighting, a comprehensive optimization model was established for index weight, and the attribute values were identified by three-parameter interval numbers.

In light of the above, this paper establishes a grey JSP model, which treats fuzzy makespan and delivery period as three- and four-parameter interval grey numbers, respectively, and defines the tardiness credibility index of jobs based on the possibility measure and necessity measure. Then, the fuzzy tardiness was determined as the optimization target and a grey mixed integer programming model was established. Finally, the proposed model was solved by an intelligent algorithm.
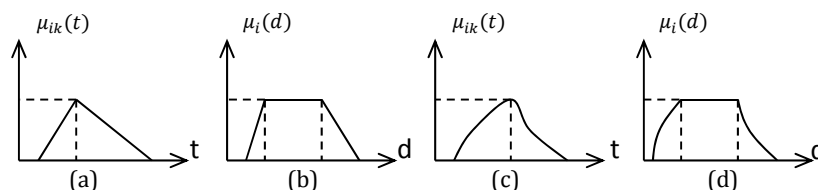

Fig. 1 (a) and (b): Linear probability functions, (c) and (d): Nonlinear probability functions

2. Grey job shop scheduling problem

2.1 Problem description

In actual scheduling, the job processing is more or less affected by some fuzzy factors. Thus, the scheduling process should be measured by probabilities. Here, the fuzzy makespan and delivery period are expressed as three- and four-parameter interval grey numbers, and the scheduling target is to optimize the mean tardiness credibility of the job. To realize the goal, it is necessary to review the definition of grey number.

Definition 1 [20] A number whose value falls in an interval rather than at a fixed point is known as the grey number and can be denoted as $\otimes$. A three-parameter interval grey number can be expressed as $a(\otimes) \in [\alpha^-, \alpha^*, \alpha^+]$ $(0 \leq \alpha^- \leq \alpha^* \leq \alpha^+)$ with α^-, α^+ and α^* being the lower limit, the upper limit and the most likely value (centre of gravity), respectively.

Remark: The three-parameter interval will degenerate into a real number if $\alpha^- = \alpha^* = \alpha^+$ and an interval grey number if $\alpha^- = \alpha^*$ or $\alpha^* = \alpha^+$. The three-parameter interval number satisfies $\int_{\alpha^-}^{\alpha^+} f(x)dx = 1$, $f_{max} = f(\alpha^*)$, and the value curve on either side of the ideal value is not necessarily linear. Hence, the membership function of the three-parameter interval number can be expressed as:

$$\mu_T(x) = \begin{cases} [x^{\frac{1}{2}} - (\alpha^-)^{\frac{1}{2}}] \cdot [(\alpha^*)^{\frac{1}{2}} - (\alpha^-)^{\frac{1}{2}}]^{-1}, & \alpha^- < x \leq \alpha^* \\ [x^{\frac{1}{2}} - (\alpha^+)^{\frac{1}{2}}] \cdot [(\alpha^*)^{\frac{1}{2}} - (\alpha^+)^{\frac{1}{2}}]^{-1}, & \alpha^* < x \leq \alpha^+ \\ 0, & x \leq \alpha^* \text{ or } x > \alpha^+ \end{cases} \tag{1}$$

Definition 2 [19] A four-parameter interval grey number can be expressed as $b(\otimes) \in [\alpha^-, \alpha^{*-}, \alpha^{*+}, \alpha^+]$ $(0 \leq \alpha^- \leq \alpha^{*-} \leq \alpha^{*+} \leq \alpha^+)$, with α^- and α^+ being the lower limit and the upper limit, respectively, and α^{*-} and α^{*+} being the most likely values of the lower limit and that of the upper limit, respectively. The two most likely values are considered as the dual centres of the interval. The four-parameter interval number satisfies $\int_{\alpha^-}^{\alpha^+} f(x)dx = 1$, $f_{max} = f(\alpha^{*-})$ or $f_{max} = f(\alpha^{*+})$.

Remark: The four-parameter interval will degenerate into a three-parameter interval grey number if $\alpha^{*-} = \alpha^{*+}$. In this case, the membership function of the four-parameter interval number can be expressed as:

$$\mu_D(x) = \begin{cases} [x^{\frac{1}{2}} - (\alpha^-)^{\frac{1}{2}}] \cdot [(\alpha^{*-})^{\frac{1}{2}} - (\alpha^-)^{\frac{1}{2}}]^{-1}, & \alpha^- < x \leq \alpha^{*-} \\ 1, & \alpha^{*-} < x \leq \alpha^{*+} \\ [x^{\frac{1}{2}} - (\alpha^+)^{\frac{1}{2}}] \cdot [(\alpha^{*+})^{\frac{1}{2}} - (\alpha^+)^{\frac{1}{2}}]^{-1}, & \alpha^{*+} < x \leq \alpha^+ \\ 0, & x \leq \alpha^- \text{ or } x > \alpha^+ \end{cases} \tag{2}$$

According to the above description, the grey JSP can be described as:

- Let $P = \{p_1, p_2, \dots, p_n\}$ be a set of n jobs, with p_i being the i-th job $(i = 1, 2, \dots, n)$.
- Let $M = \{m_1, m_2, \dots, m_m\}$ be a set of m machines, with m_j being the j-th machine $(j = 1, 2, \dots, m)$.
- Let $OP = \{op_1, op_2, \dots, op_n\}$ be a set of process sequences, with $op_i = \{op_{i1}, op_{i2}, \dots, op_{im}\}$ being the process sequence of job p_i, and op_{ik} being the machine number of the i-th job in the k-th process. If $op_{ik} = 0$, the k-th process of the i-th job is not implemented $(k = 1, 2, \dots, m)$.
- Let $T = \{a_{i1}(\otimes), a_{i2}(\otimes), \dots, a_{im}(\otimes)\}$ be the set of makespans of job p_i, with $a_{ik}(\otimes)$ being the three-parameter interval grey number indicating the makespan of the i-th job in the k-th process. $a_{ik}(\otimes) \in [T_{ik}^-, T_{ik}^*, T_{ik}^+]$, where T_{ik}^*, T_{ik}^- and T_{ik}^+ are respectively the most likely, the shortest and the longest makespans of the k-th process. The function $\mu_{ik}(t)$ describes the possibility of completing k-th process of the i-th job at time t.

- Let $D_i(\otimes) = (d_i^-, d_i^{*-}, d_i^{*+}, d_i^+)$ be a four-parameter interval grey number indicating the delivery period of the i-th job. The interval (d_i^{*-}, d_i^{*+}) is the most likely delivery period of the i-th job.

The other constraints are the same with those in the classical JSP. Because the delivery period can accurately reflect the market changes, the ultimate target of our scheduling is to find a feasible scheduling plan that minimizes the mean tardiness of the job.

2.2 Grey number operator

Suppose all jobs to the processed on the k-th machine are allocated into one sequence, in which the i-th job is preceded by the job sequence 1,2,...,i-1. Thus, the makespan of the i-th job equals the total makespan of all jobs in the preceding sequence. To obtain the total makespan, the addition of two three-interval grey numbers should be defined as follows [21].

Definition 3 Let $a_{ik}(\otimes) = [\alpha^-, \alpha^*, \alpha^+]$ and $a_{jk}(\otimes) = [\beta^-, \beta^*, \beta^+]$. Then, we have:

$$a_{ik}(\otimes) + a_{jk}(\otimes) = [\alpha^- + \beta^-, \alpha^* + \beta^*, \alpha^+ + \beta^+]$$

According to Definition 3, the grey makespan $a_i(\otimes)$ of the i-th job can be obtained as:

$$a_i(\otimes) = \sum_{q=1}^{i} a_{qk}(\otimes), \quad q = 1,2,\dots,n \tag{3}$$

Where $a_{qk}(\otimes)$ is the grey makespan of the q-th job on the k-th machine. Before obtaining the job tardiness, it is also necessary to define the reduction operator of four-parameter interval grey numbers.

Theorem 1 For the membership function (2) of four-parameter interval number, there exist $x_1 \in [\alpha_i^-, \alpha_i^{*-}]$, $x_2 \in [\alpha_j^{*+}, \alpha_j^+]$ and $x_3 = (x_1^{\frac{1}{2}} - x_2^{\frac{1}{2}})^2$ that satisfy:

$$\mu_{D_i}(x_1) = \mu_{D_j}(x_2) = \frac{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - x_3^{\frac{1}{2}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}} + (\alpha_j^{*+})^{\frac{1}{2}}}$$

Proof: If $\mu_{D_i}(x_1) = \mu_{D_j}(x_2)$, then

$$\left[x_1^{\frac{1}{2}} - (\alpha_i^-)^{\frac{1}{2}} \right] \cdot \left[(\alpha_i^{*-})^{\frac{1}{2}} - (\alpha_i^-)^{\frac{1}{2}} \right]^{-1} = \left[x_2^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} \right] \cdot \left[(\alpha_j^{*+})^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} \right]^{-1}.$$

Solving this equation, we have: $x_2 = \left(\left[(\alpha_j^{*+})^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} \right] \cdot \frac{(\alpha_i^-)^{\frac{1}{2}} - x_1^{\frac{1}{2}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}}} + (\alpha_j^+)^{\frac{1}{2}} \right)^2.$

Hence, $\frac{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - x_3^{\frac{1}{2}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}} + (\alpha_j^{*+})^{\frac{1}{2}}} = \frac{\left[(\alpha_i^-)^{\frac{1}{2}} - x_1^{\frac{1}{2}} \right] \cdot \frac{1 + [(\alpha_j^{*+})^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}}]}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}} + (\alpha_j^{*+})^{\frac{1}{2}}} = \frac{(\alpha_i^-)^{\frac{1}{2}} - x_1^{\frac{1}{2}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}}}$

Hence, there exist x_1, x_2 and $x_3 = (x_1^{\frac{1}{2}} - x_2^{\frac{1}{2}})^2$ that satisfy:

$$\mu_{D_i}(x_1) = \mu_{D_j}(x_2) = \frac{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - x_3^{\frac{1}{2}}}{(\alpha_i^-)^{\frac{1}{2}} - (\alpha_j^+)^{\frac{1}{2}} - (\alpha_i^{*-})^{\frac{1}{2}} + (\alpha_j^{*+})^{\frac{1}{2}}}$$

Q.E.D.

Definition 4 Let $b_i(\otimes) = [\alpha_i^-, \alpha_i^{*-}, \alpha_i^{*+}, \alpha_i^+]$ and $b_j(\otimes) = [\alpha_j^-, \alpha_j^{*-}, \alpha_j^{*+}, \alpha_j^+]$ be two four-parameter interval grey numbers. Through reduction operation, another four-parameter interval grey number can be obtained as $b_i(\otimes) - b_j(\otimes) = [\alpha_i^- - \alpha_j^-, \alpha_i^{*-} - \alpha_j^{*-}, \alpha_i^{*+} - \alpha_j^{*+}, \alpha_i^+ - \alpha_j^+]$.

Whereas makespan and delivery period are respectively three- and four-parameter interval grey numbers, the three-parameter interval grey number should be treated as a special type of

four-parameter interval grey number, that is, $a_{ik}(\otimes) \in [a_{ik}^-, a_{ik}^*, a_{ik}^+] \Leftrightarrow a_{ik}(\otimes) \in [a_{ik,}^- a_{ik}^*, a_{ik}^*,$ $a_{ik}^+]$. By definition 4, the grey tardiness of the i-th job in the k-th machine can be obtained as $T_i(\otimes) = a_i(\otimes) - b_i(\otimes) = (\alpha_{ik}^- - \alpha_i^+, \alpha_{ik}^{*-} - \alpha_i^{*+}, \alpha_{ik}^{*-} - \alpha_i^{*-}, \alpha_{ik}^{*+} - \alpha_i^-)$.

Definition 5 Let $b_i(\otimes) = [\alpha_i^-, \alpha_i^{*-}, \alpha_i^{*+}, \alpha_i^+]$ and $b_j(\otimes) = [\alpha_j^-, \alpha_j^{*-}, \alpha_j^{*+}, \alpha_j^+]$ be two four-parameter interval grey numbers. If $\alpha_i^- \geq \alpha_j^-, \alpha_i^{*-} \geq \alpha_j^{*-}, \alpha_i^{*+} \geq \alpha_j^{*+}$ and $\alpha_i^+ \geq \alpha_j^+$, then $b_i(\otimes) \geq b_j(\otimes)$. Similarly, if $\alpha_i^- \leq \alpha_j^-, \alpha_i^{*-} \leq \alpha_j^{*-}, \alpha_i^{*+} \leq \alpha_j^{*+}$ and $\alpha_i^+ \leq \alpha_j^+$, then $b_i(\otimes) \leq b_j(\otimes)$.

If the fuzzy tardiness $b_i(\otimes)$ of the i-th job is greater than zero, then the job production must be delayed; otherwise, the production is not delayed. Hence, the fuzzy tardiness of the job can be described by the possibility measure and the necessity measure.

2.3 Objective function

Definition 6 For the set $A, B \in F(X)(x \in X)$, suppose μ_A and μ_B are the membership functions of A and B, respectively. If $Pos_A(B) = \sup_{x \in X} min\{\mu_A(x), \mu_B(x)\}$ and $Nec_A(B) = \inf_{x \in X} max\{1 - \mu_A(x), \mu_B(x)\}$, then $Pos_A(B)$ and $Nec_A(B)$ are the possibility measures and necessity measures of B under the condition of A.

Obviously, $Pos(x \geq 0)$ and $Nec(x \geq 0)$ respectively stand for the possibility and necessity of $x \geq 0$.

Definition 7 Let $T_i(\otimes) = (\alpha_i^-, \alpha_i^{*-}, \alpha_i^{*+}, \alpha_i^+)$ be the grey tardiness of the i-th job. Then, the tardiness credibility of the i-th job can be defined as the weighted sum of possibility measures and necessity measures and denoted as $Con_i(x \geq 0)$. Therefore, $Con_i(x \geq 0) = \delta Pos(x \geq 0) + (1 - \delta)Nec(x \geq 0)(0 \leq \delta \leq 1)$.

If $0 < \delta < 1$, the following can be derived from Definition 7: If tardiness credibility of the i-th job is $Con_i(x \geq 0) = 1$, then $Nec(x \geq 0) = 1$, i.e. the job production must be delayed; If $Con_i(x \geq 0) = 0$, then $Pos(x \geq 0) = 0$, i.e. the production of the i-th job cannot be delayed.

The coefficient δ is determined by various factors, such as the decision-maker and the extent of the tardiness. Its value is negatively correlated with the conservativeness of the decision. The following properties can be obtained according to the definition of credibility:

Property 1 If the grey tardiness of the i-th job is $T_i(\otimes) = (\alpha_i^-, \alpha_i^{*-}, \alpha_i^{*+}, \alpha_i^+)(\alpha_i^- < \alpha_i^{*-} \leq \alpha_i^{*+} < \alpha_i^+)$, then the tardiness credibility $Con_i(x \geq 0)$ of the job can be expressed as:

$$Con_i(x \geq 0) = \begin{cases} 0, & \alpha_i^+ \leq 0 \\ \frac{\delta \alpha_i^+}{\alpha_i^+ - \alpha_i^{*+}}, & \alpha_i^+ > 0, \alpha_i^{*+} \leq 0 \\ \delta, & \alpha_i^{*+} > 0, \alpha_i^{*-} \leq 0 \\ \frac{\alpha_i^{*-} - \delta \alpha_i^-}{\alpha_i^{*-} - \alpha_i^-}, & \alpha_i^{*-} > 0, \alpha_i^- \leq 0 \\ 1, & \alpha_i^- > 0 \end{cases} \tag{4}$$

Let $a_{ik}(\otimes), D_i(\otimes)$ and $T_i(\otimes)$ be the grey makespan, the grey delivery period and the grey tardiness of the i-th job in the JSP, respectively. Suppose the feasible scheduling plans of all jobs are allocated to the set FS and the objective function is denoted as $f(x)$. For a known scheduling plan $x \in FS$, it is possible to establish the following grey mixed integer programming model:

$$\min_{x \in S} f(x) = \sum_{i=1}^{n} \sum_{k=1}^{m} \left\{ \omega_1 - \frac{\delta \alpha_i^+}{\alpha_i^+ - \alpha_i^{*+}} + \omega_2 \delta + \frac{\omega_3(\alpha_i^{*-} - \delta \alpha_i^-)}{\alpha_i^{*-} - \alpha_i^-} + \omega_4 \right\} m_{ijk}/n \tag{5}$$

$$s.t. \ \sum_{k=1}^{m} \sum_{i=1}^{n} m_{ikj} = 1, \ i = 1,2, \dots, n \tag{6}$$

$$\sum_{i=1}^{n} m_{ikj} \leq 1, \ k = 1,2, \dots, m; j = 1,2, \dots, n \tag{7}$$

$$a_{ikj_i}(\otimes) = \sum_{i=1}^{n} \sum_{j=1}^{j_i} a_{jk}(\otimes) m_{ijk} \quad i, j_i = 1,2, \dots, n; k = 1,2, \dots, m \tag{8}$$

$$T_i(\otimes) = a_{ikj_i}(\otimes) - D_i(\otimes) \ i, j_i = 1,2, \dots, n \tag{9}$$

$$m_{ikj} = 0,1 \quad i,j = 1,2,\dots,n; k = 1,2,\dots,m \tag{10}$$

$$\omega_q = 0,1 \quad q \in \{1,2,3,4\} \tag{11}$$

The goal of this model is to obtain the optimal scheduling plan that minimizes the value of the objective function $f(x)$, i.e. the mean tardiness credibility of all jobs. Eq. 6 specifies that each job must occupy only one position in the process sequence on each machine; Eq. 7 stipulates that each position in the process sequence on each machine can only be occupied by one job; Eq. 8 defines $a_{ikj_i}(\otimes)$, the grey makespan of the i-th job at the j_i-th position of the process sequence on the k-th machine; Eq. 9 defines the grey tardiness $T_i(\otimes)$ of the i-th job; Eq. 10 defines the indicator variable m_{ikj} ($m_{ikj} = 1$ if the i-th job occupies the j-th position in the process sequence on the k-th machine, and $m_{ikj} = 0$ if otherwise); Eq. 11 ensures that $\omega_q = 1$ if $\omega_q \leq 0$ and $\omega_q = 0$ if otherwise.

Theorem 2 If the grey tardiness of the i-th job is $T_i(\otimes) = (\alpha_i^-, \alpha_i^{*-}, \alpha_i^{*+}, \alpha_i^+)(\alpha_i^- < \alpha_i^{*-} \leq \alpha_i^{*+} < \alpha_i^+)$, with $a_j(\otimes) = (\alpha_j^-, \alpha_j^{*-}, \alpha_j^{*+}, \alpha_j^+)$ being a four-parameter interval grey number, and the makespan is nonzero, then its tardiness credibility is $Con_{T_i(\otimes)} \geq Con_{T_i(\otimes)-a_j(\otimes)}$.

Proof: Since the tardiness credibility coefficient $\delta \in [0,1]$, it is clear that $0 \leq \delta\alpha_i^+/(\alpha_i^+ - \alpha_i^{*+}) \leq \delta \leq (\alpha_i^{*-} - \delta\alpha_i^-)/(\alpha_i^{*-} - \alpha_i^-) \leq 1$ for the piecewise function of Eq. 4.

Thus, we have $0 \leq \delta\alpha_i^+/(\alpha_i^+ - \alpha_i^{*+}) \leq \delta \leq (\alpha_i^{*-} - \delta\alpha_i^-)/(\alpha_i^{*-} - \alpha_i^-) \leq 1$.

It can be seen that $T_i(\otimes) - a_j(\otimes) = (\alpha_i^- - \alpha_j^+, \alpha_i^{*-} - \alpha_j^{*+}, \alpha_i^{*+} - \alpha_j^{*-}, \alpha_i^+ - \alpha_j^-)$ $(\alpha_i^- - \alpha_j^+ \leq \alpha_i^-, \alpha_i^{*-} - \alpha_j^{*+} \leq \alpha_i^{*-}, \alpha_i^{*+} - \alpha_j^{*-} \leq \alpha_i^{*+}$ and $\alpha_i^+ - \alpha_j^- \leq \alpha_i^+)$.

Hence, $Con_{T_i(\otimes)} \geq Con_{T_i(\otimes)-a_j(\otimes)}$.

Q.E.D.

Theorem 3 In a scheduling sequence S, if several jobs exist in the process sequence of the k-th machine, the grey makespan and delivery period of the i-th job are $a_{ik}(\otimes)$ and $D_i(\otimes)$, respectively, and the grey makespan and delivery period of the j-th are $a_{jk}(\otimes)$ and $D_j(\otimes)$, respectively. Note that $a_{ik}(\otimes) = (a_{ik}^-, a_{ik}^{*-}, a_{ik}^{*+}, a_{ik}^+)$, $a_{jk}(\otimes) = (a_{jk}^-, a_{jk}^{*-}, a_{jk}^{*+}, a_{jk}^+)$, $i,j = 1,2,\dots,r$ and $k = 1,2,\dots,m$; r is the number of jobs in the process sequence of the k-th machine; i and j are respectively the positions of the i-th and j-th jobs in the process sequence of the k-th machine. If $D_i(\otimes) < D_j(\otimes)$ and $a_{ik}(\otimes) - D_i(\otimes) < a_{jk}(\otimes) - D_j(\otimes)$, then the j-th job must be processed after the i-th job.

Proof: Suppose the makespan is greater than 0 in a scheduling sequence S. Then, the mean tardiness credibility $f_k(S)$ of all jobs in the process sequence of the k-th machine can be expressed as:

$$f_k(S) = \frac{1}{r}\left[\sum_{w=1}^{i-1} Con_{T_w(\otimes)} + Con_{T_i(\otimes)} + \sum_{v=i+1}^{j-1} Con_{T_v(\otimes)} + Con_{T_j(\otimes)} + \sum_{u=j+1}^{r} Con_{T_u(\otimes)}\right] \tag{12}$$

If the j-th job is processed before the i-th job, the mean tardiness credibility $f_k(S')$ of all jobs in the process sequence of the k-th machine for the new optimal scheduling sequence S' can be expressed as:

$$f_k(S') = \frac{1}{r}\left[\sum_{w=1}^{i-1} Con_{T'_w(\otimes)} + Con_{T'_i(\otimes)} + \sum_{v=i+1}^{j-1} Con_{T'_v(\otimes)} + Con_{T'_j(\otimes)} + \sum_{u=j+1}^{r} Con_{T'_u(\otimes)}\right] \tag{13}$$

Only the i-th and the j-th jobs are interchanged, while the other jobs in the process sequence of the k-th machine are in the same positions. Thus, the grey tardiness of the u-th, v-th and w-th jobs can be expressed as: $T_u(\otimes) = T'_u(\otimes)$, $T_v(\otimes) = T'_v(\otimes)$ and $T_w(\otimes) = T'_w(\otimes)$, where $u \in [j+1, r]$, $v \in [i+1, j-1]$, $w \in [1, i-1]$, and

$$T'_j(\otimes) = \sum_{w=1}^{i-1} a_{wk}(\otimes) + a_{jk}(\otimes) - D_j(\otimes) \tag{14}$$

$$T_i(\otimes) = \sum_{w=1}^{i-1} a_{wk}(\otimes) + a_{ik}(\otimes) - D_j(\otimes) \tag{15}$$

Since $a_{ik}(\otimes) - D_j(\otimes) < a_{jk}(\otimes) - D_j(\otimes)$, we have $T'_j(\otimes) > T_i(\otimes)$. It can be seen from Theorem 2 that the tardiness credibility $Con_{T'_j(\otimes)} > Con_{T_i(\otimes)}$. Then, the grey tardiness $T'_i(\otimes)$ of the i-th job in scheduling sequence S' and that $T_j(\otimes)$ of the j-th job in scheduling sequence S can be expressed as:

$$T'_i(\otimes) = \sum_{w=1}^{i-1} a_{wk}(\otimes) + a_{jk}(\otimes) + \sum_{v=i+1}^{j-1} a_{vk}(\otimes) + a_{ik}(\otimes) - D_i(\otimes) \qquad (16)$$

$$T_j(\otimes) = \sum_{w=1}^{i-1} a_{wk}(\otimes) + a_{ik}(\otimes) + \sum_{v=i+1}^{j-1} a_{vk}(\otimes) + a_{jk}(\otimes) - D_j(\otimes) \qquad (17)$$

Since $D_j(\otimes) > D_i(\otimes)$, we have $T'_i(\otimes) > T_j(\otimes)$ according to Eqs. 16 and 17. It can be seen from Theorem 2 that the tardiness credibility $Con_{T'_i(\otimes)} > Con_{T_j(\otimes)}$. Therefore, Eqs. 12 and 13 show that the sum of tardiness credibility of the process sequence of the k-th machine is $f_k(S') > f_k(S)$. This means the scheduling sequence after exchanging the i-th job and the j-th job is no better than that before the exchange. Thus, the j-th job must be processed after the i-th job.
Q.E.D.

3. Design of hybrid grey cuckoo search

3.1 Cuckoo search

Artificial intelligence algorithm is usually used to solve job shop scheduling models. Artificial intelligence algorithms include swarm intelligence optimization algorithm, decision support algorithm [22] [23], data fusion algorithm [24] and machine learning algorithm [25], etc. Xu *et al.* [26] proposed a bat algorithm to solve the problem of two-flexible job shop scheduling. Cuckoo search (CS) [27] is a heuristic search algorithm developed by Yang and Deb in 2009. It was inspired by the obligate brood parasitism of some cuckoo species by laying their eggs in the nests of other host birds. The cuckoo algorithm has better performance than particle swarm optimization [28, 29]. Since it was established, the CS has been extensively studied by scholars around the world. In 2010, Yang and Deb applied the CS in the multi-objective solution problem [30]. In 2011, Valian *et al.* improved the CS based on feedforward neural network [31]. In the same year, Walton *et al.* proposed an improved Brown free gradient CS [32]. The basic idea of the CS comes from the breeding behaviour of cuckoo and the Lévy flight mode of birds. The Lévy flight is a random walk in which the step-lengths have a probability distribution that is heavy-tailed. When defined as a walk in a space of dimension greater than one, the steps made are in isotropic random directions. Considering its ability to avoid the local optimum trap and achieve good global optimization, the CS is adopted to solve the grey mixed integer programming model in this paper.

The CS is suitable for small scale experiments rather than optimization problems with heavy computing load. Relying solely on random walk, the search for optimal solution consumes lots of computing power, and cannot guarantee the fast convergence of the algorithm. To improve the computing performance, it is necessary to integrate the features of the problem into the CS in a particular situation. Below are the steps to create the hybrid grey cuckoo search (HGCS).

3.2 Code design

The CS is not directly applicable to the coding problem of discrete processes. Thus, the coding rule here is based on continuous processes. Suppose the position of each nest represents a feasible scheduling solution. For n jobs and m machines, the code of the feasible solution can be described by the locations of $n \times m$ nests. The code stands for a process sequence, in which each job must appear exactly m times. Taking a 4-job 3-machine problem as an example, if the position code is 132143123442, the corresponding job processing sequence should be $(J_{1,1}, J_{3,1}, J_{2,1}, J_{1,2}, J_{4,1}, J_{3,2}, J_{1,3}, J_{2,2}, J_{3,3}, J_{4,2}, J_{4,3}, J_{2,3})$, with $J_{i,j}$ being the j-th process of the i. In other words, the first job and the third job should be processed successively on the first machine; then, the sec-

ond job should be processed on the first machine, the first job should be processed on the second machine, and so on; Finally, the second job should be processed on the third machine.

3.3 Generation of initial population

In the CS, the nest location must be initialized at the start. However, the initial population is of poor quality because it is generated randomly without using any knowledge related to the problem to be solved. To improve the quality, the priority rules for minimum makespan and earliest delivery period in the JSP were introduced to develop a grey heuristic algorithm for initializing the locations of some nests. The specific steps are as follows.

For the four-parameter interval grey number $w(\otimes) = (\alpha^-, \alpha^{*-}, \alpha^{*+}, \alpha^+)$, let $sign(w(\otimes)) = \delta(\alpha^- + \alpha^{*-}) + (1-\delta)(\alpha^{*+} + \alpha^+)$ be the grey mark of $w(\otimes)$, and δ be the tardiness credibility coefficient. Then, $sign(a_{ik}(\otimes)) = \delta(a_{ik}^- + a_{ik}^*) + (1-\delta)(a_{ik}^* + a_{ik}^+)$ holds for the grey makespan $a_{ik}(\otimes) = (a_{ik}^-, a_{ik}^*, a_{ik}^+)$, and $sign(D_i(\otimes)) = \delta(d_i^- + d_i^{*-}) + (1-\delta)(d_i^{*+} + d_i^+)$ holds for the grey delivery period $D_i(\otimes) = (d_i^-, d_i^{*-}, d_i^{*+}, d_i^+)$. Thus, the minimum makespan and earliest delivery period can be expressed as $Min\{sign(a_{ik}(\otimes))\}$ and $Min\{sign(D_i(\otimes))\}$, respectively. Then, the said grey heuristic can be established through the following steps:

Step 1: Set up a set of grey markers for jobs, machines and delivery period:

$$P_k = \{sign(a_{1k}(\otimes)), sign(a_{2k}(\otimes)), \dots, sign(a_{nk}(\otimes))\}(k \in [1, m]),$$
$$M_l = \{sign(a_{l1}(\otimes)), sign(a_{l2}(\otimes)), \dots, sign(a_{lm}(\otimes))\}(l \in [1, n])$$
$$\text{and } D_q = \{sign(D_1(\otimes)), sign(D_2(\otimes)), \dots, sign(D_n(\otimes))\}.$$
$$M_l = M_l - sign(a_{lk'}(\otimes))$$

Step 2: Let $sign\left(D_q'(\otimes)\right) = Min(D_q)$, $sign(a_{ik'}(\otimes)) = Min(M_l)$ and $sign(a_{l'k}(\otimes)) = Min(P_l)$,

that is, assign the i'-th job to the k'-th machine. Then, we have $P_k = P_k - sign(a_{l'k}(\otimes))$, $D_q = D_q - sign(D_{l'}(\otimes))$, with $i = 1,2,\dots,n$ and $k = 1,2,\dots,m$.

Step 3: Repeat Step 2 until the jobs are assigned to each machine.

To diversify nest locations in the initial population, a third of the entire population was generated through the above steps, while the remaining parts were still generated randomly. The basic parameters of the CS were configured as: the dimension of the search space is m, the number of nests is n, the probability that the host bird discovers the egg laid by the cuckoo is P_a, and the maximum number of iterations is $MaxT$. After that, the initial nest locations were converted into process sequence through process coding, the value of objective function was computed corresponding to each nest location, and the optimal nest location was determined. Here, the goal is to calculate $min f(x), x \in S$, i.e. the minimum mean tardiness credibility of all jobs in the fuzzy JSP.

3.4 Generation of candidate population

First, the optimal nest location of the previous generation was preserved, and the fitness function $fitness = Q - f(x), x \in S$ of each nest location was calculated, with Q being a suitable positive number and $f(x)$ being the objective function. Then, the candidate population was generated by exploring paths and locations through Lévy flight. The direction of Lévy flight is arbitrary, and the step length obeys a heavy-tailed probability distribution. To sum up, the candidate population can be generated by:

$$x_i^{t+1} = x_i^t + \alpha \oplus Levy(\lambda), \quad i = 1,2,\dots n \tag{18}$$

where x_i^t is the location of the i-th nest in the t-th generation; α is the step length control coefficient, which is usually 1; $\oplus$ is the point-to-point multiplication; $Levy(\lambda)$ is the path of the random search of Lévy flight. The direction of Lévy flight obeys uniform distribution.

3.5 Optimal selection

The locations of all nests in the current population were improved by Theorem 3, that is, the delivery periods of the nest locations were compared. If the grey delivery period $D_i(\otimes) < D_j(\otimes)$ and $a_{ik}(\otimes) - D_i(\otimes) < a_{ik}(\otimes) - D_j(\otimes)$, the j-th job must be processed after the i-th job. Then, the selection operator of the CS was employed to compare the fitness between parent and candidate generations, and the individuals with high fitness were retained to the next generation. In other words, the algorithm calculates the maximum fitness $fitness_{max}$ of nest in the current population; if the maximum fitness of the previous generation is greater than the $fitness_{max}$, then the location of the nest with the maximum fitness in the current population must be replaced with that in the previous generation:

$$x_i^{t+1} = \begin{cases} y_i^t, f(y_i^t) \leq f(x_i^t) \\ x_i^t, f(y_i^t) > f(x_i^t) \end{cases} \tag{19}$$

The operation selection adopts the greedy strategy, which can track the optimal solution found in the evolution. This strategy prevents the degradation in the iterative process and accelerates the convergence.

3.6 Random migration

Random migration is similar to the mutation operation of the genetic algorithm. In the basic CS, the cuckoo will look for a new nest randomly after its egg was discovered by the host bird. This type of random search may fall into the local optimum trap. To solve the problem, the new nest location could be searched for based on the undiscovered nests, i.e. the crossover operation of the genetic algorithm. The crossover operation goes like this: sort N host nests randomly and save the result; repeat the operation; get a crossover step length by subtracting the results of the previous two steps; find new host nest based on the crossover step length.

Suppose the probability that the host bird discovers the cuckoo egg is P_a. If P_a is smaller than a random number obeying uniform distribution $\gamma \in [0,1]$, then the crossover operation was implemented to find new host nests, and the cross-border check was performed on the locations of the nests. The nest locations should be either retained or replaced by locations with better fitness, forming a new set of better locations. Through random migration, the relatively good nests were retained, the poor individuals were eliminated and the population was diversified.

3.7 Termination conditions

After optimal selection and random migration, the optimal nest position and the best fitness should be examined. If the iteration termination condition is satisfied (reaching the maximum number of iterations $MaxT$) or the required accuracy is attained, the global optimal value and the corresponding global optimal position should be outputted; Otherwise, the optimal selection and random migration should be performed again. Finally, the required scheduling plan should be obtained by decoding the global optimal location.

4. Results and discussion

To verify the proposed HGCS algorithm, a simulation was carried out on the classical example of $6(3) \times 6$. The example was proposed by Sakawa to simulate the scheduling problem with fuzzy makespan and delivery period. In our experiment, the HGCS and CS parameters were configured as: the number of nests was 30, the probability that the host bird discovers the cuckoo egg $P_a = 0.25$, the maximum number of iterations $MaxT = 1,000$. The simulation was carried out on Matlab 2014a (OS: Windows 7; CPU: Intel® Core™ i3-2350M 2.0GHz; Memory: 2GB).

For the $6(3) \times 6$ grey JSP, the minimum mean tardiness credibility was taken as the objective function. The grey makespan and delivery period are given in Table 1, and the process sequence is listed in Table 2.

Table 1 Grey makespan and grey delivery period

Job	Processing time a						Delivery D	Weight
	M1	M2	M3	M4	M5	M6		
J1	(5,6,13)	(1,2,3)	(2,3,4)	(2,3,4)	(3,4,5)	(3,4,5)	(20,25,35,40)	0.15
J2	(3,4,5)	(2,4,5)	(1,3,5)	(5,6,7)	(6,7,8)	(4,5,6)	(20,25,35,40)	0.15
J3	(1,2,3)	(1,2,3)	(1,2,3)	(3,4,5)	(4,5,6)	(5,6,7)	(20,28,32,40)	0.25
J4	(3,4,6)	(2,3,5)	(3,4,5)	(2,3,4)	(1,2,3)	(2,3,4)	(20,28,32,40)	0.25
J5	(2,3,4)	(4,5,6)	(2,3,4)	(1,2,3)	(2,3,4)	(3,4,5)	(30,35,40,45)	0.10
J6	(2,3,4)	(3,4,5)	(2,3,4)	(1,2,3)	(6,7,8)	(4,5,6)	(30,35,40,45)	0.10

Table 2 Process sequence

Job	Processing order
J1	1-5-2-6-4-3
J2	1-2-3-6-4-5
J3	3-6-5-4-2-1
J4	6-5-4-2-1-3
J5	6-5-4-3-2-1
J6	5-6-1-2-3-4

Table 3 Minimum mean tardiness credibility of the HGCS and the CS

δ	Algorithm	$f(x)$										mean value
		1	2	3	4	5	6	7	8	9	10	
0.1	CS	0.0491	0.0486	0.0480	0.0499	0.0483	0.0492	0.0488	0.0484	0.0493	0.0497	0.0489
	HGCS	0.0481	0.0485	0.0472	0.0480	0.0477	0.0476	0.0480	0.0479	0.0483	0.0470	0.0478
0.2	CS	0.0723	0.0786	0.0655	0.0689	0.0822	0.0755	0.0623	0.0741	0.0777	0.0814	0.0739
	HGCS	0.0564	0.0654	0.0423	0.0568	0.0601	0.0656	0.0564	0.0589	0.0641	0.0532	0.0579
0.5	CS	0.1895	0.1876	0.1910	0.1747	0.1885	0.1841	0.1678	0.1708	0.1933	0.1850	0.1832
	HGCS	0.1331	0.1230	0.1445	0.1211	0.1391	0.1509	0.1436	0.1280	0.1472	0.1370	0.1368
0.7	CS	0.2135	0.2213	0.2331	0.2210	0.2178	0.2311	0.2258	0.2361	0.2283	0.2209	0.2249
	HGCS	0.1954	0.2031	0.1986	0.1870	0.1986	0.2015	0.2364	0.2013	0.1845	0.1922	0.1999
0.8	CS	0.2548	0.2631	0.2745	0.2610	0.2511	0.2537	0.2610	0.2503	0.2587	0.2611	0.2589
	HGCS	0.2456	0.2315	0.2498	0.2356	0.2475	0.2501	0.2468	0.2468	0.2511	0.2432	0.2448

The tardiness credibility coefficient was set to 0.1, 0.2, 0.5, 0.7 and 0.8, respectively. The target value is the minimum mean tardiness credibility. The program was ran 10 times at random to obtain the mean value. The experimental results are recorded in Table 3.

As shown in Table 3, the mean target value increased with the tardiness credibility cient δ. This is mainly because the growth in the coefficient led to a significant increase in the weight of the probability measure.

With δ = 0.2, 0.5, 0.2 and 0.5, simulations were conducted respectively by the CS and the HGCS. For each value of δ, the convergence curve of the mean tardiness credibility was plotted according to one of the simulations (e.g., Fig. 2, 3, 4 and 5).

According to Table 3, the HGCS always obtained a smaller mean target value than the CS, whichever the tardiness coefficient δ. Figs. 2 to 4 reveal that the HGCS obtained better optimal target value and realized faster convergence than the CS when the curve tended to global convergence.

The objective of this paper is to minimize the average value of tardy reliability of all jobs. As is shown in Table 3, the running results are obtained by the 10 operations, whatever CS algorithm or HGCS algorithm. For example, when the operation has run six times, the value of CS algorithm is 0.1841 at δ=0.5, HGCS 0.1509. Tardiness credibility represents the weighting sum of tardiness of probability measure and certainty measure. By definition 3, we can see that the uncertainty tardiness of jobs can be well embodied by possibility measures and necessity measures. The experimental values of the algorithm are 0.1841 and 0.1509, which fully reflect the tardiness of jobs.
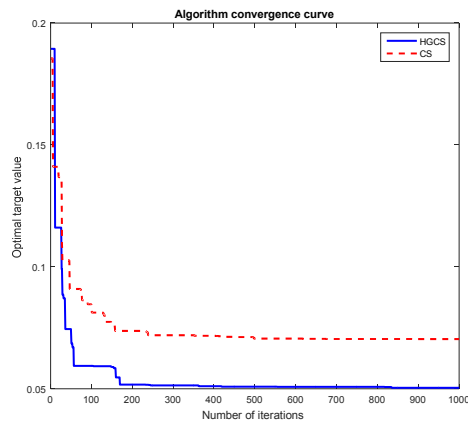
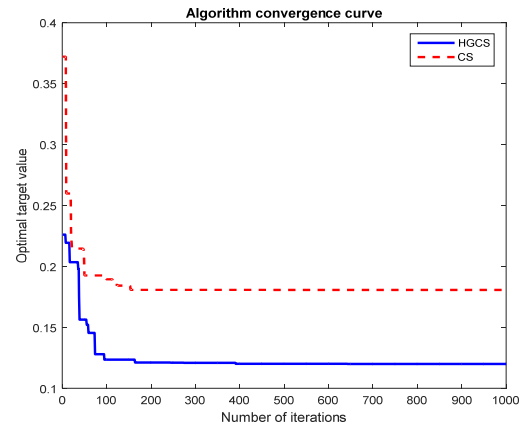**Fig. 2** Optimal target value convergence curves (δ = 0.2)

Fig. 3 Optimal target value convergence curves (δ = 0.5)
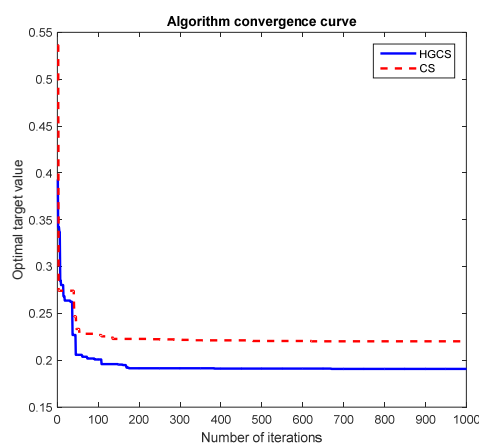

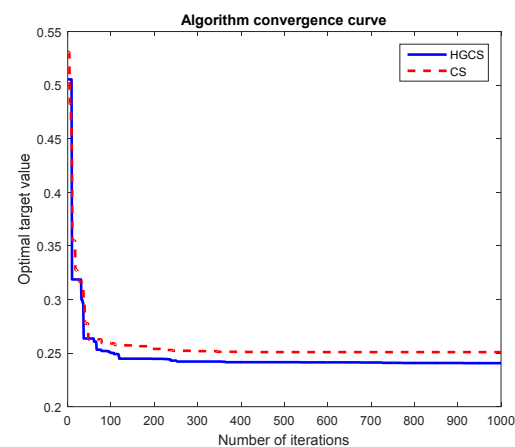**Fig. 4** Optimal target value convergence curves (δ = 0.7)

Fig. 5 Optimal target value convergence curves (δ = 0.8)

5. Conclusion

This paper proposes a grey scheduling model for the fuzzy makespan and delivery period in the JSP, which expresses the fuzzy makespan as three-parameter interval grey numbers and the fuzzy delivery period as four-parameter interval grey numbers. This approach is more generic than expressing fuzzy scheduling factors with triangular and trapezoid fuzzy numbers. On this basis, some features of optimal grey scheduling were obtained from the properties of grey numbers. These features were used to improve the basic CS into the HGCS. Then, simulations were performed on the classical example of $6(3) \times 6$. The results show that the HGCS can solve the job tardiness problem in fuzzy JSP with fuzzy makespan and delivery period. Compared to the basic CS, the HGCS achieved an excellent scheduling result. Of course, the solution presented here only applies to single objective optimization problem. The future research will tackle multi-objective optimization problems under resource constraints.

Acknowledgement

References

[1] Zhang, H., Liu, S., Moraca, S., Ojstersek, R. (2017). An effective use of hybrid metaheuristics algorithm for job shop scheduling problem, *International Journal of Simulation Modelling*, Vol. 16, No. 4, 644-657, doi: 10.2507/ IJSIMM16(4)7.400.

[2] Modrák, V., Pandian, R.S. (2010). Flow shop scheduling algorithm to minimize completion time for *n*-jobs *m*-machines problem, *Tehnički vjesnik – Technical Gazette*, Vol. 17, No. 3, 273-278.

[3] Chaudhry, I.A., Usman, M. (2017). Integrated process planning and scheduling using genetic algorithms, *Tehnički vjesnik – Technical Gazette*, Vol. 24, No. 5, 1401-1409, doi: 10.17559/TV-20151121212910.

[4] Huang, X.W., Zhao, X.Y., Ma, X.L. (2014). An improved genetic algorithm for job-shop scheduling problem with process sequence flexibility, *International Journal of Simulation Modelling*, Vol. 13, No. 4, 510-522, doi: 10.2507/ IJSIMM13(4)CO20.

[5] Ishii, H., Tada, M., Masuda, T. (1992). Two scheduling problems with fuzzy due-dates, *Fuzzy Sets and Systems*, Vol. 46, No. 3, 339-347, doi: 10.1016/0165-0114(92)90372-B.

[6] Noori-Darvish, S., Mahdavi, I., Mahdavi-Amiri, N. (2012). A bi-objective possibilistic programming model for open shop scheduling problems with sequence-dependent setup times, fuzzy processing times, and fuzzy due dates, *Applied Soft Computing*, Vol. 12, No. 4, 1399-1416, doi: 10.1016/j.asoc.2011.11.019.

[7] Gharehgozli, A.H., Tavakkoli-Moghaddam, R., Zaerpour, N. (2009). A fuzzy-mixed-integer goal programming model for a parallel-machine scheduling problem with sequence-dependent setup times and release dates, *Robotics and Computer-Integrated Manufacturing*, Vol. 25, No. 4-5, 853-859, doi: 10.1016/j.rcim.2008.12.005.

[8] Simeunović, N., Kamenko, I., Bugarski, V., Jovanović, M., Lalić, B. (2017). Improving workforce scheduling using artificial neural networks model, *Advances in Production Engineering & Management*, Vol. 12, No. 4, 337-352, doi: 10.14743/apem2017.4.262.

[9] Stanujkic, D., Magdalinovic, N., Jovanovic, R., Stojanovic, S. (2012). An objective multi-criteria approach to optimization using MOORA method and interval grey numbers, *Technological and Economic Development of Economy*, Vol. 18, No. 2, 331-363, doi: 10.3846/20294913.2012.676996.

[10] Zeng, B., Liu, S.-F., Xie, N.-M., Cui, J. (2010). Prediction model for interval grey number based on grey band and grey layer, Control and Decision, Vol. 25, No. 10, 1585-1588, doi: 10.13195/j.cd.2010.10.148.zengb.002.

[11] Luo, D., Wang, X. (2012). The multi-attribute grey target decision method for attribute value within three-parameter interval grey number, *Applied Mathematical Modelling*, Vol. 36, No. 5, 1957-1963, doi: 10.1016/ j.apm.2011.07.074.

[12] Luo, D. (2009). Decision-making methods with three-parameter interval grey number, *Systems Engineering – Theory & Practice*, Vol. 29, No. 1, 124-130, doi: 10.1016/s1874-8651(10)60033-6.

[13] Jin, F., Liu, P., Zhang, X. (2013). The multi-attribute group decision making method based on the interval grey linguistic variables weighted harmonic aggregation operators, *Technological and Economic Development of Economy*, Vol. 19, No. 3, 409-430, doi: 10.3846/20294913.2013.821685.

[14] Kamfiroozi, M.H., Aliahmadi, A., Jafari-Eskandari, M. (2012). Application of three parameter interval grey numbers in enterprise resource planning selection, *International Journal of Information, Security and Systems Management*, Vol. 1, No. 2, 72-77.

[15] Hu, Q., Zhang, W., Yu, L. (2007).The research and application of interval numbers of three parameters, *Engineering Science*, Vol. 9, No. 3, 47-51, doi: 10.3969/j.issn.1009-1742.2007.03.008.

[16] Sahu, N.K., Datta, S., Mahapatra, S.S. (2013). Decision making for selecting 3PL service provider using three parameter interval grey numbers, *International Journal of Logistics Systems and Management*, Vol. 14, No. 3, 261-297, doi: 10.1504/ijlsm.2013.052061.

[17] Yang, B., Zhao, J. (2013). Correlation coefficients of hesitant three-parameter interval grey number and their applications to clustering analysis, *Journal of Grey System*, Vol. 25, No. 2, 139-147.

[18] Song, S. (2018). Application of gray prediction and linear programming model in economic management, *Mathematical Modelling of Engineering Problems*, Vol. 5, No. 1, 46-50, doi: 10.18280/mmep.050107.

[19] Luo, D. (2009). Decision-making methods with three-parameter interval grey number, *Systems Engineering – Theory & Practice*, Vol. 29, No. 1, 124-130, doi: 10.1016/s1874-8651(10)60033-6.

[20] Hu, J., Lin, Z. (2013). Multi-criteria decision making method based on interval numbers of four parameters, *Operations Research and Management Science*, Vol. 6, 84-91, doi: 10.3969/j.issn.1007-3221.2013.06.014.

[21] Wang, N., Hu, L.-P., Li, B.-J. (2014). Study on the distance entropy model based on three-parameter interval grey numbers and its application, *Journal of Henan Agricultural University*, Vol. 48, No. 3, 386-390.

[22] Zaccone, R., Sacile, R., Fossa, M. (2017). Energy modelling and decision support algorithm for the exploitation of biomass resources in industrial districts, *International Journal of Heat and Technology*, Vol. 35, No. S1 (Special Issue), S322-S329, doi: 10.18280/ijht.35Sp0144.

[23] Dai, Y., Zhu, X., Zhou, H., Mao, Z., Wu, W. (2018) Trajectory tracking control for seafloor tracked vehicle by adaptive neural-fuzzy inference system algorithm, *International Journal of Computers Communications & Control*, Vol. 13, No. 4, 465-476, doi: 10.15837/ijccc.2018.4.3267.

[24] Wang, T.C., Xie, Y.Z. (2016). BP-GA data fusion algorithm studies oriented to smart home, *Mathematical Modelling of Engineering Problems*, Vol. 3, No. 3, 135-140.

[25] Tian, Y., Chen, W., Li, L., Wang, X., Liu, Z. (2018). Gait recognition via coalitional game-based feature selection and extreme learning machine, *NeuroQuantology*, Vol. 16, No. 2, 32-39, doi: 10.14704/nq.2018.16.2.1173.

[26] Xu, H., Bao, Z.R., Zhang, T. (2017). Solving dual flexible job-shop scheduling problem using a Bat Algorithm, *Advances in Production Engineering & Management*, Vol. 12, No. 1, 5-16, doi: 10.14743/apem2017.1.235.

[27] Yang, X.-S., Deb, S. (2010). Cuckoo search via Lévy flights, In: *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore, India, 210-214, doi: 10.1109/NABIC.2009.5393690.

[28] Liang, C.H., Zeng, S., Li, Z.X., Yang, D.G., Sherif, S.A. (2016). Optimal design of plate-fin heat sink under natural convection using a particle swarm optimization algorithm, *International Journal of Heat & Technology*, Vol. 34, No. 2, 275-280, doi: 10.18280/ijht.340217.

[29] Sen, G.D., Sharma, J., Goyal, G.R., Singh, A.K. (2017). A multi-objective PSO (MOPSO) algorithm for optimal active power dispatch with pollution control, *Mathematical Modelling of Engineering Problems*, Vol. 4, No. 3, 113-119, doi: 10.18280/mmep.040301.

[30] Yang, X.-S., Deb, S. (2013). Multiobjective cuckoo search for design optimization, *Computers & Operations Research*, Vol. 40, No. 6, 1616-1624, doi: 10.1016/j.cor.2011.09.026.

[31] Valian, E., Mohanna, S., Tavakoli, S. (2011). Improved cuckoo search algorithm for feedforward neural network training, *International Journal of Artificial Intelligence & Applications*, Vol. 2, No. 3, 36-43, doi: 10.5121/ijaia.2011.2304.

[32] Walton, S., Hassan, O., Morgan, K., Brown, M.R. (2011). Modified cuckoo search: A new gradient free optimisation algorithm, *Chaos, Solitons & Fractals*, Vol. 44, No. 9, 710-718, doi: 10.1016/j.chaos.2011.06.004.

Design, finite element analysis (FEA), and fabrication of custom titanium alloy cranial implant using electron beam melting additive manufacturing

Ameen, W.[a,b,*], Al-Ahmari, A.[a,b], Mohammed, M.K.[b], Abdulhameed, O.[a,b], Umer, U.[b], Moiduddin, K.[b]

[a]Industrial Engineering Department, King Saud University, Kingdom of Saudi Arabia
[b]Princess Fatima Alnijiris's Research Chair for Advanced Manufacturing Technology (FARCAMT Chair), Advanced Manufacturing Institute, King Saud University, Kingdom of Saudi Arabia

A B S T R A C T

Skull defect reconstruction is one of the most difficult challenges faced by the surgeons because of the complex shape of the skull. Skull defects are dramatically increasing with the increase in road accidents, tumors, and wars, thereby increasing the demand for reconstruction of skull. It is difficult to manufacture standard implants for skull defects especially for large and complex defects, due to the complexity and the difference in anatomy of skulls. Design and fabrication of custom cranial implant is required in these cases. The conventional technologies face multiple challenges in fabricating lightweight custom cranial implants closer to that of bone in terms of weight; the difference in the weight introduces stress-shielding effects onto the surrounding bone. In order to overcome this problem, several researches proposed lattice structure implants fabricated by additive manufacturing. However, lattice structure implants are difficult to remove later when some problems are encountered. This paper presents a methodology of design analysis and fabrication of solid lightweight custom cranial implant using additive manufacturing. A Case study is presented where, a custom cranial implant is designed and analysed using finite element analysis (FEA) and then fabricated using electron beam melting (EBM) additive manufacturing. The titanium alloy Ti6Al4V which is biocompatible and non-toxic is used as the implant material. The functionality, fitting, and aesthetic of the proposed design are evaluated. The results show the successful fabrication of thin custom cranial implant for skull defect reconstruction via EBM technology. The fabricated implant has sufficient strength, weight close to the weight of the removed bone portion while maintaining a good fit and aesthetics.

1. Introduction

Skull defect reconstruction is one of the most difficult surgical operations, due to the complexity of the skull shape and the difference of the skulls anatomy. The best way of treating skull defects would be autogenous bone transplantation as this will have less complications of infection, aggressive foreign body reaction, extrusion and damage to the soft tissue and skin [6]. However for the large and complex defects, the use of autologous bone reconstruction is restricted due to the limited availability of donors. Hence, there is a push towards other material implants. Titanium alloy (Ti6Al4V) is biocompatible (non-toxic and not rejected by the body), lightweight and high strength

alloy suitable for applications like medical implants [42, 43]. In general, it is difficult to manufacture standard implants for skull defects as compared to joint prosthesis because of the complexity and the difference in anatomy of the skulls. Therefore, cranial implant are fabricated on a customized and individual basis. Over the years, attempts were made to produce Ti6Al4V implants for large skull defects through manufacturing techniques like Casting, Milling and Forming. These processes invariably suffer from the problems associated with Ti6Al4V. For instance, Machining of Ti6Al4V is very hard and expensive. Casting on the other hand is cost effective but time consuming. Metal-based Additive Manufacturing has shown favorable results in custom built implant fabrication with properties as good as wrought and cast, if not better.

Additive Manufacturing (AM) is one of the latest approaches used for manufacturing products in advanced applications [1]. In AM, parts are fabricated by adding the material in a layer by layers pattern. This is the opposite of machining process, in which the material is removed or subtracted from a block to achieve the shape of the desired object [1]. In general, AM technologies are used to produce parts for high performance applications within biomedical, aerospace and automotive industries. The primary function of these technologies is to produce complicated internal features with functional design where machining and casting would require too much of lead-time, and wastage of material [2]. Medical applications generally involve complex parts and customization. In recent years AM technologies have been successfully utilized to produce various custom implants [3, 43]. Electron Beam Melting (EBM) is one of the latest AM technologies suitable for bio-medical applications [43]. In this process an electron beam melts the metal alloy powder into near net shape solid part in a layer by layer manner under controlled vacuum [4, 5].

In this research, EBM system commercialized by ARCAM AB is used to fabricate a lightweight cranial implant from Ti6Al4V ELI alloy. The fabricated implant delivers the functionality, geometric fit and aesthetics for skull defect reconstruction.

2. Literature review

Medical implants are commonly used for bone reconstructions in orthopedic, cranio-maxillofacial, dental and cosmetic surgery. The custom implant is designed and fabricated to fit the specification of patient. Design of cranial implants is the first step in skull defect reconstruction. The geometry of the implant is gathered from Computerized Tomographic (CT) scan data to achieve the fit. The design is further modified to achieve desired mechanical properties for improved performance of the implant. He *et al.* have presented a methodology in custom designed implant with the integration of Magnetic Resonance Imaging (MRI) scanning, image processing and AM technology [7]. Selection of implants materials is critical issue in craniofacial reconstructions. Alaa Kamel Abdel-Haleem *et al.* have described the advantages and disadvantages of using the rib grafts in skull and neck reconstruction [6]. Several biocompatible materials such as polyethylmethacrylate (PMMA), hydroxyapatite (HA) and Polyethylene were earlier used as the implant materials but each one of those has their own limitations [8-10].

Titanium alloy (Ti6Al4V ELI) is used in various alloys of iron, vanadium, aluminum, etc. [11]. It is used especially in aerospace, military and increasingly more in medical prosthetics [12]. Titanium alloys are one of the most widely used biocompatible material when compared to other bio-metals such as stainless steel and cobalt-chromium [13]. Ti6Al4V ELI (extra low interstitial) is the highest purity version of Ti6Al4V. It is commonly used in maxillofacial and craniofacial regions. Typically Titanium implants are fabricated as solid or mesh forms [14]. Several technologies have been used for fabrication the cranial implants. Traditionally the medical sculptors were employed to manufacture the implant based on the anatomical model using wax and clay [15]. Currently, however anatomical shape reconstructions are done using clay and computer aided design tools. Since the latter half of 1990's, Computer aided design cranial implants has been developed [16, 17]. Also a Computer Numerically Controlled (CNC) milling machine [18], direct milling method [19], and casting method are used for the fabrication of cranial implants [20].

In earlier studies, cranial defects were repaired using bulk titanium implants, which were 1.6 times more heavier than the removed bone [21, 29]. This bulky titanium implant due to the differences in young's modulus introduces stresses on the surrounding bone-implant interface known as

stress-shielding effect [22]. Researchers have tried to reduce the stress shielding effect, by introducing porous structure and making it lighter in weight, but with no clear evidence and investigation on mechanical and structural properties[23, 24]. In addition, the porous structure are difficult to remove later in case of some problem. Recently additive-manufacturing technology has shown potential for producing custom medical implants with altered mechanical properties to match the requirements of bone replacements. EBM technology has been successfully employed in the fabrication of titanium based custom design implants in orthopedic, craniofacial and maxillofacial surgeries [25-28]. This paper presents a methodology in the design, analysis and fabrication of a thin and lighter customized cranial implant with mechanical properties closer to that of bone using additive manufacturing.

3. Materials and methods

3.1 Design of the cranial defect implant

The computed tomography (CT) scan data of the patient as a Digital Imaging and Communica-tions in Medicine (DICOM) file is acquired. The CT data is then processed using MIMICS® software to generate the CAD model for the required anatomy.

Fig. 1 shows the screenshot of the MIMICS® with scan data in three orientations: axial, sagittal and coronal with a space to display the reconstructed 3D image. Segmentation of CT 2D slice images is done by selecting specific image intensities (Hounsfield units) within the region of interest. Hounsfield units (HU) is a system to measure the attenuation coefficient of tissues in CT scan images. HU are also termed CT numbers. The use of CT numbers or Hounsfield units provides an indication of the nature of the tissues. Specific tissues such as bone, skin and muscles are identified by their Hounsfield units.

Thresholding is then carried out by selecting the upper and lower threshold values of image intensities. Similarly, the pathological area is also delineated. Bone has a higher Hounsfield value when compared to skin and soft tissues as it absorbs most of the radiation.

Table 1 shows the Hounsfield values of some tissues that are commonly studied with CT images.
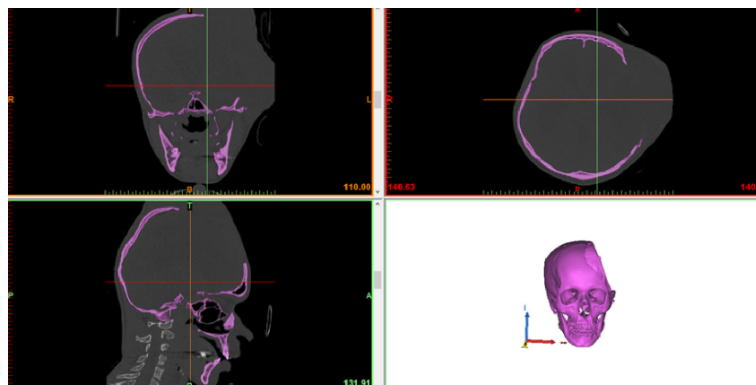
Fig. 1 CT scan images in three orientations

Table 1 Hounsfield values of tissues

Tissue	CT number (HU)
Bone	1000
Liver	40-60
White matter (brain)	46
Grey matter (brain)	43
Blood	40
Muscle	Oct-40
Kidney	30
Cerebrospinal fluid	15
Water	0
Fat	50 to -100
Air	-100

Region growing is the process by which noise is minimized and structures that are not connected in the image data are eliminated, resulting in a set of pixels that are connected within the same layer as well as with the upper and lower layers of data as seen in Fig. 2(a). By using the 3D reconstruction function, a 3D rendered model of the anatomy is generated. A skull with the defect rendered is seen Fig. 2(b). The reconstructed skull defect model was then used for implant reconstruction using 3 MATIC® software (Materialise NV, Leuven, Belgium). The steps involved during implant reconstruction include creation of a datum plane (symmetry plane) as shown in Fig. 2(c). Assuming the human body structure to be symmetric, the defect side (left) was removed from the center using the datum plane and the right side (healthy) is mirrored to reconstruct the defective area as shown in Fig. 2(d,e,f) respectively. Furthermore, the symmetry skull and the reconstructed skull with defect were merged together as shown in Fig. 2(g). The Boolean subtraction of merged symmetry skull from reconstructed skull with defect is carried out as shown in Fig. 2(h). The cranial implant produced from the subtraction process is shown in Fig. 2(i). Dimension Elite 3d printer is used to fabricate the skull and the implant prototypes as shown in Fig. 2(j,k). The fitness of the designed implant was tested visually as shown in Fig. 2(l) the result show good fitness between the fabricated model and the implant.
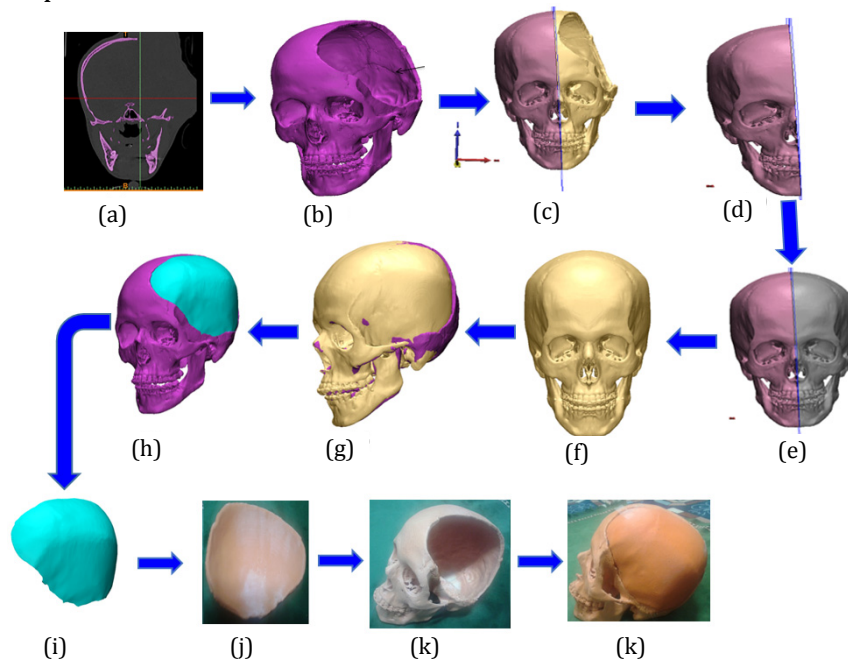
Fig. 2 Design steps of custom cranial defect implant in 3 MATIC®

Four fixation plates were designed with screw hole slots for the fixation and attachment of the screws to the cranium as illustrated in Fig. 3. The designed implant was of the same thicknesses as that of the bone portion 3.18-5.76 mm. In order to minimize the stress shielding effect, the implant thickness was reduced to bring down the weight of the Ti6Al4V ELI implant close to the bone portion of skull. The implant thickness was reduce to 0.5 mm (minimum thickness that can be produced by EBM technology) [30], by an inward offset operation on the thick cranial implant using Geomagic software.
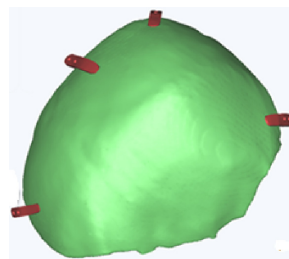
Fig. 3 Cranial implant with fixation lips

3.2 Finite element analysis

Finite element analysis (FEA) method was used to evaluate the functionality of the designed custom cranial implant. Abaqus software was used for pre- processing, solving and post-processing of the models. Table 2 shows the material properties assigned to the FE models, in which cortical bone was assigned to the skull with defect model and Arcam Ti–6Al–4V ELI was assigned to the custom cranial implant [31]. Good quality meshing with Tetrahedron elements was performed on the skull and skull implant model as shown in Fig. 4. Fine mesh was considered for the cranial implant and course mesh for the defected skull model. Table 3 shows the number of elements and nodes used during mesh generation.

Table 2 Material properties used in FE model

| Material	Young's modulus (MPa)	Poisson's ratio	Yield strength (MPa)
Cortical bone	13,700	0.3	122
Arcam Ti–6Al–4V ELI	120,000	0.3	930
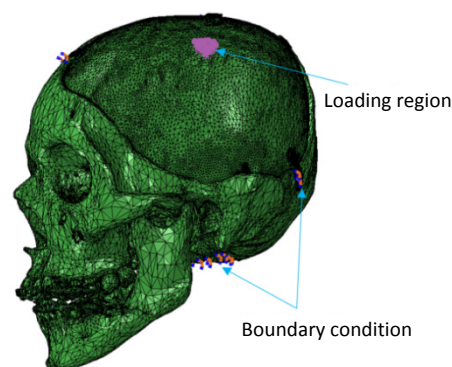
Fig. 4 Mesh generation, loads and boundary condition

A static force was applied in the small circular region (1 cm²) at the center of the implant. As recommended by the doctors the static loading simulated a relaxed person resting on a pillow. The applied force is 50 N, which corresponds to the approximate weight of the head. In addition, a further forces load of 1780N was applied which simulates the impact of a tennis ball at an average speed of 30 m/s [32-36].

Table 3 Mesh data for both skull and implant

FE Model	Number of elements	Number of nodes
Skull model	179409	368448
Implant model	117252	234532

4. Results and discussion

4.1 Finite element results

Fig. 5 shows the result of static finite element analysis with the application of 50 N force load. The maximum developed stress on the Ti6Al4V ELI implant is 1 MPa which is well below the allowable limit of material. Moreover, it is concentrated on the fixing plates of the implant. The results shows the highest strain and displacement developed on the cranial implant are 0.0000041 mm and 0.00019 mm as shown in Fig. 5(b) and Fig. 5(c), respectively. Maximum strain and displacement is also within allowable limit of material and the maximum strain is concentrated in the fixing plates and the maximum displacement is located in the middle of the implant.

The results of FE model simulating the impact of a tennis ball is shown in Fig. 6. The results shown that the maximum stress, strain and displacement are 36 MPa, 0.00014 mm and 0.0069 mm respectively. In addition, the maximum developed stress and strain is located near the fixing plates whereas the displacement is found in the middle of the implant.

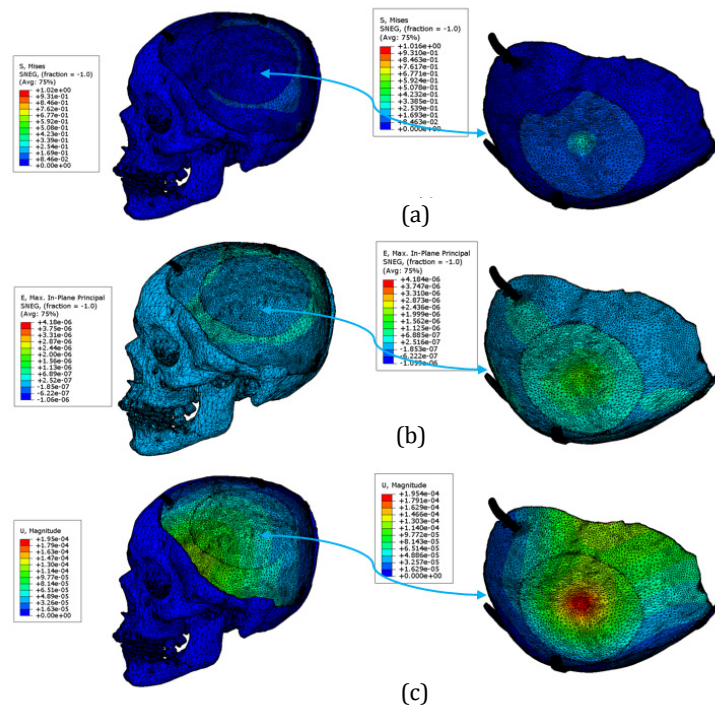**Fig. 5** Results of static loading: (a) Stress distribution, (b) Resulting strain, (c) Maximum displacement
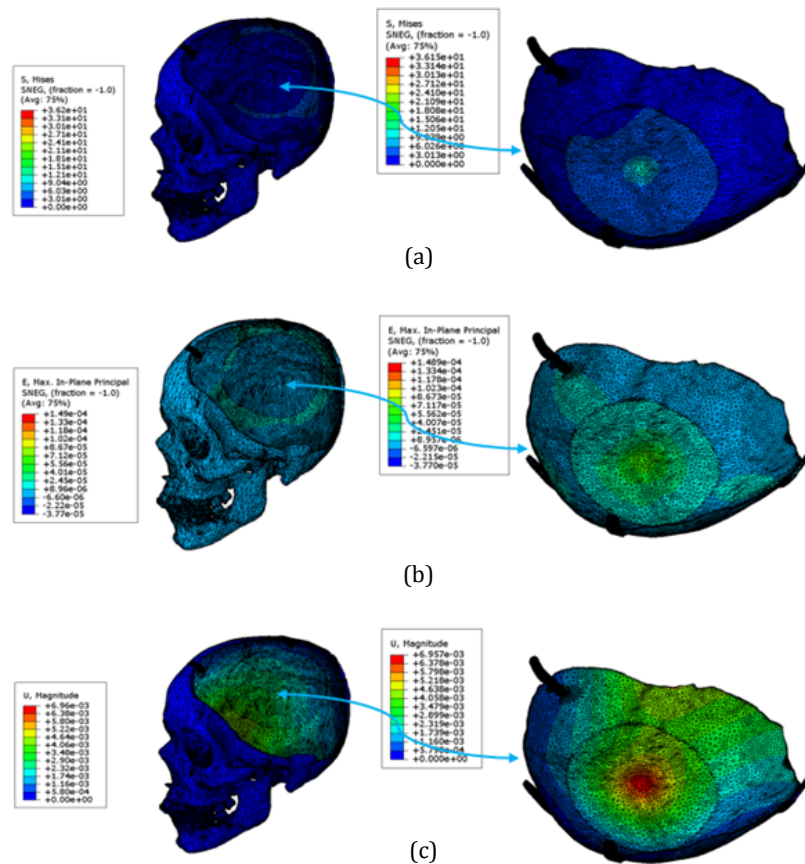

Fig. 6 Results of impact loading: (a) Stress distribution, (b) Resulting strain, (c) Maximum displacement

The results of both test loading showed that the maximum stresses, strains and displacements were within allowable limit of the Ti6Al4V ELI material, validating the designed implant with 0.5 mm thickness to be able to withstand load.

4.2 Implant fabrication results

Titanium alloy (Ti6Al4V) powder with the chemical compositions as shown in Table 4 is used as the feedstock material. The particle size analysis revealed that the size of powder particles in the range 53-107 µm with mean approximation of 75 µm as shown in Fig. 7. The morphology of the powder particles are mostly spherical in shape with little deviations in geometry. Fig. 8 illustrates the Scanning electron microscope (SEM) analysis of Ti6Al4V ELI powder using JSM-6610LV (JOEL, United States).

Table 4 Chemical composition of the Ti6Al4V ELI powder

Element	Al	V	C	Fe	O	Ti
Weight %	6.04	4.05	0.013	0.0107	0.13	Bal.

ARCAM A2 EBM machine was used to fabricate the custom Ti6Al4V ELI implant. The schematic of EBM and the Arcam setup used in this work is shown in Fig. 9 and 10, respectively [40]. EBM setup comprises of a heated tungsten filament (cathode) in a grid cup (anode) which produces the electron beam. The electrons are charged and accelerated to a kinetic energy of about 60 keV producing a maximum power of 4.8 kW with a beam spot size of 0.1-0.4 mm.

The three magnetic lenses including astigmatism lens, focus lens and deflection coils controls the direction of electron beam. The two hoppers holds the feedstock powder and the raking blade is used to spread the powder evenly over the build area. The build table moves down by one layer thickness (50 µm) and a new layer of powder is dispensed as the build progresses. Based on the set layer thickness the electron beam melts the powder in a layer by layer manner to produce the desired geometry [37,38]. The total built time for the cranial implant was approximately 8 hours with additional 3 to 4 hours for in cooling and 1 hours for post-processing (removing the sintered powder and support structures).
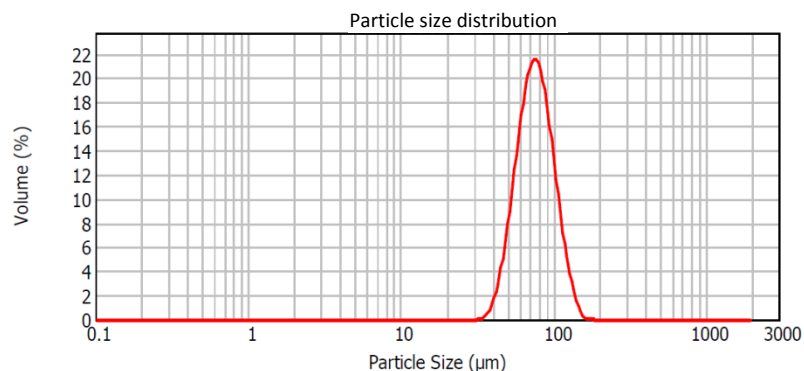

Fig. 7 Powder particle size distribution measured by laser diffraction technique
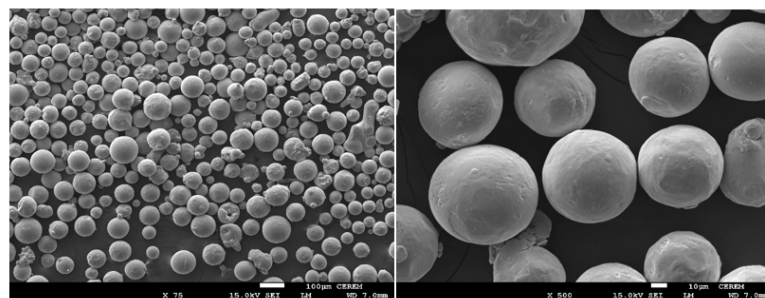
Fig. 8 Scanning electron microscope image of Ti6Al4V ELI powder particles
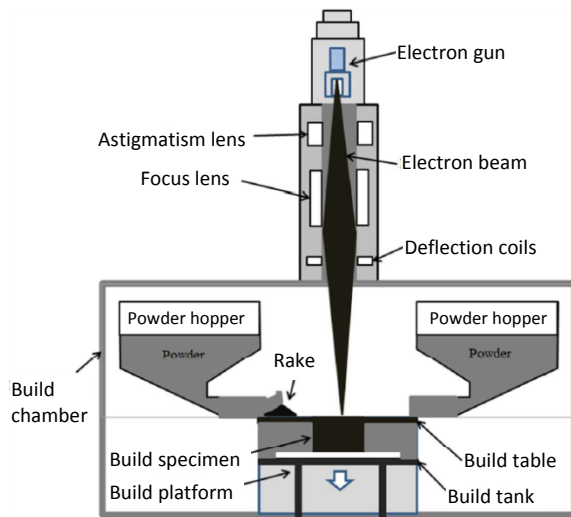
Fig. 9 Schematic of EBM process

Fig. 10 EBM setup from ARCAM
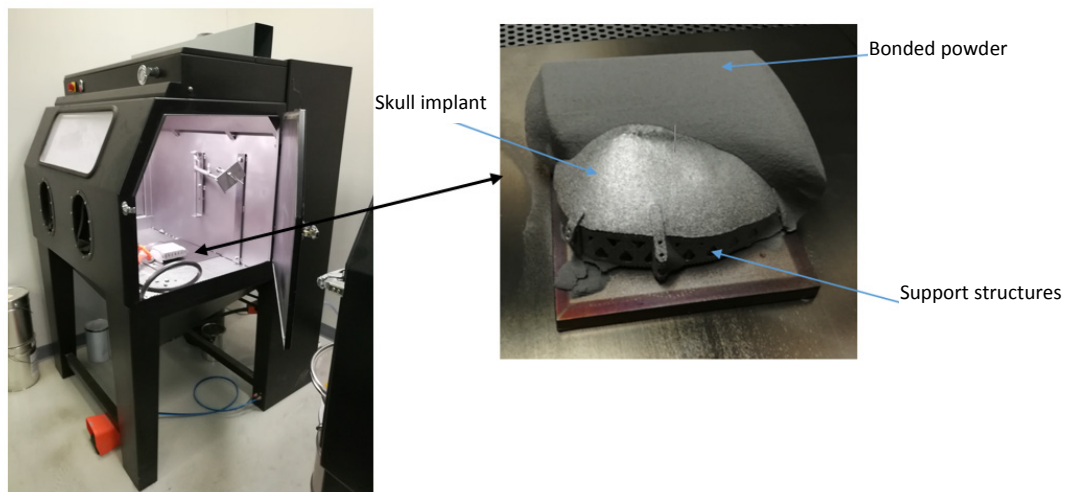
Fig. 11 Powder recovery system

The entire built process takes place under high vacuum of 0.1 Pa to 0.001 Pa. After the completion of built, fabricated parts are cooled down to room temperature under helium environment to prevent oxidation. The build envelope with fabricated custom implant is then taken to the powder recovery system (PRS) to blast the sintered powder using compressed air as shown in Fig. 11.

The support structures are then removed and the thickness of the implant is measured using the screw gauge. The thickness was found to be approximately 0.52 mm as shown in Fig. 12.

The weight of the fabricated implant is measured using digital weighing machine. The weight was found to be close to the weight of the bone portion that was removed from the skull, assuming the bone density to be ρ =210 kg m−3 [39]. The details of the fabricated implant and the equivalent bone portion are listed in Table 5.

To evaluate the accuracy of the fabricated implant, it is first scanned using ViuScan hand scanner as show in Fig. 13. Then Geomagic Qualify software is used for the 3D comparison. The Geomagic quality is one of the commonly used techniques to graphically represent the surface deviation between two models.

The EBM fabricated cranial implant was taken as test model and the original CT scan was taken as reference model. Using this software, the scanned EBM fabricated model was superimposed onto the reference CAD model with the Best Fit Alignment function using 1500 points. The differences were analyzed as positive and negative deviations. A positive deviation occurs if the test (scanned) model is larger than the reference (Original CT scan) model and a negative deviation if it is smaller.

The mean deviations and root mean square of the deviations (RMSD) between test and reference model were calculated [41]. In order to demonstrate the magnitude, location and direction of the discrepancies between them, a color coded model was developed as shown in Fig. 14.

(a)   (b)
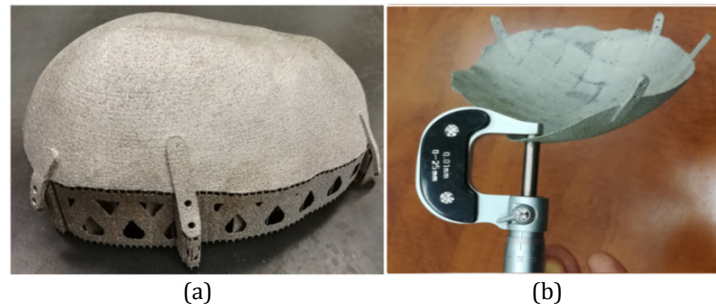
Fig. 12 (a) The fabricated implant with support structure, (b) Implant after the removal of support

Table 5 Weight details of the fabricated and the equivalent bone portion

Material	Thickness (mm)	Volume (mm³)	Weight (g)
Bone portion	Varying thickness: 3.18-5.76 mm	149114.285	31.314
Ti6Al4V ELI implant	0.52	9758.140	41.574

(a)   (b)
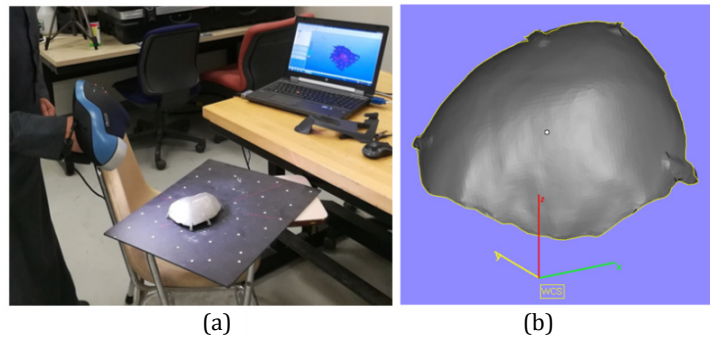
Fig. 13 (a) Implant scanning, (b) Scanned implant model
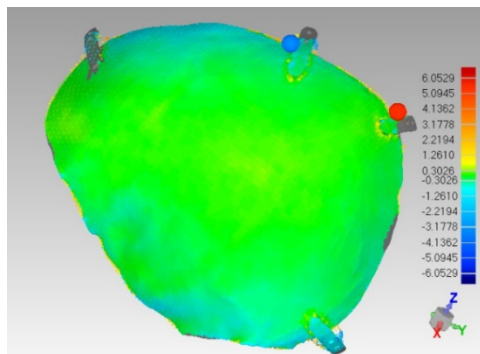
Fig. 14 The 3D comparison result

(a)   (b)

(c)
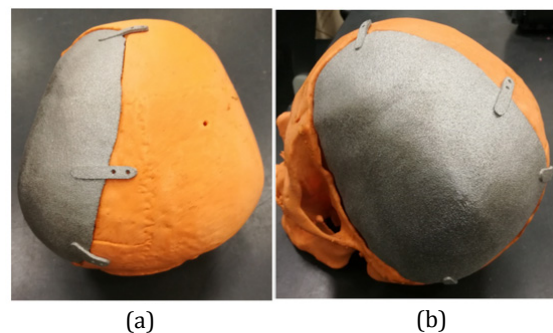
Fig. 15 Fabricated custom implant: (a) Top view, (b) Side view, (c) Front view

The results showed that the average positive and negative deviations are 0.321 mm and -0.3515 mm respectively with standard deviation 0.3246 mm and the RMSD deviation is 0.5380 mm. The fit of the fabricated implant was evaluated by fixing it on the 3D printed plastic skull model as shown in Fig. 15. The visual inspection from all the sides shows that fabricated implant successfully achieved a good fit and aesthetics.

5. Conclusion

In this study, a thin solid light weight custom cranial implant is designed, analyzed and fabricated using EBM. The craniofacial implant was designed based upon the patient CT scan images using MIMICS® and 3-MATIC®. Finite element analysis was carried out in order to assess the quality of the developed implant design under different loading conditions. The Ti6Al4V ELI implant of 0.5 mm thickness was then fabricated using EBM technology. The fitness of the fabricated implant was tested on the plastic prototype of the skull. The results showed the successful fabrication of 0.5 mm thick custom cranial implant for skull defect reconstruction via EBM technology while maintaining the structural strength requirements for reconstruction of skull defect . The main advantage of this approach for the skull defect reconstruction is the customizability, flexibility and less lead time for implant fabrication. However, this technology is still expensive and require significant amount of validation and testing before the implant is finally used for reconstruction of the defect.

Acknowledgements

References

[1] Gibson, I., Rosen, D.W., Stucker, B. (2010). *Additive manufacturing technologies: Rapid prototyping to direct digital manufacturing,* Springer, New York, USA, doi: 10.1007/978-1-4419-1120-9.
[2] Morgan, L., Lindhe, U., Harrysson, O. (2003). Rapid manufacturing with electron beam melting (EBM) – A manufacturing revolution?, In: *Proceedings of Annual International Solid Freeform Fabrication Symposium – An Additive Manufacturing Conference,* Austin, Texas, USA, 433-438.
[3] Berce, P., Chezan, H., Balc, N. (2005). The application of rapid prototyping technologies for manufacturing the custom implants, In: *Proceedings of 8th ESAFORM Conference on Material Forming,* Cluj-Napoca, Romania, 27-29.
[4] Christensen, A., Kircher, R., Lippincott, A. (2008). Qualification of electron beam melted (EBM) Ti6Al4V-ELI for orthopaedic applications, In: *Proceedings of MPMD, Materials and Processes for Medical Devices Conference,* Palm Desert, USA, 48-53.
[5] Li, X., Wang, C., Zhang, W., Li, Y. (2009). Fabrication and characterization of porous Ti6Al4V parts for biomedical applications using electron beam melting process, *Materials Letters,* Vol. 63, No. 3-4, 403-405, doi: 10.1016/j.matlet.2008.10.065.
[6] Abdel-Haleem, A.K., Nouby, R., Taghian, M. (2011). The use of the rib grafts in head and neck reconstruction, *Egyptian Journal of Ear, Nose, Throat and Allied Sciences,* Vol. 12, No. 2, 89-98, doi: 10.1016/j.ejenta.2011.08.004.
[7] He, Y., Ye, M., Wang, C. (2006). A method in the design and fabrication of exact-fit customized implant based on sectional medical images and rapid prototyping technology, *The International Journal of Advanced Manufacturing Technology*, Vol. 28, No. 5-6, 504-508, doi: 10.1007/s00170-004-2406-y.
[8] Durham, S.R., McComb, J.G., Levy, M.L. (2003). Correction of large (> 25 cm²) cranial defects with "reinforced" hydroxyapatite cement: Technique and complications, *Neurosurgery,* Vol. 52, No. 4, 842-845, doi: 10.1227/01.NEU.0000054220.01290.8E.
[9] Ducic, Y. (2002). Titanium mesh and hydroxyapatite cement cranioplasty: A report of 20 cases, *Journal of oral and maxillofacial surgery,* Vol. 60, No. 3, 272-276, doi: 10.1053/joms.2002.30575.
[10] Engstrand, T., Kihlström, L., Lundgren, K., Trobos, M., Engqvist, H., Thomsen, P. (2015). Bioceramic implant induces bone healing of cranial defects, *Plastic and Reconstructive Surgery – Global Open,* Vol. 3, No. 8, doi: 10.1097/GOX.0000000000000467.
[11] Charnley, G., Judet, T., Pirou, P., Garreau de Loubresse, C. (2000). Titanium femoral component fixation and experience with a cemented titanium prosthesis, In: Learmonth, I.D. (ed.), *Interfaces in total hip arthroplasty*, Springer, London, UK, 3-10, doi: 10.1007/978-1-4471-0477-3_1.
[12] Moiduddin, K., Darwish, S., Al-Ahmari, A., ElWatidy, S., Mohammad, A., Ameen, W. (2017). Structural and mechanical characterization of custom design cranial implant created using additive manufacturing, *Electronic Journal of Biotechnology,* Vol. 29, 22-31, doi: 10.1016/j.ejbt.2017.06.005.

[13] Niinomi, M. (2008). Mechanical biocompatibilities of titanium alloys for biomedical applications, *Journal of the Mechanical Behavior of Biomedical Materials,* Vol. 1, No. 1, 30-42, doi: 10.1016/j.jmbbm.2007.07.001.

[14] Schipper, J., Ridder, G.J., Spetzger, U., Teszler, C.B., Fradis, M., Maier, W. (2004). Individual prefabricated titanium implants and titanium mesh in skull base reconstructive surgery. A report of cases, *European Archives of Oto-Rhino-Laryngology and Head & Neck,* Vol. 261, No. 5, 282-290, doi: 10.1007/s00405-003-0686-8.

[15] Mazzoli, A., Germani, M., Raffaeli, R. (2009). Direct fabrication through electron beam melting technology of custom cranial implants designed in a PHANToM-based haptic environment, *Materials & Design,* Vol. 30, No. 8, 3186-3192, doi: 10.1016/j.matdes.2008.11.013.

[16] Goyal, S., Goyal, M. K. (2014). Restoration of large cranial defect for cranioplasty with alloplastic cranial implant material: A case report, *The Journal of Indian Prosthodontic Society,* Vol. 14, No. 2, 191-194, doi: 10.1007/s13191-012-0185-y.

[17] Lieger, O., Richards, R., Liu, M., Lloyd, T. (2010). Computer-assisted design and manufacture of implants in the late reconstruction of extensive orbital fractures, *Archives of Facial Plastic Surgery,* Vol. 12, No. 3, 186-191, doi: 10.1001/archfacial.2010.26.

[18] Joffe, J.M., Nicoll, S.R., Richards, R., Linney, A.D., Harris, M. (1999). Validation of computer-assisted manufacture of titanium plates for cranioplasty, *International Journal of Oral and Maxillofacial Surgery,* Vol. 28, No. 4, 309-313, doi: 10.1016/S0901-5027(99)80165-9.

[19] da Costa, D.D., Pedrini, H., Bazan, O. (2009). Direct milling of polymethylmethacrylate for cranioplasty applications, *The International Journal of Advanced Manufacturing Technology,* Vol. 45, No. 3-4, 318-325, doi: 10.1007/s00170-009-1978-y.

[20] da Costa, D.D., Lajarin, S.F. (2012). Comparison of cranioplasty implants produced by machining and by casting in a gypsum mold, *The International Journal of Advanced Manufacturing Technology*, Vol. 58, No. 1-4, 1-8, doi: 10.1007/s00170-011-3388-1.

[21] Saldarriaga, J.F.I., Vélez, S.C., Posada, M.D.A.C., Henao, I.E.B.B., Valencia, M.E.C.A.T. (2011). Design and manufacturing of a custom skull implant, *American Journal of Engineering and Applied Sciences,* Vol. 4, No. 1, 169-174, doi: 10.3844/ajeassp.2011.169.174.

[22] Dujovne, A.R., Bobyn, J.D., Krygier, J.J., Miller, J.E., Brooks, C.E. (1993). Mechanical compatibility of noncemented hip prostheses with the human femur, *The Journal of Arthroplasty,* Vol. 8, No. 1, 7-22, doi: 10.1016/S0883-5403(06)80102-6.

[23] Parthasarathy, J. (2014). 3D modeling, custom implants and its future perspectives in craniofacial surgery, *Annals of Maxillofacial Surgery,* Vol. 4, No. 1, 9-18, doi: 10.4103/2231-0746.133065.

[24] Cho, H.R., Roh, T.S., Shim, K.W., Kim, Y.O., Lew, D.H., Yun, I.S. (2015). Skull reconstruction with custom made three-dimensional titanium implant, *Archives of Craniofacial Surgery,* Vol. 16, No. 1, 11-16, doi: 10.7181/acfs.2015.16.1.11.

[25] Murr, L.E., Gaytan, S.M., Martinez, E., Medina, F., Wicker, R.B. (2012). Next generation orthopaedic implants by additive manufacturing using electron beam melting, *International Journal of Biomaterials,* Vol. 2012, doi: 10.1155/2012/245727.

[26] Al-Ahmari, A., Nasr, E.A., Moiduddin, K., Alkindi, M., Kamrani, A. (2015). Patient specific mandibular implant for maxillofacial surgery using additive manufacturing, In: *Proceedings of 2015 International Conference on Industrial Engineering and Operations Management (IEOM),* Dubai, UAE, 1-7, doi: 10.1109/IEOM.2015.7093788.

[27] Mazzoli, A., Germani, M., Raffaeli, R. (2009). Direct fabrication through electron beam melting technology of custom cranial implants designed in a PHANToM-based haptic environment, *Materials & Design,* Vol. 30, No. 8, 3186-3192, doi: 10.1016/j.matdes.2008.11.013.

[28] Li, X., Wang, C., Zhang, W., Li, Y. (2009). Fabrication and characterization of porous Ti6Al4V parts for biomedical applications using electron beam melting process, *Materials Letters,* Vol. 63, No. 3-4, 403-405, doi: 10.1016/j.matlet.2008.10.065.

[29] Saldarriaga, J.F.I., Vélez, S.C., Posada, M.D.A.C., Henao, I.E.B.B., Valencia, M.E.C.A.T. (2011). Design and manufacturing of a custom skull implant, *American Journal of Engineering and Applied Sciences,* Vol. 4, No. 1, 169-174, doi: 10.3844/ajeassp.2011.169.174.

[30] Syam, W.P., Al-Shehri, H.A., Al-Ahmari, A.M., Al-Wazzan, K.A., Mannan, M.A. (2012). Preliminary fabrication of thin-wall structure of Ti6Al4V for dental restoration by electron beam melting, *Rapid Prototyping Journal,* Vol. 18, No. 3, 230-240, doi: 10.1108/13552541211218180.

[31] Al-Ahmari, A., Nasr, E.A., Moiduddin, K., Anwar, S., Al Kindi, M., Kamrani, A. (2015). A comparative study on the customized design of mandibular reconstruction plates using finite element method, *Advances in Mechanical Engineering,* Vol. 7, No. 7, doi: 10.1177/1687814015593890.

[32] Tsouknidas, A., Maropoulos, S., Savvakis, S., Michailidis, N. (2011). FEM assisted evaluation of PMMA and Ti6Al4V as materials for cranioplasty resulting mechanical behaviour and the neurocranial protection, *Bio-Medical Materials and Engineering,* Vol. 21, No. 3, 139-147, doi: 10.3233/BME-2011-0663.

[33] Allen, T., Goodwill, S., Haake, S. (2008). Experimental validation of a tennis ball finite-element model for different temperatures, In: Estivalet, M., Brisson, P. (eds.), *The Engineering of Sport 7*, Springer Nature, Switzerland, 125-133, doi: 10.1007/978-2-287-09411-8_15.

[34] Chamrad, J., Marcián, P., Borák, L., Wolff, J. (2016). Finite element analysis of cranial implant, In*: Proceedings of 32nd conference with international participation computational mechanics*, Železná Ruda, Czech Republic, doi: 10.13140/RG.2.2.19870.13124.

[35] Parle, D., Ambulgekar, A., Gaikwad, K. (2015). Patient specific design and validation of cranial implants using FEA, In: *Altair Technology Conference,* Bengaluru, India, 1-9.

[36] Ridwan-Pramana, A., Marcián, P., Borák, L., Narra, N., Forouzanfar, T., Wolff, J. (2017). Finite element analysis of 6 large PMMA skull reconstructions: A multi-criteria evaluation approach, *Plos One,* Vol. 12, No. 6, doi: 10.1371/journal.pone.0179325.

[37] Koike, M., Martinez, K., Guo, L., Chahine, G., Kovacevic, R., Okabe, T. (2011). Evaluation of titanium alloy fabricated using electron beam melting system for dental applications, *Journal of Materials Processing Technology*, Vol. 211, No. 8, 1400-1408, doi: 10.1016/j.jmatprotec.2011.03.013.

[38] Karlsson, J. (2015*). Optimization of electron beam melting for production of small components in biocompatible titanium grades,* Doctoral thesis, Acta universitatis upsaliensis.

[39] Margulies, S.S., Thibault, K.L. (2000). Infant skull and suture properties: Measurements and implications for mechanisms of pediatric brain injury, *Journal of Biomechanical Engineering*, Vol. 122, No. 4, 364-371, doi: 10.1115/1.1287160.

[40] Ameen, W., Al-Ahmari, A., Mohammed, M.K., Mian, S.H. (2018). Manufacturability of overhanging holes using electron beam melting, *Metals,* Vol. 8, No. 6, 1-24, doi: 10.3390/met8060397.

[41] Al-Ekrish, A.A., Alfadda, S.A., Ameen, W., Hörmann, R., Puelacher, W., Widmann, G. (2018). Accuracy of computer-aided design models of the jaws produced using ultra-low MDCT doses and ASIR and MBIR, *International Journal of Computer Assisted Radiology and Surgery,* 1-8, doi: 10.1007/s11548-018-1809-4.

[42] Čolić, K., Sedmak, A., Legweel, K., Milošević, M., Mitrović, N., Mišković, Ž., Hloch, S. (2017). Experimental and numerical research of mechanical behaviour of titanium alloy hip implant, *Tehnički vjesnik – Technical Gazette,* Vol. 24, No. 3, 709-713, doi: 10.17559/TV-20160219132016.

[43] Ramakrishnaiah, R., Al Kheraif, A.A., Mohammad, A., Divakar, D.D., Kotha, S.B., Celur, S.L., Hashem, M.I., Vallittu, P.K., Rehman, I.U. (2017). Preliminary fabrication and characterization of electron beam melted Ti-6Al-4V customized dental implant, *Saudi Journal of Biological Sciences,* Vol. 24, No. 4, 787-796, doi: 10.1016/j.sjbs.2016.05.001.

Dynamic integration of process planning and scheduling using a discrete particle swarm optimization algorithm

Yu, M.R.[a,*], **Yang, B.**[a], **Chen, Y.**[b]

[a]School of Mechatronic Engineering, North University of China, Taiyuan, P.R. China
[b]North Automatic Control Technology Institute, Taiyuan, P.R. China

ABSTRACT

Because of the inherent relationship between process planning and scheduling, integration of process planning and scheduling (IPPS) provides a new path for further improvements of these two activities. Therefore, a novel two-phase IPPS approach is put forward in this paper. In the new method, the preplanning phase generates a process network for each job with consideration of the static shop floor status. After that, the final planning phase simultaneously creates the process plan of each job and the scheduling plan according to the current shop floor status. Based on the modified definition of IPPS and the proposed mathematical model, the IPPS problem and the dynamic IPPS problem can be solved together. Furthermore, a discrete particle swarm optimization (DPSO) algorithm is proposed to solve the IPPS optimization problem. In the DPSO algorithm, the particles update their positions by crossing with their own historical best positions (pbests) and the global best position of the population (gbest). In order to avoid local convergence, an external archive is introduced to keep more than one elite, and the gbest of each particle is randomly selected from the external archive. Furthermore, mutation operation is introduced to enhance the local search ability of DPSO algorithm. Finally, some comparative results are given to verify the efficiency and effectiveness of the proposed IPPS method and the DPSO algorithm as well as the dynamic IPPS method.

1. Introduction

In the common manufacturing systems, which transform the raw material or semi-finished product into final product using kinds of machines, some preparation activities such as materials, tools, process plans, scheduling plans and so on. In these activities, process planning and scheduling are two crucial functions which are usually carried out sequentially. In other words, the process plans, which are the outcomes of process planning, are transferred into the scheduling system, which assigns operations to specified machines at appropriate moments according to the precedence relations in the process plans, shop floor status, scheduling criteria, etc. Obviously, the effectiveness of the scheduling results should be strongly dependent on the process plans. However, in the past years, these two activities are often executed sequentially, in which the scheduling plans are generated separately according to the fix process plans obtained by the process planning.

However, when considering the disturbances in manufacturing process such as arrival of urgent jobs, due date change and machine breakdown, the traditional method, in which process planning and scheduling are separately treated, seems to be inadequate for following reasons [1]:

- Traditionally, the process plans are generated by process planners under some ideal assumptions, for example, the resources in the shop floor are always available, which is unrealistic in a real manufacturing environment;
- The conventional process planning methods provide deterministic process plans to scheduling system, which ignores the possibility of improvement for scheduling with alternative process plans;
- Because of the time delay between process planning phase and scheduling phase, even if the dynamic shop floor status is considered in advance, it may change greatly when the schedule plan is executed, thus the generated optimal process plans may become suboptimal or even invalid;
- In traditional way, the process planning generates optimal plans with the consideration of process plan criterion, and the scheduling works in a similar manner, conflicts may appear in such a separate way.

To overcome these shortages, Chryssolouris *et al.* [2] first proposed the concept of integration of process planning and scheduling (IPPS). When process planning and scheduling are integrated, the manufacturing system can respond promptly to disruptions. As a result, the resource utilization and the productivity will be improved.

In the following sections, some literature reviews on IPPS are given in Section 2, and Section 3 describes the materials and methods. Then the discrete particle swarm optimization algorithm is presented in Section 3, which is followed by the application of DPSO in optimizing the proposed IPPS problem in Section 4, which is followed by some experiment results in Section 5 and conclusions in Section 6.

2. Literature review

Ever since the concept of IPPS was proposed by Chryssolouris *et al.* [2], a lot of research papers on IPPS could be found. Based on the descriptions in some existing publications on IPPS [1, 3-6], the IPPS methods can be divided into three categories: non-linear process planning (NLPP), closed loop process planning (CLPP) and distributed process planning (DPP). These IPPS approaches and related contributions are described below.

NLPP: The NLPP method attempts to generate as many process plans for each part as possible under ideal shop floor status with consideration of the process flexibility, sequence flexibility [7] and operation flexibility. Here, the process flexibility represents the possibility of machining a specified feature using alternative processes or series of processes, hereafter, use *macro-level plan* to stand for process or a serial of processes for machining a feature. And the sequence flexibility stands for possibility of different sequences of the selected macro-level plans. Finally, the operation flexibility means that a specified process may be arranged to different machines. After that, different priority levels are assigned to these process plans according to the optimization objectives of process planning, e.g. total machining cost/time, etc. And then, these process plans will be tested in the scheduling system according to their priority levels.

CLPP: In the CLPP method, the process plans are always feasible with respect to the current scheduling environment because they are generated just before the jobs are released to the shop floor, in the other words, the status of each resource in the shop floor is known to the process planner. Some of the CLPP approaches first generate off-line plans according to the static shop floor status, and then make some necessary on-line refinement on the basis of the availability of resources on the shop floor at the subsequent scheduling phase.

DPP: The DPP method performs both process planning and scheduling simultaneously in a hierarchic manner. In most of the DPP methods, they can be divided into two phases, which are preplanning and final planning. Some other DPP approaches may have three phases: preplan-

ning, pair planning and final planning. Similarly, at the preplanning phase, the process planning function recognizes the features and the feature relationships of each job, and determines the macro-level plan for each feature. Meanwhile, the scheduling function estimates the required machine capabilities. Then, at the final planning phase (which can be further divided into pair planning and final planning in some DPP methods), both the final process plan of each part and the scheduling plan are generated simultaneously, in this way, the integration occurs at the moment that the machining processes and the available resources are matched.

Both the NLPP and CLPP methods are one-way information flows as shown in Fig. 1, in which process planning is still executed before scheduling. In this way, the objectives of process planning and scheduling may be conflicted. Therefore, Both NLPP and CLPP are incapable of finding the global optimal solution. With this in mind, the DPP method, which performs the technical and capacity-related planning tasks simultaneously, may be the only one IPPS method that integrates the functions of process planning and scheduling. However, in most of the DPP methods [8, 9], the macro-level plan of each feature is determined in the preplanning stage; and the selected macro-level plans for all features are used for matching the operation capabilities of the available resources in the final planning stage. It means that the flexibility of the final planning phase is reduced because of ignoring the process flexibility.

As mentioned before, optimization algorithm is another research focus on IPPS because of its NP-hard characteristic [11, 12]. Kinds of optimization algorithms, such as genetic algorithm (GA) [13], particle swarm optimization (PSO) algorithm [8], etc. have been used for optimizing the IPPS problems. However, most of these algorithms are designed for one of the three existing IPPS methods. Therefore, the performances of those optimization algorithms may be restricted by the corresponding IPPS methods.

As is known to all, the basic purpose of IPPS is handling the inevitable disturbances in the manufacturing processes by treating process planning and scheduling as a whole. However, even in the newest publications on IPPS, dynamic IPPS is still less mentioned in the literatures for its complexity. For example, Zhang and Wong [14] proposed an object-coding genetic algorithm for IPPS, and Wang *et al.* [15] addressed a systematic approach for optimizing the IPPS problems in sustainable machining. The models in these two papers are still designed for static IPPS, in which the results of IPPS are final process plans and schedules. It is worth mentioning that even the optimal solution is obtained using integrated method, this solution may also be deteriorated in the execution stage because of the inevitable disturbances. Therefore, dynamic IPPS is necessary for further improving the performances of the manufacturing system.

To sum up, an effective IPPS method and the corresponding optimization algorithm are still necessary to be developed to get full integration of process planning and scheduling. Therefore, this paper proposes a novel IPPS method to overcome the shortages of the three existing IPPS methods. Based on the modified definition of IPPS, the mathematical model of the proposed IPPS method is presented, which is also suitable for dynamic IPPS method. Finally, a discrete particle swarm optimization is developed for optimizing the proposed IPPS problem.

Fig. 1 Information flows of: (a) NLPP and (b) CLPP [10]

3. Materials and methods

3.1 The proposed IPPS method

As shown in Fig. 2, the proposed IPPS method is a two-phase approach, which also contains preplanning and final planning. Be different from the existing DPP methods, at the preplanning phase, instead of determining the macro-level plan for each feature, the process network, which is widely used in representing the flexibility of process plan [16, 17], of each part to be machined is generated according to the static resources. A process network types of nodes, which are

starting node, intermediate node and ending node, more details for the process network can be found in the reference of Ho *et al.* [17]. Then the process networks of all the parts to be machined are stored and will be used in the final planning phase. With consideration of optimization criteria of both process planning and scheduling, an optimal plan, which includes the process plans of all parts to be machined and the scheduling plan, is generated according to the current shop floor status. Of course, an effective optimization algorithm is necessary because of the complexity of the problem.

After that, the optimal plan is released to the shop floor immediately. Suppose the makespan of the optimal plan is T and current time is t_0. Because disturbances such as machine breakdown, rush order and order cancellation may occur during the time interval t_0 to $t_0 + T$, a dynamic IPPS method, in which both the process plans of the jobs and the schedule are changeable, is presented to handle the disruptions.
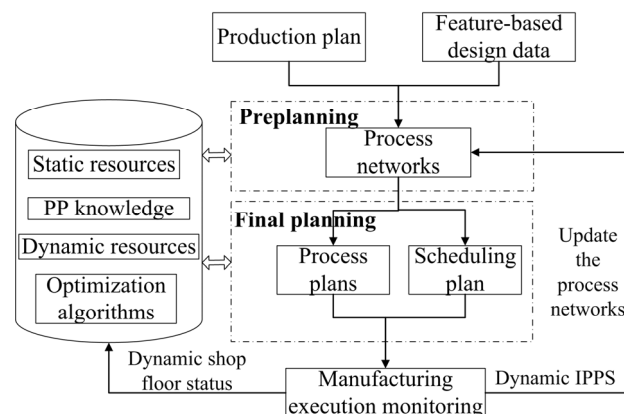
Fig. 2 The structure of the proposed IPPS method (PP stands for process planning)

3.2 Mathematical model of the proposed IPPS method

In order to solve the IPPS and the dynamic IPPS in a uniform manner, the definition of IPPS [18] is modified as follows.

Given a set of n jobs, at the time t, the number of the unfinished features of each part is N_{if}, the number of the available machine at time t is M_t. considering the flexibility of process, sequence and operation, determine the macro-level plan for each unfinished feature and the corresponding machine (machine set) as well as the sequences of the operations on each machine. Meanwhile, the precedence constraints in each part are satisfied and some optimization objectives can be achieved.

Based on the proposed IPPS method and the modified definition of IPPS, the mathematical model of the proposed IPPS method is founded. In this paper, the optimization objectives of process planning and scheduling are minimum machining time and minimum makespan, separately. Firstly, some reasonable assumptions and notations are given below:

- Jobs are independent. Job preemption is not allowed and each machine can handle only one job at any moment.
- The different operations of the same job cannot be machined simultaneously.
- The setup time and the transport time are negligible or included in the processing time.

M_t the total number of the available machines at the time t;
N the total number of jobs to be machined at the time t;
N_{if} the number of the unfinished features of the i-th job at the time t;
N_{ijo} the number of alternative macro-level plans of the j-th feature of i-th job;
O_{ijk} the k-th alternative macro-level plan of the j-th feature of i-th job;
M_{ijkl} the l-th alternative machine (machine set) of the macro-level plan O_{ijk};
t_{ijkl} the machining time of macro-level plan O_{ijk} on machine l;
c_{ijkl} the completion time of macro-level plan O_{ijl} on machine k;

$$X_{ijk} = \begin{cases} 1 & \text{the } k\text{-th alternative macro-level plan is selected for } j\text{-th feature of } i\text{-th job} \\ 0 & \text{othervise} \end{cases}$$

$$Y_{ijkl} = \begin{cases} 1 & \text{the } l\text{-th machine is selected for the macro-level plan } O_{ijkl} \\ 0 & \text{othervise} \end{cases}$$

Objectives:

- Minimizing the total machining time:

$$F_1 = \text{Min} \sum_{i=1}^{N} \sum_{j=1}^{N_{if}} \sum_{k=1}^{N_{ijo}} \sum_{l=1}^{M_t} t_{ijkl} \times Y_{ijkl} \times X_{ijk} \tag{1}$$

- Minimizing the makespan:

$$F_2 = Min\{Makespan = Max\{c_{ijkl} \times X_{ijk} \times Y_{ijkl}\} \tag{2}$$

Subject to:

- The j-th feature of the i-th job can only choose one macro-level plan from its alternatives:

$$\sum_{k=1}^{N_{ijo}} X_{ijk} \tag{3}$$

- The macro-level plan O_{ijk} can only choose one machine (machine set) from its alternatives:

$$\sum_{k=1}^{N_{ijo}} X_{ijk} \tag{4}$$

Other constraints can be found in Li *et al.* [19]. It is worth mentioning that the numbers of the available machines and the jobs to be processed as well as the unfinished features of each job are varied as time goes by. Therefore, this mathematical model is also suitable for the dynamic IPPS method. And these kinds of information of dynamic shop floor status are provided by the manufacturing execution system.

3.3 Particle swarm optimization algorithm

The proposed IPPS method involves in four tasks, which are selection of macro-level plan for each unfinished feature, selection of machine (machine set) for each selected macro-level plan, sequencing the selected macro-level plans and determining the starting time of each process on its corresponding machine. It is a typical NP-hard problem and much more complex to solve than that of existing IPPS methods, which are partially involved in the four referred tasks. Particle swarm optimization (PSO) algorithm has been widely used in many optimization problems since it was proposed in 1995 [20]. However, the continuous characteristic of PSO algorithm restricts its applications in combinatory optimizations problems. Some researchers [21, 22] used modified encoding methods to transform the scheduling problems into continuous versions, and then the PSO can be used to solve these scheduling problems. However, they are incapable of solving the optimization problem of the proposed IPPS method since its complexity. Therefore, this paper proposes a discrete particle swarm optimization (DPSO) algorithm for the proposed IPPS method.

Standard PSO algorithm

In most of the researches, the standard PSO refers to a modified PSO algorithm [23], in which an inertia weight is introduced to improve the optimizing ability. In the PSO algorithm, a swarm, which contains a number of particles with certain positions and velocities, is initialized randomly. Then each particle dynamically adjusts its velocity and position according to its own and the population's experiences. This can be explained as follows:

$$V_{id}^{t+1} = \begin{cases} \omega V_{id}^t + c_1 r_1 (pbest_{id} - X_{id}^t) + c_2 r_2 (gbest_d - X_{id}^t) & V_{dmin} \leq V_{id}^{t+1} \leq V_{dmax} \\ V_{dmax} & V_{id}^{t+1} > V_{dmax} \\ V_{dmin} & V_{id}^{t+1} < V_{dmin} \end{cases} \qquad (5)$$

$$X_{id}^{t+1} = \begin{cases} X_{id}^t + V_{id}^{t+1} & -X_{dmax} \ll X_{id}^{t+1} \ll X_{dmax} \\ X_{dmax} & X_{id}^{t+1} > X_{dmax} \\ -X_{dmax} & X_{id}^{t+1} < -X_{dmax} \end{cases} \qquad (6)$$

V_{id}^{t+1} is the velocity of the d-th dimension of the i-th particle at the time t, and it means the distance to be traveled in the d-th dimension from its current position. ω is the inertial weight used for regulating the trade-off between the global exploration and local exploration abilities of the swarm. c_1 and c_2 are the acceleration constants, r_1 and r_2 are two random functions within the range of [0,1]. $pbest_{id}$ represents the d-th dimension of the i-th particle's best historical position, and $gbest_d$ represents the d-th dimension of the swarm's global best position. The particles update their positions and velocities according to Eq. 5 and Eq. 6, and the $pbest$ and $gbest$ are also updated according to the fitness function. Thus all the particles move to the global best solution to finish the search process.

Discrete PSO algorithm

It can be found that the position and the velocity in standard PSO are both continuous variables from Eq. 5 and Eq. 6, which are meaningless for the IPPS optimization problems. This paper proposes a discrete version of PSO, in which the particles update their positions according to the following equation:

$$X_i^{t+1} = X_i^{t+1} \otimes P_{ibest}^t \otimes P_{gbest}^t \qquad (7)$$

X_i^{t+1} and X_i^t are the positions of the i-th particle in current and the next iteration, P_{ibest}^t is the historical best position of the i-th particle, and P_{gbest}^t is the best position of the population in the tth iteration. $\otimes$ stands for the crossover operator. Eq. 7 shows that the particles adjust their positions according to their own and the population's experiences, which maintains the advantages of the standard PSO algorithm. The discrete particle swarm optimization algorithm can be explained as follow according to Fig. 3:

- The particle X_i^t and its *pbest* P_{ibest}^t are considered as two parents P_1 and P_2; O_1 and O_2 are two offspring of them. Select the better one of these two offspring for the next crossover, e.g. O_2 is selected;
- O_2 and P_{gbest}^t, which is the *gbest* of X_i^t, are treated as two new parents P_1' and P_2'; O_1' and O_{21}' are two offspring of them. Select the better one of these two offspring to be used as the new position of X_i^t, that is X_i^{t+1}.

Fig. 3 Illustration of discrete particle swarm optimization algorithm

Generally, the *gbest* of the swarm is unique. Under this condition, each particle will cross with that *gbest*, which may lead to local convergence. For this reason, this paper introduces an external archive to keep the elites, which are the first N_{ea} (N_{ea} the size of the external archive) parti-

cles with lowest fitness values, generated in each iteration. Then each particle randomly selects an individual from the external archive as its *gbest* in each iteration, and the members of the external archive are also updated iteratively.

4. The application of discrete particle swarm optimization (DPSO) in optimizing the proposed IPPS problem

When applying the DPSO algorithm to optimize the IPPS problem, some necessary explanations and definitions should be given in advance. The information of a process network can also be listed in a table. Table 1 lists the information contained in four process networks of four jobs, in which '999' means that machine cannot finish the process, and O_{1211}-O_{1212} means the macro-level plan O_{121} contains two processes. It is worth mentioning that different macro-level plans of the same feature may have different machining time on the same machine. For example, the machining time of O_{111} and O_{112} on machine M_1 is 3 and 5 units, separately. The machine priority level for a macro-level plan is defined as follows. *Machine priority level*: For a macro-level plan, the machine priority level (1>2>3>4>5, etc.) is determined by the machining time of the machine or machine set. The longer the machining time of a machine is, the higher its machine priority level will be. And if the machining time of two machines (machine sets) for the same macro-level plan are equal, the bigger the number of the machine is, the higher its machine priority level will be.

Table 1 Information contained in the four process networks of four jobs

Jobs	Features	Alternative macro-level plans	Alternative machines and machining time				
			M_1	M_2	M_3	M_4	M_5
J$_1$	F$_{11}$ (before F$_{12}$)	O$_{111}$	999	3	4	999	4
		O$_{112}$	5	999	3	4	999
		O$_{113}$	3	4	999	999	4
	F$_{12}$	O$_{1211}$-O$_{1212}$	2/999	3/4	999/3	999/4	999/999
		O$_{1221}$-O$_{1222}$	3/2	999/3	4/999	4/3	3/999
	F$_{13}$	O$_{131}$	4	5	999	3	999
		O$_{132}$	999	999	3	4	5
		O$_{133}$	3	4	3	5	3
J$_2$	F$_{21}$	O$_{211}$	6	999	5	999	999
		O$_{212}$	999	6	999	6	5
	F$_{22}$	O$_{2211}$-O$_{2212}$	3/4	999/2	4/3	999/999	3/999
	F$_{23}$ (before F$_{24}$)	O$_{231}$	5	3	999	4	999
		O$_{232}$	4	999	4	5	5
		O$_{233}$	4	4	5	999	3
	F$_{24}$	O$_{241}$	5	3	4	999	4
		O$_{242}$	4	999	5	999	6
J$_3$	F$_{31}$	O$_{3111}$-O$_{3112}$-O$_{3113}$	999/3/999	4/5/3	4/4/6	999/4/5	5/999/4
	F$_{32}$	O$_{321}$	4	999	5	3	4
		O$_{322}$	5	5	4	6	4
	F$_{33}$ (before F$_{34}$)	O$_{331}$	4	6	4	5	999
		O$_{332}$	3	5	3	4	999
		O$_{333}$	999	4	999	3	5
	F$_{34}$	O$_{3411}$-O$_{3412}$	999/4	4/5	5/3	3/999	999/999
		O$_{342}$	10	10	9	999	999
J$_4$	F$_{41}$	O$_{411}$	4	999	5	999	4
	F$_{42}$ (before F$_{43}$)	O$_{421}$	3	4	999	3	5
		O$_{422}$	5	3	4	999	999
		O$_{423}$	3	4	999	4	3
	F$_{43}$	O$_{4311}$-O$_{4312}$-O$_{4313}$	5/3/999	4/5/4	5/999/6	999/5/999	4/6/5
	F$_{44}$	O$_{441}$	5	6	5	999	5
		O$_{442}$	6	999	5	6	4
		O$_{4431}$-O$_{4432}$	999/4	2/3	4/999	999/999	3/4

Table 2 Machine priority levels of the alternative machines

| Jobs | Features | Alternative macro-level plans | Machine priority level(1>2>3>4>5) | | | | |
			1	2	3	4	5
J_1	F_{11}	O_{111}	$M_2(3)$	$M_3(4)$	$M_5(4)$	$M_1(999)$	$M_4(999)$
	(before	O_{112}	$M_3(3)$	$M_4(4)$	$M_1(5)$	$M_2(999)$	$M_5(999)$
	F_{12})	O_{113}	$M_1(3)$	$M_2(4)$	$M_5(4)$	$M_3(999)$	$M_4(999)$
	F_{12}	O_{1211}-O_{1212}	$M_1+M_3(5)$	$M_1+M_2(6)$	$M_1+M_4(6)$	$M_2+M_3(7)$	$M_2(7)$
		O_{1221}-O_{1222}	$M_1(5)$	$M_5+M_1(5)$	$M_5+M_2(6)$	$M_4(7)$	$M_4+M_2(7)$
	F_{13}	O_{131}	$M_4(3)$	$M_1(4)$	$M_2(5)$	$M_3(999)$	$M_5(999)$
		O_{132}	$M_3(3)$	$M_4(4)$	$M_5(5)$	$M_1(999)$	$M_2(999)$
		O_{133}	$M_1(3)$	$M_3(3)$	$M_5(3)$	$M_2(4)$	$M_4(5)$
J_2	F_{21}	O_{211}	$M_3(5)$	$M_1(6)$	$M_2(999)$	$M_4(999)$	$M_5(999)$
		O_{212}	$M_5(5)$	$M_2(6)$	$M_4(6)$	$M_1(999)$	$M_3(999)$
	F_{22}	O_{2211}-O_{2212}	$M_1+M_2(5)$	$M_5+M_2(5)$	$M_1+M_3(6)$	$M_1(7)$	$M_3(7)$
	F_{23}	O_{231}	$M_2(3)$	$M_4(4)$	$M_1(5)$	$M_3(999)$	$M_5(999)$
	(before	O_{232}	$M_1(4)$	$M_3(4)$	$M_4(5)$	$M_5(5)$	$M_2(999)$
	F_{24})	O_{233}	$M_5(3)$	$M_1(4)$	$M_2(4)$	$M_3(5)$	$M_4(999)$
	F_{24}	O_{241}	$M_2(3)$	$M_3(4)$	$M_5(4)$	$M_1(5)$	$M_4(999)$
		O_{242}	$M_1(4)$	$M_3(5)$	$M_5(6)$	$M_2(999)$	$M_4(999)$
J_3	F_{31}	O_{3111}-O_{3112}-O_{3113}	$M_1+M_2(10)$	$M_2+M_4(11)$	$M_2(12)$	$M_1+M_5(12)$	$M_3(14)$
	F_{32}	O_{321}	$M_4(3)$	$M_1(4)$	$M_5(4)$	$M_3(5)$	$M_2(999)$
		O_{322}	$M_3(4)$	$M_5(4)$	$M_1(5)$	$M_2(5)$	$M_4(6)$
	F_{33}	O_{331}	$M_1(4)$	$M_3(4)$	$M_4(5)$	$M_2(6)$	$M_5(999)$
	(before	O_{332}	$M_1(3)$	$M_3(3)$	$M_4(4)$	$M_2(5)$	$M_5(999)$
	F_{34})	O_{333}	$M_4(3)$	$M_2(3)$	$M_5(5)$	$M_1(999)$	$M_3(999)$
	F_{34}	O_{3411}-O_{3412}	$M_4+M_3(6)$	$M_4+M_1(7)$	$M_2+M_3(7)$	$M_3(8)$	$M_2(9)$
		O_{342}	$M_3(9)$	$M_1(10)$	$M_2(10)$	$M_4(999)$	$M_5(999)$
J_4	F_{41}	O_{411}	$M_1(4)$	$M_5(4)$	$M_3(4)$	$M_2(999)$	$M_4(999)$
	F_{42}	O_{421}	$M_1(3)$	$M_4(3)$	$M_2(4)$	$M_5(5)$	$M_3(999)$
	(before	O_{422}	$M_2(3)$	$M_3(4)$	$M_1(5)$	$M_4(999)$	$M_5(999)$
	F_{43})	O_{423}	$M_1(3)$	$M_5(3)$	$M_2(4)$	$M_4(4)$	$M_3(999)$
	F_{43}	O_{4311}-O_{4312}-O_{4313}	$M_1+M_2(12)$	$M_2(13)$	$M_1+M_5(13)$	$M_1+M_3(14)$	$M_5(15)$
	F_{44}	O_{441}	$M_5(4)$	$M_3(5)$	$M_1(6)$	$M_4(6)$	$M_2(999)$
		O_{442}	$M_1(5)$	$M_3(5)$	$M_5(5)$	$M_2(6)$	$M_4(999)$
		O_{4431}-O_{4432}	$M_2(5)$	$M_2+M_1(6)$	$M_2+M_5(6)$	$M_5(7)$	$M_5+M_1(7)$

Based on the definition of machine priority level, the machine priority levels of the machines shown in Table 1 can be obtained as shown in Table 2. As referred before, the process network is generated based on static shop floor status in the preplanning phase, thus in the final planning phase, the status of the shop floor as well as the machine priority levels should be updated.

4.1 Encoding and decoding

Encoding refers to problem mapping, which translates a problem to a particle. For the jobs listed in Table 2, a possible particle which contains three sections is shown in Fig. 4. The feature string, whose length equals to the total number of the features of all jobs, represents the manufacturing sequence of the features. For example, '1, 2, 3' represent the features of job 1, and '4, 5, 6, 7' represent the features of job 2, etc. The MLP string stands for the selected macro-level plans for corresponding features, e.g. the second 2 in the MLP string corresponds to the number 9 in the feature string. It means that the second feature of the third job (F_{32}) has selected its second alternative macro-level plan (O_{322}). The machine string represents the machine priority levels of the selected machines, e.g. the macro-level plan referred before (O_{322}) has selected the first prior machine (M_3) from its alternatives because the value of the corresponding position in the machine string is 1.

Feature string	3	1	2	4	9	10	12	15	8	11	6	7	5	13	14
MLP string	1	3	2	1	2	3	1	2	1	2	1	2	1	2	1
Machine string	2	1	2	2	1	1	2	2	1	2	2	1	2	3	2

Fig. 4 A particle for the four jobs listed in Table 2 (MLP stands for macro-level plan)

Decoding means solution generation, which translates a particle to a solution of the optimization problem. In the proposed IPPS method, the solution contains the process plan of each job and the scheduling plan. Based on the encoding method referred before, the decoding process is described as follows (take the particle shown in Fig. 4 for example):

Step 1: Translate the feature string to the features machining sequence of each job, for example, the four jobs' feature machining sequence are '3, 1, 2', '4, 6, 7, 5', '9, 10, 8, 11' and '12, 15, 13, 14', separately; meanwhile, the machining sequence of all the features is also determined, that is F_{13}-F_{11}-F_{12}-F_{21}-F_{32}-F_{33}-F_{41}-F_{44}-F_{31}-F_{34}-F_{23}-F_{24}-F_{22}-F_{42}-F_{43};

Step 2: Translate the MLP string to the selected macro-level plans for corresponding features;

Step 3: Translate the machine string to the selected machines for corresponding macro-level plans;

Step 4: Combine the results of steps 1, 2 and 3 to generate the process plan for each job.

Step 5: According to the results of steps 1, 2 and 3, the features machining sequence, selected macro-level plans and corresponding machines are known, and then the decoding method proposed by Zhang *et al.* [24] is introduced to decode the particle to an active scheduling plan.

4.2 Fitness function

In this paper, the optimizing objectives of process planning and scheduling are minimum total machining time and minimum makespan, respectively. Although these two objectives have the same dimension, their orders of magnitude may vary enormously. Thus the normalization of the data is still necessary. Let $D_i(T_i, M_i)$ $(1 \leq i \leq N)$ stand for the assessment of the solution corresponding to the i-th particle, in which N is the number of the particles in the swarm. T_i and M_i represent the total machining time and the makespan, respectively. The procedure of normalization can be explained as follows:

$$T_i' = \frac{T_i - \min\{T_j, 1 \leq j \leq N\}}{\max\{T_j, 1 \leq j \leq N\} - \min\{T_j, 1 \leq j \leq N\}} \tag{8}$$

$$M_i' = \frac{M_i - \min\{M_j, 1 \leq j \leq N\}}{\max\{M_j, 1 \leq j \leq N\} - \min\{M_j, 1 \leq j \leq N\}} \tag{9}$$

Thus, two new arrays of total machining time and makespan are obtained, which are $[T_1', T_2', \ldots, T_N']$ and $[M_1', M_2', \ldots, M_N']$, respectively. Then the final objective function of the IPPS optimization problem as well as the fitness function can be obtained as follow:

$$F = \omega_1 T_i' + \omega_2 M_i' \tag{10}$$

Where ω_1 and ω_2 stand for the weights of total machining time and makespan, respectively. Meanwhile, the relations $0 \leq \omega_1 \leq 1$, $0 \leq \omega_2 \leq 1$ and $\omega_1 + \omega_2 = 1$ should be satisfied.

4.3 Particle updating strategy

In the DPSO algorithm, the particles update their positions by crossing with their own historical best positions (*pbests*) and the global best position of the population (*gbest*), and a mutation operator is designed to enhance the local search ability of the algorithm. Based on the encoding method referred before, the crossover and mutation operations are designed for the IPPS problem.

Crossover

As shown in Fig. 4, a particle contains three different strings: feature string, MLP string and machine string. The crossover operator acts on them separately. Thus three crossover operations are needed. The procedure of the crossover operation for the feature string is described as follows according to Fig. 5.
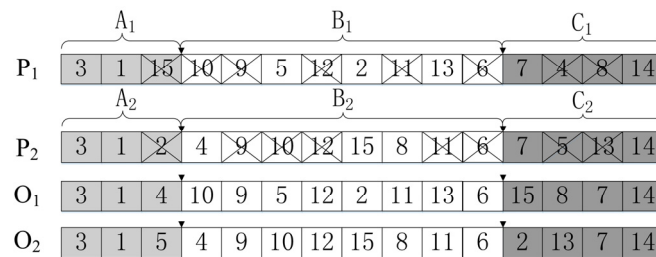
Fig. 5 The crossover operation for the feature string

Step 1: P_1 and P_2 are the feature strings of two particles, and generate two empty offspring: O_1 and O_2;

Step 2: Randomly select two different crossover points to divide P_1 to three sections: A_1, B_1 and C_1; this process is done to P_2 at the same positions of crossover points;

Step 3: Copy the elements in B_1 (B_2) into the same positions of O_1 (O_2);

Step 4: Delete the existing elements of O_1 (O_2) in P_2 (P_1), and then copy the remaining elements in P_2 (P_1) to the remaining empty positions in O_1 (O_2).

The typical two-point crossover operator is introduced and implemented on the MLP string and an example of this crossover operation for the MLP string is presented in Fig. 6. The crossover operation for machine string has the same procedure with that of MLP string.

It is worth mentioning that since the number of the alternative macro-level plans of different features may be different, the ranges of the values of different bits in MLP string may be different. It means that infeasible particles may be generated by the crossover operation, and a refinement procedure is necessary, which will be presented in Section *Refinement*.
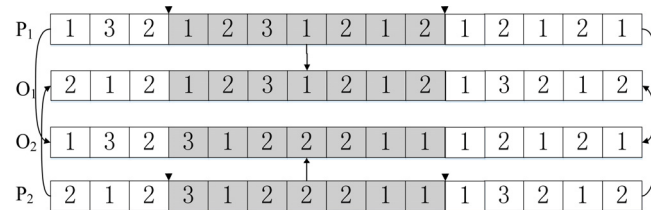
Fig. 6 The crossover operation for the MLP string

Mutation

Mutation operation is introduced in the DPSO algorithm to enhance its local search ability. Since the values of the elements in the MLP string are restricted by the values of corresponding elements in the feature string, the mutation operations should be sequentially conducted for the feature string as well as MLP string and the machine string. In order to save space, a partial particle is taken for example, and the procedure for mutation operations can be described as follows according to Fig. 7.

Step 1: Mutate the feature string: randomly select two different positions and exchange their elements in the feature string and the MLP string as well as the machine string;

Step 2: Mutate the MLP string: randomly select a bit from the MLP string and change its value within the range of that bit according to the value of its corresponding bit in the feature string;

Step 3: Mutate the machine string: randomly select a bit from the machine string and change its value.

Fig. 7 The mutation operation for a partial particle

Refinement

As referred before, the crossover operation may generate infeasible particles, in which the values of the elements in the MLP string may exceed their value ranges. Additionally, the mutation operation and the random initialization may also generate invalid particles, in which the precedence relationships among features may be dissatisfied. The procedure of the refinement for the infeasible particles is described as follows:
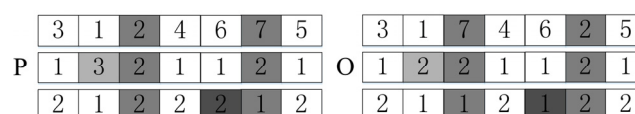
Step 1: The constraint adjustment method of Li *et al.* [25] is introduced to adjust the feature string. In order to maintain the corresponding relationships among features, macro-level plans and machines, the adjustment operation is simultaneously implemented on the three strings. It means that if two elements in the feature string are swapped, the elements of corresponding positions in both the MLP string and the machine string should be swapped, simultaneously.

Step 2: Check the values of the elements in the MLP string. If the value of an element exceeds its value range, randomly choose a value within its value range to replace its previous value. For example, the numbers of the alternative macro-level plans of the features F_{12} and F_{13} are 2 and 3, separately. Suppose after crossover operation, the values of the elements in the MLP string correspond to these two features are 3 and 2, respectively. It means that the value of the element in the MLP string corresponds to the feature F_{12} has exceeded its value range [1, 2], and then the value of that element may be randomly replaced by 1 or 2.

4.4 The procedure of applying DPSO in optimizing the proposed IPPS problem

Based on the explanations before, the procedure of applying DPSO in optimizing the proposed IPPS problem is described as follows:

Step 1: Parameter setting and initialization: set the parameters and randomly generate the initial swarm, iteration number $i = 1$;

Step 2: Refinement: refine the infeasible particles using the refinement method presented in Section *Refinement*;

Step 3: Decoding: decode each particle to a solution, which contains the process plan of each job and the scheduling plan, using the decoding method referred before;

Step 4: Evaluate the fitness of each particle and sort the particles according to their fitness values in ascending order, then the first N_{ea} particles are selected for updating the members of external archive (EA);

Step 5: Update EA: if $i = 1$, insert the particles selected in step 4 into EA directly; otherwise, combine the particles selected in step 4 and the existing particles in the EA, and sort these $2N_{ea}$ particles according to their fitness values in ascending order, then the first N_{ea} particles are preserved in EA;

Step 6: Update the *pbest* and *gbest* of each particle, in which the *gbest* of each particle is randomly selected from the members of EA;

Step 7: If $i \leq Iter_{max}$ ($Iter_{max}$ is the maximum iteration number), go to step 8; otherwise, go to step 9;

Step 8: Update the positions of the particles using the updating strategy (crossover and mutation) presented in Section 5. 3, $i = i+1$, and go to step 2;

Step 9: Output the best solution: decode the best particle in EA, whose fitness value is the lowest, to a solution, which includes the process plan of each job and the scheduling plan.

5. Results and discussion

As referred before, the proposed IPPS method has combined the advantages of the three existing IPPS methods, which are NLPP, CLPP and DPP methods. Additionally, based on the modified definition and the mathematical model of IPPS, dynamic scheduling can be extended to dynamic IPPS for effectively handling the disruptions. Besides, a DPSO algorithm is proposed for optimizing the IPPS problem. In this section, the proposed DPSO algorithm is compared with some other

intelligent algorithms to show its effectiveness, and two test cases are designed to verify the effectiveness of the proposed IPPS method.

5.1 Case 1

Case 1 is designed to verify the efficiency and effectiveness of the proposed DPSO algorithm. Because there is no existing test instance, which is very similar with the IPPS problem in this paper, in order to compare the proposed algorithm with other intelligent algorithms, some classical benchmarks of flexible job-shop scheduling problem (FJSP), which are widely introduced in related researches, are used for instead. The sources of the benchmarks and the testing environment are given below:

Kacem data: five representative instances which cover a wide range from 4×5 to 15×10 taken from the works of Kacem *et al.* [26, 27] are used in the experiment without the consideration of release date. The instances are 4×5 problem (I_1), 8×8 problem (I_2), 10×7 problem (I_3), 10×10 problem (I_4) and 15×10 problem (I_5), and the details of them can be found in literatures [26, 27].

BRdata: the BRdata is consists of 10 test problems from Brandimarte [28] which were randomly generated using uniform distributions. In the selected BRdata, the number of machines ranges from 4 to 15, the number of jobs ranges from 10 to 20, the number of operations for each job ranges from 5 to 15, and the number of operations for all jobs ranges from 55 to 240.

The programs are implemented using Matlab, running on a PC with 2.2GHz CPU and 4GB RAM. The main parameters are: the population size is $10 \times n$, where n is the number of jobs. The maximum number of iteration is $10 \times n \times m$, where m is the number of machines. The crossover probability $P_c = 1$, which means that each particle should cross with its own *pbest* and its selected *gbest*. Since the selected benchmarks are often used for testing multi-objective FJSP, and the multi-objective optimization is not the focus of this paper, the objectives are combined into a scalar function with equal weight of each objective.

Firstly, the proposed DPSO algorithm is applied on Kacem data and the results are compared with several other methods in recent publications, which include PDABC [29], MOGA [30], HPSO [31] and PSO/LS [32]. The comparative results are shown in Table 3. The three objectives are F_1 (makespan), F_2 (total workload) and F_3 (workload of the critical machine). The Gantt charts of the optimal solutions of these instances obtained by DPSO are shown in Fig. 8.

Table 3 Comparative results of the five instances of Kacem data

Instance	PDABC			MOGA			HPSO			PSO/LS			DPSO		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
I_1	11	32	10	11	32	10	11	32	10	16	32	8	11	32	1
（4×5）	12	32	8	11	34	9				16	33	7			
	13	33	7	12	32	8									
I_2	14	77	12	15	81	11	14	77	12				14	77	1
（8×8）	15	75	12	15	75	12									
	16	73	13	16	73	13									
I_3	12	61	11							16	60	12	11	62	1
（10×7）	11*	63*	11*							15	61	11			
	12	60	12							15*	62*	10*			
I_4	8	41	7	8	42	5	7*	43*	6*	8	41	7	7	43	5
（10×10）	7	43	5	7	42	6				7*	44*	5*			
	8	42	5	8	41	7				7	42	6			
				7*	45*	5*									
I_5	12*	91*	11*	11	91	11	11*	93*	11*	23*	91*	11*	11	91	1
（15×10）	11*	93*	11*	12	95	10									
				11	98	10									

* Means that the solution is worse than the solution obtained by the proposed DPSO algorithm
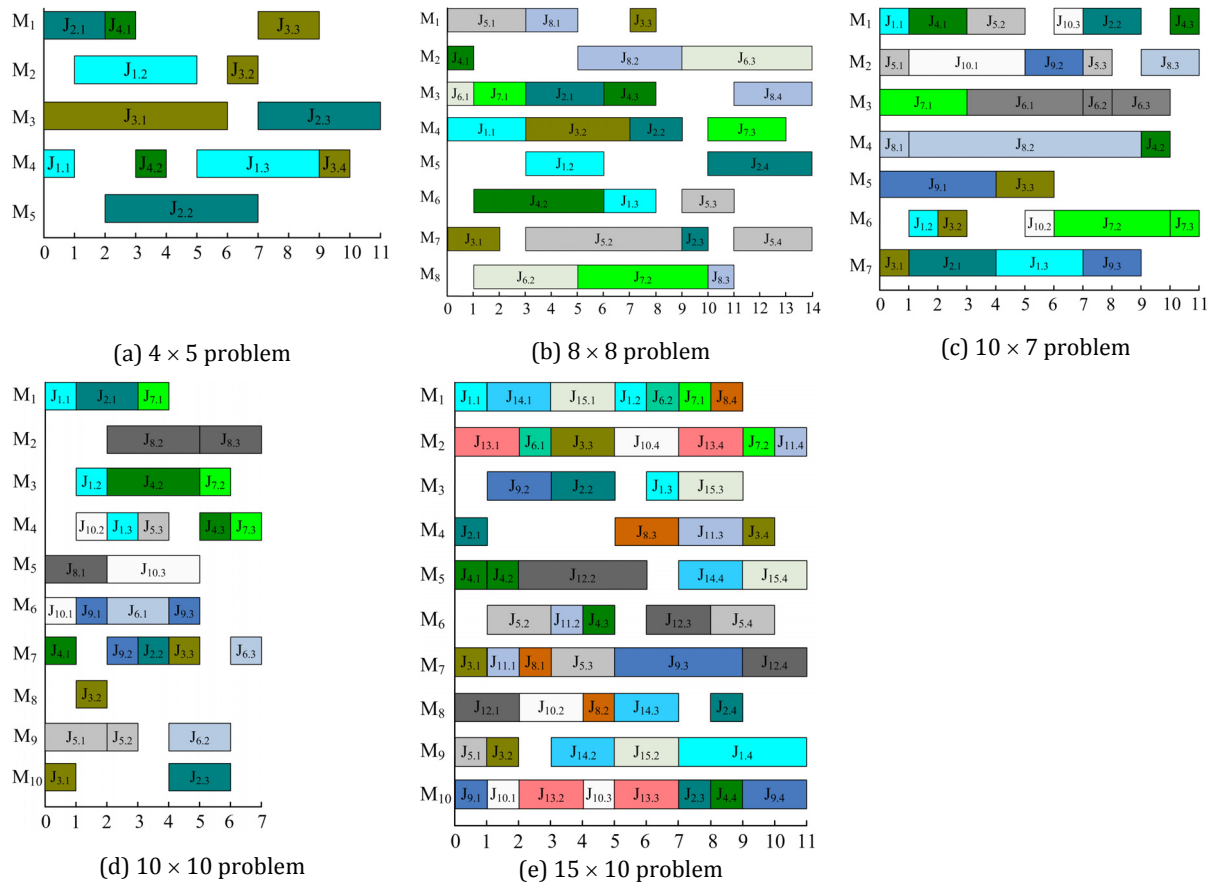
(a) 4 × 5 problem

(b) 8 × 8 problem

(c) 10 × 7 problem

(d) 10 × 10 problem

(e) 15 × 10 problem

Fig. 8 Gantt charts of the optimal solutions for (a) I_1, (b) I_2, (c) I_3, (d) I_4, and (e) I_5

Secondly, the proposed DPSO algorithm is applied on the BRdata, and the experiment results are compared with those of other two algorithms in recent publications, which are artificial immune (AI) [33] and efficient search (ES) [34], the comparative results are listed in Table 4.

From Table 3, it can be found that the results of the proposed DPSO algorithm are no worse (most of them are better) than those of other methods. Similarly, from the comparative results shown in Table 4, it is obvious that most of the solutions obtained by the DPSO algorithm are better than those of the AI algorithm, meanwhile, the CPU times are also reduced a lot. Although the CPU times of DPSO algorithm are longer than those of ES, the quality of the solutions has been greatly improved. In a word, the proposed DPSO algorithm stands out for its efficiency and effectiveness.

Table 4 Comparative results of the ten instances of BRdata

Instance	AI				ES				DPSO			
	F_1	F_2	F_3	T_{CPU}/S	F_1	F_2	F_3	T_{CPU}/S	F_1	F_2	F_3	T_{CPU}/S
MK01	40	171	36	97.21	42*	162*	42*	4.78	40	167	36	17.93
MK02	26*	154*	26*	103.46	28*	155*	28*	3.02	26	150	26	19.24
MK03	204*	1207*	204*	247.37	204*	852*	204*	26.14	204	850	204	56.41
MK04	60*	403*	60*	152.07	68*	372*	67*	17.74	60	372	60	45.87
MK05	173*	686*	173*	171.95	177*	702*	177*	8.26	172	687	172	92.54
MK06	63*	470*	56*	245.62	75*	431*	67*	18.79	58	429	55	89.46
MK07	140	695	140	161.92	150*	717*	150*	5.68	140	695	140	124.62
MK08	523*	2723*	523*	392.25	523	2524	523	67.67	523	2524	523	148.44
MK09	312*	2591*	306*	389.71	311*	2374*	299*	77.76	307	2282	299	263.74
MK10	214*	2121*	206*	384.54	227*	1989*	221*	122.52	197	2029	197	296.47

* Means that the solution is worse than the solution obtained by the proposed DPSO algorithm

5.2 Case 2

The efficiency and the effectiveness of the proposed DPSO algorithm have been verified by the experiment results in case 1. This section designs a test case to prove the effectiveness of the proposed IPPS method. Suppose the four jobs listed in Table 2 are released to the shop floor and all the machines in the shop floor are idle and available at the moment $t = 0$. Under this assumption, in the NLPP method, the first priory process plan for each part is feasible and will be selected at the scheduling phase. Thus the optimal process plan for a job obtained by the CLPP method will be the same as the process plan selected for that job in the NLPP method. Without loss of generality, in the DPP method, the first alternative macro-level plan is selected for each feature in the preplanning phase. The three existing IPPS methods and the proposed IPPS method are used to solve the IPPS problem listed in Table 2 under the referred assumption, separately. The comparative results are shown in Table 5 (process plans) and Fig. 9 (scheduling plans), respectively, in which T represents the total machining time. According to the comparative results we can find that the proposed IPPS method outperforms the DPP method because the DPP method ignores the process flexibility (possibility of machining the same feature with alternative macro-level plans). Furthermore, as referred before, actually, the NLPP and CLPP methods perform the process planning and scheduling sequentially, the optimization objectives of process planning and scheduling may be conflicted. For example, although the total machining time has reduced from 73 to 71 (about 2.74 %), the makespan time has increased from 22 to 27 (about 22.72 %), which means that the proposed IPPS method also does better than the NLPP and CLPP methods.

Table 5 Process plans of the three existing methods and the proposed IPPS method

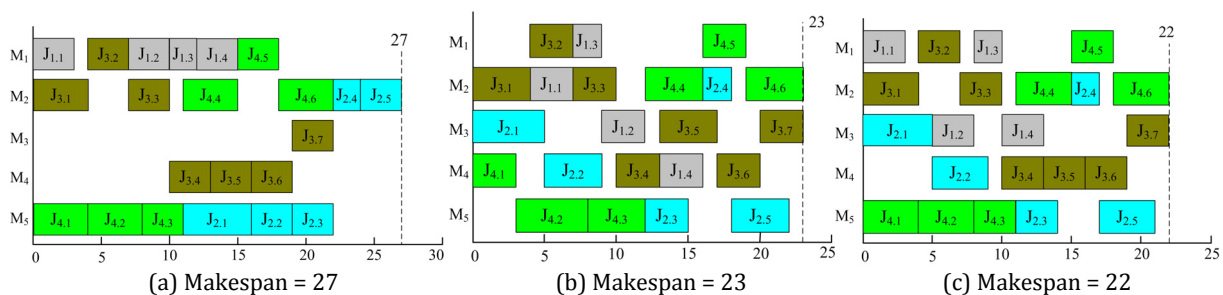| Job	NLPP (CLPP)	DPP	The proposed IPPS
J_1	$O_{113}(M_1)$-$O_{1221}(M_1)$-$O_{1222}(M_1)$ -$O_{133}(M_1)$	$O_{111}(M_2)$-$O_{1211}(M_1)$-$O_{1212}(M_3)$- $O_{131}(M_4)$	$O_{113}(M_1)$-$O_{132}(M_3)$-$O_{1211}(M_1)$ -$O_{1212}(M_3)$
J_2	$O_{212}(M_5)$-$O_{233}(M_5)$-$O_{2211}(M_5)$ -$O_{2212}(M_2)$-$O_{241}(M_2)$	$O_{211}(M_3)$-$O_{231}(M_4)$-$O_{2211}(M_5)$- $O_{2212}(M_2)$-$O_{241}(M_5)$	$O_{211}(M_3)$-$O_{231}(M_4)$-$O_{2211}(M_5)$ -$O_{2212}(M_2)$-$O_{241}(M_5)$
J_3	$O_{3111}(M_2)$-$O_{3112}(M_1)$-$O_{3113}(M_2)$ -$O_{321}(M_4)$-$O_{333}(M_4)$-$O_{3411}(M_4)$ -$O_{3412}(M_3)$	$O_{3111}(M_2)$-$O_{3112}(M_1)$-$O_{3113}(M_2)$- $O_{321}(M_4)$-$O_{331}(M_3)$-$O_{3411}(M_4)$- $O_{3412}(M_3)$	$O_{3111}(M_2)$-$O_{3112}(M_1)$-$O_{3113}(M_2)$ -$O_{321}(M_4)$-$O_{333}(M_4)$-$O_{3411}(M_4)$ -$O_{3412}(M_3)$
J_4	$O_{441}(M_5)$-$O_{411}(M_5)$-$O_{423}(M_5)$ -$O_{4311}(M_2)$-$O_{4312}(M_1)$-$O_{4313}(M_2)$	$O_{421}(M_4)$-$O_{441}(M_5)$-$O_{411}(M_5)$- $O_{4311}(M_2)$-$O_{4312}(M_1)$-$O_{4313}(M_2)$	$O_{442}(M_5)$-$O_{411}(M_5)$-$O_{423}(M_5)$ -$O_{4311}(M_2)$-$O_{4312}(M_1)$-$O_{4313}(M_2)$
T	71	75	73

Fig. 9 Gantt charts of the schedules (a) NLPP and CLPP, (b) DPP and (c) the proposed IPPS method

5.3 Case 3

As referred before, based on the modified definition and the mathematical model of IPPS, dynamic scheduling can be extended to dynamic IPPS for effectively handling the disruptions occurred in the execution process. In this section, two typical dynamic events, which are machine breakdown and arrival of urgent jobs, are designed to verify the effectiveness of the dynamic IPPS method.

Machine breakdown

The optimal scheduling plan obtained by the proposed IPPS method is released to the shop floor. Suppose that at the time $t = 5$ the machine M_4 is broken-down with a relatively long repair time. Under this condition, traditionally, reactive scheduling methods often transfer the affected oper-

ations to their alternative machines for instead, which is the so called route changing rescheduling [35]. In the route changing rescheduling method, neither the alternative macro-level plan of each feature nor the operation sequence of each job is unchangeable. However, in the dynamic IPPS method, the process plans are also changeable, which may improve the flexibility of rescheduling. The comparative results are shown in Table 6 (process plans) and Fig. 10 (scheduling plans), respectively. From the results we can find that both the total machining time and makespan are reduced by the dynamic IPPS method.

Table 6 Process plans of the route changing rescheduling strategy and the dynamic IPPS method

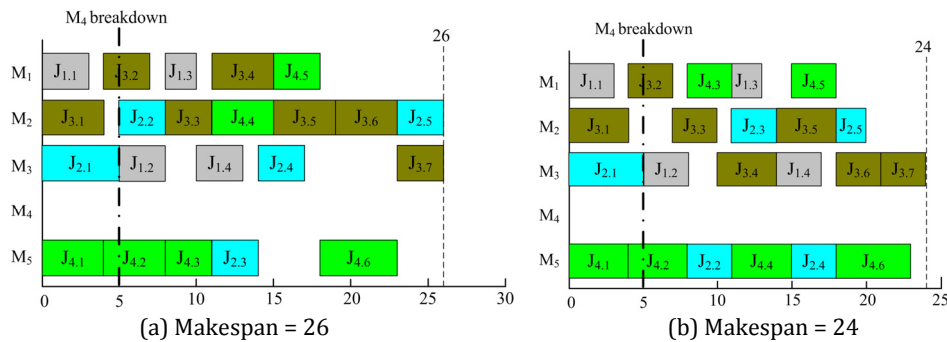| Job	Route changing rescheduling	Dynamic IPPS method
J_1	$O_{113}(M_1)$-$O_{132}(M_3)$-$O_{1211}(M_1)$-$O_{1212}(M_3)$	$O_{113}(M_1)$-$O_{132}(M_3)$-$O_{1211}(M_1)$-$O_{1222}(M_3)$
J_2	$O_{211}(M_3)$-$O_{231}(M_2)$-$O_{2211}(M_5)$ -$O_{2212}(M_3)$-$O_{241}(M_2)$	$O_{211}(M_3)$-$O_{233}(M_5)$-$O_{241}(M_2)$ -$O_{2211}(M_5)$-$O_{2212}(M_2)$
J_3	$O_{3111}(M_2)$-$O_{3112}(M_1)$-$O_{3113}(M_2)$-$O_{321}(M_1)$- $O_{333}(M_2)$-$O_{3411}(M_2)$-$O_{3412}(M_3)$	$O_{3111}(M_2)$-$O_{3112}(M_1)$-$O_{3113}(M_2)$-$O_{322}(M_3)$ -$O_{3411}(M_2)$-$O_{3412}(M_3)$ -$O_{332}(M_3)$
J_4	$O_{442}(M_5)$-$O_{411}(M_5)$-$O_{423}(M_5)$-$O_{4311}(M_2)$ -$O_{4312}(M_1)$-$O_{4313}(M_5)$	$O_{442}(M_5)$-$O_{411}(M_5)$-$O_{423}(M_1)$-$O_{4311}(M_5)$ -$O_{4312}(M_1)$-$O_{4313}(M_5)$
T	76	74

Fig. 10 Gantt chart of route changing rescheduling method (a) and the dynamic IPPS method (b)

Arrival of urgent jobs

The optimal scheduling plan obtained by the proposed IPPS method is released to the shop floor. Suppose two urgent jobs are released to the shop floor at the moment $t = 5$. The information of these two jobs is listed in Table 7. Generally, the urgent jobs should be completed as soon as possible. The comparative results are shown in Fig. 11, in which the dynamic IPPS method also outperforms the route changing rescheduling method.

From the above comparative results, it is obvious that the dynamic IPPS method is much more suitable than other rescheduling methods because it can increase the flexibility of rescheduling. Of course, the robustness and the stability of the dynamic IPPS method should be further studied.
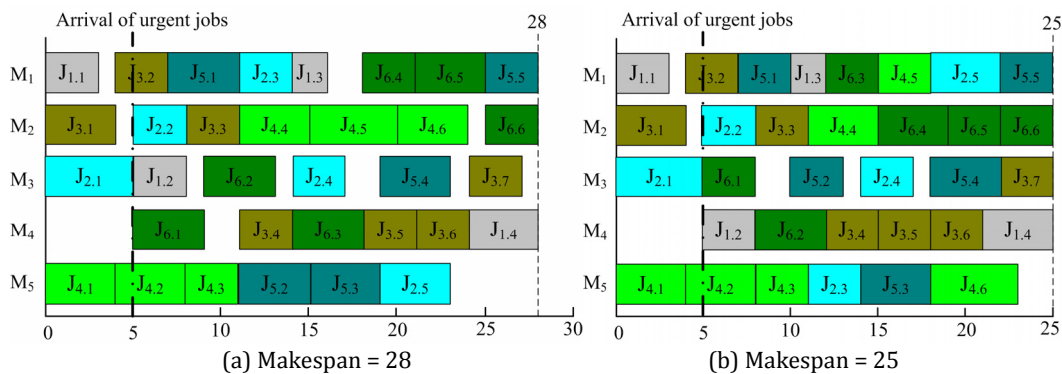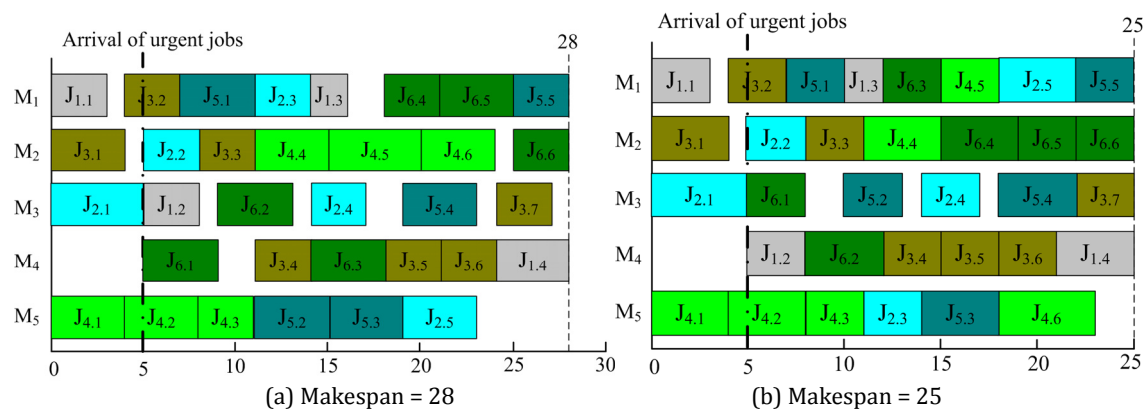

Fig. 11 Gantt charts of the route changing rescheduling method (a) and the dynamic IPPS method (b)

Table 7 Information of the two urgent jobs

Job	Feature	Alternative macro-level plan	Alternative machines and machining time				
			M_1	M_2	M_3	M_4	M_5
J_5	F_{51}	O_{511}	4	6	999	5	4
	(before F_{52})	O_{512}	5	999	3	4	999
		O_{513}	999	4	5	999	6
	F_{52}	O_{521}	5	6	999	5	4
		O_{522}	999	5	4	5	999
	F_{53}	O_{531}	3	5	4	999	4
		O_{532}	4	999	3	4	999
		O_{533}	999	4	999	5	3
	F_{54}	O_{5411}-O_{5412}	4/3	999/4	4/999	5/4	999/999
		O_{542}	999	6	7	999	7
J_6	F_{61}	O_{611}	4	999	5	4	999
		O_{612}	5	4	999	6	5
	F_{62}	O_{621}	999	5	4	6	5
		O_{622}	4	3	999	5	4
	F_{63}	O_{631}	5	4	5	4	999
	(before F_{64})	O_{632}	4	999	3	999	5
		O_{633}	4	4	5	6	999
	F_{64}	O_{6411}-O_{6412}-O_{6413}	3/4/999	999/4/3	4/999/5	5/4/4	999/999/4

Fig. 11 Gantt charts of the route changing rescheduling method (a) and the dynamic IPPS method (b)

6. Conclusion

This paper proposes a novel IPPS method which has combined the advantages of the three existing IPPS methods. Based on the modified definition of IPPS, the mathematical model of the proposed IPPS method is presented. This model is also suitable for the dynamic IPPS method, which can greatly improve the flexibility of rescheduling. Furthermore, a discrete version of particle swarm optimization algorithm is put forward to solve the optimization problem of the proposed IPPS method. In the DPSO algorithm, the particles update their positions by crossing with their own historical best positions (*pbests*) and the global best position of the population (*gbest*). Therefore, the continuous PSO algorithm can be used for the combinatory optimization problems such as optimization of the IPPS problems. In order to avoid local convergence, an external archive is introduced to keep more than one elite, and the *gbest* of each particle is randomly selected from the members of the external archive. Finally, the mutation operation is introduced to enhance the local search ability of the DPSO algorithm. As shown in the experiments and results of case 1, the proposed DPSO algorithm stands out from several typical intelligent optimization algorithms for its efficiency and effectiveness. Besides, the PDSO is designed for the proposed IPPS problem, which involves in four tasks as referred before, and the process planning, scheduling or the three existing IPPS methods are partially of the proposed IPPS problem, therefore, the DPSO algorithm could also be used to solve those problems.

In this work, we select the total machining time to judge the process plan and the makespan for scheduling plan, to simplify, we have combined these two objectives into one weighted function. However, other objectives such as total machining cost, total tardiness and utilization of machines may also be considered to complete the proposed IPPS method. It is no doubt that the Pareto-based optimization approaches will make this work more perfect, which will be studied in our future work. Furthermore, the robustness and the stability of the dynamic IPPS method should be further studied as well.

Acknowledgements

References

[1] Phanden, R.K., Jain, A., Verma, R. (2011). Integration of process planning and scheduling: A state-of-the-art review, *International Journal of Computer Integrated Manufacturing*, Vol. 24, No. 6, 517-534, doi: 10.1080/0951192x.2011.562543.

[2] Chryssolouris, G., Chan, S., Cobb, W. (1984). Decision making on the factory floor: An integrated approach to process planning and scheduling, *Robotics and Computer-Integrated Manufacturing*, Vol. 1, No. 3-4, 315-319, doi: 10.1016/0736-5845(84)90020-6.

[3] Yu, M., Zhang, Y., Chen, K., Zhang, D. (2015). Integration of process planning and scheduling using a hybrid GA/PSO algorithm, *The International Journal of Advanced Manufacturing Technology,* Vol. 78, No. 1-4, 583-592, doi: 10.1007/s00170-014-6669-7.

[4] Tan, W., Khoshnevis, B. (2000). Integration of process planning and scheduling – A review, *Journal of Intelligent Manufacturing*, Vol. 11, No. 1, 51-63, doi: 10.1023/a:1008952024606.

[5] Li, X., Gao, L., Zhang, C., Shao, X. (2010). A review on integrated process planning and scheduling, *International Journal of Manufacturing Research*, Vol. 5, No. 2, 161-180, doi: 10.1504/IJMR.2010.031630.

[6] Zhang, H., Liu, S., Moraca, S., Ojstersek, R. (2017). An effective use of hybrid metaheuristics algorithm for job shop scheduling problem, *International Journal of Simulation Modelling*, Vol. 16, No. 4, 644-657, doi: 10.2507/IJSIMM 16(4)7.400.

[7] Huang, X.W., Zhao, X.Y., Ma, X.L. (2014). An improved genetic algorithm for job-shop scheduling problem with process sequence flexibility, *International Journal of Simulation Modelling*, Vol. 13, No. 4, 510-522, doi: 10.2507/IJSIMM13(4)CO20.

[8] Guo, Y.W., Li, W.D., Mileham, A.R., Owen, G.W. (2009). Optimisation of integrated process planning and scheduling using a particle swarm optimisation approach, *International Journal of Production Research*, Vol. 47, No. 14, 3775-3796, doi: 10.1080/00207540701827905.

[9] Leung, C.W., Wong, T.N., Mak, K.L., Fung, R.Y.K. (2010). Integrated process planning and scheduling by an agent-based ant colony optimization, *Computers & Industrial Engineering*, Vol. 59, No. 1, 166-180, doi: 10.1016/j.cie.2009.09.003.

[10] Zhang, H.-C., Merchant, M.E. (1993). IPPM – A prototype to integrate process planning and job shop scheduling functions, *CIRP Annals*, Vol. 42, No. 1, 513-518, doi: 10.1016/S0007-8506(07)62498-6.

[11] Dai, M., Tang, D., Xu, Y., Li, W. (2014). Energy-aware integrated process planning and scheduling for job shops, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, Vol. 229, No. 1, 13-26, doi: 10.1177/0954405414553069.

[12] Galzina, V., Lujić, R., Šarić, T. (2012). Adaptive fuzzy particle swarm optimization for flow-shop scheduling problem, *Tehnički vjesnik - Technical Gazette,* Vol. 19, No. 1, 151-157.

[13] Chaudhry, I.A., Usman, M. (2017). Integrated process planning and scheduling using genetic algorithms, *Tehnički vjesnik – Technical Gazette*, Vol. 24, No. 5, 1401-1409, doi: 10.17559/TV-20151121212910.

[14] Zhang, L., Wong, T.N. (2015). An object-coding genetic algorithm for integrated process planning and scheduling, *European Journal of Operational Research*, Vol. 244, No. 2, 434-444, doi: 10.1016/j.ejor.2015.01.032.

[15] Wang, S., Lu, X., Li, X.X., Li, W.D. (2015). A systematic approach of process planning and scheduling optimization for sustainable machining, *Journal of Cleaner Production*, Vol. 87, No. 914-929, doi: 10.1016/j.jclepro.2014.10.008.

[16] Kim, Y.K., Park, K., Ko, J. (2003). A symbiotic evolutionary algorithm for the integration of process planning and job shop scheduling, *Computers & Operations Research*, Vol. 30, No. 8, 1151-1171, doi: 10.1016/S0305-0548(02)00063-1.

[17] Ho, Y.-C., Moodie, C.L. (1996). Solving cell formation problems in a manufacturing environment with flexible processing and routeing capabilities, *International Journal of Production Research*, Vol. 34, No. 10, 2901-2923, doi: 10.1080/00207549608905065.

[18] Guo, Y.W., Li, W.D., Mileham, A.R., Owen, G.W. (2009). Applications of particle swarm optimisation in integrated process planning and scheduling, *Robotics and Computer-Integrated Manufacturing*, Vol. 25, No. 2, 280-288, doi: 10.1016/j.rcim.2007.12.002.

[19] Li, X., Gao, L., Shao, X., Zhang, C., Wang, C. (2010). Mathematical modeling and evolutionary algorithm-based approach for integrated process planning and scheduling, *Computers & Operations Research*, Vol. 37, No. 4, 656-667, doi: 10.1016/j.cor.2009.06.008.

[20] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization, In: *Proceedings of ICNN'95 – International Conference on Neural Networks,* Perth, Australia, 1942-1948, doi: 10.1109/ICNN.1995.488968.

[21] Lin, T.-L., Horng, S.-J., Kao, T.-W., Chen, Y.-H., Run, R.-S., Chen, R.-J., Lai, J.-L., Kuo, I.-H. (2010). An efficient job-shop scheduling algorithm based on particle swarm optimization, *Expert Systems with Applications*, Vol. 37, No. 3, 2629-2636, doi: 10.1016/j.eswa.2009.08.015.

[22] Sha, D.Y., Hsu, C.-Y. (2006). A hybrid particle swarm optimization for job shop scheduling problem, *Computers & Industrial Engineering*, Vol. 51, No. 4, 791-808, doi: 10.1016/j.cie.2006.09.002.

[23] Shi, Y., Eberhart, R. (1998). A modified particle swarm optimizer, In: *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence*, Anchorage, USA, 69-73, doi: 10.1109/ICEC.1998.699146.

[24] Zhang, G., Gao, L., Shi, Y. (2011). An effective genetic algorithm for the flexible job-shop scheduling problem, *Expert Systems with Applications*, Vol. 38, No. 4, 3563-3573, doi: 10.1016/j.eswa.2010.08.145.

[25] Li, W.D., Ong, S.K., Nee, A.Y.C. (2004). Optimization of process plans using a constraint-based tabu search approach, *International Journal of Production Research*, Vol. 42, No. 10, 1955-1985, doi: 10.1080/0020754031 0001652897.

[26] Kacem, I., Hammadi, S., Borne, P. (2002). Pareto-optimality approach for flexible job-shop scheduling problems: Hybridization of evolutionary algorithms and fuzzy logic, *Mathematics and Computers in Simulation*, Vol. 60, No. 3-5, 245-276, doi: 10.1016/S0378-4754(02)00019-8.

[27] Kacem, I., Hammadi, S., Borne, P. (2002). Approach by localization and multiobjective evolutionary optimization for flexible job-shop scheduling problems, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews),* Vol. 32, No. 1, 1-13, doi: 10.1109/tsmcc.2002.1009117.

[28] Brandimarte, P. (1993). Routing and scheduling in a flexible job shop by tabu search, *Annals of Operations Research*, Vol. 41, No. 3, 157-183, doi: 10.1007/bf02023073.

[29] Li, J.-Q., Pan, Q.-K., Gao, K.-Z. (2011). Pareto-based discrete artificial bee colony algorithm for multi-objective flexible job shop scheduling problems, *The International Journal of Advanced Manufacturing Technology*, Vol. 55, No. 9-12, 1159-1169, doi: 10.1007/s00170-010-3140-2.

[30] Wang, X., Gao, L., Zhang, C., Shao, X. (2010). A multi-objective genetic algorithm based on immune and entropy principle for flexible job-shop scheduling problem, *The International Journal of Advanced Manufacturing Technology*, Vol. 51, No. 5-8, 757-767, doi: 10.1007/s00170-010-2642-2.

[31] Xia, W., Wu, Z. (2005). An effective hybrid optimization approach for multi-objective flexible job-shop scheduling problems, *Computers & Industrial Engineering*, Vol. 48, No. 2, 409-425, doi: 10.1016/j.cie.2005.01.018.

[32] Moslehi, G., Mahnam, M. (2011). A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local search, *International Journal of Production Economics*, Vol. 129, No. 1, 14-22, doi: 10.1016/j.ijpe.2010.08.004.

[33] Bagheri, A., Zandieh, M., Mahdavi, I., Yazdani, M. (2010). An artificial immune algorithm for the flexible job-shop scheduling problem, *Future Generation Computer Systems*, Vol. 26, No. 4, 533-541, doi: 10.1016/j.future.2009.10.004.

[34] Xing, L.-N., Chen, Y.-W., Yang, K.-W. (2009). An efficient search method for multi-objective flexible job shop scheduling problems, *Journal of Intelligent Manufacturing*, Vol. 20, No. 3, 283-293, doi: 10.1007/s10845-008-0216-z.

[35] He, W., Sun, D.-H. (2013). Scheduling flexible job shop problem subject to machine breakdown with route changing and right-shift strategies, *The International Journal of Advanced Manufacturing Technology*, Vol. 66, No. 1-4, 501-514, doi: 10.1007/s00170-012-4344-4.

An integral algorithm for instantaneous uncut chip thickness measuring in the milling process

Li, Y.[a], Yang, Z.J.[a], Chen, C.[b, *], Song, Y.X.[a], Zhang, J.J.[a], Du, D.W.[c]

[a]College of Mechanical Science and Engineering, Jilin University, Changchun, Jilin, P.R. China
[b]College of Engineering, China Agricultural University, Beijing, P.R. China
[c]Institution of Oceangraphic Instrumentation, Shandong Academy of Science, Qingdao, P.R. China

A B S T R A C T

Instantaneous uncut chip thickness (IUCT) calculation is an essential work for dynamic cutting force prediction accurately in milling process. This study presents an integral algorithm in polar coordinate system for measuring the thickness of transient uncut chip. The milling trajectory, cycloidal motion, is adopted in the formulation. Both milling continuity and cutter run-out are also considered in this model. The developed model offers a methodology for calculating the IUCT precisely. Furthermore, a series of simulations are carried out under different processing parameters. The results suggest that increasing both the feed per tooth and number of teeth can surge the width of IUCT slightly, but decrease with smaller cutter radius. The milling force simulations are validated by the experiment results measured in the reference and compared with classical approximate method, showing the proposed IUCT model providing good applications in instantaneous milling force predictions.

© 2018 CPE, University of Maribor. All rights reserved.

1. Introduction

Milling is an important processing method in the field of machining. Milling force that originates in the cutter–workpiece interface is an important factor that affects vibrations, milling efficiency, surface quality and the chatter stability of CNC tools [1-3]. Various analytical prediction models for cutting force have been established to improve the accuracy and reliability of milling force [4-6]. These models differ considerably, but the cutting force and IUCT are closely related [7, 8]. Besides, the force prediction accuracy is also important for monitoring of work conditions, such as tool life, surface roughness, even the machining stability of the milling process [9, 10]. Martellotti was first to use an approximated formula to calculate chip thickness. The simplified tooth path is considered as circular and lack of tooth eccentricity or tooth run-out [11]. The IUCT expression of the tooth path is $T_t(\alpha) = f_{rN} \cdot \sin \alpha$, where f_{rN} is feed per tooth and α is the instantaneous angular position of the tooth, which is widely used and researched [12, 13]. S. Spiewak then proposed an improved model of calculating IUCT in milling process. He facilitated stepwise and orderly increases in model sophistication until a desirable level of performance was achieved. The application results indicates that the speed, accuracy and reliability of monitoring and control potentially have been all improved, while the calculation process is rather complex and time consuming [14]. Another classical model is the one proposed by Li *et al.* in 2001. Instead of using a numerical method, they calculated the thickness of instantaneous uncut chip with Taylor's series by analyzing true tooth trajectories [15]. In Kumanchik's model, an analytic

expression for uncut chip thickness in milling process was formulated while considering a series of impact factors, such as the cycloidal motion of teeth, run-out, and uneven teeth spacing [16]. Some people then expressed interest in the model through making improvements or modifications in the calculations. The main contribution of Ge Song and his co-authors is their investigation on the accurate positions of geometric points in the profile of IUCT with different level of cut width. This study also adopted the iterative algorithm to improve the accuracy of the thickness of uncut chip at the chip cross-section [17]. Chip thickness in circular interpolation was examined in 2008. Saï *et al.* found an opposing trend between chip thickness and the radius of circular trajectory. The regions or movements of the cutter should be also considered [18]. N. Grossi improved the trochoidal motion formula by introducing run-out values by two variables- the distance (d) between geometric centers of two adjacent cutters and the instantaneous cutter angle α [19]. However, the present methods for measuring IUCT ignored the true path shape of the cutter tooth. Most used models entail a circular tool-path approximation evenly. As mentioned earlier, cutting force is primarily a function of IUCT in an analytical cutting force approach. These simplified models could not achieve satisfied accuracy definitely. Besides, Milling is a continuous process and the approximate circumferential model considers the IUCT as discrete line that connects the tool center to the current tooth's cutting edge which fails to describe the transient chip formation and also limits the application scope of formulas.

This study introduces an integral algorithm in polar coordinate system for measuring the thickness of transient uncut chip. The milling cutter trajectory, cycloidal motion, is adopted in the novel formula, in which the IUCT is associated with feed rate per teeth, number of teeth and cutter radius. Both the effect of milling continuity and the cutter run-out are also considered in this method. Furthermore, different processing parameters are used to obtain a series of IUCT. The simulation milling forces are validated by the experiments and the classical method, showing the advantages and accuracy of using the new proposed method. The nomenclature is given in the Appendix A.

2. Milling tooth path and equation

Researchers investigated the path of cutter tip during milling. Most of them assumed the tooth path of cutting points as circular [14]. However, the true path shape of the cutter should be considered as a trochoid, which is a more reasonable approach [1, 15]. Fig. 1(a) describes movement in the direction of a straight line of the milling cutter. The radius of the cutter is assumed as r, where in the teeth are evenly distributed along the circumference. When the milling cutter is in clockwise rotation, the total teeth are labeled as $1,2,3,…,i,…,N$. The speed of motor spindle is given as n (r/min).

The feed rate of the workpiece is labeled as v_f (mm/min). Thus, milling feed rate can be calculated using the Eqs. 1 and 2 [20]:

$$v_f = f_{rN} \cdot N \cdot n \tag{1}$$

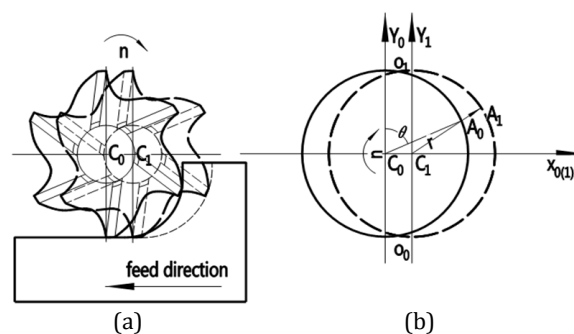$$n = \frac{1000 \cdot v}{\pi \cdot D} \tag{2}$$

Fig. 1 Model for milling cutter rotation: (a) milling along the liner direction; (b) milling tooth trajectory in Cartesian coordinate system

Variable f_{rN} is feed distance per tooth for one cycling, v is the rate of cutting, D is cutter diameter, which is twice the radius of the milling tool. The reasonable parameters can be given according to machining experience.

Cutter movement consists of its own rotation and rectilinear motion along the cutting direction. The feed distance of the tool center varies with the rotation angle of the cutter. Thus, feed distance for per angle f_{rad} (mm/rad) can be deduced from Eqs. 1 and 2:

$$f_{rad} = \frac{f_{rN} * N}{2\pi} \tag{3}$$

As shown in Fig. 1(b), the initial center of the tool is labeled as C_0, which coincides with the origin of the Cartesian coordinate system. The direction of the X-axis is opposite to the feed path of the workpiece and the Y-axis is normal to the X-axis. The Z-axis is perpendicular to the XOY coordinate plane and has the same direction as the cut depth. The equations represent the trajectories of arbitrary cutting point Q, namely i tooth. The tip of $i + 1$ tooth can be expressed as in [15]:

$$\begin{aligned} x_i &= f_{rz}\varphi_i + r\sin(\varphi_i - \omega t) \\ y_i &= r\cos(\varphi_i - \omega t) \end{aligned} \tag{4}$$

φ_i, the instantaneous angle of the cutter, is measured with the positive Y-axis as reference. Positive clockwise is the correct direction of the milling cutter, where $\varphi_i = \varphi_0 + (i - 1)$. θ and φ_0 are the initial angle of the i tooth. The angle between i tooth and $i +1$ tooth can be calculated as:

$$\theta = 2\pi/N \tag{5}$$

3. Integral algorithm for instantaneous uncut chip thickness

3.1 Trajectory equation in polar form

Firstly, we transform the trajectory equation, deduced in Section 2, into the polar coordinate form. The polar point is coincident with origin C_0. The polar axis along the direction of X-axis positive can be found in the Cartesian coordinate system. Counterclockwise direction is assumed positive in polar angle α. The details of transformation are expressed as Eq. 6:

$$\begin{cases} x_i = \rho \cdot \cos\alpha \\ y_i = \rho \cdot \sin\alpha \end{cases} \tag{6}$$

The polar equation of the trajectory of i tooth can be represented by the following equation:

$$\rho_0^2 = r^2 \tag{7}$$

The polar equation of the route of $i + 1$ tooth can be written as Eq. 8:

$$(\rho_1 \cdot \cos\alpha - f_{rad} \cdot \varphi_i)^2 + (\rho_1 \cdot \sin\alpha)^2 = r^2 \tag{8}$$

Where ρ_0 and ρ_1 are respectively the polar radius of path i and $i + 1$ tooth, which are denoted as curve 'C_0A_0' and curve 'C_1A_1' in Fig. 1(b). We assume that $m = f_{rad} \cdot \varphi_i$ can be solved and simplified easily.

$$\rho_1 = m \cdot \cos\alpha \pm \sqrt{r^2 - (m \cdot \sin\alpha)^2} \tag{9}$$

It is obvious that feed distance is smaller than the cutter radius [13]. Therefore, we determine the following formula as reasonable when $m \ll r$:

$$m \cdot \cos\alpha - \sqrt{r^2 - (m \cdot \sin\alpha)^2} < 0 \tag{10}$$

Thus,

$$\rho_1 = m \cdot cos\alpha + \sqrt{r^2 - (m \cdot sin\alpha)^2} \tag{11}$$

where $\alpha \in (\alpha_0, \alpha_1)$ and α_0, α_1 are the polar angle formed at point A0 and A1, respectively, as shown in Fig. 1(b). The Y-axis positive axis is the polar axis where the teeth is cut into and cut out from the workpiece.

3.2 Analytical equation based on integral algorithm

Each cutter tooth is periodically engaged in the workpiece during the milling process. Thus, the IUCT is determined by the shape of two adjacent teeth trajectory when the milling process keeps steady, as shown in Fig. 2(a). The cuttings produced by each tooth are continuous and shaped by the cutting teeth. Chip thickness varies with instantaneous position angle in the cutting plane and the trajectories of the milling cutter [16]. According to the transformation and discussion in Section 3.1, Eqs. 7 and 11 represented the milling trajectories of the cutter tip attached to two adjacent teeth in the polar coordinate system. To improve calculation accuracy, the continuity of each milling tooth in one cycle should not be ignored. The uncut chip region between two adjacent teeth is divided into small regions, as shown in Fig. 2(b).
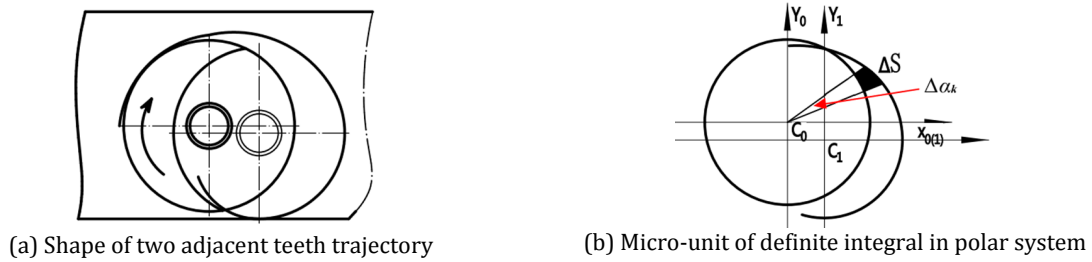

(a) Shape of two adjacent teeth trajectory   (b) Micro-unit of definite integral in polar system

Fig. 2 Micro-unit of definite integral for true milling trajectory

The milling field is taken as the microcell unit and labeled as ΔS for ease of calculation. ΔS can be expressed by Eq. 12:

$$\Delta S = \Delta S_1 - \Delta S_0 \tag{12}$$

Following unit milling angle α_k, the area is taken as micro-unit of the definite integral. Eq. 13 can be described according to the area equation of the curved fan:

$$S = \lim_{\Delta\alpha_k \to 0} \sum_{k=1}^{n} \left(\frac{1}{2}\rho_1^2 \Delta\alpha_k - \frac{1}{2}\rho_0^2 \Delta\alpha_k \right) \tag{13}$$

Thus,

$$S = \int_{\delta_1}^{\delta_2} \left(\frac{1}{2}\rho_1^2 - \frac{1}{2}\rho_0^2 \right) d\alpha \tag{14}$$

where $\delta_1 \in \left(\frac{\alpha_0 \cdot 180°}{\pi}, \frac{\alpha_1 \cdot 180°}{\pi} \right)$ and $\delta_2 = \delta_1 + 1°$. The aim is to balance the calculation efficiency and accuracy. As written and discussed in the preceding section, the uncut chip area in the milling plane can be calculated using Eq. 14, which fully considers the continuity and trochoid of each tooth path. The uncut chip area can be calculated through Eq. 15:

$$S = \frac{m^2}{2} \cdot \int_{\delta_1}^{\delta_2} \cos(2 \cdot \alpha)\, d\alpha + m \cdot \int_{\delta_1}^{\delta_2} \cos\alpha \cdot \sqrt{r^2 - (m \cdot \sin\alpha)^2}\, d\alpha \tag{15}$$

where angle α is measured with its polar axis along the direction of *X*-axis positive. The above equation is transformed into connect the instantaneous angle of cutter φ_i and the area of uncut chip thickness, where

$$\alpha = \begin{cases} \dfrac{\pi}{2} - \varphi_i, & 0 < x < \dfrac{\pi}{2} \\ -(\varphi_i - \dfrac{\pi}{2}), & \dfrac{\pi}{2} \le x < \pi \end{cases} \tag{16}$$

The definite integral expression in φ_i is taken as integration variable:

$$S = \frac{m^2}{2} \cdot \int_{\delta_1}^{\delta_2} \cos(2 \cdot \varphi_i)\, d\varphi_i - m \cdot \int_{\delta_1}^{\delta_2} \sin\varphi_i \cdot \sqrt{r^2 - (m \cdot \cos\varphi_i)^2}\, d\varphi_i \tag{17}$$

Given that the instantaneous angle of cutter φ_i varies in small range, the corresponding arc length and radius can be regarded as approximately vertical. Thus, the shape of the integral section can be assumed as rectangle. The arc length for per radian can be calculated by Eqs. 18 and 19:

$$l_1 = \frac{\pi \cdot r}{180} \tag{18}$$

$$l_2 = \int_{\delta_1}^{\delta_2} \sqrt{\rho(\alpha)^2 + \overline{\rho(\alpha)^2}} \, d\alpha \tag{19}$$

where l_1 and l_2 represent per unit angle arc length of trajectory C_0 and C_1, respectively. Several milling parameters are adopted, shown in Fig. 3, to research length variation with rotation angle.

The length of milling path arc for each unit has a small range when specific milling parameters are given according to the Eqs. 18 and 19. The radius of the cutter significantly affects the arc length of milling path, as shown in Figs. 3(b) and 3(c). Thus, the mean value calculated according to specific radius is reasonable. The variation of uncut chip thickness with feed distance is 0.1 and 0.2. The upper and lower extremes of arc length obtained by Eq. 19 are used for calculation. Thus, the arc length mean value is employed in the calculation to improve the universality and accuracy of the proposed equation. The final expression of the uncut chip thickness is given as:

$$T_n(\alpha) = \frac{S}{\bar{l}} \tag{20}$$

$$\bar{l} = \frac{l_1 + l_2}{2} \tag{21}$$

S is integral area for per unit angle expressed as Eq. 17. The corresponding milling force based on the proposed analytical equation is deduced and calculated in the following section for the model verification.
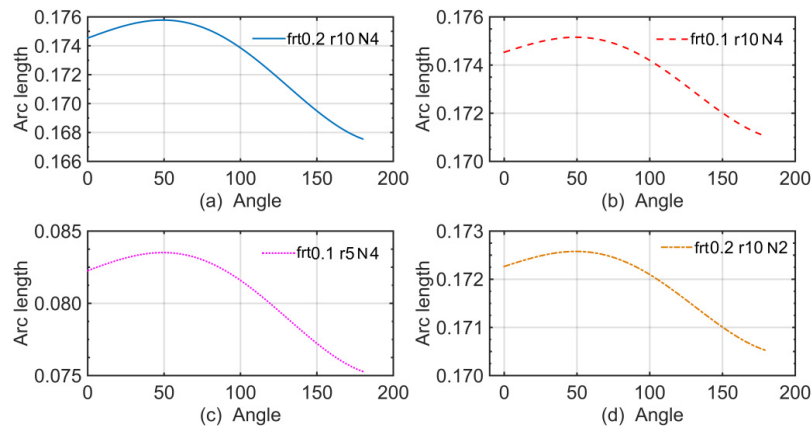

Fig. 3 Arc length for per milling unit with different milling parameters:
(a) f_{rt} = 0.2 mm/tooth, r = 10 mm, N = 4; (b) f_{rt} = 0.1mm/tooth, r = 10 mm, N = 4;
(c) f_{rt} = 0.1 mm/tooth, r = 5 mm, N = 4; (d) f_{rt} = 0.2 mm/tooth, r = 10 mm, N =2

4. Milling force calculation based on ICUT

The dynamic cutting force, according to the analytical mode proposed in reference [13], can be expressed as follows:

$$\begin{cases} F_{tj} = K_{tc}t(\alpha)z + K_{te}z \\ F_{rj} = K_{rc}t(\alpha)z + K_{re}z \\ F_{aj} = K_{ac}t(\alpha)z + K_{ae}z \end{cases} \tag{22}$$

where z is the contact length of cutting edge. During the milling process, both the shearing force of the primary shear zone and the rubbing force of the tertiary deformation zone exist. The mill-

ing force results depend on the instantaneous uncut chip thickness on the one hand, and on the other hand, material shear characteristics and the workpiece surface friction status are also closely related. The IUCT, $t(\alpha)$, can be calculated as the integral algorithm proposed in Section 3. The specific cutting force coefficients, $K_{ic}(i = t/r/a)$ and $K_{ie}(i = t/r/a)$ in the equation, are determined by the average cutting force correction test [21].

The Eq. 22 can be mapped along the X, Y, and Z-directions in the Cartesian coordinate system using the following matrix transition:

$$\begin{pmatrix} dF_X \\ dF_Y \\ dF_Z \end{pmatrix} = \begin{pmatrix} -\cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & -\cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dF_{tj} \\ dF_{rj} \\ dF_{aj} \end{pmatrix} \tag{23}$$

The cutting force calculation process can be seen in following flowchart, Fig. 4. The computer programs in Matlab R2012b are carried out to obtain the specific milling force in direction $X/Y/Z$ in the following section.
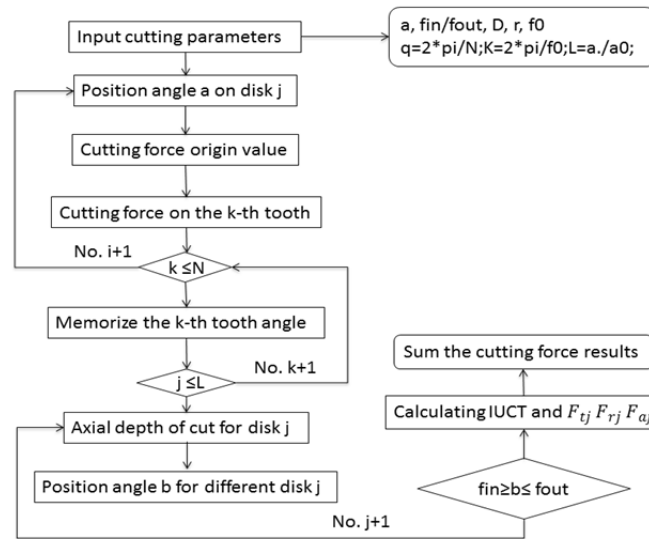

Fig. 4 Cutting force calculation flowchart

5. Results and discussion

5.1 Simulation and experiment analysis

Fig. 5 shows the measured milling forces in reference [21] and the simulations with the proposed integral method for a validation. The comparisons display that the trend of milling forces is roughly same in direction X, Y and Z. The chosen cutting conditions are listed in Table 1.
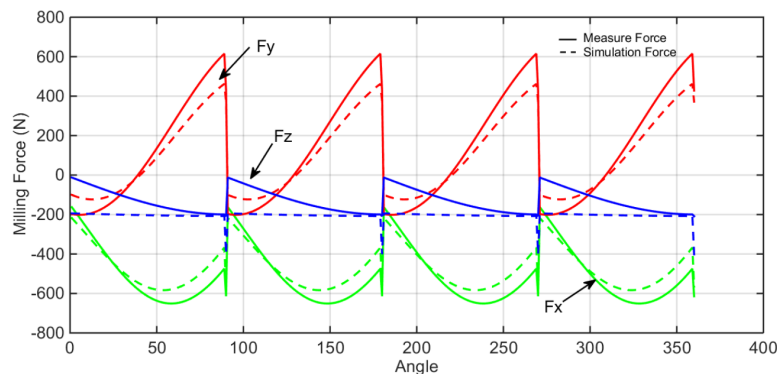

Fig. 5 Measured (from Yucesan and Altintas [21]) and predicted cutting forces for half immersion milling; $N = 4$, $f_{rt} = 0.051$ mm/tooth, $a_p = 5.08$ mm, $v = 30$ m/min

Table 1 Cutting conditions for single flute

Item	Parameter
Spindle speed	2500 rev/min
Cutting rate	30 m/min
Feed rate per teeth	0.051 mm/tooth
Depth of cut	5.08 mm
Cutter diameter	100 mm
Number of teeth	4
Helix angle	30°
Rake angle	10°
Material	AI 7050
Cutting force coefficients	In reference [21]

The measured forces from the reference are slightly larger than the calculations. It is suspect that the differences come from the specific cutting force coefficients which have large impacts on the simulations. During the milling process, cutting forces were generated from shearing force coming from chip formation and ploughing force at the flank of the cutting edge which are two main components of the cutting force coefficients [11, 19].

Thus, some specific milling tests should be carried out for comparisons when conditions permitted. In this work we just chose several specific cutting force coefficients for a comparison. Both the shear and edge cutting force coefficients are assigned according to the references [12, 21]. The developed integral IUCT model is used for milling force calculation compared with the results obtained by the approximate model $T_t(\alpha) = f_{rN} \sin \alpha$ with one single flute, as shown in Fig. 6.

The specific cutting condition is list in Table 2. The milling force with proposed integral model in X/Y/Z directions does not conform to those with approximate model through the whole cutting range, especially the milling force in horizontal machining surface. It is noted that the forces in the direction of X and Y are affected by the cutter tooth trajectory obviously. It is supposed that the trochoid motion of the tooth path causes a phase lag in the IUCT [15]. The maximum appears at 117°, not at 90°, where the milling force peak, in the *X*-direction, is about 250 N which is slightly larger than that calculated by approximate model. The integral algorithm reflects the continuity of milling process that is considered as the reason why the values are different at specific rotation angle. The proposed integral model for milling force prediction has good accuracy. In particular, it provides a better reference for predicting the rotation angle location of the maximum cutting force. While there are fine distinctions in Z direction and the fluctuations are also very little. It is suggested that cutting force in Z direction does nothing but depth of cut. However, the proposed integral equation ignores the effect of angle at cutting in and cutting out position. The milling force at 0° and 180°may not be accurate enough during the milling process. In addition, the analytic equation merely considers feeding movement as a straight line. But other milling conditions are not included such as the cutter centre follows a circular motion. The variation of IUCT may differ, as well as the calculation model, which should be explored further.
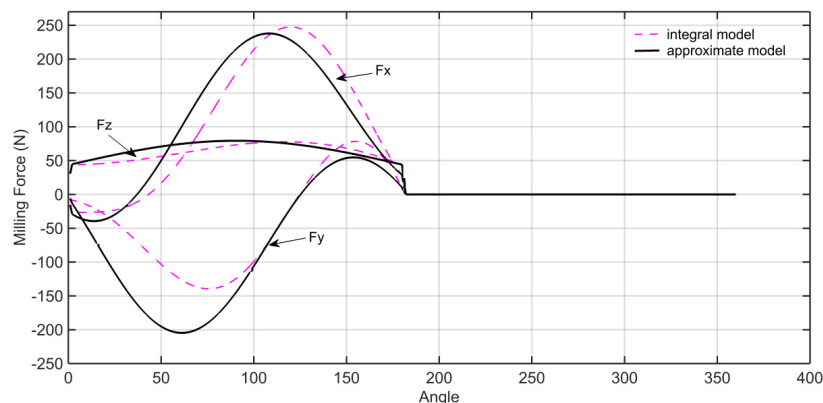

Fig. 6 Comparison of milling force results with different calculation models

Table 2 Cutting conditions for single flute

Item	Parameter
Cutting speed (m/min)	100m/min
Feed rate per teeth (mm/z)	0.06mm/tooth
Depth of cut(mm)	2mm
Cutter diameter(mm)	16mm
Number of teeth	1
Helix angle	0°
Rake angle	6°
Material	AISI 1045
Cutting force coefficients	In reference [12]

5.2 Effect of milling parameters on chip thickness

A variety of milling parameters, such as number of teeth, feed distance per tooth and milling cutter radius, are included in the analytical equation, which have a significant impact on the thickness of instantaneous uncut chip. Thus, in the following section the influence of the parameters is analysed. The results are shown in Fig. 7 and Fig. 8. Fig. 7(a) shows that when the feed rate for per tooth increases from 0.1 mm to 0.2 mm, the IUCT increases at the same time.
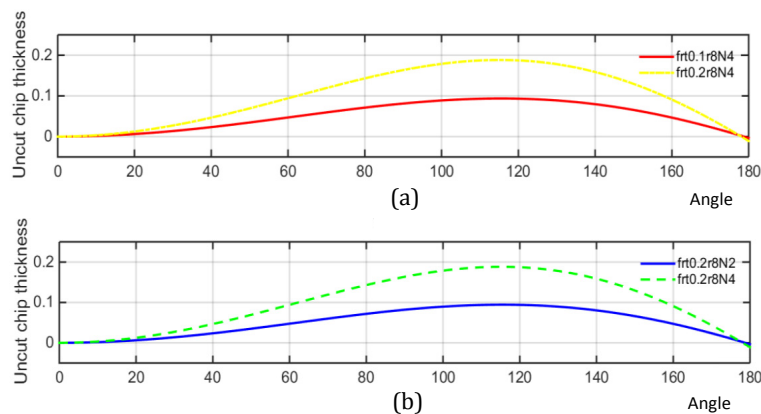

Fig. 7 (a) Effect of different feeding distance on uncut chip thickness (f_{rt} = 0.1 mm/tooth and 0.2 mm/tooth, r = 8 mm, N = 4); (b) Effect of different number of tooth on uncut chip thickness (f_{rt} = 0.2 mm/tooth, r = 8 mm, N = 2/4)

The width value of IUCT becomes bigger with more cutter teeth, as shown in Fig.7 (b). It is believed that this mainly comes from more teeth milling simultaneously, which induce in thermo-visco-plastic strength changes between adjacent teeth in the material. It can be deduced that it is in favour of micro milling process and improving roughness of the machined surface with multi-teeth cutter.

   Fig. 8 shows the effect of cutter radius on the IUCT obtained by simulating under different radius. It is shown that the ICUT decrease when the cutter radius increase from 5 mm to 8 mm. It is supposed that small radius tools could improve the metal removing rate that is more suitable for machining precise part with small volume. It is noticeable that the results in Figs. 7 and 8 show that the maximum value of the IUCT appears at the rotation angle of 117° approximatively, which are consistent with the milling force results described in Fig. 6.
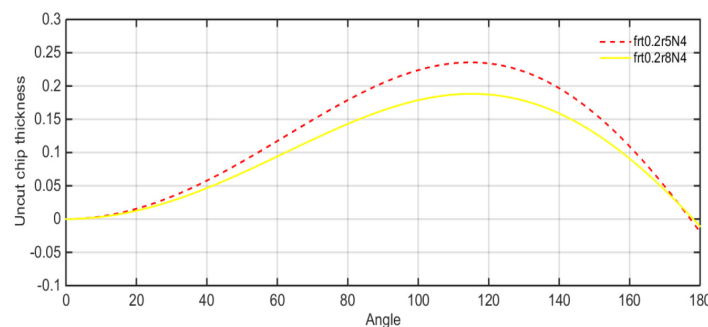

Fig. 8 Effect of different radius of milling cutter on uncut chip thickness (f_{rt} = 0.2mm/tooth, r = 5 mm and 8 mm, N = 4)

6. Conclusion

A method based on polar integral algorithm for calculating the thickness of transient uncut chip thickness is proposed for the cutting force predictions in milling process. The analytical equation is also suitable for engineering application. The main findings of the research:

- True milling trajectory considered as trochoid motion was established in the polar coordinates. Both milling continuity and milling trajectory are considered when developing a new model for IUCT calculations. The new developed model offers a methodology for calculating the IUCT precisely.
- The milling forces obtained by experiment in reference [21] and simulations with the developed IUCT model are compared in the paper. The integral model proves to be more reasonable in terms of the location where the peak appears. Results indicate that the outcomes followed by the polar integral algorithm are close to the reality.
- Different process parameters are simulated and compared according to the results calculated by proposed model. The comparisons suggest that increasing both the feed per tooth and number of teeth can surge the width of IUCT slightly, but decrease with smaller cutter radius.

Acknowledgement

References

[1] Omar, O.E.E.K., El-Wardany, T., Ng, E., Elbestawi, M.A. (2007). An improved cutting force and surface topography prediction model in end milling, *International Journal of Machine Tools and Manufacture,* Vol. 47, No. 7-8, 1263-1275, doi: 10.1016/j.ijmachtools.2006.08.021.

[2] Matsubara, A., Ibaraki, S. (2009). Monitoring and control of cutting forces in machining processes: A review, *Monitoring and Control of Cutting Forces in Machining Processes,* Vol. 3, No. 4, 445-456, doi: 10.20965/ijat.2009.p0445.

[3] Mwinuka, T.E., Mgwatu, M.I. (2015). Tool selection for rough and finish CNC milling operations based on tool-path generation and machining optimisation, *Advances in Production Engineering & Management*, Vol. 10, No. 1, 18-26, doi: 10.14743/apem2015.1.189.

[4] Milfelner, M., Cus, F., Balic, J. (2005). An overview of data acquisition system for cutting force measuring and optimization in milling, *Journal of Materials Processing Technology,* Vol. 164-165, 1281-1288, doi: 10.1016/j.jmatprotec.2005.02.146.

[5] Liu, X.-W., Cheng, K., Webb, D., Luo, X.-C. (2002). Improved dynamic cutting force model in peripheral milling. Part I: Theoretical model and simulation, *The International Journal of Advanced Manufacturing Technology,* Vol. 20, No. 9, 631-638, doi: 10.1007/s001700200200.

[6] Ikua, B.W., Tanaka, H., Obata, F., Sakamoto, S. (2001). Prediction of cutting forces and machining error in ball end milling of curved surfaces -I theoretical analysis, *Precision Engineering,* Vol. 25, No. 4, 266-273, doi: 10.1016/S0141-6359(01)00077-0.

[7] Qu, S., Zhao, J., Wang, T., Tian, F. (2015). Improved method to predict cutting force in end milling considering cutting process dynamics, *The International Journal of Advanced Manufacturing Technology,* Vol. 78, No. 9-12, 1501-1510, doi: 10.1007/s00170-014-6731-5.

[8] Han, Z., Jin, H., Fu, H. (2015). Cutting force prediction models of metal machining processes: A review, In*: Proceedings of 2015 International Conference on Estimation, Detection and Information Fusion (ICEDIF),* Harbin, China, 323-328, doi: 10.1109/ICEDIF.2015.7280216.

[9] Gradišek, J., Kalveram, M., Weinert, K. (2004). Mechanistic identification of specific force coefficients for a general end mill, *International Journal of Machine Tools and Manufacture,* Vol. 44, No. 4, 401-414, doi: 10.1016/j.ijmachtools.2003.10.001.

[10] Saric, T., Simunovic, G., Simunovic, K. (2013). Use of neural networks in prediction and simulation of steel surface roughness, *International Journal of Simulation Modelling,* Vol. 12, No. 4, 225-236, doi: 10.2507/IJSIMM12(4)2.241.

[11] Martellotti, M.E. (1945). An analysis of the milling process, Part II: Down milling, *Transactions of ASME,* Vol. 67, No. 4, 233-251.

[12] Gonzalo, O., Beristain, J., Jauregi, H., Sanz, C. (2010). A method for the identification of the specific force coefficients for mechanistic milling simulation, *International Journal of Machine Tools and Manufacture,* Vol. 50, No. 9, 765-774, doi: 10.1016/j.ijmachtools.2010.05.009.

[13] Budak, E., Altintaş, Y., Armarego, E.J.A. (1996). Prediction of milling force coefficients from orthogonal cutting data, *Journal of Manufacturing Science and Engineering*, Vol. 118, No. 2, 216-224, doi: 10.1115/1.2831014.

[14] Spiewak, S. (1995). An improved model of the chip thickness in milling, *CIRP Annals,* Vol. 44, No. 1, 39-42, doi: 10.1016/s0007-8506(07)62271-9.

[15] Li, H.Z., Liu, K., Li, X.P. (2001). A new method for determining the undeformed chip thickness in milling, *Journal of Materials Processing Technology,* Vol. 113, No. 1-3, 378-384, doi: 10.1016/S0924-0136(01)00586-6.

[16] Kumanchik, L.M., Schmitz, T.L. (2007). Improved analytical chip thickness model for milling, *Precision Engineering,* Vol. 31, No. 3, 317-324, doi: 10.1016/j.precisioneng.2006.12.001.

[17] Song, G., Li, J., Sun, J. (2013). Approach for modeling accurate undeformed chip thickness in milling operation, *The International Journal of Advanced Manufacturing Technology,* Vol. 68, No. 5-8, 1429-1439, doi: 10.1007/s00170-013-4932-y.

[18] Saï, L., Bouzid, W., Zghal, A. (2008). Chip thickness analysis for different tool motions: For adaptive feed rate, *Journal of Materials Processing Technology*, Vol. 204, No. 1-3, 213-220, doi: 10.1016/j.jmatprotec.2007.11.094.

[19] Grossi, N., Sallese, L., Scippa, A., Campatelli, G. (2015). Speed-varying cutting force coefficient identification in milling, *Precision Engineering,* Vol. 42, 321-334, doi: 10.1016/j.precisioneng.2015.04.006.

[20] Fu, G. (2000). Basic knowledge of metal cutting, In: Shi, Q.F, Wang, L. (eds.), *Metal cutting manual*, (Third edition), Shanghai Science and Technology Press, Shanghai, 1-31, (in Chinese).

[21] Altintas, Y. (2012). *Manufacturing automation: Metal cutting mechanics, machine tool vibrations, and CNC design,* (Second edition), Cambridge University Press, New York, USA, doi: 10.1017/CBO9780511843723.

[22] Moufki, A., Dudzinski, D., Le Coz, G. (2015). Prediction of cutting forces from an analytical model of oblique cutting, application to peripheral milling of Ti-6Al-4V alloy, *The International Journal of Advanced Manufacturing Technology,* Vol. 81, No. 1-4, 615-626, doi: 10.1007/s00170-015-7018-1.

[23] Armarego, E.J.A., Deshpande, N.P. (1993). Force prediction models and CAD/CAM software for helical tooth milling processes. II. Peripheral milling operations, *International Journal of Production Research,* Vol. 31, No. 10, 2319-2336, doi: 10.1080/00207549308956860.

Appendix A

Nomenclature

$T_t(\alpha)$	IUCT calculated by traditional model
φ	Instantaneous cutter angle
f_{rN}	Feed per tooth
r	Radius of milling cutter
N	Number of cutter tooth
n	Spindle speed
v_f	Feed rate of the workpiece
v	Rate of cutting
D	Cutter diameter
f_{rad}	Feed per rad
φ_i	Instantaneous cutter angle of the *i*-th tooth
θ	Cutter pitch angle
φ_0	Initial angle of the cutter
α	Instantaneous cutter angle in polar system
$\rho_i (i = 0,1)$	Polar radius of trajectory for *i*-th tooth
ΔS	Microcell unit of milling field
α_k	Unit milling angle
$\delta_i (i = 0,1)$	Lower and upper limit of integral
S	IUCT area of unit milling angle
$l_i (i = 1,2)$	Arc length of trajectory per radian
$\bar{l}$	Mean value of arc length
$T_n(\alpha)$	IUCT calculated by proposed integral model
z	Cutting edge contact length
$K_{ic} (i = t/r/a)$	Tangential/radial/axial shearing force coefficients
$K_{ie} (i = t/r/a)$	Tangential/radial/axial edge force coefficients
a_P	Depth of cut
$dF_i (i = x/y/z)$	Instantaneous cutting force in *X*-, *Y*- and *Z*-direction
fin/fout	Entry/ Exit angle

Change-point estimation for repairable systems combining bootstrap control charts and clustering analysis: Performance analysis and a case study

Yang, Z.J.[a], Du, X.J.[a], Chen, F.[a,*], Chen, C.H.[a,*], Tian, H.L.[a], He, J.L.[a]

[a]School of Mechanical Science and Engineering, Jilin University, Changchun, P.R. China

A B S T R A C T

Complex repairable systems with bathtub-shaped failure intensity will normally go through three periods in the lifecycle, which requires maintenance policies and management decisions accordingly. Therefore, the accurate estimation of change points of different periods has great significance. This paper addresses the challenge of change-point estimation in failure processes for repairable systems, especially for sustained and gradual processes of change. The paper proposes a sectional model composed of two non-homogeneous Poisson processes (NHPPs) to describe the bathtub-shaped failure intensity. In order to obtain the accurate change-point estimator, a novel hybrid method is developed combining bootstrap control charts with the sequential clustering approach. Through Monte Carlo simulations, the proposed change-point estimation method is compared with two powerful estimation procedures in various conditions. The results suggest that the proposed method performs effective and satisfactory for failure processes with no limits of distributions, changing ranges and sampling schemes. It especially provides higher precision and lower uncertainty in detecting small shifts of change. Finally, a case study analysing real failure data from a heavy-duty CNC machine tool is presented. The parameters of the proposed NHPP model are estimated. The change point of the early failure period and the random failure period is also calculated. These findings can contribute to determining the burn-in time in order to improve the reliability of the machine tool.

© 2018 CPE, University of Maribor. All rights reserved.

1. Introduction

In lifecycle reliability analysis, the failure patterns of complex repairable systems can be generally categorised as early failure period, random failure period and wear-out failure period, which have various failure mechanisms and performances. The failure intensity during the lifecycle usually follows a bathtub-shaped curve in practice. Monitoring the changing trend and determining the change point for different periods will provide reasonable guidance for health management and decision-making in time, such as early failure elimination experiment, maintenance strategy, etc., which also enables the improvement of reliability and efficiency.

Statistical process control (SPC) charts have been widely used in the process monitoring. When a signal is detected to be out-of-control, it may indicate a sudden change in quality or a transition from one state to the next. Therefore, engineers could be initiated to trace back the cause based on the signal and make appropriate adjustments in time. Many researches have been done for change-point estimation considering various change patterns in multivariate en-

vironment based on SPC theory. Amiri and Allahyari [1] made a literature review for the existing methods to detect the real time of change in control charts. Pignatiello and Samuel [2] adopted the maximum likelihood estimator (MLE) for a step change in CUSUM and EWMA control charts. Perry *et al.* [3, 4] considered various quality characteristics for change point estimation of different change types involving linear trend and monotonic change. Unknown-parameter change-point models were also defined for identifying changes in multivariate process [5-7].

In addition to these traditional approaches, artificial neural network (ANN), clustering techniques and other learning based methods have been combined with SPC charts for monitoring of complicated systems [8]. As for ANN methods, Amiri *et al.* [9] developed a probabilistic ANN procedure to estimate the change points of quality characteristics following normal distributions. Ahmadzadeh *et al.* [10, 11] combined ANN with multivariate exponentially weighted moving average (MEWMA) control charts to estimate the actual change points in multivariate process. Maleki *et al.* [12] proposed an ANN-based model to detect step changes considering the correlation between multivariate or attribute quality characteristics. Change point estimators were also identified based on ANN in five change patterns [13]. When it comes to clustering techniques, fuzzy-statistical clustering approach was adopted to estimate change points in various control charts with fixed or variable sample sizes [14, 15]. The step change, linear trend change and monotonic change are identified using clustering approach by [16-18], respectively. Other learning based methods, such as support vector machines, have also been incorporated into SPC to solve the change-point estimation problems [19, 20].

Despite the literature has paid large attention to the SPC methods applying to change-point detection in quality monitoring, they are mostly designed for identifying step changes in the process with a specific distribution. The study of change-point estimation in reliability monitoring is rarely considered, especially for the failure processes of complex repairable systems. Considering the characteristics of complex repairable systems, the study of this problem is unique in that:

- Due to the long life cycle and high reliability, the sample size of failure data is usually very small. The historical data that can be used for reference are also not good enough owing to different operating environments and failure mechanisms.
- The failure process always consists of multiple periods. Accordingly, the real-time monitoring of reliability is essential in order to adjust maintenance strategies in time. Otherwise, it may result in decreased productivity, increased cost, and even some catastrophic damage.
- The change pattern prefers to be identified as a trend showing a bathtub-shaped curve instead of a step change. An improved SPC chart is required to be sensitive to the gradual and sustained change.
- Since the change point simply detected from SPC charts could be relatively rough [10], a change-point model based on MLE can be developed to help improve the accuracy and stability of the estimation. It makes the existing SPC charts subjected to some specific distribution not applicable.

Therefore, the paper first develops a sectional NHPP model to describe the bathtub-shaped failure intensity in the lifecycle. The model considers systems with minimal maintenance especially subject to early failures, it is able to flexibly fit time-ordered failure data in the early failure period and the random failure period. Then, a novel hybrid estimation method integrating a bootstrap control chart with a sequential clustering approach is proposed. The method is superior in real-time monitoring the gradual and sustained trend of change. Moreover, based on MLE with strict statistical deduction, the proposed change-point estimation method not only improves the calculation efficiency, but also achieves good accuracy even in the case of small samples.

This paper is structured as follows. The proposed model is established in Section 2, whose bathtub-shaped failure intensity function is also discussed in this part. The following section integrates clustering method with SPC to explain the change-point estimation procedure. In Section 4, the performance of the proposed approach is assessed by a series of simulations, and a numerical application is presented. Finally, conclusions are made in Section 5.

2. The bathtub-shaped failure intensity model

An NHPP can be adopted to describe the failures of complex repairable systems under minimal maintenance strategy. The strategy assumes that maintenance time can be ignored and the system will only return to the same state as it was right before the failure occurred, which is more rational for reliability modelling of complex repairable systems. In practical work, maintenance actions often involve only some parts of the system, so the overall state of the system is not affected. Based on this, our bathtub-shaped failure intensity model is proposed under this assumption and set up from the NHPP.

The NHPP is also called the Weibull process [21] when it has a failure intensity function as

$$\omega(t) = \lambda\beta t^{\beta-1}, t \geqslant 0, \lambda > 0, \beta > 0 \tag{1}$$

where λ is the intensity parameter, β is the shape parameter, and t is the operating time.

Following the NHPP, the number of failures during the time interval $(t_1, t_2]$ equals to

$$W(t_1, t_2) = E(N(t_2) - N(t_1)) = \int_{t_1}^{t_2} \omega(\mu)d\mu \tag{2}$$

The same can be inferred that $W(0, t) = E(N(t))$ represents the expected number of failures through time interval $[0, t]$, thereby the cumulative failure intensity function is given by

$$W(t) = \int_0^{t_0} \omega(\mu)d\mu = \lambda t^\beta \tag{3}$$

Then the corresponding reliability function is derived as

$$R(t) = e^{-\lambda t^\beta} \tag{4}$$

As well as the cumulative density function

$$F(t) = 1 - R(t) = 1 - e^{-\lambda t^\beta} \tag{5}$$

And the probability density function

$$f(t) = \frac{dF(t)}{dt} = \omega(t)R(t) = \lambda\beta t^{\beta-1} e^{-\lambda t^\beta} \tag{6}$$

However, a single NHPP model can only illustrate the situation that failure intensity and failure time are strictly monotonic, which is unable to describe the non-monotonic trends of different life stages. Therefore, a sectional model of multiple NHPPs is required to fit the bathtub-shaped curve in order to describe the rules of failures in different failure periods, which could help to obtain the change point.

2.1 Two sectional NHPP model

Experts have developed various models to describe complex and diverse data distributions [22], most of them are constructed assuming that systems are non-repairable or 'repair as new'. It is not appropriate for the concern in our case. Based on this, we propose a sectional model involving two NHPPs, representing the early failure period and the random failure period, respectively.

$$\omega(t) = \begin{cases} \omega_1(t) = \lambda_1\beta_1 t^{\beta_1-1}, & 0 \leqslant t < t_0 \\ \omega_2(t) = \lambda_2\beta_2 t^{\beta_2-1}, & t_0 \leqslant t < \infty \end{cases} \tag{7}$$

We divide the operating time t of equipment into two intervals, t_0 on behalf of the division point. Where λ_1, β_1 denote the intensity parameter and the shape parameter of the failure intensity function for early failure period, λ_2, β_2 are for random failure period, when $\beta_1 = \beta_2$, the model will degenerate into a single NHPP, that is the reason $\beta_1 \neq \beta_2$ is defined.

The sectional NHPP model has a cumulative failure intensity function with the expression

$$W(t) = \begin{cases} W_1(t) = \lambda_1 t^{\beta_1}, & 0 \leqslant t < t_0 \\ W_2(t) = \lambda_1 t_0^{\beta_1} + \lambda_2 t^{\beta_2} - \lambda_2 t_0^{\beta_2}, & t_0 \leqslant t < \infty \end{cases} \tag{8}$$

When $0 \leqslant t < t_0$, it is called the early failure period with the cumulative failure intensity $W(t) = W_1(t)$, when $t_0 \leqslant t < \infty$, it is called the random failure period with the cumulative failure intensity $W(t) = W_2(t)$, where $W_1(t)$ and $W_2(t)$ are two-parameter NHPP models, respectively.

2.2 Characterisation of bathtub-shaped failure intensity

For the failure intensity functions of the two sectional NHPP model, different values of the shape parameter β are used to denote different periods of the product's life. As shown in Fig. 1, when $\beta < 1$, the failure intensity decreases with time t, it means that failures will happen more and more infrequently with system ageing. When $\beta > 1$, the failure intensity increases as t progresses, it means that the failure rate will rise with system ageing. When $\beta = 1$, the failure intensity is a constant.

Accordingly, the cumulative failure intensity, which is also known as the cumulative failure number, has a change pattern seen in Fig. 1. Through further mathematical derivation, the conclusion agrees with the trend of failure intensity, it indicates that the growth rate of cumulative failure number will first reduce with time in early failure period, when it comes to random failure period, the growth rate will maintain a steady state, after that it will continue to rise during wear-out failure period.

In view of our proposed sectional model, it is aimed at describing the changing trend of the early failure period and the random failure period. Through setting the constraints that failure intensity and cumulative failure intensity are both continuous at t_0, the accurate change point t_0 could be obtained. which is expressed as

$$\begin{cases} \omega_1(t_0) = \omega_2(t_0) \\ W_1(t_0) = W_2(t_0) \end{cases} \tag{9}$$

Substitute Eq. 7 and Eq. 8 in Eq. 9

$$\begin{cases} \lambda_1\beta_1 t_0^{\beta_1-1} = \lambda_2\beta_2 t_0^{\beta_2-1} \\ \lambda_1 t_0^{\beta_1} = \lambda_1 t_0^{\beta_1} + \lambda_2 t_0^{\beta_2} - \lambda_2 t_0^{\beta_2} \end{cases} \tag{10}$$

Since the cumulative failure intensity function is much in evidence to be continuous at t_0 with the sectional two NHPP modeling, the change point t_0 can be obtained as

$$t_0 = \left(\frac{\lambda_1\beta_1}{\lambda_2\beta_2}\right)^{\frac{1}{\beta_2-\beta_1}} \tag{11}$$

At this point, the bathtub-shaped failure intensity and its change point can be achieved with our proposed sectional model, which can be adopted to assist the failure process monitoring.
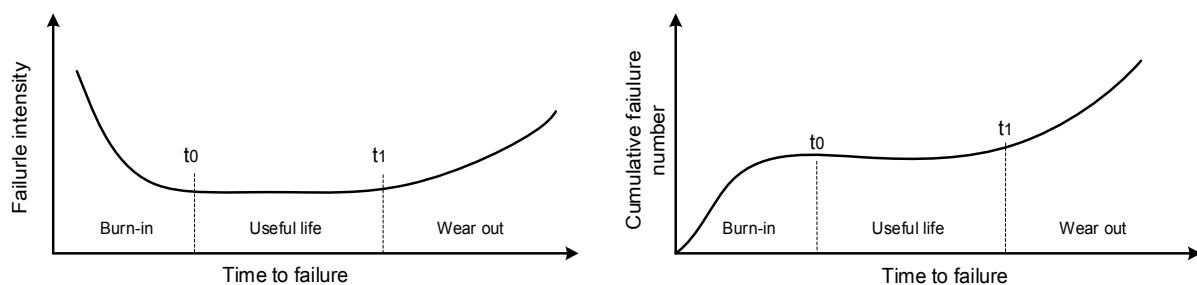

Fig. 1 Bathtub-shaped curves of the failure intensity and corresponding cumulative failure number

3. Change-point estimation combined SPC with sequential clustering

For the purpose of estimating the exact change point, a specific bootstrap control chart to monitor the performance of repairable systems is established. The objective is to track possible transition points by detecting all out-of-control signals during the failure process. These candidate points help to classify data patterns into the two phases of early failure period and random failure period, respectively. Then, the clustering techniques are adopted to eliminate interference data within out-of-control signals in order to identify the optimum transition point. Finally, the accurate change point can be estimated by fitting the two sections of our proposed failure intensity model with the segmented observation data.

The proposed approach integrates the advantages of SPC method and clustering analysis. Considering the characteristics of SPC, two possible states are first predetermined as in-control and out-of-control, it will simplify the clustering model with a known number of clusters. Meanwhile, the data is monitored in time series, the order of points are preserved for the iteration to search for the optimal clusters. At last, cluster settings will only be examined when an out-of-control signal is detected. The number of iterations is dramatically reduced to save computing time. Therefore, the SPC method is beneficial for improving efficiency for change-point estimation. Additionally, the intervention of sequential clustering approach is for benefits of lower interference and better accuracy, because the SPC chart in this study is required to be more sensitive to sustained shifts and gradual changing trends rather than sudden changes or random noises.

3.1 The bootstrap control chart

Traditional control charts [23], like standard Shewhart charts, can only be strictly applied to the normal distribution by monitoring the shifts of mean and variance. As to those modified CUSUM charts and EWMA charts [24], they are always set up for some specific distributions. In our case, the failure process is modelled with a sectional NHPP, it's difficult to find the applicable control charts and corresponding methods to establish its control limits. Therefore, a bootstrap control chart with no limits of any specific distribution is developed.

The particular SPC chart is established based on Monte Carlo simulation [25], which can be implemented as the following steps shown in Fig. 2.

Step 1: Determine the stable values of parameter λ_s and β_s. They can be defined by experience or MLEs with observations collected in Phase I, which represents a stable state in SPC theory.

Step 2: Generate bootstrap random variables $t_{i1}, t_{i2}, \dots, t_{in}$ from the NHPP model with parameters λ_s and β_s in size n. n equals to the data volume of Phase II, which is the subsequent monitoring process based on the estimated control limits. The method from [26] is improved to generate random variables of failure times $t_{i1}, t_{i2}, \dots, t_{in}$ following the NHPP sequentially.

Step 3: Calculate the MLEs from $t_{i1}, t_{i2}, \dots, t_{in}$ to obtain λ_i and β_i of the NHPP models.

Step 4: Obtain the p-quantile W_{pi} with $W_{pi} = (-(1/\lambda_i) \ln p)^{1/\beta_i}$, where W_{pi} represents $100 p$-th percentile of interest.

Step 5: Repeat the steps from Step.2 to Step.4 for $B (B > 1000)$ times to obtain B groups of p-quantiles. Order them from the smallest to the largest as $W_{p1} < W_{p2} < \cdots < W_{pB}$. Then the centre line (CL) is yielded as a mathematical expectation of acquired data, the lower control limit (LCL) is the i-th p-quantile W_{pi}, and the upper control limit (UCL) is the $(B - i)$-th value as $W_{p(B-i)}$, where $i = \alpha B/2$, representing that there're i estimators beyond the control limits. The false alarm risk α indicates the probability when the system is diagnosed to be out of control while it's actually in control.
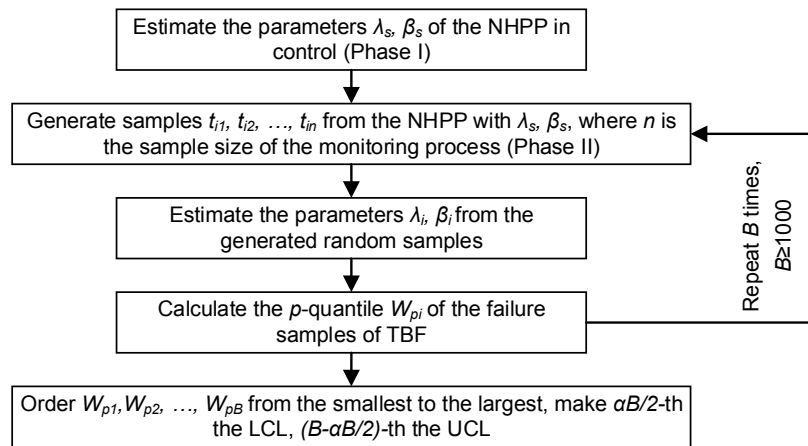
Fig. 2 Procedure to establish the bootstrap control chart

After control limits are obtained, the bootstrap control chart can be established through online monitoring. Time between failures (TBF) observations are plotted after each failure, and process shifts can be detected under some suitable runs rules.

Take an instance as shown in Fig. 3, we assume there are 200 observations in total, the first 100 points in Phase I are supposed to be in control. Then the control limits, including UCL, LCL and CL, are calculated following the steps from Step 1 to Step 5. When it's converted to Phase II, the other 100 points are plotted on the established control chart, which should have been detected to be out of control as we expected. However, the fact in Fig. 3 shows that several observations (red solid points) in Phase I are beyond the control limits, meanwhile, a large number of observations (black solid points) in Phase II are within the control limits. It infers that the proposed control chart has a problem of random interference, the accuracy of detection will be reduced and it would be hard to determine which signal is the optimal transition point. In view of this problem, the sequential clustering approach is introduced to combine with the proposed SPC method.

3.2 The sequential clustering approach

The purpose of introducing clustering approach is to remove the disturbance in SPC and extract the optimal point $\hat{\tau}$, it can detect whether systems convert to another failure period and divide observations in two phases.

In the proposed sequential clustering approach, two possible clusters are defined: for the i-th out-of-control signal O_i, all observations before O_i are classified to the in-control cluster, and all observations after O_i are automatically classified to the out-of-control cluster. In the interest of obtaining the optimal transition point $\hat{\tau}$, first of all, the clustering model with applicable validity indices is developed. Then, the objective function of each cluster setting according to time series is examined. At last, the optimal transition point $\hat{\tau}$ is obtained by optimising the objective function. Two validity indices and the objective function from [16] are illustrated and improved as follows.

Clusters within variation

A cluster within variation V_{within} is expressed as the distance between observations and their cluster centres, given by C_{in} of in-control cluster centre and C_{out} of out-of-control cluster centre, as shown in Fig. 4. In consideration of Phase I and Phase II based on the specific bootstrap control chart, C_{in} in Phase I equals to the in-control mean. C_{in} in Phase II is substituted as the expectation of the change-point model, which is also defined as the CL of established control chart.

Therefore, V_{within} in Phase I is

$$V_{within} = \sum_{i=1}^{\tau} (\overline{X}_i - C_{in})^2 + \sum_{i=\tau+1}^{n} (\overline{X}_i - C_{out})^2 \tag{12}$$

where $\overline{X}_i$ is the average value of i-th subgroup of observations, and

$$C_{in} = \sum_{i=1}^{\tau} \frac{\overline{X_i}}{\tau}, \quad C_{out} = \sum_{i=\tau+1}^{n} \frac{\overline{X_i}}{n-\tau} \tag{13}$$

What's more, V_{within} in Phase II is

$$V_{within} = \sum_{i=1}^{\tau} (\overline{X_i} - CL)^2 + \sum_{i=\tau+1}^{n} (\overline{X_i} - C_{out})^2 \tag{14}$$

Clusters between variation

A cluster between variation is expressed as the distance between cluster centres and the total cluster centre of all observations C_T, as shown in Fig. 4. Similar with the situation of clusters within variation V_{within}, C_T is substituted as the CL of bootstrap control chart in Phase II, so as to obtain the expression of $V_{between}$ in Phase I as

$$V_{between} = \tau(C_{in} - C_T)^2 + (n-\tau)(C_{out} - C_T)^2 \tag{15}$$

where $C_T = \sum_{i=1}^{n} \overline{X_i}/n$ and $V_{between}$ in Phase II is

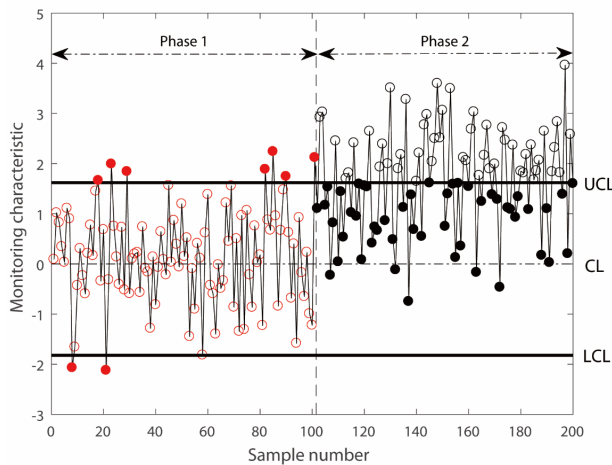$$V_{between} = \tau(C_{in} - CL)^2 + (n-\tau)(C_{out} - CL)^2 \tag{16}$$

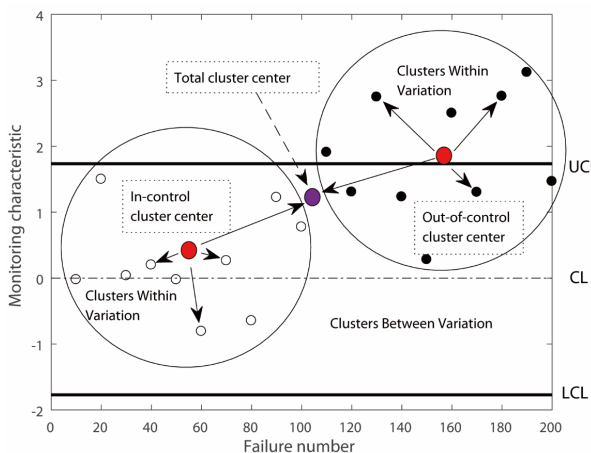Fig. 3 An example of established bootstrap control chart

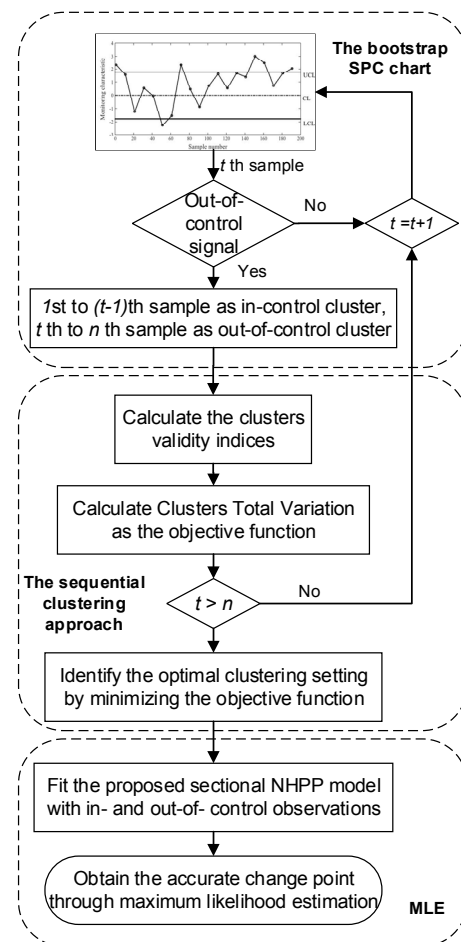Fig. 4 Illustration of clusters within and between variation

Fig. 5 Change-point estimation procedure

The objective function

Clusters total variation V_{total} is defined as the objective function. It integrates the within and between variations for the purpose of valuating different cluster settings more completely and getting the optimal clusters more efficiently. The objective function has the expression as

$$V_{total} = V_{within} + V_{between} \tag{17}$$

For the classification in Phase I, it is derived from Eq. 12 and Eq. 15

$$V_{total} = \sum_{i=1}^{\tau} (\bar{X}_\iota - C_{in})^2 + \sum_{i=\tau+1}^{n} (\bar{X}_\iota - C_{out})^2 - \tau(C_{in} - C_T)^2 + (n-\tau)(C_{out} - C_T)^2 \tag{18}$$

What's more, for the classification in Phase II, Eq. 14 and Eq. 16 is combined as

$$V_{total} = \sum_{i=1}^{\tau} (\bar{X}_\iota - CL)^2 + \sum_{i=\tau+1}^{n} (\bar{X}_\iota - C_{out})^2 - \tau(C_{in} - CL)^2 + (n-\tau)(C_{out} - CL)^2 \tag{19}$$

To find out the optimal transition point $\hat{\tau}$ and assign observations to in- and out-of-control clusters, the proposed objective function should be minimised.

$$\hat{\tau} = argmin\ V_{total}(\tau) \tag{20}$$

3.3 Change-point estimation procedure

The overall structure of proposed change-point estimation procedure is summarised as Fig. 5. First of all, the bootstrap SPC chart is constructed, possible combinations of in- and out-of-control clusters are classified by each out-of-control signal. Then, the sequential clustering approach is adopted, proposed validity indices are examined sequentially for each clustering setting and the objective function is optimised to obtain the best assignment of observations. Finally, the best assignment of observations is used to fit the proposed two sectional NHPP model, the more precise change-point estimator could be calculated through maximum likelihood estimation.

As for the part of maximum likelihood estimation, let $f(t_1, t_2, \dots, t_{\hat{\tau}}, \dots, t_n)$ denotes the joint probability density of failure times $0 < t_1 < \cdots < t_{\hat{\tau}} < \cdots < t_n$, where n is the random number of failures. Among them, $t_1, t_2, \dots, t_{\hat{\tau}}$ belong to the optimal in-control cluster, and $t_{\hat{\tau}}, t_{\hat{\tau}+1}, \dots, t_n$ belong to the optimal out-of-control cluster. The joint probability density can be constructed with a failure intensity $\omega(t)$ in Eq. 7 and a cumulative failure intensity $W(t)$ in Eq. 8.

$$f(t_1, t_2, \dots t_{\hat{\tau}}, \dots, t_n) = \begin{cases} (\lambda_1 \beta_1)^{\hat{\tau}} \prod_{i=1}^{\hat{\tau}} t_i^{\beta_1 - 1} e^{-\lambda_1 t_{\hat{\tau}}^{\beta_1}}, & 0 < t_1 < t_2 < \cdots < t_{\hat{\tau}} \\ (\lambda_2 \beta_2)^{n-\hat{\tau}} \prod_{i=\hat{\tau}+1}^{n} t_i^{\beta_2 - 1} e^{-\lambda_2 t_n^{\beta_2} + \lambda_2 t_0^{\beta_2} - \lambda_1 t_0^{\beta_1}}, & t_{\hat{\tau}} < \cdots < t_n \end{cases} \tag{21}$$

The log-likelihood function can be obtained as

$$\ln(L) = \hat{\tau}(\ln\lambda_1 + \ln\beta_1) + (\beta_1 - 1)\sum_{i=1}^{\hat{\tau}} \ln t_i - \lambda_1 t_{\hat{\tau}}^{\beta_1} + (n - \hat{\tau})(\ln\lambda_2 + \ln\beta_2)$$
$$+ (\beta_2 - 1)\sum_{i=\hat{\tau}+1}^{n} \ln t_i - \lambda_2 t_n^{\beta_2} + \lambda_2 t_0^{\beta_2} - \lambda_1 t_0^{\beta_1} \tag{22}$$

Differentiate log-likelihood functions with respect to $\lambda_1, \beta_1, \lambda_2, \beta_2$, making the equations result in zero, the MLEs of parameters will yield from

$$\begin{cases} \dfrac{\partial \ln(L)}{\partial \lambda_1} = \dfrac{\hat{\tau}}{\lambda_1} - t_{\hat{\tau}}{}^{\beta_1} - t_0{}^{\beta_1} = 0 \\[2mm] \dfrac{\partial \ln(L)}{\partial \beta_1} = \dfrac{\hat{\tau}}{\beta_1} + \sum_{i=1}^{\hat{\tau}} \ln t_i - \lambda_1 t_{\hat{\tau}}{}^{\beta_1} \ln(t_{\hat{\tau}}) - \lambda_1 t_0{}^{\beta_1} \ln(t_0) = 0 \\[2mm] \dfrac{\partial \ln(L)}{\partial \lambda_2} = \dfrac{n - \hat{\tau}}{\lambda_2} - t_n{}^{\beta_2} - t_0{}^{\beta_2} = 0 \\[2mm] \dfrac{\partial \ln(L)}{\partial \beta_2} = \dfrac{n - \hat{\tau}}{\beta_2} + \sum_{i=\hat{\tau}+1}^{n} \ln t_i - \lambda_2 t_n{}^{\beta_2} \ln(t_n) + \lambda_2 t_0{}^{\beta_2} \ln(t_0) = 0 \end{cases} \tag{23}$$

With the constraint of Eq. 11 the MLE of change point is obtained as

$$t_0 = argmax\ L(t) \tag{24}$$

4. Performance analysis and a case study

In this section, the performance of the proposed change-point estimation approach is evaluated and compared with some existing methods through Monte Carlo simulations. To demonstrate feasibility and accuracy in industrial applications, the proposed two sectional NHPP model with the estimation process is applied to analyse the field data of a heavy-duty CNC machine tool.

4.1 Performance comparison of the proposed approach with existing estimation procedures

Three series of simulations considering phases I and II of different distributions with various changing shifts are conducted. In these simulations, the evaluation results of the proposed objective function are compared with the results of methods in [16, 27].

The scheme of simulation study is constructed as follows.

Series 1 simulation

In each simulation run, samples of size $m = 1$ are generated from a normal distribution with $\mu = 0$ and $\sigma = 1$ up to the real change point $\tau = 100$ for Phase I. The bootstrap control chart is constructed following the procedure in Section 3.1 with these in-control samples. Then, for Phase II, with a shift in $\mu = 0.5, 1, 1.5, 2, 3$, respectively, samples are generated and monitored on the established control chart. At last, the proposed indices are calculated to estimate the change point. For each adjusted parameter, average change point and corresponding standard error of 10000 simulation runs are computed. The results are tabulated in Table 1 with the comparing results from [16, 27].

Conclusions can be drawn that the proposed method along with the other two comparative methods have quite close monitoring results, which are also consistent with the real change point. Meanwhile, with the increase of the parameter adjustment value, the precision of the determined out-of-control signal is improved. Among them, Ghazanfari *et al.*'s method works a little better than the other two, especially for small shifts. Despite the proposed method's slightly inferior performance with the change-point estimates, it is more stable and the results obtained are more accurate and reliable considering the much smaller standard errors.

Table 1 Average change-point estimates and associated standard errors in Series 1 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	104.45	100.39	99.94	99.7	99.61
	$std(\tilde{\tau})$	23.07	7.15	3.93	3.71	3.49
Ghazanfari *et al.* [16]	$\tilde{\tau}$	103.24	100.15	99.69	99.53	99.33
	$std(\tilde{\tau})$	23.28	7.81	5.97	5.45	5.95
Proposed method	$\tilde{\tau}$	104.59	102.71	101.66	101.33	101.1
	$std(\tilde{\tau})$	17.98	3.44	1.44	0.85	0.37

Series 2 simulation

In the second series of simulation, generate samples of size $m = 1$ from an Exponential distribution with $\lambda = 2$ as the first 100 observations in Phase I. Then, define the altering parameters set as $\lambda + \delta$ ($\delta = 0.5, 1, 1.5, 2, 3$) for the next 100 samples. The change point is estimated for 10000 times in the case of each parameter. The results of average change point and standard error are listed in Table 2 to compare the performance of three methods.

As given in Table 2, the performances of three methods are essentially in agreement with *Series 1 simulation* for Normal distributions. By comparing the results, it illustrates that the proposed method performs best when the shift size of the parameter is small, while its estimated change point tends to be stable under increasing parameter adjustments.

Table 2 Average change-point estimates and associated standard errors in Series 2 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	113.33	104.69	102.54	101.73	101.05
	$std(\tilde{\tau})$	16.81	6.89	4.66	4.02	2.51
Ghazanfari *et al.* [16]	$\tilde{\tau}$	112.83	104.51	102.6	101.76	101.02
	$std(\tilde{\tau})$	15.97	6.8	4.21	3.81	3.23
Proposed method	$\tilde{\tau}$	109.15	104.95	103.96	103.64	103.2
	$std(\tilde{\tau})$	13.65	6.51	4.97	4.53	3.58

Series 3 simulation

At last, simulations are executed for the situation that failure process following Normal distributions with a small amount of samples. Subgroups of size $m = 4$ for Phase I are generated from a Normal distribution with $\mu = 100$ and $\sigma = 5$ until the real change point $\tau = 10$. Then from another Normal distribution with a shift in μ as $\mu = \mu + \sigma \times \delta$ ($\delta = 0.5, 1, 1.5, 2, 3$), samples with a size of 10 for Phase II are generated. The simulation procedure for each parameter μ is repeated over 10000 times, and the estimated results are compared with other methods [16, 27] in Table 3.

The simulation considers the situation when there is no sufficient amount of observations. We can tell from the results that our proposed method still behaves well with a small sample size for all levels of alterations. To be more specific, it is more closed to the real change point compared with the other two methods [16, 27] especially for small shift sizes of parameter μ. Even if the results for larger shift sizes are slightly rougher than the others, the standard errors of the proposed method are much smaller, which means the performance is more stable and the obtained results are considerably consistent.

Table 3 Average change-point estimates and associated standard errors in Series 3 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	13.51	11.54	9.76	10.39	10.12
	$std(\tilde{\tau})$	3.78	2.37	1.43	0.9	0.39
Ghazanfari *et al.* [16]	$\tilde{\tau}$	10.38	10.04	10	10	10
	$std(\tilde{\tau})$	4.1	1.56	0.59	0.27	0.05
Proposed method	$\tilde{\tau}$	10	10.95	11.02	11.02	11
	$std(\tilde{\tau})$	3.55	1.15	0.43	0.16	0.03

4.2 Numerical application of a heavy-duty CNC machine tool

The case of a heavy-duty CNC machine tool from some factory is studied. The heavy-duty CNC machine tool is a typical kind of complex repairable system, it has the features such as small batch of production, long service life, and high maintenance cost. Therefore, the concern of the factory is focused on ensuring the reliability of the service period and reducing maintenance cost. To achieve this, in practice of engineering, a maintenance policy of 'minimal repair' is always implemented to minimize the impact of maintenance, burn-in is also a common mean to improve the reliability and maximize the useful life.

Table 4 Field data of TTF and TBF of the heavy-duty CNC machine tool (Unit: hour)

No.	TTF	TBF	No.	TTF	TBF	No.	TTF	TBF	No.	TTF	TBF
1	10	10	8	480.5	43.5	15	747.08	18.5	22	1831.17	375
2	178	160	9	523.5	38	16	848.58	99.5	23	1905.17	72
3	230	50	10	576.5	50	17	919.17	68.5	24	2060.17	149
4	276.5	44.5	11	618.5	40	18	1084.17	163	25	2336.17	262
5	297	17.5	12	649.5	29	19	1227.17	107.5	26	2429.67	71.5
6	385.5	77	13	683	24.5	20	1376.17	147	27	2548.67	111.5
7	435	42.5	14	727.08	42	21	1454.17	74	28	2745.17	182

Based on these facts, this case study aims to estimate the accurate change point of early failure period and random failure period in order to determine the optimal burn-in time of the product. Then take account of the effect of maintenance and the lack of failure data, our proposed method is particularly suitable for this situation. The field monitoring was conducted for 3000 hours since it initially finished production and came into early failure testing, the time to failure (TTF) data of the total 28 failures were collected and listed in Table 4.

Choose TBF as monitoring characteristic, the bootstrap control chart is constructed. Compared with the traditional Weibull probability plot, the SPC chart maintains a time sequence of TBFs and realizes real-time condition monitoring. As shown in Fig. 6, basically the first 15 data are below the LCL ignoring the randomness of the observed point 2 and point 6, it illustrates that failures happened frequently and the machine tool was still in the early failure phase. After the 15th data, observations mainly kept stable around the CL with no signal outside the control limit, which is interpreted as the state converted into the random phase with stable failure intensity.

However, the error caused by unsteadiness and uncertainty of samples, such as a sudden change at point 2 and point 6, may affect the accuracy and reliability of SPC monitoring. Then the sequential clustering approach is adopted to obtain the accurate transition point for the optimal clustering setting. Following the procedure in Section 3.2, the validity indices are calculated each time when the observation is above the LCL, and the proposed objective function turns out to be minimized at point 17, which has been marked with a red circle in Fig. 6.

As a result, the observations are divided into two optimal groups taking 17th point as the boundary. Substitute them into two sections of the proposed NHPP model as Eq. 21 and the MLEs of parameters are obtained as $\lambda_1 = 0.8045, \beta_1 = 0.3819, \lambda_2 = 5.7262 \times 10^{-6}, \beta_2 = 1.8648$.

Substitute the results above in Eq. 11 to calculate the change point t_0, which connects the early failure period and the random failure period. The result is $t_0 = 1.0161 \times 10^3 \, h$.
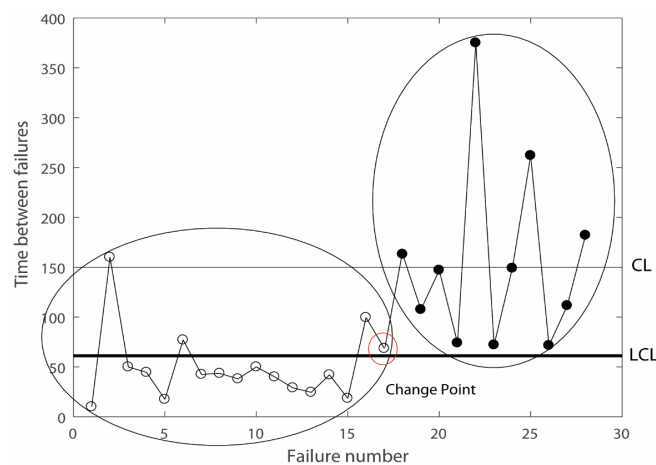

Fig. 6 Bootstrap control charts with the optimal clusters

Table 5 Definitions of goodness-of-fit indices

Index	Expression	Variable
K-S	$\max\left\{\max\limits_{1\leq i\leq M}(\dfrac{i}{M}-\widehat{U}_l),\max\limits_{1\leq i\leq M}(\widehat{U}_l-\dfrac{(i-1)}{M})\right\}$	L: The log-likelihood function,
A-D	$-\dfrac{\left\{\sum_{i=1}^{M}(2i-1)[\ln\widehat{U}_l+\ln(1-\widehat{U}_{M+1-i})]\right\}}{M}-M$	$t_1, t_2, \dots t_n$: The failure times, k: The number of parameters in the model, $M = n-1$,
C-V	$\dfrac{1}{12M}+\sum\limits_{i=1}^{M}[\widehat{U}_l-\dfrac{(2i-1)}{2M}]^2$	$\beta = \dfrac{n}{\sum_{i=1}^{n-1}\ln(\frac{t_n}{t_i})}$,
AIC	$-2\ln(L)+2k$	$\widehat{U}_l = (\frac{t_i}{t_n})^\beta, i=1,\cdots,M.$

Table 6 Comparing results of the goodness-of-fit test

Method	Change point	Parameter estimator	K-S	A-D	C-V	AIC
Original model	-	(0.0796, 0.7405)	0.1849 (0.20)	1.0727 (1.33)	0.2196 (0.23)	313.982
Proposed model	Phase 1 (1:17)	(0.8045, 0.3819)	0.1502 (0.28)	0.3964 (1.36)	0.0452 (0.24)	191.0966
	Phase 2(18:28)	(5.7262e-06, 1.8648)	0.184 (0.34)	0.3928 (1.39)	0.0489 (0.25)	160.5379

After obtaining the estimated change point and parameters corresponding to the optimal in-control and out-of-control clusters, the goodness-of-fit test is adopted. The Kolmogorov-Smirnov (K-S) statistic, the Anderson-Darling (A-D) statistic, the Cramér–von Mises (C-V) statistic, and the Akaike information criterion (AIC) are used as goodness-of-fit indices, which are defined in Table 5. If the calculated indices have smaller values, then the model is accepted as a better fitting result of the data. The results of the proposed two sectional NHPP model and the original NHPP model are listed in Table 6 for comparison. It's easy to tell that the values of three statistics and AIC of our new model are all smaller than the original model, the results indicate that our new model together with the change-point estimation approach can hence provide a better fit to the lifetime data, the estimated change point of the early failure period and the random failure period results in more consistent with the engineering practice, the practicality and accuracy of the proposed method are verified.

For the new products just involved in manufacturing work, they are more sensitive to the early failures, and their early phase in lifecycle contains a large proportion of failures as well, for the design, manufacturing or assembling reason. Burn-in is a popular process to eliminate early failures and improve the reliability. However, if the burn-in lasts too long, the useful life of products will be reduced. Otherwise, the reliability will decrease and the maintenance cost will increase. In practical execution, our proposed approach can present a simple and effective way to detect the changing trend of reliability and identify the optimal burn-in time, which will be of great value.

5. Conclusion

The paper proposed a change-point estimation method to approximate the failure process in lifecycle and estimated the change points of different failure periods for complex repairable systems. In the proposed method, a sectional model involving two NHPP functions was first developed to describe the bathtub-shaped failure intensity. Then, a bootstrap SPC chart was proposed in order to monitor changing trends of reliability and estimate the real time of change point. By introducing a sequential clustering approach in the control chart, the random interference caused by traditional SPC is eliminated and the calculation efficiency is improved. Three series of simulation considering the processes with different distributions, changing ranges and sampling schemes were conducted. The proposed method was proved to have a considerably good per-

formance, especially for the small shifts. To illustrate the use of the method in practice, a case study of a heavy-duty CNC machine tool mainly affected by early failures was also conducted. The feasibility and the accuracy of the method are verified. The estimated change point of early failure period and random failure period can give guidance for the identification of burn-in time.

The main advantages of the proposed method lie in its flexibility, accuracy and practical value for estimating the change point considering the characteristics of complex repairable systems. The developed model has a straightforward form to flexibly describe the bathtub-shaped failure pattern. The estimation approach has no limit to any specific distribution, it is convenient to achieve real-time monitoring of the gradual and sustained changing trend of system status. The change point estimator is accurate even if the sample size of failure data is small. However, our method is motivated by the specific field of complex repairable systems with certain failure patterns, its application scope requires further study and validation. The model could also be modified for estimating multiple change points, such as the point t_1 of random failure period and wear-out failure period (as shown in Fig. 1) in order to determine the optimal replacement time. In our future study, we will take different maintenance strategies into consideration, meanwhile, new compounding models for identifying multiple change points will also be investigated.

Acknowledgement

References

[1] Amiri, A., Allahyari, S. (2012). Change point estimation methods for control chart postsignal diagnostics: A litera-ture review, *Quality and Reliability Engineering International*, Vol. 28, No. 7, 673-685, doi: 10.1002/qre.1266.

[2] Pignatiello Jr, J.J., Samuel, T.R. (2001). Estimation of the change point of a normal process mean in SPC applica-tions, *Journal of Quality Technology*, Vol. 33, No. 1, 82-95, doi: 10.1080/00224065.2001.11980049.

[3] Perry, M.B., Pignatiello Jr, J.J. (2006). Estimation of the change point of a normal process mean with a linear trend disturbance in SPC, *Quality Technology & Quantitative Management*, Vol. 3, No. 3, 325-334, doi: 10.1080/168437 03.2006.11673118.

[4] Perry, M.B., Pignatiello Jr, J.J., Simpson, J.R. (2007). Change point estimation for monotonically changing Poisson rates in SPC, *International Journal of Production Research*, Vol. 45, No. 8, 1791-1813, doi: 10.1080/002075406 00622449.

[5] Hawkins, D.M., Zamba, K.D. (2005). Statistical process control for shifts in mean or variance using a change point formulation, *Technometrics*, Vol. 47, No. 2, 164-173, doi: 10.1198/004017004000000644.

[6] Zamba, K.D., Hawkins, D.M. (2006). A multivariate change-point model for statistical process control, *Techno-metrics,* Vol. 48, No. 4, 539-549, doi: 10.1198/004017006000000291.

[7] Holland, M.D., Hawkins, D.M. (2014). A control chart based on a nonparametric multivariate change-point model, *Journal of Quality Technology*, Vol. 46, No. 1, 63-77, doi: 10.1080/00224065.2014.11917954.

[8] Weese, M., Martinez, W., Megahed, F.M., Jones-Farmer, L.A. (2016). Statistical learning methods applied to pro-cess monitoring: An overview and perspective, *Journal of Quality Technology*, Vol. 48, No. 1, 4-24, doi: 10.1080/ 00224065.2016.11918148.

[9] Amiri, A., Niaki, S.T.A., Moghadam, A.T. (2015). A probabilistic artificial neural network-based procedure for variance change point estimation, *Soft Computing*, Vol. 19, No. 3, 691-700, doi: 10.1007/s00500-014-1293-x.

[10] Ahmadzadeh, F. (2009). Change point detection with multivariate control charts by artificial neural network, *The International Journal of Advanced Manufacturing Technology*, Vol. 97, No. 9-12, 3179-3190, doi: 10.1007/s001 70-009-2193-6.

[11] Ahmadzadeh, F., Lundberg, J., Strömberg, T. (2013). Multivariate process parameter change identification by neural network, *The International Journal of Advanced Manufacturing Technology*, Vol. 69, No. 9-12, 2261-2268, doi: 10.1007/s00170-013-5200-x.

[12] Maleki, M.R., Amiri, A., Mousavi, S.M. (2015). Step change point estimation in the multivariate-attribute process variability using artificial neural networks and maximum likelihood estimation, *Journal of Industrial Engineering International*, Vol. 11, No. 4, 505-515, doi: 10.1007/s40092-015-0117-7.

[13] Ghiasabadi, A., Noorossana, R., Saghaei, A. (2013). Identifying change point of a non-random pattern on $\bar{X}$ control chart using artificial neural networks, *The International Journal of Advanced Manufacturing Technology*, Vol. 67, No. 5-8, 1623-1630, doi: 10.1007/s00170-012-4595-0.

[14] Zarandi, M.H.F., Alaeddini, A. (2010). A general fuzzy-statistical clustering approach for estimating the time of change in variable sampling control charts, *Information Sciences*, Vol. 180, No. 16, 3033-3044, doi: 10.1016/j.ins.2010.04.017.

[15] Alaeddini, A., Ghazanfari, M., Nayeri, M.A. (2009). A hybrid fuzzy-statistical clustering approach for estimating the time of changes in fixed and variable sampling control charts, *Information Sciences*, Vol. 179, No. 11, 1769-1784, doi: 10.1016/j.ins.2009.01.019.

[16] Ghazanfari, M., Alaeddini, A., Niaki, S.T.A., Aryanezhad, M.-B. (2008). A clustering approach to identify the time of a step change in Shewhart control charts, *Quality and Reliability Engineering International*, Vol. 24, No. 7, 765-778, doi: 10.1002/qre.925.

[17] Kazemi, M.S., Bazargan, H., Yaghoobi, M.A. (2014). Estimating the drift time for processes subject to linear trend disturbance using fuzzy statistical clustering, *International Journal of Production Research*, Vol. 52, No. 11, 3317-3330, doi: 10.1080/00207543.2013.872312.

[18] Zarandi, M.H.F., Najafi, S. (2015). A type-2 fuzzy-statistical clustering approach for estimating the multiple change points in a process mean with monotonic change, *The International Journal of Advanced Manufacturing Technology*, Vol. 77, No. 9-12, 1751-1765, doi: 10.1007/s00170-014-6570-4.

[19] Kazemi, M.S., Kazemi, K., Yaghoobi, M.A., Bazargan, H. (2016). A hybrid method for estimating the process change point using support vector machine and fuzzy statistical clustering, *Applied Soft Computing*, Vol. 40, 507-516, doi: 10.1016/j.asoc.2015.11.021.

[20] He, S., Jiang, W., Deng, H. (2018). A distance-based control chart for monitoring multivariate processes using support vector machines, *Annals of Operations Research*, Vol. 263, No. 1-2, 191-207, doi: 10.1007/s10479-016-2186-4.

[21] Ebeling, C.E. (2004). *An introduction to reliability and maintainability engineering*, Tata McGraw-Hill Education, New Delhi, India.

[22] Jiang, R. (2015). *Introduction to quality and reliability engineering*, Springer, New York, USA, doi: 10.1007/978-3-662-47215-6.

[23] Jensen, W.A., Jones-Farmer, L.A., Champ, C.W., Woodall, W.H. (2006). Effects of parameter estimation on control chart properties: A literature review, *Journal of Quality Technology*, Vol. 38, No. 4, 349-364, doi: 10.1080/00224065.2006.11918623.

[24] Zhang, M., Megahed, F.M., Woodall, W.H. (2014). Exponential CUSUM charts with estimated control limits, *Quality and Reliability Engineering International*, Vol. 30, No. 2, 275-286, doi: 10.1002/qre.1495.

[25] Nichols, M.D., Padgett, W.J. (2006). A bootstrap control chart for Weibull percentiles, *Quality and Reliability Engineering International*, Vol. 22, No. 2, 141-151, doi: 10.1002/qre.691.

[26] Zhang, Y., Qiu, J., Liu, G., Zhao, Z. (2015). Fault sample generation for virtual testability demonstration test subject to minimal maintenance and scheduled replacement, *Mathematical Problems in Engineering*, Vol. 2015, Article ID 645047, doi: 10.1155/2015/645047.

[27] Samuel, T.R., Pignatiello Jr., J.J., Calvin, J.A. (1998). Identifying the time of a step change with $\overline{X}$ control charts, *Quality Engineering*, Vol. 10, No. 3, 521-527, doi: 10.1080/08982119808919166.

Multi-objective optimization for delivering perishable products with mixed time windows

Wang, X.P.[a,b,*], Wang, M.[b], Ruan, J.H.[c] , Li, Y.[a]

[a]Institute of Systems Engineering, Dalian University of Technology, Dalian, Liaoning, P.R. China
[b]School of Business, Dalian University of Technology, Panjin, Liaoning, P.R. China
[c]College of Economics and Management, Northwest A&F University, Yangling, Shanxi, P.R. China

A B S T R A C T

Perishable products generally have a short shelf life, and the freshness often depends on the postharvest time. The freshness of perishable products can ensure better customer satisfaction. Owing to the deterioration of perishable goods, the complexity of the corresponding vehicle routing problem (VRP) increases, because time delay will lead to serious costs. In this study, we are concerned with not only time-sensitive spoilage rates with mixed time windows, but also the delay costs in delivering perishable products. This study proposes a multi-objective VRP optimization model with mixed time windows and perishability (MO-VRPMTW-P) to minimize the distribution costs and maximize the freshness of perishable products. Then, in view of the fresh products orders space and time characteristics, we propose a heuristic algorithm (ST-VNSGA) composed of a variable neighbourhood search (VNS) method and a genetic algorithm (GA) considering the spatio-temporal (ST) distance to solve the complex multi-objective problem. The solution algorithms are evaluated through a series of experiments. We illustrate the performance and efficiency comparisons of ST-VNSGA with the method without spatio-temporal strategy algorithm and NSGA-II algorithm. It is demonstrated that the proposed ST-VNSGA algorithm can lead to a substantial decrease in the computation time and major improvements in solutions quality, thus revealing the efficiency of considering the spatio-temporal strategy with mixed time windows.

1. Introduction

The distribution of perishable products can be abstracted as a vehicle routing problem (VRP) [1]. It has been recognized that managing perishable products distribution, such as the distribution of vegetables, milk, meat, and flowers, is a difficult problem [2]. Perishable products generally have a short lifecycle, and the value or quality of perishable products decreases rapidly once they are produced, they will continue to decay while being delivered [3]. The life of perishable products depends on time. However, perishable products freshness affects customer satisfaction [4]. Owing to deterioration of perishable goods, distributors are increasingly adopting delivery strategies to fulfil their orders. Timely delivery of perishable food affects not only the delivery operator's cost [5], but also the satisfaction of customers. Furthermore, with the recent rapid development in fresh e-commerce, the characteristics of orders tend to be of smaller lot-size and higher frequency, which has increased the complexity of distribution problems.

However, traditional production scheduling focuses on minimizing the transportation cost of perishable goods without explicitly considering the time-sensitive freshness of perishable products and the serious delay costs in mixed time windows [6-9]. In particular, in fresh e-commerce, the delivery of perishable products is different from traditional transportation, and creates new challenges. In the literature concerned explicitly with freshness, Chen *et al.* [10] proposed a non-linear mathematical model to consider VRP with time windows (VRPTW) for perishable food products. The objective of this model was to maximize the expected total profit of the supplier. Osvald and Stirn [11] formulated a VRPTW with time-dependent travel-times (VRPTWTD). The model considered the impact of perishability as part of the overall distribution costs. Hsu *et al.* [12] focused on the randomness of the perishable food delivery process and presented a stochastic VRPTW model to obtain optimal delivery routes. These papers mainly considered the impact of perishability affects the overall distribution costs, however, they didn't take into account customer satisfaction with regard to freshness.

Some scholars [13-19] researched perishable products from the supply chain network perspective. For example, Amorim *et al.* [15] explored a multi-objective framework that integrated production and distribution perishable goods planning problems. However, most studies have concentrated on conventional VRPs, paying little attention to designing an effective method combined with time-sensitive freshness and delay costs for the multi-objective perishable good distribution problem. Meanwhile, different fresh delivery orders have specific space and time characteristics. It is therefore critical to design an effective and efficient delivery method so that the supplier can ensure the freshest products are delivered in a cost-effective and timely manner.

In this study, we both consider vehicle time-dependent travel and perishable products freshness under mixed time window constraints, where freshness has a minimum level that the customer can accept. We establish a multi-objective VRP optimization model with mixed time windows and perishability (MO-VRPMTW-P) to minimize the distribution costs and maximize the freshness of perishable products. We consider the time-sensitive freshness of perishable products, and the high serious delay cost in mixed time windows. Thus, a company can reduce costs and achieve a higher level of customer satisfaction with regard to freshness. Previous algorithms mainly considered the customer spatial location relationships, but did not take time and space characteristics constraints into account. Then, considering the obvious space and time characteristics of the fresh food distribution task, we design a heuristic algorithm (ST-VNSGA) composed of a local variable neighbourhood search algorithm (VNS) and a global genetic algorithm (GA) that considers spatio-temporal distance to solve this multi-objective problem, hence, provide the decision maker with a whole set of equally efficient solutions.

The remainder of this paper is organized as follows. In Section 2, we establish an optimization model to minimize the total costs and maximize the freshness of the products with mixed time windows. The solution of the model and a suitable algorithm are presented in Section 3. In Section 4, several examples are tested to verify the effectiveness and efficiency of the proposed model and algorithms for MO-VRPMTW-P. Finally, the paper concludes with a summary and an outlook on further research topics in Section 5.

2. Problem statement and mathematical model

2.1 Problem statement

A perishable product distribution network is a complex system that presents many challenges. We propose a mathematical model the MO-VRPMTW-P that considers the time-sensitive spoilage rates of perishable products, aiming at achieving the following objectives:

- Minimum total costs, which contain fixed costs, transportation costs, damage costs and penalty costs;
- Maximum the freshness of perishable products.

We assume that the distribution system includes a distribution centre and multiple customers. All vehicles must leave and return to the distribution centre. The distribution centre has

sufficient capacity to complete all tasks. The transportation costs between the customers depend on travel distance. Each vehicle can travel at most one route per time period. Each customer can be served by only one vehicle. The demand and time window of the customers are known. These perishable products need to be sent to customers. If not, this usually means higher costs for the operator, who must pay a penalty for the loss. We consider only the transit time, ignoring loading and unloading time. Each vehicle has a capacity constraint.

2.2 Mathematical model

The optimization model for the MO-VRPMTW-P is presented to analysis the proposed problem. The parameters and variables of the models are defined in Table 1. Table 2 shows the list of abbreviations.

Table 1 The parameters and variables of the models

Parameters	Explanation of the parameters
N	A set of customer notes, $N = \{n\|n = 1,2, \dots, \|N\|\}$ represents customers
N^+	A set of depot and customer notes, $N^+ = \{n\|n = 1,2, \dots, \|N\|\}, n = 0$ represents depot
K	A set of vehicles, $K = \{k\|k = 1,2, \dots, \|K\|\}$
$[e_i, l_i]$	e_i and l_i are the starting and ending time of the time window at customer i, respectively, $i \in N$
Q	The maximum capacity of the vehicle
d_{ij}	The distance between node i and node j, $i, j \in N^+, i \neq j$
q_i	The demand of the customer i that vehicle k service, $i \in N$
r_0	The minimum freshness level that the customer can accept
T	The life cycle of perishable products
v	The speed of vehicle k
f	The unit fixed costs of the vehicle
c_0	The average cost of travelling
y_{ik}	Binary that takes the value 1 if customer i is assigned to vehicle k
t_{ij}	The transportation time between node i and node j, $i, j \in N^+, i \neq j$
t_{ik}	The time vehicle k arrives at customer i, $i \in N$
$r(t_{ik})$	The perishable product freshness level when customer i is serviced by vehicle k
x_{ijk}	Binary that takes the value 1 if vehicle k transports between node i and node j. $i, j = 1,2, \dots, n$

Table 2 The list of abbreviations

Abbreviation	Full names
MO-VRPMTW-P	A multi-objective VRP optimization model with mixed time windows and perishability
ST-VNSGA	A heuristic algorithm composed of a local variable neighbourhood search algorithm and a global genetic algorithm that considers spatio-temporal distance
VNSGA	A heuristic algorithm composed of a local variable neighbourhood search algorithm and a global genetic algorithm
ST-NSGA-II	A heuristic algorithm that NSGA-II algorithm considers spatio-temporal distance

Analysis of objective functions

The total costs Z_1 are composed of fixed costs C_1, transportation costs C_2, damage costs C_3, and penalty costs P. To describe the characteristics of perishable products, we cite the freshness factor from the literature [20]. The objectives specific expressions are described as follows:

- Fixed costs and transportation costs

The vehicle fixed costs C_1 generally include the depreciation costs, maintenance costs, and so on. Transportation costs C_2 depend on the distance vehicle travelled. They are expressed as below:

$$C_1 = \sum_{j \in N \setminus \{0\}} \sum_{k \in K} f x_{0jk} \tag{1}$$

$$C_2 = \sum_{(i,j) \in E} c_0 d_{ij} \sum_{k \in K} x_{ijk} \tag{2}$$

- Damage costs

The loss of quality in the transportation of the perishable products is a significant cost for companies. The product quality decay influences network designs [21-22]. We select perishable milk as the main study object. The goods have specific time storage characteristics. The rate of corruption is mainly exponential with time [20, 23-25], $r = e^{-\phi t}$, where r is the freshness of perishable product quality, and ϕ is the shrinkage factor, which is related to the type of goods. If the goods are sensitive to time, the value is relatively small; otherwise, the value is large. The exponential damage percentage changes over time.

Thus, the perishable product damage costs can be expressed as:

$$C_3 = w \sum_{i \in N} \sum_{k \in K} q_{ik}(1 - r(t_{ik})) \, y_{ik} \tag{3}$$

w represents the unit value of the perishable products. $r(t_{ik})$ is the freshness of customer i, the calculation expressions are: $r(t_i) = e^{-\phi(t_{ik} - t_{0k})}$. t_{0k} is the start time of vehicle k from the distribution centre, $k = 1, 2, \ldots, |K|$.

- Mixed time window penalty costs

To improve the quality of distribution services, customers require the distributor arrives within specified time windows; if not, they need to pay the corresponding wait or delay costs. The quality of fresh goods is sensitive to time. Any vehicle that arrives early has to wait until the beginning of the time windows. Any vehicle that arrives late will incur a penalty, and the delay costs of damaged goods are serious. Therefore, this study uses mixed time windows to measure fresh good distribution penalty costs. Thus, the penalty function can be shown in the formula:

$$P_i(t_{ik}) = \begin{cases} \alpha q_{ik}(t_{ik} - LT_i) & T_{ie} < t_{ik} < ET_i \\ 0 & ET_i < t_{ik} < LT_i \\ q_{ik}(t_{ik} - LT_i)^\beta & LT_i < t_{ik} < T_{il} \\ q_{ik} & T_{ie} > t_{ik} \ , t_{ik} > L_i \end{cases} \tag{4}$$

$P_i(t_{ik})$ represents penalty costs if vehicle k transgresses the time window of customer i. The lower bound T_{ie} represents the earliest arrival time that a customer can endure when a service starts earlier than ET_i. Similarly, the upper bound T_{il} represents the latest arrival time that the customer can endure when the service starts later than LT_i. Here, α represents the penalty coefficient that is within the actual time but earlier than the optimal satisfactory time and β represents the penalty coefficient that is within the actual time but later than the optimal satisfactory time. $\alpha > 1, \beta > 1$. The corresponding penalty function is shown in Fig.1.
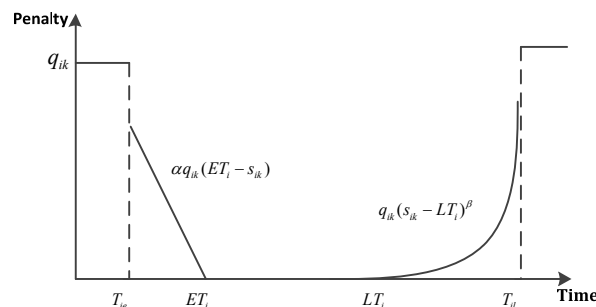

Fig. 1 Relationship between arrival time, time-windows and penalty cost

- Average freshness

The products average freshness is related to the quality and freshness factor. Thus, the average freshness is defined as:

$$Z_2 = \sum_{i \in N} \sum_{k \in K} q_{ik} r(t_{ik}) y_{ik} \bigg/ \sum_{i=0}^{N} q_{ik} \tag{5}$$

Then, the mathematical model MO-VRPMTW-P can be given as follows. Objective functions:

$$maxZ_1 = \sum_{j \in N \setminus \{0\}} \sum_{k \in K} f x_{0jk} + \sum_{(i,j) \in E} c_0 d_{ij} \sum_{k \in K} x_{ijk} + w \sum_{i \in N} \sum_{k \in K} q_{ik} \left(1 - e^{-\phi(t_{ik}-t_{0k})}\right) y_{ik}$$
$$+ \sum_{(i \in N)} P_i(t_{ik}) \tag{6}$$

$$minZ_2 = \sum_{i \in N} \sum_{k \in K} q_{ik} e^{-\phi(t_{ik}-t_{0k})} y_{ik} \Bigg/ \sum_{i=0}^{N} q_{ik} \tag{7}$$

Constraint conditions:

$$\sum_{j \in N} \sum_{k \in K} x_{0jk} \leq K \tag{8}$$

$$\sum_{j \in N} x_{0jk} = \sum_{j \in N} x_{j0k} \leq 1, \quad \forall k \in K \tag{9}$$

$$\sum_{i \in N^+} \sum_{j \in N^+} x_{ijk} q_i \leq Q, \quad \forall k \in K \tag{10}$$

$$\sum_{i \in N^+} \sum_{k \in K} x_{ijk} = 1, \quad \forall j \in N \tag{11}$$

$$\sum_{i \in N^+} \sum_{k \in K} x_{jik} = 1, \quad \forall j \in N \tag{12}$$

$$r(t_{ik}) = e^{-\phi(t_{ik}-t_{0k})}, \quad \forall i \in N, k \in K \tag{13}$$

$$r(t_{ik}) \geq r_0, \quad \forall i \in N \tag{14}$$

$$t_{jk} = (t_{ik} + d_{ij}/v_0)x_{ijk}, \quad \forall i,j \in N^+, k \in K \tag{15}$$

$$x_{ijk} \in \{0,1\}, \quad i,j \in N^+, k \in K \tag{16}$$

In the first objective function Eq. 6, total costs are minimized, namely: the fixed costs, transportation costs, damage costs, and penalty costs. In the second objective function Eq. 7, the average remaining freshness of products to be delivered is maximized. Eq. 8 is the number of vehicles constraint. Eq. 9 states that each vehicle should leave and return to distribution centre. Eq. 10 is a vehicle capacity constraint. Eqs. 11 and 12 represent that each customer should be serviced, and each can only be serviced once time. Specially, Eq.13 defines freshness function of the perishable products, and Eq. 14 ensures the lowest level of freshness that the customer can accept. Eq.15 defines the time that vehicle k takes from leaving customer i to arrival at customer j. Eq. 16 represents that vehicle k serves customer i before customer j.

3. Used methods

The MO-VRPMTW-P problem is an NP-hard problem. Multiple objectives need to be optimized at the same time. In the combination optimization problem, GA is an efficient global optimal algorithm and VNS is an efficient local search algorithm. In this work, we combine the improved GA with the VNS algorithm. Traditional algorithms mainly consider the customer spatial location relationship; however, they rarely take orders the time and space characteristic constraints of orders into consideration. Since the orders have obvious ST characteristics, we use the k-means method to cluster the nodes to obtain the initial solution considering the ST strategy in the first stage. Then, in the second stage, we adopt VNSGA to optimize the distribution route.

3.1 Generate initial solution

Calculating spatio-temporal distance

In the process of delivery, each perishable order has a corresponding demand. Considering orders with spatio-temporal distance may solve the problem more effectively than just considering the distance. So, we use the definition of ST distance from the literature [26] to cluster orders.

Use D_{ij}^{ST} to denote the ST distance. D_{ij}^{S}, D_{ij}^{T} represent the Euclidean distance and temporal distance between customer i and j, respectively. The transportation time of the points is related to the Euclidean distance, which means $D_{ij}^{S} = t_{ij}$. Here, $[a, b]$ and $[c, d]$ are the time windows of customer i and j, the specific arrival time at customer j is $t' \in (a', b')$, $a' = a + t_{ij}$, $b' = b + t_{ij}$. Use $Sav_{ij}(t')$ to denote the saved time when vehicle arrives at customer j at the moment t'. Here, A is the maximum window, and K_1, K_2, K_3 are parameters related to time.

$$Sav_{ij}(t') = \begin{cases} k_2 t' + k_1 d - (k_1 + k_2)c & t' < c \\ -k_1 t' + k_1 d & c \leq t' \leq d \\ -k_3 t' + k_3 d & t' > d \end{cases} \qquad (17)$$

A greater $Sav_{ij}(t')$ means a smaller spatial distance. $D_{ij}^{T}(t')$ is defined as:

$$D_{ij}^{T}(t') = k_1 A - Sav_{ij}(t') \quad t' \in (a', b') \qquad (18)$$

Temporal distance $\overline{D_{ij}}$ is defined as:

$$\begin{aligned} \overline{D_{ij}^{T}} &= \int_{a'}^{b'} D_{ij}^{T}(t')dt'/(b' - a') \\ &= k_1 A \\ &\quad - \int_{\min(a',c)}^{\min(b',c)} (k_2 t' + k_1 d - (k_1 + k_2)c)dt' \\ &\quad + \int_{\min(\max(a',c),d)}^{\max(\min(b',d),c)} (-k_1 t' + k_1 d)dt' \\ &\quad + \int_{\min(a',d)}^{\min(b',d)} (k_3 t' + k_3 d)dt'/(b' - a') \end{aligned} \qquad (19)$$

We take the maximum distance as the temporal distance.

$$D_{ij}^{T} = \max(\overline{D_{ij}^{T}}, \overline{D_{ji}^{T}}) \qquad (20)$$

The ST distance D_{ij}^{ST} is related to D_{ij}^{S} and D_{ij}^{T}. α_1, α_2 are the D_{ij}^{S} and D_{ij}^{T} weight coefficients, respectively. $\alpha_1 + \alpha_2 = 1$. The ST distance can be expressed as follow:

$$D_{ij}^{ST} = \alpha_1 \left(\frac{D_{ij}^{S} - \min\limits_{m,n \in C, m \neq n}(D_{mn}^{S})}{\max\limits_{m,n \in C, m \neq n}(D_{mn}^{S}) - \min\limits_{m,n \in C}(D_{mn}^{S})} \right) + \alpha_2 \left(\frac{D_{ij}^{T} - \min\limits_{m,n \in C}(D_{mn}^{T})}{\max\limits_{m,n \in C, m \neq n}(D_{mn}^{T}) - \min\limits_{m,n \in C}(D_{mn}^{T})} \right) \qquad (21)$$

Construct initial solution

After the calculation of the ST distance, we apply k-means method to cluster the orders and construct initial solution. Here, k is the number of vehicles. The orders are divided into k clusters, and the clustering centre $z_i(z_1, z_2, \ldots z_k)$ of each cluster is o_i. The cluster k is defined as follow:

$$Min \sum_{j=1}^{k} \sum_{i \in z_i/\{o_i\}} D_{io_j}^{ST} \qquad (22)$$

$k = max \sum_{i \in N} d_i/Q$. $\sum d_i$ represents the total demand of the largest distribution order, and $D(i, o_j)$ represents the distance of the sub-order i to the cluster centre z_i.

3.2 Optimization solution based on VNSGA

We combine the improved GA and the VNS to solve the multi-objective problem. Owing to the different mechanisms used in population search and local search, we use two different fitness functions in the selection operations. We apply non-dominance ranking and crowding distance sorting in the NSGA-II method as a global strategy and adopt an external archive to maintain the process of evolution. Then, VNS realizes the dynamic search. The flow chart is shown in Fig. 2.
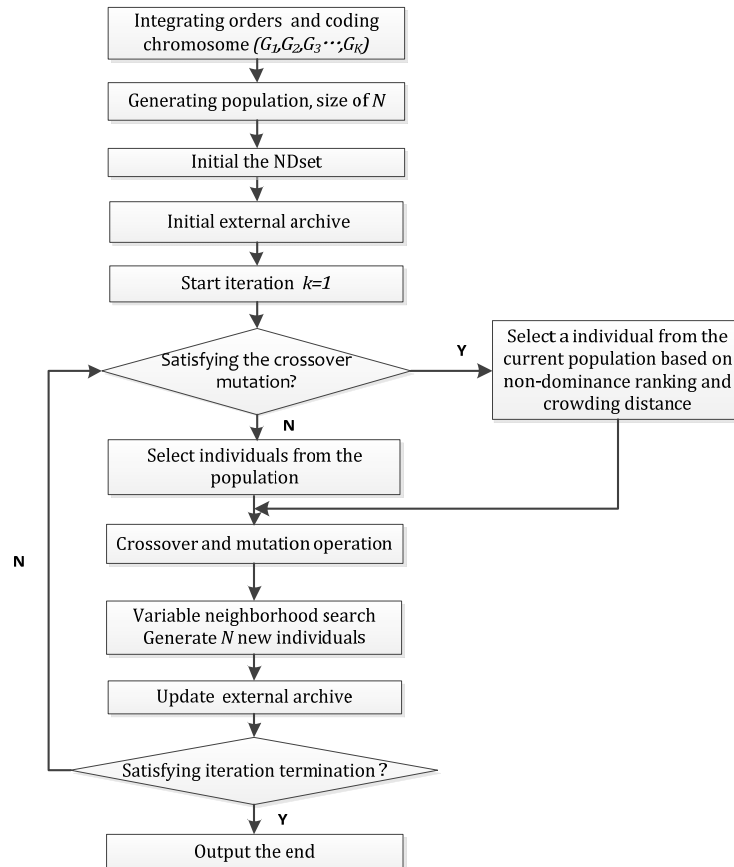

Fig. 2 Flowchart of the solution algorithm

Fitness function

We adopt two different fitness functions in population and local search operation selection.

First, in the population selection operation, we propose the ranking and crowding degree method based on the Pareto dominance. $fit_1(x)$ is the population selection fitness. Use $rank(x)$ to denote the relationship of Pareto dominance, $rank(x) < rank(y)$ indicates that x dominates y. The crowding distance is $Crowding - distance(x)$. The first keyword is ascending in a particular order and the second key sort is descending according to the crowding distance.

$$fit_1(x) = (rank(x), Crowding - distane(x)) \tag{23}$$

Second, in the local search selection operation, we select a better solution from the neighbourhood. $S(x)$ is the number of solutions x dominated. $W(x)$ is the number of solutions x dominate in storage pool. $fit_2(x)$ is the local search selection operation fitness.

$$fit_2(x) = \frac{1 + S(x)}{1 + W(x)} \tag{24}$$

4. Results and discussion

4.1 Data description

In this section, we use numerical experiments to demonstrate the efficiency and advantages of applying our heuristic algorithm. We adopt the instances developed by Solomon for VRPTW. Six different instance types are considered: R1, R2, C1, C2, RC1, and RC2.

Suppose the distributed perishable product is milk. Set the shrinkage factor as $\phi = 1/200$. $\alpha = 1.5, \beta = 1.5, \alpha_1 = 0.5, \alpha_2 = 0.5$. These experiments were performed on a personal computer with Intel® CoreTM i5-4460 CPU at 2.40 GHz and 8.00 GB of RAM. The computation run time unit is seconds. The stopping criterion is set to $Max_t = 300$. The parameters are listed in Table 3.

Table 3 Parameters of the experiments.

Parameter	Meaning	Value
v	The speed of the vehicle (km/h)	30
T	The lifecycle of the perishable product (h)	24
f	The unit fixed costs per of the vehicle (yuan)	50
Q	The maximum capacity of the vehicle (kg)	300
c_0	The average cost of travelling (yuan/km)	2.5
r_0	The minimum freshness level that the customer can accept	0.75
w	The unit value of the perishable products(yuan/kg)	30
K	A set of vehicles	50

4.2 Effectiveness of considering spatio-temporal distance

Consider the time and space characteristics of the order, we propose a method based on the spatio-temporal metrics to verify the strategy of spatio-temporal distance. We define the method that considers the customer spatio-temporal location relationship as ST-VNSGA and the method that does not consider the customer spatial location as VNSGA. Comparison results between ST-VNSGA and VNSGA are given in Table 4.

Table 4 Comparison results between ST-VNSGA and VNSGA

| Case	ST-VNSGA			VNSGA			Gap		
	Time	Cost	Freshness	Time	Cost	Freshness	Time	Cost	Freshness
	(s)	(yuan)	(%)	(s)	(yuan)	(%)	(%)	(%)	(%)
R101_25	43	771.58	93.4	73	791.24	92.1	41.10	2.48	1.41
R101_50	79	1403.31	91.7	144	1449.73	88.6	45.14	3.20	3.50
R101_100	127	2649.43	88.1	278	2784.37	85.7	54.31	4.84	2.80
C101_25	22	185.36	94.5	26	191.11	92.9	15.38	3.00	1.72
C101_50	44	367.98	92.8	54	379.45	91.7	18.52	3.02	1.20
C101_100	79	730.91	92.3	99	758.69	90.0	20.2	3.66	2.56
RC101_25	29	799.72	93.3	49	819.43	91.6	40.82	2.41	1.86
RC101_50	52	1360.98	91.4	94	1398.21	89.0	44.68	2.66	2.70
RC101_100	91	2556.67	87.5	189	2678.85	84.9	51.85	4.56	3.06
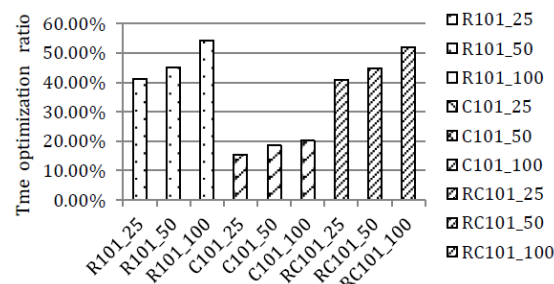
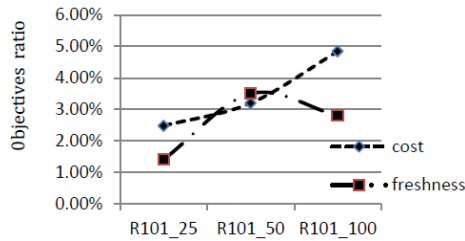**Fig. 3** Time optimization rate with the spatio-temporal strategy
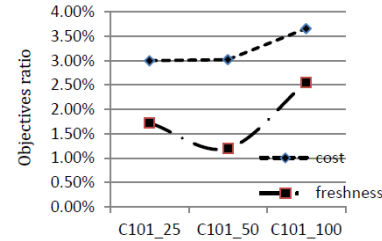
Fig. 4 R class case objectives ratio
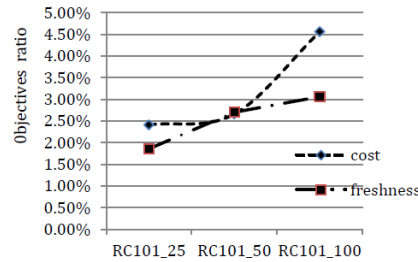

Fig. 5 C class case objectives ratio

Fig. 6 RC class case objectives ratio

Table 4 lists the computation results of ST-VNSGA and VNSGA with regard to run time, cost and freshness, as well as the optimization gap. Fig. 3 displays the optimization rate of the run time with regard to R, C, and RC classes. As can be seen from Table 4, ST-VNSGA can obtain a better solution in less time compared to VNSGA. For example, the ST-VNSGA run times of R101_25, R101_50, and R101_100 are 43, 79, and 127, respectively. The VNSGA run times are 73, 144, and 278, and the improvement rates of run time are 41.10 %, 45.14 %, and 54.31 %, respectively. The ST strategy can add the customer to the path where the distance is as close to the customer as possible. Thus, the ST strategy can both effectively reduce the search scope and reach a better solution faster.

Fig. 4 shows R class optimization rate in cost and freshness. Similarly, Fig. 5, Fig. 6 show the results for the C class and RC class, respectively. Conclude from Table 4 and Figs. 4–6 , the cost and freshness of ST-VNSGA-P are also optimized to some extent; for example, the cost of R101_100 reduces to 4.84 % and the freshness of RC101_100 increases to 3.06 %, which proves that the considered spatio-temporal approach has certain guiding significance.

At the same time, Table 4 and Fig. 3 indicate that the proposed algorithm run time is reduced greatly compared to VNSGA, especially in R and RC problems, where the run time is reduced by more than 50 %. R class problem optimization rates are 41.10 %, 45.14 %, and 54.31 %, RC class problem optimization rates are 40.82 %, 44.68 %, and 51.85 %, respectively. The customer points in C class almost appear as an aggregated distribution. The improvement in spatio-temporal distance is not obvious. However, the points in R and RC class randomly spread, so the efficiency of ST-VNSGA solution significantly improved. Thus, the strategy proposed in this study is more suitable for the optimization of the disperse region.

Meanwhile, we can see from Fig. 3 that considering spatio-temporal distance has good potential in solving large-scale VRPs. For example, RC class problem run time optimization rates are 40.82 %, 44.68 %, and 51.85 %, all successively increasing. Therefore, the ST strategy optimizes obviously effect on large-scale VRPs.

To validate the ST strategy optimization effect on the fresh product distribution model and algorithm, we compared without ST strategy (VNSGA) with ST-VNSGA in different order environments. Six numerical examples were created for testing. The impact of the ST strategy on run time and objective values is listed in Table 5.

As indicated by Table 5, comparing the run time rate R101_100 (54.31 %) with R201_100 (46.15 %), C101_100 (20.2 %) with C201_100 (15.46 %), and RC101_100 (51.85 %) with RC201_100 (42.15 %), ST-VNSGA excels VNSGA algorithm in run time optimization, especially with the narrow time windows. In VNSGA, clustering and optimization proceed at the same time. ST-VNSGA optimizes the procedure after clusters. The strict time window constraint interferences the progress clustering. The larger time windows are, the constraints are smaller.

Table 5 Comparison results of different orders environment

| Case	ST-VNSGA			VNSGA			Gap		
	Time (s)	Cost (yuan)	Freshness (%)	Time (s)	Cost (yuan)	Freshness (%)	Time (%)	Cost (%)	Freshness (%)
R101_100	127	2649.43	88.1	278	2784.37	85.7	54.31	4.84	2.80
R201_100	161	1756.43	85.8	299	1802.98	83.1	46.15	2.58	3.25
C101_100	79	730.91	92.3	99	758.69	90.0	20.2	3.66	2.56
C201_100	82	368.84	90.1	97	380.16	88.8	15.46	2.98	1.46
RC101_100	91	2556.67	87.5	189	2678.85	84.9	51.85	4.56	3.06
RC201_100	199	1794.49	84.1	344	1934.89	80.5	42.15	7.26	4.47

We conclude that time windows have a major impact on the delivered perishable products; moreover, the cost and freshness of ST-VNSGA increase at different rates; for example, the cost of R201_100 decreases by 7.26 % at most, and the freshness of RC201_100 increases by 4.47 %.

4.3 Effectiveness of ST-VNSGA

For a better analysis of the effectiveness of the proposed heuristic algorithm combined GA and VNS, we compare the NSGA-II algorithm (ST-NSGA-II) with ST-VNSGA. Comparison results between ST-VNSGA and ST-NSGA-II are provided in Table 6.

Table 6 Comparison results between ST-VNSGA and ST-NSGA-II

Case	ST-VNSGA			VNSGA			Gap		
	Time (s)	Cost (yuan)	Freshness (%)	Time (s)	Cost (yuan)	Freshness (%)	Time (%)	Cost (%)	Freshness (%)
R101_25	43	771.58	93.4	38	789.35	91.9	13.16	2.25	1.63
R101_50	79	1403.31	91.7	69	1427.97	88.9	14.49	1.7	3.15
R101_100	127	2649.43	88.1	98	2801.66	86.3	29.59	5.43	2.09
C101_25	22	185.36	94.5	19	189.26	93.8	15.79	2.06	0.75
C101_50	44	367.98	92.8	37	383.81	90.1	18.91	4.12	3.00
C101_100	79	730.91	92.3	65	762.49	88.9	21.54	4.14	3.82
RC101_25	29	799.72	93.3	23	811.18	90.1	26.09	1.41	3.55
RC101_50	52	1360.98	91.4	43	1451.15	87.4	20.93	6.21	4.58
RC101_100	91	2556.67	87.5	72	2720.21	83.1	26.39	6.01	5.29
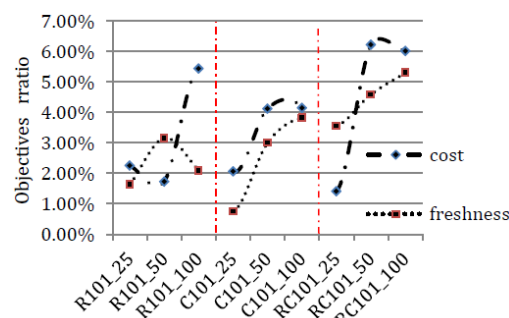
Fig. 7 Objective ratio between ST-VNSGA and ST-NSGA-II under different order environments

We can further derive the validity of ST-VNSGA in Table 6. Result shows that for the run time in R, C, and RC classes, ST-VNSGA takes on an increasing format, for example, R101_25 (13.16 %), R101_50 (14.49 %), and R101_100 (29.59 %). Because the VNS algorithm needs to search mass neighbourhood structures, which increases the run time dramatically, and the ranking based on Pareto dominance further increases the operation time. However, we can see that ST-VNSGA ameliorates the quality of satisfactory solutions; for example, the cost optimal rate of RC101_100 is 6.01 %, the freshness optimal rate of RC101_100 is 5.29 %, indicating that the local search strategy has excellent optimization in seeking the best solution. ST-VNSGA has an

advantage in solving multiple objectives, which shows that the proposed algorithm is effective and efficient.

The objective value improves between ST-VNSGA and ST-NSGA-II with regard to cost and freshness, as shown in Fig. 7. In terms of the algorithm solving efficiency, the ST-VNSGA has a significant advantage, especially with the growing number of customers.

4.4 Contrastive analysis of results

According to the contrastive analysis of optimization results, we have found several findings:

- Compared with the method using spatial clustering, the strategy that considers the spatio-temporal distance distribution can achieve better solutions in a shorter period of time, especially in the medium or large-scale distribution problem where it can reach 54.31 % (Table 4, R101_100). This also shows that the algorithm has good potential in solving large-scale VRPs. The strategy proposed in this paper, ST-VNSGA, has obvious advantages for dispersed customer distributions. The results also provide effective decision support to solve the fresh distribution practical problem.
- In fresh product distribution, the heuristic algorithm calculation run time with narrow time window constraints is better than the algorithm based on spatial distance in accordance with the actual needs of customers. In addition, it can save the cost of logistics and improve the service with regard to freshness to a certain extent.
- From the results (Table 6), we can see that ST-VNSGA ameliorates the quality of satisfactory solutions, has excellent optimization in seeking the best solution, and has advantages in solving the multiple objectives, especially with the growing number of customers, with regard to cost and freshness, which show that ST-VNSGA is effective and efficient.

5. Conclusion

In this study, we established the MO-VRPMTW-P model to minimize the distribution costs and maximize the freshness of perishable products. We considered the time-sensitive freshness of perishable products and the high cost of delay in mixed time windows. Then, in view of the fresh product order space and time characteristics, we designed a heuristic algorithm that considers spatio-temporal distance (ST-VNSGA) to solve the fresh product distribution problem. Several numerical examples were presented to demonstrate the effectiveness and efficiency of the proposed algorithm. It was demonstrated that these algorithms can lead to a substantial decrease in run time and major improvements in solution quality, which reveals the importance of considering a spatio-temporal strategy with mixed time windows.

It is worth noting that some areas require improvement; for example, we will focus on interference management in the process of perishable product distribution in the next step.

Acknowledgement

References

[1] Hwang, H.-S. (1999). A food distribution model for famine relief, *Computers & Industrial Engineering*, Vol. 37, No. 1-2, 335-338, doi: 10.1016/S0360-8352(99)00087-X.

[2] Teunter, R.H., Flapper, S.D.P. (2003). Lot-sizing for a single-stage single-product production system with rework of perishable production defectives, *OR Spectrum*, Vol. 25, No. 1, 85-96, doi: 10.1007/s00291-002-0105-3.

[3] Amorim, P., Almada-Lobo, B. (2014). The impact of food perishability issues in the vehicle routing problem, *Computers & Industrial Engineering*, Vol. 67, 223-233, doi: 10.1016/j.cie.2013.11.006.

[4]   Song, B.D., Ko, Y.D. (2016). A vehicle routing problem of both refrigerated- and general-type vehicles for perishable food products delivery, *Journal of Food Engineering*, Vol. 169, 61-71, doi: 10.1016/j.jfoodeng.2015.08.027.

[5]   Hu, H., Zhang, Y., Zhen, L. (2017). A two-stage decomposition method on fresh product distribution problem, *International Journal of Production Research,* Vol. 55, No. 16, 4729-4752, doi: 10.1080/00207543.2017.1292062.

[6]   Tarantilis, C.D., Kiranoudis, C.T. (2001). A meta-heuristic algorithm for the efficient distribution of perishable foods, *Journal of Food Engineering*, Vol. 50, No. 1, 1-9, doi: 10.1016/S0260-8774(00)00187-4.

[7]   Tarantilis, C.D., Kiranoudis, C.T. (2002). Distribution of fresh meat, *Journal of Food Engineering*, Vol. 51, No. 1, 85-91, doi: 10.1016/S0260-8774(01)00040-1.

[8]   Zhang, G., Habenicht, W., Spieß, W.E.L. (2003). Improving the structure of deep frozen and chilled food chain with tabu search procedure, *Journal of Food Engineering*, Vol. 60, No. 1, 67-79, doi: 10.1016/S0260-8774(03)00019-0.

[9]   Faulin, J. (2003). Applying MIXALG procedure in a routing problem to optimize food product delivery, *Omega*, Vol. 31, No. 5, 387-395, doi: org/10.1016/S0305-0483(03)00079-3.

[10] Chen, H.-K., Hsueh, C.-F., Chang, M.-S. (2009). Production scheduling and vehicle routing with time windows for perishable food products, *Computers & Operations Research*, Vol. 36, No. 7, 2311-2319, doi: 10.1016/j.cor.2008.09.010.

[11] Osvald, A., Zadnik Stirn, L. (2008). A vehicle routing algorithm for the distribution of fresh vegetables and similar perishable food, *Journal of Food Engineering*, Vol. 85, No. 2, 285-295, doi: 10.1016/j.jfoodeng.2007.07.008.

[12] Hsu, C.-I., Hung, S.-F., Li, H.-C. (2007). Vehicle routing problem with time-windows for perishable food delivery, *Journal of Food Engineering*, Vol. 80, No. 2, 465-475, doi: 10.1016/j.jfoodeng.2006.05.029.

[13] Ruan, J., Shi, Y. (2016). Monitoring and assessing fruit freshness in IOT-based e-commerce delivery using scenario analysis and interval number approaches, *Information Sciences*, Vol. 373, 557-570, doi: 10.1016/j.ins.2016.07.014.

[14] Rong, A., Akkerman, R., Grunow, M. (2011). An optimization approach for managing fresh food quality throughout the supply chain, *International Journal of Production Economics*, Vol. 131, No. 1, 421-429, doi: 10.1016/j.ijpe.2009.11.026.

[15] Amorim, P., Günther, H.-O., Almada-Lobo, B. (2012). Multi-objective integrated production and distribution planning of perishable products, *International Journal of Production Economics*, Vol. 138, No. 1, 89-101, doi: 10.1016/j.ijpe.2012.03.005.

[16] Prindezis, N., Kiranoudis, C.T., Marinos-Kouris, D. (2003). A business-to-business fleet management service provider for central food market enterprises, *Journal of Food Engineering*, Vol. 60, No. 2, 203-210, doi: 10.1016/S0260-8774(03)00041-4.

[17] Hasani, A., Zegordi, S.H., Nikbakhsh, E. (2012). Robust closed-loop supply chain network design for perishable goods in agile manufacturing under uncertainty, *International Journal of Production Research*, Vol. 50, No. 16, 4649-4669, doi: 10.1080/00207543.2011.625051.

[18] Govindan, K., Jafarian, A., Khodaverdi, R., Devika, K. (2014). Two-echelon multiple-vehicle location-routing problem with time windows for optimization of sustainable supply chain network of perishable food, *International Journal of Production Economics*, Vol. 152, 9-28, doi: 10.1016/j.ijpe.2013.12.028.

[19] Claassen, G.D.H., Gerdessen, J.C., Hendrix, E.M.T., van der Vorst, J.G.A.J. (2016). On production planning and scheduling in food processing industry: Modelling non-triangular setups and product decay, *Computers & Operations Research*, Vol. 76, 147-154, doi: 10.1016/j.cor.2016.06.017.

[20] Pahl, J., Voß, S. (2014). Integrating deterioration and lifetime constraints in production and supply chain planning: A survey, *European Journal of Operational Research*, Vol. 238, No. 3, 654-674, doi: 10.1016/j.ejor.2014.01.060.

[21] de Keizer, M., Akkerman, R., Grunow, M., Bloemhof, J.M., Haijema, R., van der Vorst, J.G.A.J. (2017). Logistics network design for perishable products with heterogeneous quality decay, *European Journal of Operational Research,* Vol. 262, No. 2, 535-549, doi: 10.1016/j.ejor.2017.03.049.

[22] Albrecht, W., Steinrücke, M. (2018). Coordinating continuous-time distribution and sales planning of perishable goods with quality grades, *International Journal of Production Research,* Vol. 56, No. 7, 2646-2665, doi: 10.1080/00207543.2017.1384584.

[23] Devapriya, P., Ferrell, W., Geismar, N. (2017). Integrated production and distribution scheduling with a perishable product, *European Journal of Operational Research*, Vol. 259, No. 3, 906-916, doi: 10.1016/j.ejor.2016.09.019.

[24] Moon, I., Lee, S. (2000). The effects of inflation and time-value of money on an economic order quantity model with a random product life cycle, *European Journal of Operational Research,* Vol. 125, No. 3, 588-601, doi: 10.1016/S0377-2217(99)00270-2.

[25] Chen, J.-M., Lin, C.-S. (2002). An optimal replenishment model for inventory items with normally distributed deterioration, *Production Planning & Control*, Vol. 13, No. 5, 470-480, doi: 10.1080/09537280210144446.

[26] Qi, M., Lin, W.-H., Li, N., Miao, L. (2012). A spatiotemporal partitioning approach for large-scale vehicle routing problems with time windows, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 48, No. 1, 248-257, doi: 10.1016/j.tre.2011.07.001.

Game theoretic analysis of supply chain based on mean-variance approach under cap-and-trade policy

He, L.F.[a], **Zhang, X.**[a,*], **Wang, Q.P.**[b,*], **Hu, C.L.**[c,*]

[a]College of Management and Economics, Tianjin University, Tianjin, P.R. China
[b]School of Management Science and Engineering, Hebei University of Economics and Business, Shijiazhuang, P.R. China
[c]College of Intelligent Manufacturing, Tianjin Sino-German University of Applied Sciences, Tianjin, P.R. China

A B S T R A C T

In recent years, carbon emission problem occurred by carbon dioxide as one of the main greenhouse gases, has become the focus due to its great influence on human life. With the increase of consumers' low-carbon consciousness, this paper studies the supply chain which consists of a single supplier and a single manufacturer in presence of market low-carbon preference. First, we establish the mean-variance analysis model. Second, we study the optimal decisions of channel members considering the risk factor in three situations: traditional supply chain without emission reduction, individual emission reduction by manufacturer and supply chain collaborative emission reduction. Finally, the equilibrium results are demonstrated by numerical studies. The results show that chain members' profits are not only affected by their own risk-aversion level, but also by other chain member's risk aversion level. More important is that there exist the optimal carbon emission reduction level and profits in system collaborative emission reduction. The research makes operation mechanism of low-carbon supply chain clearer and provides a theoretical reference for supply chain members on pricing and investment strategy of emission reduction.

1. Introduction

In recent years, the quick boosting of industry economy has brought many environmental problems, such as the greenhouse effect that threats human health seriously. In this regard, it is universally accepted that massive carbon emission is the first dominant factor resulting in greenhouse effect Dinan [1], so many countries attempt to use scientific approaches to control carbon emissions. In Kyoto Protocol, the idea of 'cap-and-trade' policy was therefore designed to control the emission of greenhouse gases, which is a significant step to mitigate the negative effect of carbon emission. It is noteworthy that controlling carbon emission in the production process is the key tache of managing low-carbon supply chain. Thereby, a lot of manufacturing enterprises pay attention to low-carbon production and implement effective emission reduction measures. Generally, controlling carbon emission is of great significance to promote the sustainable development of environment.

A number of researches on low-carbon supply chain have focused on performance evaluation and management, such as, Yin *et al.* [2], Chen [3]. For in-depth research, the studies on low carbon supply chain can be divided into three categories: carbon footprint distribution, e.g. Wang *et al.* [4], Yang *et al.* [5], production and operation decision, e.g. *Wang et al.* [6], Nie *et al.* [7], Zhao

et al. [8]; supply chain structure under carbon emission constraints, e.g. Cholette *et al.* [9], Du *et al.* [10]. For ones interested in low-carbon supply chain, please refer to Jharkharia [11] to learn more progress.

One of main mathematical concepts to characterize our study is the use of mean variance. Chen *et al.* [12] studied the basic inventory model based on the newsvendor model by applying mean variance model. Wu *et al.* [13] researched newsvendor model with mean-variance analysis considering stockout cost, and the results reveal that the risk-averse newsvendor may even order more products than the risk-neutral. Ray *et al*. [14] conducted the research on purchasing strategy by mean-variance analysis considering disruption risk, which is the first time applying this approach to manage supply chain under disruption. Considering different risk attitudes, Choi *et al.* [15] investigated the newsvendor problem by utilizing mean-variance approach. Similarly, Yamaguchi *et al.* [16] developed mean-variance model, discussed the optimal order quantity following the three risk attitudes: risk-neutral attitude, risk-averse attitude and risk-prone attitude. Zhuo *et al.* [17] explored the two-echelon supply chain with option contracts under the mean-variance framework. More information about supply chain with mean-variance, please see Chiu *et al*. [18].

There is an increasing body of literature on cap-and-trade policy, which is relevant to the present study. In 1997 the cap-and-trade system was put forward in the Kyoto Protocol, the cap is proposed by government, and green organization has become a supplier of carbon emission. Different from traditional supply chain, the carbon emission permit can be traded via carbon market so that the carbon emission is utilized and controlled effectively. Considering stochastic demand, Zhang *et al.* [19] addressed the optimal production strategy of manufacturer in cap-and-trade system. Furthermore, they investigated purification efficiency in the case of multi-time. Du *et al.* [20] explored the low carbon supply chain consisting of one manufacturer who results in massive carbon emission and single emission permit supplier in a single period. Du *et al*. [21] discussed the carbon emission policy by utilizing the Stackelberg game in cap-and-trade system, furthermore, investigated the impact of cap on manufacturer, supplier and supply chain. Du *et al*. [22] constructed optimal production model for manufacturer considering environmental and preference with fixed cap and price of emission permit. According to the sizes of trade price and purification cost, the manufacturer decides whether to invest purification cost or trade emission permits via market for maximizing the profit. Zhao *et al*. [23] researched the problem of coordination mechanism and design of supply chain consisting of a sing manufacturer and a sing retailer considering the cap of carbon emission in a low carbon environment. Furthermore, Yuan *et al*. [24] studied the coordination of supply chain with revenue sharing contract in cap-and-trade system. In view of quick response strategy under newsvendor setting, Lee *et al*. [25] investigated the optimal pricing decisions of the company and policy maker who proposed the cap and trading price.

Our study is also related to consumers' preference to low-carbon product, such as Ma *et al*. [26] and Wang *et al*. [27]. Similarly, Zhang [28] assumed that demand of products is a linear function with product prices and carbon emission and explored the optimal decisions of channel members in three situations: traditional supply chain without emission reduction, manufacturer's individual emission reduction and collaborative emission reduction. Similar researches are also conducted in He *et al*. [29] and He *et al*. [30]. However, the risk factor of channel members wasn't considered in the literature. Choi *et al*. [15] discussed maximum profit of supply chain under risk free constraint and risk constraint, and the result showed that supply chain risk actively affects the decision-making of supply chain members, so it cannot be ignored. Differently, this paper simultaneously considers the mean and variance to investigate the problem of low carbon supply chain, which makes the research more consistent with the reality, and ensure rigor of the research problem.

Comprehensively, the extant literature on low carbon supply chain mainly focus on optimization in firm and supply chain level but without incorporating risk concerns, which is exactly the gap we attempt to fill. The Mean-variance model considers both mean and variance to capture risk concerns, which is close to the reality. However, the existing literature on low carbon supply chain solely takes mean into account. In response to the enhancement of people's awareness of

low carbon, countries are also actively taking action to reduce carbon emission, such as, the designing cap-and-trading policy. To echo above problems, this paper exploits the mean-variance analysis to construct the game of supply chain under cap-and-trade policy, then the optimal decisions of chain members are explored and some numerical examples are provided to illustrate the models.

2. Preliminaries and notation

The necessary assumptions used in this paper are listed as follows:

- The market information is symmetrical and complete, and the market can completely be cleared. Thereby, there is no excess supply or excess demand, and the Pareto optimal of equilibrium price is ensured.
- Assume a unit material from the supplier can produce a unit product by the manufacturer.
- According to the study of Abdallah *et al.* [31], we assume that the government has implemented a more market-oriented economic means, namely, cap-and-trade system. In order to implement the strategy fairer and effectiveness, the government sets the carbon cap of per unit product which is assumed as λ_g. According to the international practice and combined with China's national conditions, the carbon trading price p_c is allowed in the carbon emission trading market.
- Assume the demand is linearly decreasing in retail price but increasing in the low-carbon preference with random interruption. We therefore give the liner function $D = \tilde{\alpha} - \beta p + \gamma e$, where α is intrinsic demand, ξ the random variable, e the emission reduction per unit product and $\tilde{\alpha} = \alpha + \xi$. β and γ are the demand responsiveness in price and low-carbon preference, respectively. For the ease of computation but without loss of generality, we assume that ξ follows normal distribution $N(0, \sigma^2)$. Avoiding the trival case, let $\alpha > \beta(c_s + c_m)$.
- According to the standard hypothesis in classical model, namely, there is a quadratic function relationship between cost and R & D investment), we use $C(e) = \theta e^2/2$ to represent emission reduction cost, where θ is the parameter of emission reduction cost. Zhang [28] and Raz *et al.* [32] also use the function to represent the relationship between carbon emission reduction level and emission reduction cost.

The notations used in this paper are summarized in Table 1.

Table 1 Notation summarization

Notation	Meaning
p	Unit retail price
p_c	Unit trading price of carbon emission
D	The demand of product
c_m	Unit cost of manufacturer
c_s	Unit cost of supplier
w	Unit wholesale price
e_0	Initial carbon emission of unit product
λ_g	Carbon emission cap of unit product
θ	Parameter of emission reduction cost
A_m	The risk-averse level of manufacturer
A_s	The risk-averse level of supplier

3. Optimal decision of supply chain considering risk attitude

3.1 Optimal decision-making in traditional supply chain without emission reduction

In cap-and-trade system, if the manufacturer want to keep his traditional production without emission reduction, he must purchase carbon permits to meet the cap set by the government. In this situation, the manufacturer's profit is:

$$\pi_m = (p - w - c_m)(\alpha + \xi - \beta p) + p_c(\lambda_g - e_0)(\alpha + \xi - \beta p)$$

The first term in the right hand side (RHS) of above function represents the sales revenue of manufacturer, and the second denotes the cost for purchasing carbon emission permit. In addition, we assume that the cap proposed by the government is less than the initial carbon emission of per unit product $\lambda_g < e_0$.

Accordingly, the supplier's problem is:

$$\pi_s = (w - c_s)(\alpha + \xi - \beta p)$$

The expected profits of chain members are expressed as follows

$$E\pi_m = (p - w - c_m)(\alpha - \beta p) + p_c(\lambda_g - e_0)(\alpha - \beta p) \tag{1}$$

$$E\pi_s = (w - c_s)(\alpha - \beta p) \tag{2}$$

Then, the variance profit functions of manufacturer and supplier are shown as follows

$$Var_{\pi_m} = E[\pi_m - E\pi_m]^2 = [p - w - c_m + p_c(\lambda_g - e_0)]^2 \sigma^2 \tag{3}$$

$$Var_{\pi_s} = E[\pi_s - E\pi_s]^2 \sigma^2 \tag{4}$$

If the supply chain members want to get the optimal profits, which means that the mean and variance of profit must be simultaneously considered and maximize the mean of profit while minimize the variance. Ray *et al*. [14] established the mean-variance analysis model to study the disruption risk of supply chain. Moreover, the risk of supply chain was introduced into the model, which ensured the scientific of the research problem. Similarly, Wu *et al*. [13] introduced the mean-variance model into the newsvendor model considering shortage cost, and drew the different conclusion compared with the previous studies. Therefore, we establish the mean-variance model based on $MV: E\pi - A\sqrt{Var}$, where $A(0 < A \leq 1)$ represents the risk aversion level of supply chain members. Finally, we get the utility function of the manufacturer as follows

$$MV: U_m = [p - w - c_m + p_c(\lambda_g - e_0)](\alpha - \beta p) - A_m[p - w - c_m + p_c(\lambda_g - e_0)]\sigma \tag{5}$$

The utility of manufacturer consists of two parts, i.e. fixed utility and risk utility. Let the revenue of per unit product multiply by the demand show the fixed utility of manufacturer. The second part represents the risk utility of manufacturer, where A_m denotes the risk aversion level of the manufacturer and σ shows the variation of demand.

Accordingly, the utility function of supplier is expressed as follow

$$MV: U_s = (w - c_s)(\alpha - \beta p) - A_s(w - c_s)\sigma \tag{6}$$

Similarly, the first term in the RHS of above function shows the fixed utility of supplier by letting the revenue of per material multiply by demand. The second term expresses the risk utility of supplier, where A_s represents the risk aversion level of the supplier.

According to the Eqs. 5 and 6, we find that the profits of chain members in risk averse supply chain are less than that in risk-neutral supply chain.

The second partial derivative of Eq. 5 with the retail price takes the form:

$$\frac{d^2 U_m}{dp^2} = -2\beta < 0$$

Let the first partial derivative of Eq. 5 with respect to the retail price equal zero, then we can get the optimal response function of p as follow

$$p = \frac{\alpha - A_m \sigma}{2\beta} + \frac{\omega + c_m - p_c(\lambda_g - e_0)}{2} \tag{7}$$

Observing Eq. 7, we find that the retail price is decreasing in manufacturer's risk aversion level. The reason is that a low retail price of the product will attract some of the lower price sensitive consumers, and promote consumption, then risk of fluctuations in the demand is reduced. Finally, the manufacturer's profits have been ensured.

Substituting Eq. 7 into the utility function of supplier Eq. 6, then we get the second partial derivative in terms of w

$$\frac{d^2 U_s}{dw^2} = -\beta < 0$$

Let the first partial derivative in w equal zero, we can get the optimal wholesale price

$$w^{1*} = \frac{\alpha + A_m\sigma - 2A_s\sigma}{2\beta} - \frac{c_m - c_s - p_c(\lambda_g - e_0)}{2} \tag{8}$$

In Eq. 8, the wholesale price decreases in supplier's risk aversion level. The reason why this phenomenon happens can be explained as the risk-averse supplier reduces the wholesale price to attract manufacturer to order more material for maximizing her profit. Meanwhile, the wholesale price is increasing in manufacturer's risk attitude in that the risk-averse manufacturer reduces the retail price in order to stimulate demand and accordingly needs to increase order quantity from supplier, which in turn encourage supplier charging manufacturer higher wholesale price.

Now we can rewrite the function of p as follows

$$p^{1*} = \frac{3\alpha - A_m\sigma - 2A_s\sigma + \beta[c_m + c_s - p_c(\lambda_g - e_0)]}{4\beta} \tag{9}$$

3.2 Optimal decision-making in emission reduction by manufacturer

In this section we consider the fact that emission permit purchasing cost is considerable in the traditional production mode in cap-and-trade system, which seriously impact on the profit of manufacturer. In addition, consumers are willing to pay high price for low-carbon products with low-carbon preference. Considering above situation, manufacturer tries to change the traditional production mode and starts to conduct low-carbon production. Suppose that one unit product reduces e_1 units carbon emission after manufacturer applies low-carbon effort and the surplus carbon emission can't be used in the next period. Hence, conducting low-carbon production results in associated cost $\frac{1}{2}\theta e_1^2$. Manufacturer's profit can be expressed as follows

$$\pi_m = (p - w - c_m)(\alpha + \xi - \beta p + \gamma e_1) + p_c(\lambda_g - e_{0+}e_1)(\alpha + \xi - \beta p + \gamma e_1) - \frac{1}{2}\theta e_1^2$$

The first term represents sales revenue, the second term the carbon trading cost (if buying permits) or revenue (if selling permits) and the last term carbon emission reduction cost.

Accordingly, the supplier's profit can be described as below

$$\pi_s = (w - c_s)(\alpha + \xi - \beta p + \gamma e_1)$$

The expected profits of chain members are expressed as follows

$$E\pi_m = (p - w - c_m)(\alpha - \beta p + \gamma e_1) + p_c(\lambda_g - e_0 + e_1)(\alpha - \beta p + \gamma e_1) - \frac{1}{2}\theta e_1^2 \tag{10}$$

$$E\pi_s = (w - c_s)(\alpha - \beta p + \gamma e_1) \tag{11}$$

According to the Eqs. 3 and 4, we get the utility functions of chain members as follows

$$MV: U_m = (\alpha - \beta p + \gamma e_1 - A_m\sigma)[p - w - c_m + p_c(\lambda_g - e_0 + e_1)] - \frac{1}{2}\theta e_1^2 \tag{12}$$

$$MV: U_s = (\alpha - \beta p + \gamma e_1 - A_s\sigma)(w - c_s) \tag{13}$$

The first partial derivatives of U_m with respect to the retail price and carbon emission reduction level as follows

$$\frac{\partial U_m}{\partial p} = -2\beta p + \beta[w + c_m - p_c(\lambda_g - e_o + e_1)] + \alpha + \gamma e_1 - A_m\sigma \tag{14}$$

$$\frac{\partial U_m}{\partial e_1} = \gamma[p - w - c_m + p_c(\lambda_g - e_o + e_1)] + p_c(\alpha - \beta p + \gamma e_1 - A_m \sigma) - \theta e_1 \qquad (15)$$

Solving the Hessian matrix of the Eq. 12 in terms of the retail price and carbon emission reduction level yields

$$H^1 = \begin{bmatrix} -2\beta & \gamma - \beta p_c \\ \gamma - \beta p_c & 2p_c \gamma - \theta \end{bmatrix}$$

When $2\beta(\theta - 2p_c\gamma) > (\gamma - \beta p_c)^2$, there are the optimal solutions of Eq. 12. Let the Eqs. 14 and 15 equal zero, then the optimal decisions of manufacturer can be expressed as follows

$$p = \frac{[\theta - p_c(\gamma + \beta p_c)](\alpha - A_m\sigma) - (\gamma^2 + \beta p_c\gamma - \beta\theta)[w + c_m - p_c(\lambda_g - e_0)]}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2} \qquad (16)$$

$$e_1 = \frac{(\gamma + \beta p_c)[\alpha - A_m\sigma - \beta(w + c_m - p_c(\lambda_g - e_0))]}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2} \qquad (17)$$

Substituting the Eqs. 16 and 17 into the Eq. 13, we can rewrite the utility function of supplier as below

$$U_s = \frac{w - c_s}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2}\{\beta\theta[\alpha - \beta(w + c_m - p_c(\lambda_g - e_0))] + [\beta\theta - (\gamma + \beta p_c)^2]A_m\sigma \\ -[2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]A_s\sigma\} \qquad (18)$$

Solving the first derivative and second derivative with wholesale price of Eq. 18, we can obtain

$$\frac{\partial U_s}{\partial w} = \frac{w - c_s}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2}\{\alpha - \beta(w + c_m - p_c(\lambda_g - e_0)) + [\beta\theta - (\gamma + \beta p_c)^2]A_m\sigma \\ -[2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]A_s\sigma\} - \frac{\beta^2\theta}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2} \qquad (19)$$

$$\frac{\partial^2 U_s}{\partial w^2} = -\frac{2\beta^2\theta}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2} < 0$$

Let Eq. 19 equal zero, the optimal wholesale price is expressed as below

$$w^{2*} = \frac{\alpha}{2\beta} - \frac{c_m - c_s - p_c(\lambda_g - e_0)}{2} + \frac{[\beta\theta - (\gamma + \beta p_c)^2]A_m\sigma - [2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]A_s}{2\beta^2\theta} \qquad (20)$$

Finally, substituting the Eq. 20 into Eqs. 16 and 17, we rewrite the optimal solutions as below

$$p^{2*} = -\frac{\gamma(\gamma + \beta p_c) - \beta\theta}{2\beta^2\theta[2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]}\{\beta[\alpha + \beta(c_m + c_s - p_c(\lambda_g - e_0))] \\ + [\alpha + \beta(c_m + c_s - p_c(2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2))]A_m\sigma \\ - [2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]A_s\sigma\} + \frac{[\theta - p_c(\gamma + \beta p_c)](\alpha - A_m\sigma)}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2} \qquad (21)$$

$$e_1 = \frac{\gamma + \beta p_c}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2}\{\frac{\alpha - \beta[c_m + c_s - p_c(2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2)]}{2} \\ - \frac{[(3\beta\theta - (\gamma + \beta p_c)^2)]\sigma - [2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2]A_s\sigma}{2\beta\theta}\} \qquad (22)$$

3.3 Optimal decision-making in channel collaborative emission reduction

In order to improve the market competitiveness and get more profits, manufacturer conducts low carbon production considering consumers' preference of low carbon products. Then the market demand has been promoted with the behavior of manufacturers' emission reduction. Finally, the profit of supplier has been improved with more materials ordered by manufacturer. Thereby, the rational supplier proposes that he is willing to cooperate with manufacturer and

bear a part of carbon emission reduction cost. Under the mode of cooperative emission reduction, what is the changes in decision-making between supply chain members? Does the collaborative emission reduction between the supply chain members maximize the economic benefits of each other? This section analyzes these problems.

First, we get the chain members' profits as follows.

$$\pi_m = (p - w - c_m)D + p_c(\lambda_g - e_0 + e)D - \frac{1}{2}(1 - \delta)\theta e^2$$

$$\pi_s = (w - c_s)D - \frac{1}{2}\delta\theta e^2$$

Where the parameter δ denotes the emission reduction cost sharing proportion of supplier.

The expected profits of chain members are expressed as follows

$$E\pi_m = [p - w - c_m + p_c(\lambda_g - e_0 + e)](\alpha - \beta p + \gamma e) - \frac{1}{2}(1 - \delta)\theta e^2 \qquad (23)$$

$$E\pi_s = (w - c_s)(\alpha - \beta p + \gamma e) - \frac{1}{2}\delta\theta e^2 \qquad (24)$$

The first term of RHS in Eq. 23 represents the revenues consisting of both sales revenue and carbon emission trading revenue or cost and the second term the carbon emission reduction cost undertaken by manufacturer. Similarly, the first term of Eq. 24 expresses the sales revenue, and the last term the carbon emission reduction cost undertaken by supplier.

Then we can get the utility functions of chain members as follows

$$MV\colon U_m = [p - w - c_m + p_c(\lambda_g - e_0 + e)](\alpha - \beta p + \gamma e - A_m\sigma) - \frac{1}{2}(1 - \delta)\theta e^2 \qquad (25)$$

$$MV\colon U_s = (w - c_s)(\alpha - \beta p + \gamma e - A_s\sigma) - \frac{1}{2}\delta\theta e^2 \qquad (26)$$

Similarly, we get the first partial derivatives of U_m with respect to the retail price and carbon emission reduction level as follows

$$\frac{\partial U_m}{\partial p} = -2\beta p + \beta[w + c_m - p_c(\lambda_g - e_o)] + \alpha + \gamma e_1 - A_m\sigma \qquad (27)$$

$$\frac{\partial U_m}{\partial e} = \gamma[p - w - c_m + p_c(\lambda_g - e_o + e)] + p_c(\alpha - \beta p + \gamma e - A_m\sigma) - (1 - \delta)\theta e \qquad (28)$$

Solving the Hessian matrix of the Eq. 25 in terms of the retail price and carbon emission reduction level, we get

$$H^2 = \begin{bmatrix} -2\beta & \gamma - \beta p_c \\ \gamma - \beta p_c & 2p_c\gamma - (1 - \delta)\theta \end{bmatrix}$$

When $2\beta[(1 - \delta)\theta - 2p_c\gamma] > (\gamma - \beta p_c)^2$, there are the optimal solutions of Eq. 25. Let the Eqs. 27 and 28 equal zero, the optimal decisions of manufacturer are expressed as follows

$$p = \frac{[(1 - \delta)\theta - p_c(\gamma + \beta p_c)](\alpha - A_m\sigma) - [\gamma^2 + \beta p_c\gamma - \beta(1 - \delta)\theta][\lambda_g - e_o + e]}{2\beta[(1 - \delta)\theta - 2p_c\gamma] - (\gamma - \beta p_c)^2} \qquad (29)$$

$$e = \frac{(\gamma + \beta p_c)[\alpha - A_m\sigma - \beta(w + c_m - p_c(\lambda_g - e_0))]}{2\beta[(1 - \delta)\theta - 2p_c\gamma] - (\gamma - \beta p_c)^2} \qquad (30)$$

Substituting the Eqs. 29 and 30 into the Eq. 26, we can rewrite the utility function of supplier as below

$$U_s = \frac{w - c_s}{2\beta(\theta - 2p_c\gamma) - (\gamma - \beta p_c)^2}\{\beta\theta(1 - \delta)[\lambda_g - e_0] + [\beta\theta(1 - \delta) - (\gamma + \beta p_c)^2]A_m\sigma$$

$$[\beta\theta(1 - \delta) - (\gamma + \beta p_c)^2]A_m\sigma\} - \frac{1}{2}\delta\theta\frac{(\gamma + \beta p_c)^2[\alpha - A_m\sigma - \beta(w + c_m - p_c(\lambda_g - e_0))]^2}{[2\beta(\gamma - \beta p_c) - (\gamma - \beta p_c)^2]^2} \qquad (31)$$

Solving the second derivative with wholesale price of Eq. 31, we can get

$$\frac{d^2 U_s}{dw^2} = -\frac{2\beta^2\theta(1-\delta)}{2\beta[(1-\delta\theta)-2p_c\gamma]-(\gamma-\beta p_c)^2} - \frac{\beta^2\theta\delta(\gamma+\beta p_c)^2}{[2\beta((1-\delta\theta)-2p_c\gamma)-(\gamma-\beta p_c)^2]^2} < 0$$

Therefore, the optimal wholesale price is generated by letting the first derivative with wholesale price of Eq. 31 equal zero

$$w^{2*} = \frac{\alpha}{2\beta} - \frac{c_m - c_s - p_c(\lambda_g - e_0)}{2} + \frac{[\beta\theta - (\gamma+\beta p_c)^2]A_m\sigma - [2\beta(\theta-2p_c\gamma)-(\gamma-\beta p_c)^2]A_s}{2\beta^2\theta}$$

$$+ \frac{[(2\beta((1-\delta)\theta-2p_c\gamma)-(\gamma-\beta p_c)^2)^2 - \beta\theta(2\beta((1-\delta)\theta-2p_c\gamma)-(\gamma-\beta p_c)^2) + 2\beta^2\theta^2\delta(1-\delta)]A_m\sigma}{\beta^2\theta[2(2\beta((1-\delta)\theta-2p_c\gamma)-(\gamma-\beta p_c)^2)+\delta(\gamma+\beta p_c)^2]} \quad (32)$$

Finally, substituting the Eq. 32 into the Eqs. 29 and 30, we get

$$p^{3*} = \frac{[(1-\delta)\theta - p_c(\gamma+\beta p_c)](\alpha - A_m\sigma) - [\gamma^2 + \beta p_c\gamma - \beta(1-\delta)\theta][w^{3*}+c_m-p_c(\lambda_g-e_o)]}{2\beta[(1-\delta)\theta-2p_c\gamma]-(\gamma-\beta p_c)^2} \quad (33)$$

$$e = \frac{(\gamma+\beta p_c)[\alpha - A_m\sigma - \beta(w^{3*}+c_m-p_c(\lambda_g-e_0))]}{2\beta[(1-\delta)\theta-2p_c\gamma]-(\gamma-\beta p_c)^2} \quad (34)$$

It is difficult to elegantly express the p^{3*} and e in explicit format. Thereby, we substitute w^{3*} into Eqs. 29 and 30 to represent the equilibrium results. Further research on the problem will be carried out in the numerical analysis section.

4. Computational study and discussion

In this section, we conduct a series of numerical computation to demonstrate previous models and obtain some managerial insights.

Proposition 1: The carbon emission reduction level in chain members' reduction cooperation is greater than that in emission reduction by manufacturer solely.

The result reveals that the reduction cooperation of chain members can not only meet the low carbon preference of consumers, but promote the environment sustainable development. Additionally, the carbon emission reduction level is increasing in supplier's risk-averse level, while decreasing in manufacturer's risk-averse level, as shown in Fig. 1 and Fig. 2, respectively. the reason resulting in this phenomenon can be explained like that the risk-averse manufacturer wouldn't like to input more costs to conduct low-carbon production, while the risk-averse supplier is just opposite.
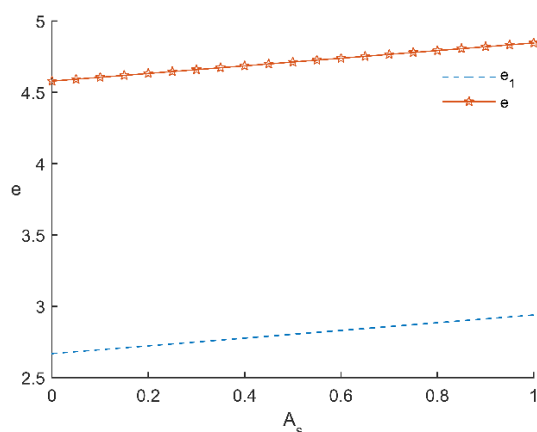

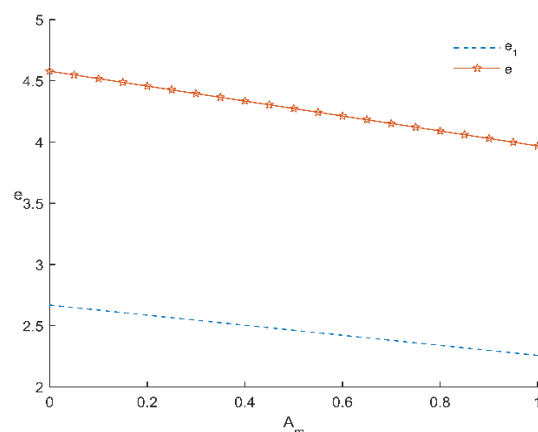**Fig. 1** The carbon emission reduction level with versus A_s variation

Fig. 2 The carbon emission reduction level with versus A_m variation

Proposition 2: The retail price is decreasing in risk aversion level of supply chain members.

As shown in Figs. 3 and 4, we can explain the phenomenon in this way. The manufacturer reduces retailer price to attract price sensitive consumers for getting high profit. Furthermore, observing Fig. 3, we find there are different relationships among the three retail prices with the
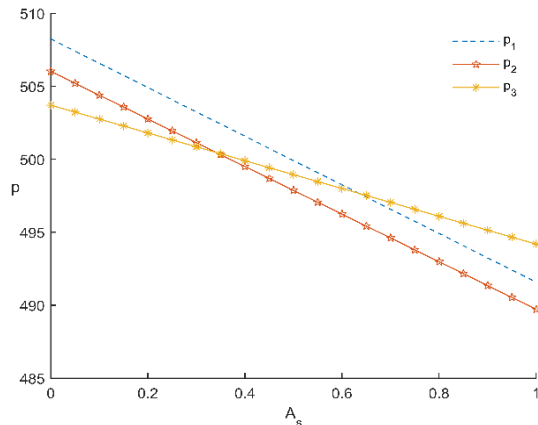
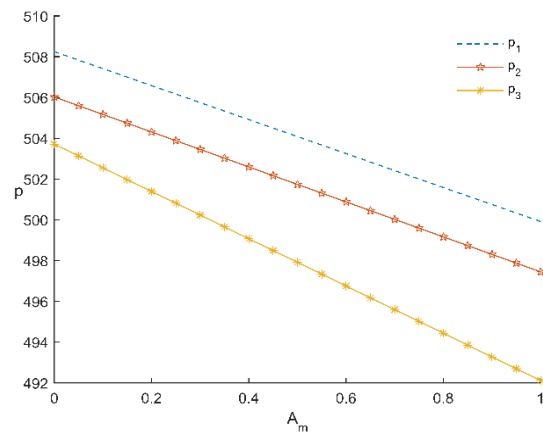**Fig. 3** The retail price with versus A_s variation

Fig. 4 The retail price with versus A_m variation

variation of supplier's risk aversion level. While the retail price in collaborative emission reduction is the lowest price no matter how manufacturer's risk aversion level changes, as shown in Fig. 4.

Proposition 3: The wholesale price is closely related to the risk aversion level of chain members.

It is decreasing in supplier's risk aversion level, while it shows opposite result in manufacturer's risk aversion level, as shown in Figs. 5 and 6. The reason resulting in this phenomenon is that the risk-averse supplier reduces the wholesale price in order to promote the order of manufacturer. For risk-averse manufacturer, as we described aforesaid, he reduces retail price aim to stimulate consumption with more order occurred. Finally, the upstream supplier takes this chance to improve wholesale price.
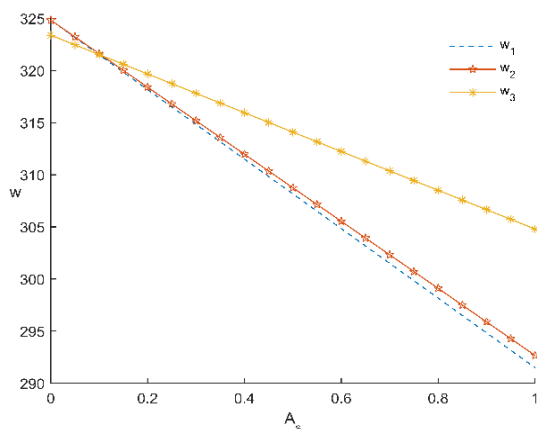

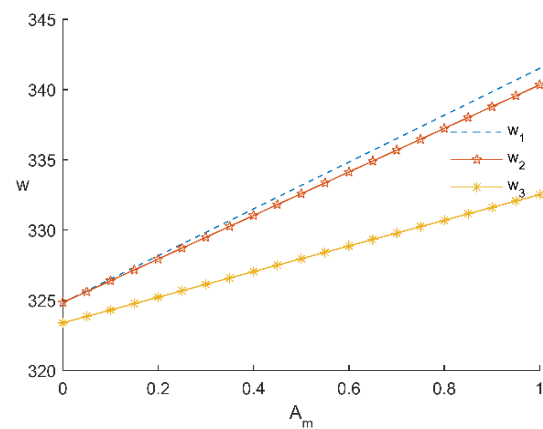**Fig. 5** The wholesale price with versus A_s variation

Fig. 6 The wholesale price with versus A_m variation

Proposition 4: The profits of chain members are decreasing in their own risk aversion levels. In addition, there are always the optimal profits of chain members in emission reduction cooperation when supplier keeps the certain risk aversion level.

Observing Figs. 7, 8, 9, and 10, the chain members all get the optimal profits in collaborative emission reduction when the risk aversion level of supplier is less than 0.4. Additionally, risk-averse chain members' profits are increasing in their own risk aversion levels.

the cause of this kind of appearance is that risk-averse chain members decreases the prices (wholesale price or retail price) to stimulate consumption for maximizing the profits, nevertheless, the profits are still decreasing.
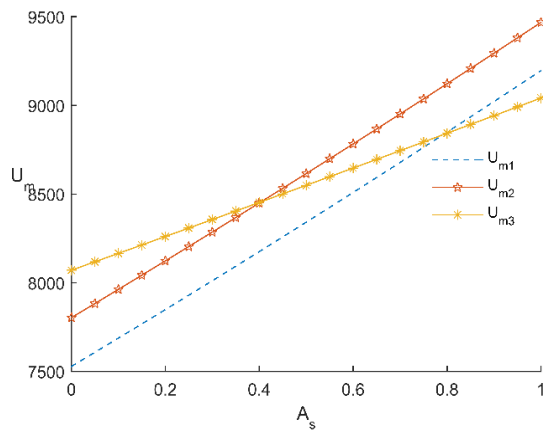
Fig. 7 The profit of manufacturer with versus A_s variation
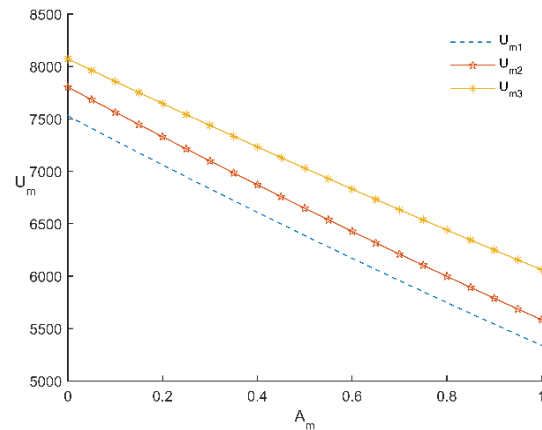
Fig. 8 The profit of manufacturer with A_m variation
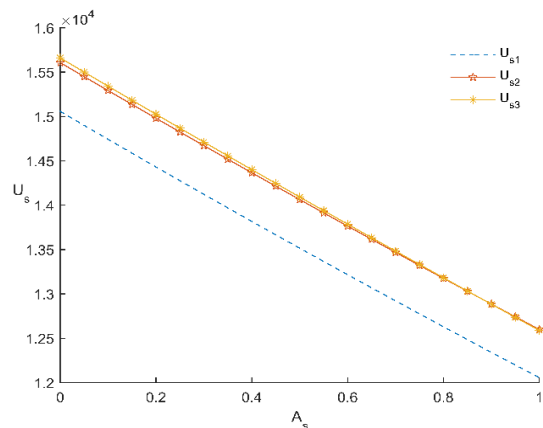
Fig. 9 The profit of supplier with versus A_s variation
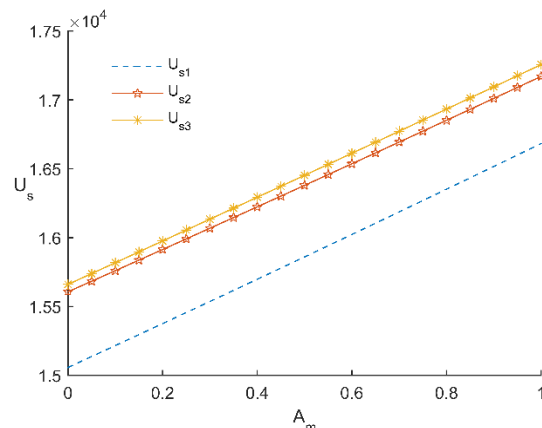
Fig. 10 The profit of supplier with A_m variation

Proposition 5: The profits of chain members are increasing in the cap imposed by government.

The cap proposed by the government directly impact on the economic benefits of the supply chain members in the risk-neutral supply chain. With increase of the cap, the economic benefits of supply chain members have been improved, and it is more beneficial to the chain members to cooperate with each other about carbon emission. The result also reveals another information that the cap provided by the government can be directly converted into economic income for the chain members. As shown in Table 2.

The parameters are summarized in Table 3.

Table 2 The profits of chain members

λ_g	U_m^1	U_m^2	U_m^3	U_s^1	U_s^2	U_s^3
5	7528.8	7803.5	8071.2	15057.5	15607.0	15660.4
6	7600.2	7877.6	8147.4	15200.4	15755.1	15809.1
7	7672.0	7952.0	8224.0	15344.0	15903.9	15958.4
8	7744.1	8026.7	8301.0	15488.3	16053.5	16108.5

Table 3 Parameter

Parameter	α	β	γ	c_m	c_s	θ	σ	p_c	e_0	δ	λ_g
Value	200	0.3	0.4	10	8	80	10	3	10	0.4	5

5. Conclusion

Considering consumers' low-carbon preference and the risk concerns of chain members, this study exploits mean-variance approach to explore the optimal decisions of supply chain agents. Three different cases have been discussed: traditional supply chain without emission reduction, emission reduction by manufacturer individually and chain members' collaborative emission reduction.Some meaningful conclusions are obtained through analytical analysis and computa-

tional study. The results show the emission reduction level in collaborative emission reduction is greater than that by manufacturer individually. Moreover, there is an optimal profits of supply chain members in collaborative emission reduction, which indicates that in the cap-and-trade system, collaborative emission reduction of chain members can not only increases the profits of chain members, but also improves the emission reduction level. Additionally, the optimal decisions are affected by both manufacturer's risk level and supplier's risk level. The research provides a cooperative emission-reduction way for chain members constrained by low-carbon policy.

The significance of this paper is reflected in three aspects. First, considering the chain member's risk, we explore the equilibrium results in three cases as described aforesaid, and figure out the operational mechanism for a risk-aversion supply chain in the cap-and-trade policy. Second, we study the influence of risk attitude on the profits of chain members, which reveals that the level of risk aversion plays an important role in the decision-making of supply chain members. The research compensates the defects of existing literature that assumes supply chain members are risk-natural. Finally, we consider the risk of chain members and construct mean-variance analysis model to let the research close to the reality.

The proposed models can be extended in several ways. For instance, extending the model to information asymmetry, which is common and important in reality. Furthermore, the study may be extended to more-than-two-tier supply chain. Additionally, studying the low carbon supply chain in multiple periods may be also a potential direction in the future.

Acknowledgment

Reference

[1] Dinan, T. (2008). *Policy options for reducing CO$_2$ emissions,* Washington, D.C., The Congress of the United States, Congressional Budget Office.

[2] Yin, Z., Zhang, X. (2014). A review of the research on low carbon supply chain in the context of open economy, *Inquiry Into Economic Issues*, Vol. 9, 154-159, doi: 10.3969/j.issn.1006-2912.2014.09.025.

[3] Chen, J. (2012). Study on supply chain management in a low-carbon era, *Journal of Systems & Management*, Vol. 21, Vol. 6, 721-728, doi: 10.3969/j.issn.1005-2542.2012.06.002.

[4] Wang, L., Zheng, Y., Gao, Y. (2014). Study on multi stage carbon footprint optimization of manufacturing supply chain, *Journal of the Party School of the Central Committee of the C.P.C.*, Vol. 18, No. 4, 92-97, doi: 10.14119/j.cnki.zgxb.2014.04.057.

[5] Yang, C., Li, X., Shao, J. (2013). Optimization of carbon footprint of reverse supply chain in complex uncertain environment, *Low-carbon Economy*, Vol. 4, doi: 10.13529/j.cnki.enterprise.economy.2013.04.007.

[6] Wang, W.-B., Deng, W.-W., Bai, T., Da, Q.-L., Nie, R. (2016). Design the reward-penalty mechanism for reverse supply chains based on manufacturers' competition and carbon footprint constraints, *Journal of Industrial Engineering and Engineering Management*, Vol. 30, No. 2, 188-194, doi: 10.13587/j.cnki.jieem.2016.02.023.

[7] Nie, J.-J., Wang, T., Zhao, Y.-X., Zahng, L.-N. (2015). Collecting strategies for the closed-loop supply chain with remanufacturing in the constraint of carbon emission, *Journal of Industrial Engineering and Engineering Management*, Vol. 29, No. 3, 249-256, doi: 10.13587/j.cnki.jieem.2015.03.028.

[8] Zhao, D.-Z., Yuan, B.-Y., Xia, L.-J., Xie, X.-P. (2014). Dynamic game study in supply chain with manufacturers' competition under the constraint of productions' emission, *Industrial Engineering & Management*, Vol. 1, No. 8, 65-71, doi: 10.3969/j.issn.1007-5429.2014.01.011.

[9] Cholette, S., Venkat, K. (2009). The energy and carbon intensity of wine distribution: A study of logistical options for delivering wine to consumers, *Journal of Cleaner Production*, Vol. 17, No. 16, 1401-1413, doi: 10.1016/j.jclepro.2009.05.011.

[10] Du, S., Ma, F., Fu, Z., Zhu, L., Zhang, J. (2011). Game-theoretic analysis for an emission-dependent supply chain in a 'cap-and-trade' system, *Annals of Operations Research*, Vol. 228, No. 1, 135-149, doi: 10.1007/s10479-011-0964-6.

[11] Das, C., Jharkharia, S. (2018). Low carbon supply chain: A state-of-the-art literature review, *Journal of Manufacturing Technology Management*, Vol. 29, No. 2, 398-428, doi: 10.1108/JMTM-09-2017-0188.

[12] Chen, F., Federgruen, A. (2000). *Mean-variance analysis of basic inventory models,* Working Paper, Columbia Business School, New York, USA.

[13] Wu, J., Li, J., Wang, S., Cheng, T.C.E. (2008). Mean-variance analysis of the newsvendor model with stockout cost, *Omega*, Vol. 37, No. 3, 724-730, doi: 10.1016/j.omega.2008.02.005.

[14] Ray, P., Jenamani, M. (2016). Mean-variance analysis of sourcing decision under disruption risk, *European Journal of Operational Research*, Vol. 250, No. 2, 679-689, doi: 10.1016/j.ejor.2015.09.028.

[15] Choi, T.-M., Li, D., Yan, H. (2008). Mean–variance analysis of a single supplier and retailer supply chain under a returns policy, *European Journal of Operational Research*, Vol. 184, No. 1, 356-376, doi: 10.1016/j.ejor.2006.10.051.

[16] Yamaguchi, S., Goto, H., Kusukawa, E. (2017). Mean-variance analysis for optimal operation and supply chain coordination in a green supply chain, *Industrial Engineeering & Management Systems*, Vol. 16, No. 1, 22-43, doi: 10.7232/iems.2017.16.1.022.

[17] Zhuo, W., Shao, L., Yang, H. (2018). Mean-variance analysis of option contracts in a two-echelon supply chain, *European Journal of Operational Research*, Vol. 271, No. 2, 535-547, doi: 10.1016/j.ejor.2018.05.033.

[18] Chiu, C.-H., Choi, T.-M. (2016). Supply chain risk analysis with mean-variance models: A technical review, *Annals of Operations Research*, Vol. 240, No. 2, 489-507, doi: 10.1007/s10479-013-1386-4.

[19] Zhang, J.-J., Nie, T.-F., Du, S.-F. (2011). Optimal emission-dependent production policy with stochastic demand, *International Journal of Society Systems Science*, Vol. 3, No. 1-2, 21-39, doi: 10.1504/IJSSS.2011.038931.

[20] Du, S., Ma, F., Fu, Z., Zhu, L., Zhang, J. (2011). Game-theoretic analysis for an emission-dependent supply chain in a 'cap-and-trade' system, *Annals of Operations Research*, Vol. 228, No. 1, 135-149, doi: 10.1007/s10479-011-0964-6.

[21] Du, S., Zhu, L., Liang, L., Ma, F. (2013). Emission-dependent supply chain and environment-policy-making in the 'cap-and-trade' system, *Energy Policy*, Vol. 57, 61-67, doi: 10.1016/j.enpol.2012.09.042.

[22] Du, S., Hu, L., Song, M. (2014). Production optimization considering environmental performance and preference in the cap-and-trade system, *Journal of Cleaner Production*, Vol. 112, Part 2, 1600-1607, doi: 10.1016/j.jclepro.2014.08.086.

[23] Zhao, D., Wang, B., Xu, C. (2014). Study on the coordination mechanism of supply chain considering the restriction of product carbon emissions, *FORECAST,* Vol. 33, No. 5, 76-80.

[24] Yuan, J., Ma, J., Yang, W. (2016). Revenue-sharing contract for supply chain under a cap and trade system, In: *Proceedings of 2016 International Conference on Logistics, Informatics and Service Sciences (LISS),* Sydney, Australia, 1-6, doi: 10.1109/LISS.2016.7854442.

[25] Lee, J., Lee, M.L., Park, M. (2018). A newsboy model with quick response under sustainable carbon cap-n-trade, *Sustainability*, Vol. 10, No. 5, doi: 10.3390/su10051410.

[26] Ma, C., Liu, X., Zhang, H., Wu, Y. (2016). A green production strategies for carbon-sensitive products with a carbon cap policy, *Advances in Production Engineering & Management*, Vol. 11, No. 3, 216-226, doi: 10.14743/apem2016.3.222.

[27] Wang, Q., He, L. (2018). Managing risk aversion for low-carbon supply chains with emission abatement outsourcing, *International Journal of Environmental Research and Public Health,* Vol. 15, No. 2, 367-387, doi: 10.3390/ijerph15020367.

[28] Zhang, J. (2016). *The game analysis of low carbon supply chain collaboration in emission reduction research and development,* Master thesis, Hunan University, Changsha, Hunan Province, China.

[29] He, L., Zhao, D., Xia, L. (2015). Game theoretic analysis of carbon emission abatement in fashion supply chains considering vertical incentives and channel structures, *Sustainability*, Vol. 7, No. 4, 4280-4309, doi: 10.3390/su7044280.

[30] He, L., Hu, C., Zhao, D., Lu, H., Fu, X., Li, Y. (2016). Carbon emission mitigation through regulatory policies and operations adaptation in supply chains: Theoretic developments and extensions, *Natural Hazards*, Vol. 84, Supplement 1, 179-207, doi: 10.1007/s11069-016-2273-5.

[31] Abdallah, T., Farhat, A., Diabat, A., Kennedy, S. (2012). Green supply chains with carbon trading and environmental sourcing: Formulation and life cycle assessment, *Applied Mathematical Modelling*, Vol. 36, No. 9, 4271-4285, doi: 10.1016/j.apm.2011.11.056.

[32] Raz, G., Druehl, C.T., Blass, V. (2013). Design for the environment: Life-cycle approach using a newsvendor model, *Production and Operations Management*, Vol. 22, No. 4, 940-957, doi: 10.1111/poms.12011.

Decision-making strategies in supply chain management with a waste-averse and stockout-averse manufacturer

Jian, M.[a]**, Wang, Y.L.**[a,*]

[a]School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, P.R. China

ABSTRACT

Behavioral preferences is an important factor that affects the decision-making strategies of enterprises. Usually, the behavioral preferences will lead to decision-making that deviates from profit maximization. In this study, we investigate the influence of a dominant manufacturer's behavioral preferences on decision-making and subsequent impact on profits. This study looks at the profits of the manufacturer, retailer and the system as a whole. We construct a two-stage supply chain involving a retailer and a manufacturer who may have risk-neutral (RN), stockout-aversion (SA), waste-aversion (WA), and stockout- and waste-aversion (SW) preferences. Through a comparison and analysis of the four cases, we find that the manufacturer's wholesale price increases (decreases) with the SA (WA) coefficient, while the retailer's order quantity is completely the opposite. The manufacturer's wholesale price is the highest in the WA model, followed by the RN, SA and SW models, in that order. The retailer's order quantity is the largest and smallest in the SA and WA models, respectively, while the size of the order quantity between the RN and SW models depends on the ratio m (the ratio of the SA to the WA). Moreover, we also explore the changing trends of the decision-making and profits of the participants and the system profit with the degree of SA and WA, comparing the profits of the four cases.

1. Introduction

The supply chain decision-making is a very important problem in supply chain (SC) management and therefore is also a very difficult problem for enterprises. In practice, many retailers must look at overordering cost, underordering cost and predicted demand to determine their ordering decisions before a sales season [1]. In addition, the manufacturer also must decide the pricing strategy for its products according to its own business interests. Traditional research on decision-making strategies are based mainly on risk neutrality, as in [2-3]. However, we know from behavioral economics that decision-makers in the real world may have behavioral preferences, which is also an important reason for the existence of "decision bias" [1]. Therefore, it is more practical to study the decision-making strategies of decision-makers by integrating behavior preferences into the SC.

In the existing literature, behavioral preferences research focuses mainly on loss aversion and risk aversion. A study of loss aversion by Niu *et al.* [4] considers the sustainability of the with alternative power structures, and the impact of the supplier's attitudes to loss on the 's sustainability and profitability. Choi [5] studies the newsboy problem, where the decision-maker is loss-averse under carbon emission constraints. When there is competition between manufacturers, research by Li and Li [6] show an ordering decision problem in the case of a supply disrup-

tion. Xu *et al.* [7] study the optimal order quantity in the loss-averse newsvendor model with backordering. In the two-period game, research by Yang and Xiao [8] study the pricing, ordering and quality of retailers' service decisions when loss-averse consumers are sensitive to service quality. They also construct a SC coordination mechanism.

A study of risk aversion by Giri [9] simultaneously considers the case of a retailer ordering from a manufacturer with random output and a manufacturer with a fixed output and studies the inventory decision problem of a risk-averse retailer. Zhu *et al.* [10] consider the global optimization problem when the manufacturer has two sales channels. Wu *et al.* [11] study the influence of capacity uncertainty on the order decision of risk-averse managers. Additionally, Oh *et al.* [12] consider the impact of uncertainty in service cost on the product pricing decision when the service provider of the downstream SC is risk-averse. Egging *et al.* [13] discuss the influence of risk aversion on gas investment decisions. Xiao *et al.* [14] consider the retailer's service and pricing decisions when both system members and consumers are risk-averse. Zhou *et al.* [15] study a coordination problem when chain members and service provider are risk-averse. Zheng *et al.* [16] study the pricing decisions of two competitive shipping companies, one of which is risk-averse. Liu [17] analyzes the influence of risk aversion on the decision-making and expected profits of dual-channel members when information asymmetry exists in the risk aversion. For more research on risk aversion, see [18-20].

However, there are few studies available on the decision-making strategy problems that arise when decision makers have SA or WA preferences. Schweitzer and Cachon [1] study the ordering decision when the decision-maker in the downstream SC has either a SA or WA preferences. In contrast to the previous literature cited, in this paper, we introduce the SA and WA coefficients under the manufacturer-led Stackelberg game. We also consider that the manufacturer may have RN, SA, WA, and SW preferences. This paper mainly studies the following three problems:

- How will the four different behavioral preferences of the manufacturer affect the decision-making and expected profits of the chain members, as well as the expected profit of the system?
- How will the change in the manufacturer's SA and WA preferences affect the decisions of the chain members?
- How will the change in the manufacturer's SA and WA preferences affect the expected profits of the chain members and system?

To solve these problems, we build a Stackelberg game model composed of a retailer and a manufacturer, who is the leader and assume the demand is random. The retailer pursues its own expected profit maximization by making order decisions, while the manufacturer pursues its own expected utility maximization by setting wholesale pricing. We compare the decision-making behavior and expected profits of the chain members, and the expected profit of the system, under the four different behavioral preferences. At the same time, we also analyze the influence of the change in the SA and WA coefficients on the decision-making and expected profits of the chain members, as well as the expected profit of the system. The above conclusions can be used to some extent to guide decision-makers on how to adjust their decision-making strategies according to the manufacturer's SA and WA preferences.

This paper is organized as follows. Section 2 provides the assumptions and notations of the model. Section 3 presents the model analysis under the four kinds of behavioral preferences. Section 4 offers the comparison and analysis of the four cases. Section 5 provides the numerical study, and Section 6 summarizes the full text.

2. Assumptions and notations

2.1 Model assumptions

The mathematical model presented in this paper is based on the following assumptions:

Assumption 1: The entire SC consists of a manufacturer and a retailer, in which the manufacturer is the leader and the retailer is the follower.

Assumption 2: The cumulative distribution function and probability density function of random demand D are $F(x)$ and $f(x)$, respectively, and the expected demand is $E(x) = \mu$.

Assumption 3: Because market demand D is random, the retailer may overorder or underorder. To simplify the model, we ignore any salvage value or shortage loss. This assumption is common in the existing literature (e.g., [1, 4, 16, 20]) and does not have any effect on the conclusions.

Assumption 4: The manufacturer may have four behavioral preferences: RN, SA, WA, and SW.

Assumption 5: $p > w > c$.

2.2 Notations

The meanings of the other symbols used in this article are in Table 1.

Table 1 The listing of notations ($j \in \{RN, WA, SA, SW\}$)

Symbol	Meaning	Symbol	Meaning
w	Wholesale price	c	Production cost
p	Retail price	q	Order quantity
β	WA coefficient	λ	Stockout-aversion coefficient
RN	Risk-neutral	WA	Waste-averse
SA	Stockout-averse	SW	Stockout- and waste-averse
SC	Supply chain	π_R^j	The retailer's expected profit in case j
π_S^j	The supplier's expected profit in case j	π_C^j	The expected profit of supply chain in case j
U_S^j	The supplier's expected utility in case j	$*$	Optimal value

3. Model analysis

In this section, we will study the optimal decisions of the participants under the four behavioral preferences models. In the manufacturer-led Stackelberg game, the manufacturer has priority decision-making power and first determines the wholesale price w to maximize its expected utility. Then, the retailer determines its order quantity q in response to the manufacturer's decision-making and the predicted market demand D.

The expected profits of the manufacturer and the retailer are as follows:

$$\pi_S^j = (w - c)q \tag{1}$$

$$\pi_R^j(q) = pEmin(q, D) - wq = (p - w)q - p \int_0^q F(x)dx \tag{2}$$

where $j \in \{RN, WA, SA, SW\}$. Superscripts RN, WA, SA and SW indicate that the manufacturer has risk-neutral, waste-aversion, stockout-aversion and stockout- and waste-aversion preferences, respectively.

Case 1 With reference to Schweitzer and Cachon [1], when the manufacturer has SW preferences, the manufacturer's expected utility function is

$$U_S^{SW}(w) = wq - cq - E\{\beta(q - D)^+ - \lambda(D - q)^+\}$$

$$= (w - c + \lambda)q - \lambda\mu + (\beta + \lambda) \int_0^q F(x)dx \tag{3}$$

In the Eq. 3, wq is the product sales income of the manufacturer, cq is the production cost of the product, $\beta(q - D)^+$ is the waste penalty loss and $\lambda(D - q)^+$ is the shortage penalty loss. Among these variables, higher values of coefficients β and λ indicate higher degrees of WA and SA, respectively, by the manufacturer.

The inverse induction method is used to solve the model. From Eq. 2, we have $\frac{\partial^2 \pi_R^{SW}}{\partial q^2} = -pf(q) < 0$; then, π_R^{SW} is concave in q. Let $\frac{\partial \pi_R^{SW}}{\partial q} = 0$, and we can obtain the following:

$$q^{SW*} = F^{-1}\left(\frac{p-w}{p}\right) \tag{4}$$

By substituting q^{SW*} into Eq. 3, the manufacturer's expected utility optimization problem is

$$\max_w U_S^{SW} = (w - c + \lambda)q^{SW*} - \lambda\mu + (\beta + \lambda)\int_0^{q^{SW*}} F(x)dx \tag{5}$$

Let $\frac{\partial U_S^{SW}}{\partial q} = 0$, and we can obtain the following:

$$F^{-1}\left(\frac{p-w}{p}\right) - \frac{(w - c + \lambda)(p - v) + (p - w)(\lambda + \beta)}{(p-v)^2 f\left[F^{-1}\left(\frac{p-w}{p}\right)\right]} = 0 \tag{6}$$

Eq. 6 is too complex to obtain intuitionistic results, in order to gain more valuable conclusions, we follow prior literature [10, 20] and assume that the random demand variable obeys a uniform distribution, that is, $x \sim U(0, B)$. Through Eqs. 4 and 6, we can now obtain the following:

$$w^{SW*} = \frac{p(p + c + \beta)}{2p + \beta + \lambda}, \ q^{SW*} = \frac{B(p - c + \lambda)}{2p + \beta + \lambda} \tag{7}$$

By substituting Eq. 7 into Eqs. 1 and 2, we can obtain the following:

$$\pi_S^{SW*} = \frac{B(p - c + \lambda)[p(p + \beta) - c(p + \beta + \lambda)}{(2p + \beta + \lambda)^2}, \pi_R^{SW*} = \frac{Bp(p - c + \lambda)^2}{2(2p + \beta + \lambda)^2} \tag{8}$$

$$\pi_C^{SW*} = \pi_S^{SW*} + \pi_R^{SW*} = \frac{B(p - c + \lambda)[p(3p + 2\beta + \lambda) - c(3p + 2\beta + 2\lambda)]}{2(2p + \beta + \lambda)^2} \tag{9}$$

Case 2 When the manufacturer has only WA preferences, that is, $\lambda = 0$, then the optimal decisions and expected profits of the participants are shown in the third row of Table 2.

Case 3 When the manufacturer has only SA preferences, that is, $\beta = 0$, then the optimal decisions and expected profits of the participants are shown in the fourth row of Table 2.

Case 4 When the manufacturer has RN preferences, that is, $\lambda = 0$ and $\beta = 0$, then the optimal decisions and expected profits of the participants are shown in the fifth row of Table 2.

Table 2 Optimal decisions and expected profits of the manufacturer and retailer ($j \in \{RN, WA, SA, SW\}$)

Preferences	q^{j*}	w^{j*}	π_S^{j*}	π_R^{j*}
RN	$\dfrac{B(p-c)}{2p}$	$\dfrac{p+c}{2}$	$\dfrac{B(p-c)^2}{4p}$	$\dfrac{B(p-c)^2}{8p}$
WA	$\dfrac{B(p-c)}{2p+\beta}$	$\dfrac{p(p+c+\beta)}{2p+\beta}$	$\dfrac{B(p-c)^2(p+\beta)}{(2p+\beta)^2}$	$\dfrac{Bp(p-c)^2}{2(2p+\beta)^2}$
SA	$\dfrac{B(p-c+\lambda)}{2p+\lambda}$	$\dfrac{B(p+c)}{2p+\lambda}$	$\dfrac{B(p-c+\lambda)[p^2-c(p+\lambda)]}{(2p+\lambda)^2}$	$\dfrac{Bp(p-c+\lambda)^2}{2(2p+\lambda)^2}$
SW	$\dfrac{B(p-c+\lambda)}{2p+\beta+\lambda}$	$\dfrac{B(p+c+\beta)}{2p+\beta+\lambda}$	$\dfrac{B(p-c+\lambda)[(p-c)(p+\beta)-c\lambda]}{(2p+\beta+\lambda)^2}$	$\dfrac{Bp(p-c+\lambda)^2}{2(2p+\beta+\lambda)^2}$

Table 3 Optimal expected profit of the SC

π_C^{RN*}	$\dfrac{3B(p-c)^2}{8p}$
π_C^{WA*}	$\dfrac{B(p-c)^2(3p+2\beta)}{2(2p+\beta)^2}$
π_C^{SA*}	$\dfrac{B(p-c+\lambda)[p(3p+\lambda)-c(3p+2\lambda)]}{2(2p+\lambda)^2}$
π_C^{SW*}	$\dfrac{B(p-c+\lambda)[p(3p+2\beta+\lambda)-c(3p+2\lambda+2\beta)]}{2(2p+\lambda+\beta)^2}$

4. Comparison and analysis of the four behavioral preferences mode

This section mainly analyzes the conclusions obtained in the previous section. We analyze the influence of the change in SA coefficient λ and WA coefficient β on the decisions and expected profits of the participants, as well as the expected profit of the system. Additionally, the decisions and expected profits of the participants and the expected profit of the system are compared and analyzed under the four cases. The equilibrium results in the four cases are summarized in Tables 2 and 3.

We first analyze the effect of λ and β on the q^{SW*} and w^{SW*}. From Eq. 7, we can obtain

$$\frac{\partial q^{SW*}(\lambda,\beta)}{\partial \lambda} = \frac{B(p+\beta+c)}{(2p+\beta+\lambda)^2} > 0, \frac{\partial q^{SW*}(\lambda,\beta)}{\partial \beta} = -\frac{B(p+\lambda-c)}{(2p+\beta+\lambda)^2} < 0 \qquad (10)$$

$$\frac{\partial w^{SW*}(\lambda,\beta)}{\partial \lambda} = -\frac{p(p+\beta+c)}{(2p+\beta+\lambda)^2} < 0, \frac{\partial w^{SW*}(\lambda,\beta)}{\partial \beta} = \frac{p(p+\lambda-c)}{(2p+\beta+\lambda)^2} > 0 \qquad (11)$$

Therefore, we can obtain Proposition 1.

Proposition 1 (1) The retailer's order quantity increases with the λ, while it decreases with the β. (2) The manufacturer's wholesale price decreases with the λ, while it increases with the β.

Proposition 1 shows that the SA (WA) preferences motivates (weakens) the retailer's ordering enthusiasm. As λ (β) increases, the order quantity will increase (decrease). On the other hand, the dominant manufacturer will adjust the price of the product according to the ordering decisions of the retailer in order to maximize its expected utility. So as λ (β) increases, the wholesale price will decrease (increase).

Next, we will analyze the influence of λ and β on the expected profits of the participants. From Table 2, we can obtain

$$\frac{\partial \pi_R^{SW*}(\lambda,\beta)}{\partial \beta} = -\frac{Bp(p-c+\lambda)^2}{(2p+\beta+\lambda)^3} < 0, \frac{\partial \pi_R^{SW*}(\lambda,\beta)}{\partial \lambda} = \frac{Bp(p-c+\lambda)(p+c+\beta)}{(2p+\beta+\lambda)^3} > 0 \qquad (12)$$

$$\frac{\partial \pi_S^{SA*}(\lambda)}{\partial \lambda} = -\frac{B\lambda(p+c)^2}{(2p+\lambda)^3} < 0, \frac{\partial \pi_S^{WA*}(\beta)}{\partial \beta} = -\frac{B\beta(p-c)^2}{(2p+\beta)^3} < 0 \qquad (13)$$

$$\frac{\partial \pi_S^{SW*}(\lambda,\beta)}{\partial \beta} = \frac{B(p-c+\lambda)[p(\lambda-\beta)+c(\lambda+\beta)]}{(2p+\beta+\lambda)^3} \qquad (14)$$

$$\frac{\partial \pi_S^{SW*}(\lambda,\beta)}{\partial \lambda} = \frac{B(p+c+\beta)[p(\beta-\lambda)-c(\lambda+\beta)]}{(2p+\beta+\lambda)^3} \qquad (15)$$

From Eq. 14, there are $\frac{\partial \pi_S^{SW*}}{\partial \beta} > 0$ when $\lambda > \lambda_1$ and $\frac{\partial \pi_S^{SW*}}{\partial \beta} < 0$ when $0 \leq \lambda < \lambda_1$, where $\lambda_1 = \frac{\beta(p-c)}{p+c}$. From Eq. 15, there are $\frac{\partial \pi_S^{SW*}}{\partial \lambda} > 0$ when $\beta > \beta_1$ and $\frac{\partial \pi_S^{SW*}}{\partial \lambda} < 0$ when $0 \leq \beta < \beta_1$, where $\beta_1 = \frac{\lambda(p-c)}{p+c}$.

Through the above analysis, we can obtain the following proposition.

Proposition 2 (1) The retailer's expected profit decreases and increases with the β and the λ, respectively. (2) When the manufacturer has only WA (SA) preferences, the manufacturer's expected profit is decreases with the β and the λ. (3) When the manufacturer has SW preferences, the relationship between the manufacturer's expected profit and the β (λ) is determined by the λ (β): there is a threshold λ_1 (β_1), and when the value of λ (β) is large, that is, $\lambda > \lambda_1$ ($\beta > \beta_1$), the manufacturer's expected profit increases with the β (λ); conversely, the manufacturer's expected profit decreases with the β (λ).

As λ (β) increases, the retailer's expected profit increases (decreases), which is due to an increase (decrease) in order volume. When the manufacturer has only WA (SA) preferences, it is understandable that the manufacturer's expected profit decreases as the order quantities de-

crease. However, the manufacturer's expected profit decreases as order volume increases, which seems counterintuitive; this is because the gains from the increase in order quantity are not sufficient to compensate for the loss caused by the reduction in product price. In addition, the relationship between manufacturer's expected profit and the λ (β) is related to the β (λ) in the SW model.

From Table 3, we can obtain

$$\frac{\partial \pi_C^{SW*}(\lambda, \beta)}{\partial \lambda} = \frac{B(p + c + \beta)[p(\beta + p) - c(p + \lambda + \beta)]}{(2p + \beta + \lambda)^3} \tag{16}$$

$$\frac{\partial \pi_C^{SW*}(\lambda, \beta)}{\partial \beta} = -\frac{B(p - c + \lambda)[p(\beta + p) - c(p + \lambda + \beta)]}{(2p + \beta + \lambda)^3} \tag{17}$$

From Eq. 16, there are $\frac{\partial \pi_C^{SW*}}{\partial \lambda} \geq 0$ when $\beta \geq \beta_2$ and $\frac{\partial \pi_C^{SW*}}{\partial \lambda} < 0$ when $0 \leq \beta < \beta_2$, where $\beta_2 = \frac{c(p+\lambda)-p^2}{p-c}$. From Eq. 17, there are $\frac{\partial \pi_C^{SW*}}{\partial \beta} \geq 0$ when $\lambda \geq \lambda_2$ and $\frac{\partial \pi_C^{SW*}}{\partial \beta} < 0$ when $0 < \lambda < \lambda_2$, where $\lambda_2 = \frac{(p-c)(p+\beta)}{c}$.

Since the decisions and profits of the participants must be positive, the following inequalities are established:

$$\begin{cases} p(p + \beta) - c(p + \beta + \lambda) > 0 \\ p^2 - c(p + \lambda) > 0 \end{cases} \tag{18}$$

$$\frac{\partial \pi_C^{SA*}}{\partial \lambda} = \frac{B(p + c)[p^2 - c(p + \lambda)]}{(2p + \lambda)^3} > 0, \frac{\partial \pi_C^{WA*}}{\partial \beta} = -\frac{B(p - c)^2(p + \beta)}{(2p + \lambda)^3} < 0 \tag{19}$$

From the above analysis, we can obtain Proposition 3.

Proposition 3 (1) When the manufacturer has only SA (WA) preferences, the expected profit of SC increases (decreases) with the λ (β). (2) When the manufacturer has SW preferences, the relationship between the expected profit of SC and the β (λ) is determined by λ(β). There is a threshold λ_2 ($\beta > \beta_2$), and when the value of λ (β) is large, that is, $\lambda > \lambda_2$ ($\beta > \beta_2$), the expected profit of SC increases with the β (λ); in contrast, the expected profit of SC decreases with the β (λ).

Proposition 3 shows that the system profit is increasing (decreasing) with the λ (β) in the SA (WA) model. Because for the whole system, the order quantity increases (decreases) with the increase (decrease) in λ (β), which can (cannot) satisfy the potential market demand as much as possible. In addition, if λ (β) is larger, there will be a risk of overordering or underordering in the SW model. Therefore, reducing (increasing) the order amount is beneficial to increase the profit of the system.

From Eqs. 10 and 11, we can obtain

$$\frac{\partial w^{SW*}(\lambda, \beta)}{\partial \lambda} < 0 \Rightarrow \left. \begin{cases} w^{SW*}(\lambda, \beta) < w^{SW*}(0, \beta) = w^{WA*}(\beta) \\ w^{SW*}(\lambda, 0) = w^{SA*}(\lambda) < w^{SA*}(0) = w^{RN*} \end{cases} \right\}$$

$$\frac{\partial w^{SW*}(\lambda, \beta)}{\partial \beta} > 0 \Rightarrow \left. \begin{cases} w^{SW*}(\lambda, \beta) < w^{SW*}(\lambda, 0) = w^{SA*}(\lambda) \\ w^{SW*}(0, \beta) = w^{WA*}(\beta) > w^{WA*}(0) = w^{RN*} \end{cases} \right\}$$

$$\Rightarrow w^{SW*} < w^{SA*} < w^{RN*} < w^{WA*}$$

$$\frac{\partial q^{SW*}(\lambda, \beta)}{\partial \lambda} > 0 \Rightarrow \left. \begin{cases} q^{SW*}(\lambda, \beta) > q^{SW*}(0, \beta) = q^{WA*}(\beta) \\ q^{SW*}(\lambda, 0) = q^{SA*}(\lambda) > q^{SA*}(0) = q^{RN*} \end{cases} \right\}$$

$$\frac{\partial q^{SW*}(\lambda, \beta)}{\partial \beta} < 0 \Rightarrow \left. \begin{cases} q^{SW*}(\lambda, \beta) < q^{SW*}(\lambda, 0) = q^{SA*}(\lambda) \\ q^{SW*}(0, \beta) = q^{WA*}(\beta) < q^{WA*}(0) = q^{RN*} \end{cases} \right\}$$

$$\Rightarrow q^{WA*} < q^{SW*} \text{ or } q^{RN*} < q^{SA*}$$

Let $m = \lambda/\beta$, then

$$q^{SW*} - q^{RN*} = \frac{B\beta[c(1+m) + p(m-1)]}{2p(2p + m\beta + \beta)} \tag{20}$$

From Eq. 20, there are $q^{SW*} > q^{RN*}$ when $m > m_1$ and $q^{SW*} < q^{RN*}$ when $0 \leq m < m_1$, where $m_1 = \frac{p-c}{p+c}$.

From the above analysis, we can obtain Proposition 4.

Proposition 4

(1) $w^{SW*} < w^{SA*} < w^{RN*} < w^{WA*}$; (2) $\begin{cases} q^{WA*} < q^{RN*} < q^{SW*} < q^{SA*}, m > m_1 \\ q^{WA*} < q^{SW*} < q^{RN*} < q^{SA*}, 0 \leq m < m_1 \end{cases}$

where $m_1 = \frac{p-c}{p+c}$.

Proposition 4 indicates that (1) the manufacturer's wholesale price is the highest in the WA model, followed by the RN, SA and SW models, in that order. (2) As shown in Fig. 1, the retailer's order quantity is the largest and the smallest in the SA and WA models respectively. However, the size of the order quantity between the RN and SW models depends on the ratio m. There is a threshold m_1, and when the ratio m is large, that is, $m > m_1$, the order quantity is larger in the SW model; otherwise, the order quantity is larger in the RN model.
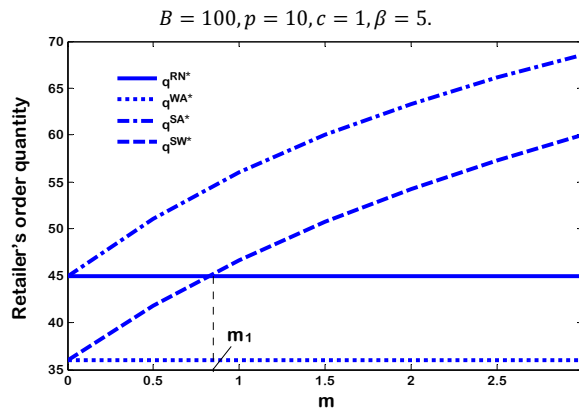

Fig. 1 The effect of ratio m on the retailer's order quantity

From the Eq. 12, we can obtain the following:

$$\frac{\partial \pi_R^{SW*}(\lambda, \beta)}{\partial \beta} < 0 \Rightarrow \begin{cases} \pi_R^{SW*}(\lambda, \beta) < \pi_R^{SW*}(\lambda, 0) = \pi_R^{SA*} \\ \pi_R^{SW*}(0, \beta) = \pi_R^{WA*}(\beta) < \pi_R^{WA*}(0) = \pi_R^{RN*} \end{cases}$$

$$\frac{\partial \pi_R^{SW*}(\lambda, \beta)}{\partial \lambda} > 0 \Rightarrow \begin{cases} \pi_R^{SW*}(\lambda, \beta) > \pi_R^{SW*}(0, \beta) = \pi_R^{WA*}(\beta) \\ \pi_R^{SW*}(\lambda, 0) = \pi_R^{SA*}(\lambda) > \pi_R^{SA*}(0) = \pi_R^{RN*} \end{cases}$$

$$\Rightarrow \pi_R^{WA*} < \pi_R^{RN*} \text{ or } \pi_R^{SW*} < \pi_R^{SA*}$$

Let $m = \lambda/\beta$, then

$$\pi_R^{SW*} - \pi_R^{RN*} = \frac{4Bp^2(p - c + m\beta)^2 - B(p - c)^2(2p + \beta + m\beta)^2}{8p(2p + m\beta + \beta)^2} \tag{21}$$

From Eq. 21, there are $\pi_R^{SW*} > \pi_R^{RN*}$ when $m > m_1$ and $\pi_R^{SW*} < \pi_R^{RN*}$ when $0 \leq m < m_1$, where $m_1 = \frac{p-c}{p+c}$.

Proposition 5 $\begin{cases} \pi_R^{WA*} < \pi_R^{RN*} < \pi_R^{SW*} < \pi_R^{SA*}, m > m_1 \\ \pi_R^{WA*} < \pi_R^{SW*} < \pi_R^{RN*} < \pi_R^{SA*}, 0 \leq m < m_1 \end{cases}$

As shown in Fig. 2(b), Proposition 5 shows that the retailer's expected profit is the highest and lowest in the SA and WA models respectively. However, the size of the expected profit between the RN and SW models depends on the ratio m. There is a threshold m_1, and when the ratio m is large, that is, $m > m_1$, the expected profit is larger in the SW model; conversely, the expected profit is larger in the RN model.
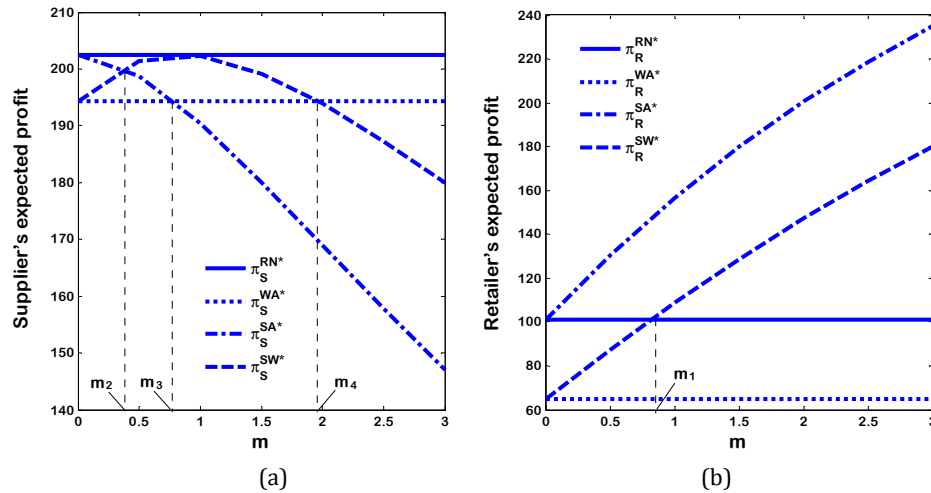
Fig. 2 The effect of ratio m on the expected profit of manufacturer (a) and retailer (b)

Proposition 6
$$\begin{cases} \pi_S^{WA*} < \pi_S^{SW*} < \pi_S^{SA*} < \pi_S^{RN*}, \ 0 \le m < m_2 \\ \pi_S^{WA*} < \pi_S^{SA*} < \pi_S^{SW*} < \pi_S^{RN*}, \ m_2 < m < m_3 \\ \pi_S^{SA*} < \pi_S^{WA*} < \pi_S^{SW*} < \pi_S^{RN*}, \ m_3 < m < m_4 \\ \pi_S^{SA*} < \pi_S^{SW*} < \pi_S^{WA*} < \pi_S^{RN*}, \ m > m_4 \end{cases}$$

where $m_2 = \dfrac{-2p^2 - c(2p+\beta) + \sqrt{4p^3(p+\beta) + 4cp^2(2p+\beta) + c^2(4p^2+\beta^2)}}{2\beta(p+c)}$, $m_3 = \dfrac{p(p-c)}{p^2+c(p+\beta)}$, $m_4 = \dfrac{(2p+\beta)(p-c)}{p^2+c(p+\beta)}$.

The proof for Preposition 6 is in Appendix A.

As shown in Fig. 2(a), Proposition 6 indicates that the manufacturer's expected profit is highest in the RN model. However, the expected profit difference between the *WA*, *SW* and *SA* models depends on the ratio m. There are three thresholds, m_2, m_3 and m_4, where $m_2 < m_3 < m_4$. When $0 \le m < m_2$, the expected profit is the highest and lowest in the *SA* and *WA* models, respectively. When $m_2 < m < m_3$, the expected profit is the highest and lowest in the *SW* and *WA* models, respectively. When $m_3 < m < m_4$, the expected profit is the highest and lowest in the *SW* and *SA* models, respectively. When $m > m_4$, the expected profit is the highest and lowest in the *WA* and *SA* models, respectively.

Proposition 7
$$\begin{cases} \pi_C^{WA*} < \pi_C^{SW*} < \pi_C^{RN*} < \pi_C^{SA*}, \ 0 \le m < m_5 \\ \pi_C^{WA*} < \pi_C^{RN*} < \pi_C^{SW*} < \pi_C^{SA*}, \ m_5 < m < m_6 \\ \pi_C^{WA*} < \pi_C^{RN*} < \pi_C^{SA*} < \pi_C^{SW*}, \ m > m_6 \end{cases}$$

where $m_5 = \dfrac{p-c}{p+c}$ and $m_6 = \dfrac{2p^2 + p\beta - 2c(3p+\beta) + \sqrt{p^2(2p+\beta)^2 + 4cp(2p^2+p\beta-\beta^2) + 4c^2(p^2+\beta^2)}}{4c\beta}$.

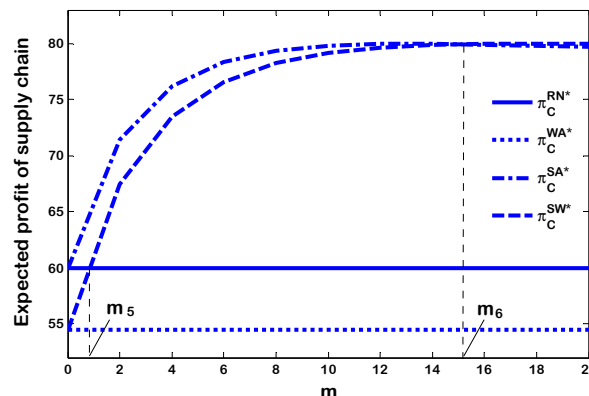The proof for Preposition 7 is in Appendix A.

Fig. 3 The effect of ratio m on the expected profit of SC

As shown in Fig. 3, Proposition 7 shows that the expected profit of SC is lowest in the WA model. However, the expected profit difference between the RN, SW and SA models depends on the ratio m. There are two thresholds, m_5 and m_6, where $m_5 < m_6$. When $0 \leq m < m_5$, the expected profit is the highest and lowest in the SA and SW models, respectively. When $m_5 < m < m_6$, the expected profit is the highest and lowest in the SA and RN models, respectively. When $m > m_6$, the expected profit is the highest and lowest in the SW and RN models, respectively.

5. Results and discussion

A numerical simulation was used to intuitively show the impact of the λ and the β on the decisions and expected profits of the chain members and the expected profit of the system. At the same time, we also want to verify the correctness of the above conclusions. Therefore, this section uses a numerical simulation method for further analysis. Without the loss of generality, the parameters are set as follows: $B = 100, p = 10, c = 1$.

From Fig. 4(a), we can see that the manufacturer's wholesale price is decreasing with the λ and increasing with the β. It can be seen from Fig. 4(b) that the retailer's order quantity is increasing with the λ and decreasing with the β.

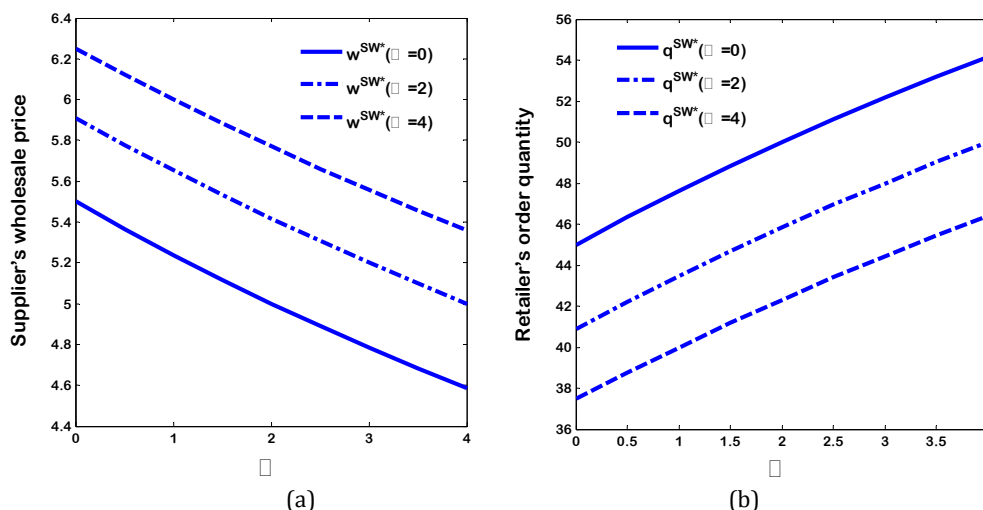(a)   (b)

Fig. 4 The effect of λ and β on wholesale price (a), and order quantity (b)

From Fig. 5(a), we can see that the retailer's expected profit is increasing with the λ and decreasing with the β. Fig. 5(b) shows that when the manufacturer has only SA (WA) preferences, the manufacturer's expected profit is decreasing with the λ (β).

From Figs. 6(a) and 7(a), we can see that when the manufacturer has SW preferences, the relationship between the expected profit of the manufacturer (SC) and the λ depends on the β. If β is large, that is, $\beta = 4.0$, then the expected profit is increasing with λ; in contrast, if β is small, that is, $\beta = 0.5$, then the expected profit is decreasing with λ.

Figs. 6(b) and 7(b) show that the relationship between the expected profit of the manufacturer (SC) and the β depends on the λ. If λ is large, that is, $\lambda = 3.0$, then the expected profit is increasing with β; in contrast, if λ is small, that is, $\lambda = 0.4$, then the expected profit is decreasing with β.

Fig. 8 shows that when the manufacturer has only the SA (WA) preferences, the expected profit of the SC is increasing (decreasing) with the λ (β).
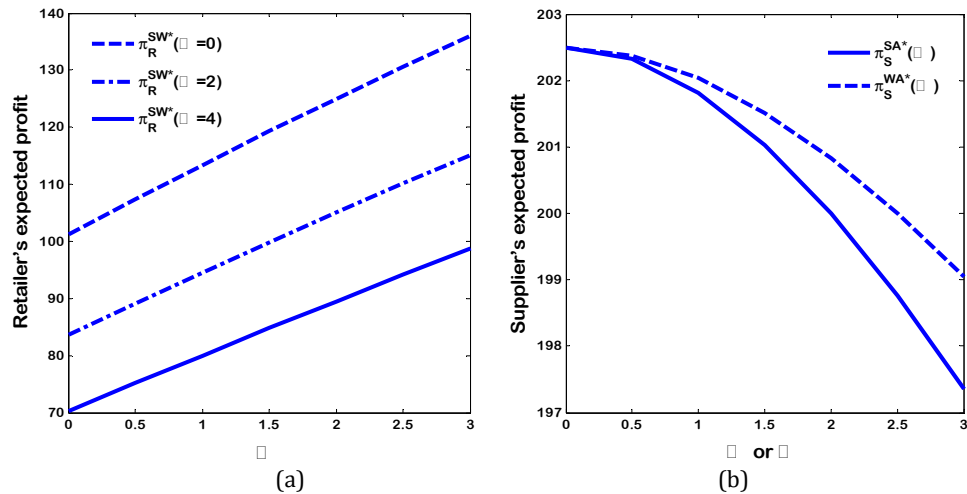
(a)   (b)

Fig. 5 The effect of λ and β on the expected profits of retailer (a) and manufacturer (b)
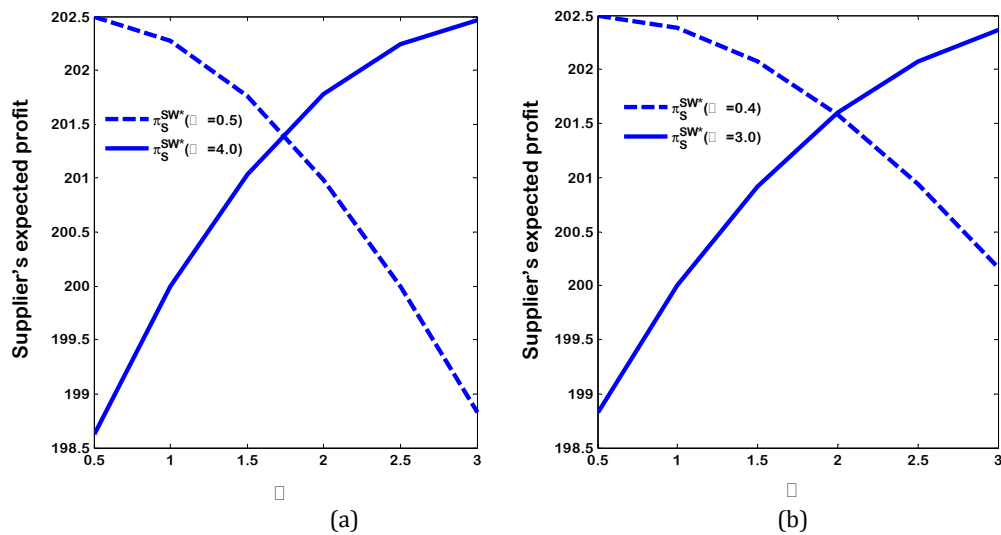
(a)   (b)

Fig. 6 The effect of λ and β on the expected profit of the manufacturer

(a)   (b)
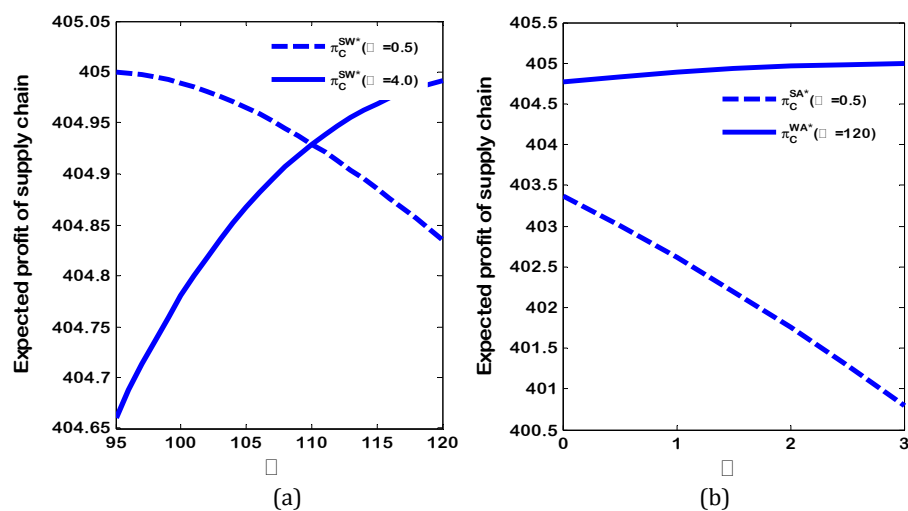
Fig. 7 The effect of λ and β on the expected profit of the SC
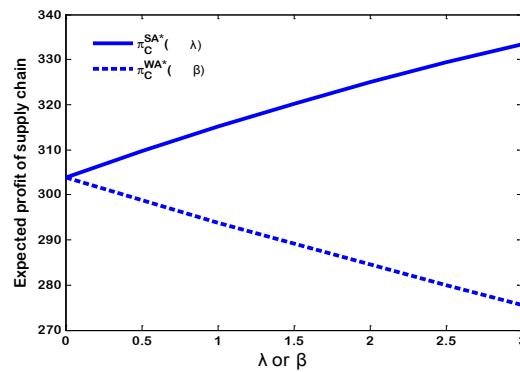
Fig. 8 The effect of λ or β on the expected profit of the SC

6. Conclusion

This paper studies the influence of a dominant manufacturer's behavioral preferences on the decision-making and expected profits of the chain members, as well as the expected profit of the system. The behavioral preferences investigated include RN, SA, WA and SW preferences. Through the analysis of this paper, we can draw the following conclusions:

- The manufacturer's wholesale price is increasing (decreasing) with the SA (WA) coefficient, while the retailer's order quantity (expected profit) is the opposite;
- When the manufacturer has only the SA (WA) preferences, the manufacturer's expected profit decreases with the SA (WA) coefficient;
- When the manufacturer has the SW preferences, the relationship between the expected profit of the manufacturer (SC) and the SA coefficient depends on the WA coefficient. The relationship between the expected profit of the manufacturer (SC) and the WA coefficient depends on the SA coefficient;
- The manufacturer's wholesale price is the highest in the WA model, followed by the RN, SA and SW models, in that order;
- The retailer's order quantity (expected profit) is the largest and smallest in the SA and WA models, respectively, while the size of the order quantity (expected profit) between the RN and SW models depends on the ratio m, and there is a ratio threshold of m_1;
- The manufacturer's expected profit is the largest in the RN model. However, the size of the expected profit among the WA, SW and SA models depends on the ratio m, and there are three ratio thresholds of m_2, m_3 and m_4, where $m_2 < m_3 < m_4$;
- The expected profit of the SC is the lowest in the WA model. However, the size of the expected profit among the RN, SW and SA models depends on the ratio m, and there are two ratio thresholds of m_5 and m_6, where $m_5 < m_6$.

Acknowledgement

References

[1] Schweitzer, M.E., Cachon, G.P. (2000). Decision bias in the newsvendor problem with a known demand distribtion: Experimental evidence, *Management Science*, Vol. 46, No. 3, 404-420, doi: 10.1287/mnsc.46.3.404. 12070.

[2] Jian, M., Fang, X., Jin, L.-Q., Rajapov, A. (2015). The impact of lead time compression on demand forecasting risk and production cost: A newsvendor model, *Transportation Research Part E: Logistics and Transportation Review,* Vol. 84, 61-72, doi: 10.1016/j.tre.2015.10.006.

[3] Zhu, X.D., Li, B.Y., Wang, Z. (2017). A study on the manufacturing decision-making and optimization of hybrid-channel supply chain for original equipment manufacturer, *Advances in Production Engineering & Management,* Vol. 12, No. 2, 185-195, doi: 10.14743/apem2017.2.250.

[4] Niu, B., Chen, L., Zhang, J. (2017). Sustainability analysis of supply chains with fashion products under alterntive power structures and loss-averse supplier, *Sustainability,* Vol. 9, No. 6, 1-19, <u>doi: 10.3390/su9060995</u>.

[5] Choi, S. (2018). A loss-averse newsvendor with cap-and-trade carbon emissions regulation, *Sustainability,* Vol. 10, No. 7, 1-12, <u>doi: 10.3390/su10072126</u>.

[6] Li, X., Li, Y. (2016). On the loss-averse dual-sourcing problem under supply disruption, *Computers & Operations Research,* Vol. 100, 301-313, <u>doi: 10.1016/j.cor.2016.12.011</u>.

[7] Xu, X., Wang, H., Dang, C., Ji, P. (2017). The loss-averse newsvendor model with backordering, *International Journal of Production Economics,* Vol. 188, 1-10, <u>doi: 10.1016/j.ijpe.2017.03.005</u>.

[8] Yang, D., Xiao, T. (2017). Coordination of a supply chain with loss-averse consumers in service quality, *International Journal of Production Research,* Vol. 55, No. 12, 3411-3430, <u>doi: 10.1080/00207543.2016.1241444</u>.

[9] Giri, B.C. (2011). Managing inventory with two suppliers under yield uncertainty and risk aversion, *International Journal of Production Economics,* Vol. 133, No. 1, 80-85, <u>doi: 10.1016/j.ijpe.2010.09.015</u>.

[10] Zhu, L., Ren, X., Lee, C., Zhang, Y. (2017). Coordination contracts in a dual-channel supply chain with a risk-averse retailer, *Sustainability,* Vol. 9, No. 11, 1-21, <u>doi: 10.3390/su9112148</u>.

[11] Wu, M., Zhu, S.X., Teunter, R.H. (2013). The risk-averse newsvendor problem with random capacity, *European Journal of Operational Research,* Vol. 231, No. 2, 328-336, <u>doi: 10.1016/j.ejor.2013.05.044</u>.

[12] Oh, S., Rhodes, J., Strong, R. (2016). Impact of cost uncertainty on pricing decisions under risk aversion, *European Journal of Operational Research,* Vol. 253, No. 1, 144-153, <u>doi: 10.1016/j.ejor.2016.02.034</u>.

[13] Egging, R., Pichler, A., Kalvø, Ø.I., Walle-Hansen, T.M. (2017). Risk aversion in imperfect natural gas markets, *European Journal of Operational Research,* Vol. 259, No. 1, 367-383, <u>doi: 10.1016/j.ejor.2016.10.020</u>.

[14] Xiao, T., Choi, T.M., Yang, D., Cheng, T.C.E. (2012). Service commitment strategy and pricing decisions in retail supply chains with risk-averse players, *Service Science,* Vol. 4, No. 3, 236-252, <u>doi: 10.1287/serv.1120.0021</u>.

[15] Zhou, Y.-W., Li, J., Zhong, Y. (2018). Cooperative advertising and ordering policies in a two-echelon supply chain with risk-averse agents, *Omega,* Vol. 75, 97-117, <u>doi: 10.1016/j.omega.2017.02.005</u>.

[16] Zheng, W., Li, B., Song, D.-P. (2017). Effects of risk-aversion on competing shipping lines' pricing strategies with uncertain demands, *Transportation Research Part B: Methodological,* Vol. 104, 337-356, <u>doi: 10.1016/j.trb.2017.08.004</u>.

[17] Liu, M., Cao, E., Salifou, C.K. (2016). Pricing strategies of a dual-channel supply chain with risk aversion, *Transportation Research Part E: Logistics and Transportation Review,* Vol. 90, 108-120, <u>doi: 10.1016/j.tre.2015.11.007</u>.

[18] Downward, A., Young, D., Zakeri, G. (2016). Electricity retail contracting under risk-aversion, *European Journal of Operational Research,* Vol. 251, No. 3, 846-859, <u>doi: 10.1016/j.ejor.2015.11.040</u>.

[19] Ohmura, S., Matsuo, H. (2016). The effect of risk aversion on distribution channel contracts: Implications for return policies, *International Journal of Production Economics,* Vol. 176, 29-40, <u>doi: 10.1016/j.ijpe.2016.02.019</u>.

[20] Li, B., Chen, P., Li, Q., Wang, W. (2014). Dual-channel supply chain pricing decisions with a risk-averse retailer, *International Journal of Production Research,* Vol. 52, No. 23, 7132-7147, <u>doi: 10.1080/00207543.2014.939235</u>.

Appendix A

Proof of Proposition 6

Let $m = \lambda/\beta$; from Eq. 13 and Table 2, we can obtain

$$
\left.
\begin{array}{l}
\dfrac{\partial \pi_S^{SA*}(\lambda)}{\partial \lambda} < 0 \Rightarrow \pi_S^{SA*}(\lambda) < \pi_S^{SA*}(0) = \pi_S^{RN*} \\[2mm]
\dfrac{\partial \pi_S^{WA*}(\beta)}{\partial \beta} < 0 \Rightarrow \pi_S^{WA*}(\beta) < \pi_S^{WA*}(0) = \pi_S^{RN*} \\[2mm]
\pi_S^{SW*} - \pi_S^{RN*} = -\dfrac{B\beta^2[c(1+m)-p(1-m)]^2}{4p(2p+\beta+m\beta)^2} < 0
\end{array}
\right\} \tag{A1}
$$

$$
\Rightarrow max\{\pi_S^{SA*}, \pi_S^{WA*}, \pi_S^{SW*}, \pi_S^{RN*}\} = \pi_S^{RN*}
$$

$$
\pi_S^{SW*} - \pi_S^{SA*} = -\frac{B\beta^2(p-c+m\beta)\{p[p(2m-1)+m\beta^2]+c[p+2mp+m\beta(1+m)]\}}{(2p+m\beta)^2(2p+m\beta+\beta)^2} \tag{A2}
$$

$$
\pi_S^{WA*} - \pi_S^{SA*} = \frac{B(p-c)^2(p+\beta)}{(2p+\beta)^2} - \frac{B(p-c+m\beta)[p^2-c(p+m\beta)]}{(2p+m\beta)^2} \tag{A3}
$$

$$
\pi_S^{SW*} - \pi_S^{WA*} = \frac{B(p-c+m\beta)[(p-c)(p+\beta)-cm\beta]}{(2p+\beta+m\beta)^2} - \frac{B(p-c)^2(p+\beta)}{(2p+\beta)^2} \tag{A4}
$$

From Eqs. A1 to A4, we can obtain Proposition 6. Where m_2, m_3, and m_4 satisfy Eqs. A2, A3, and A4, respectively.

Additionally, $m_3 = \frac{p(p-c)}{p^2+c(p+\beta)}$, $m_2 = \frac{-2p^2-c(2p+\beta)+\sqrt{4p^3(p+\beta)+4cp^2(2p+\beta)+c^2(4p^2+\beta^2)}}{2\beta(p+c)}$, and $m_4 = \frac{(2p+\beta)(p-c)}{p^2+c(p+\beta)}$.

Proof of Proposition 7

Let $m = \lambda/\beta$; from Eqs. 19 and Table 3, we can obtain

$$\pi_C^{WA*} - \pi_C^{SW*} = \frac{B(p-c)^2(3p+2\beta)}{2(2p+\beta)^2} - \frac{B(p-c+m\beta)[p(3p+2\beta+m\beta)-c(3p+2m\beta+2\beta)]}{2(2p+m\beta+\beta)^2}$$

$$\frac{\partial(\pi_C^{WA*}-\pi_C^{SW*})}{\partial m} = -\frac{B\beta(p+c+\beta)[p(p+\beta)-c(p+m\beta+\beta)]}{(2p+m\beta+\beta)^3} < 0$$

$$\Rightarrow max\{\pi_C^{WA*}-\pi_C^{SW*}\} = \left(\pi_C^{WA*}-\pi_C^{SW*}\right)_{m=0} = 0 \Rightarrow \pi_C^{WA*} < \pi_C^{SW*} \tag{A5}$$

$$\left. \begin{array}{l} \frac{\partial \pi_C^{SA*}(\lambda)}{\partial \lambda} > 0 \Rightarrow \pi_C^{SA*}(\lambda) > \pi_C^{SA*}(0) = \pi_C^{RN*} \\ \frac{\partial \pi_C^{WA*}(\beta)}{\partial \beta} < 0 \Rightarrow \pi_C^{WA*}(\beta) < \pi_C^{WA*}(0) = \pi_C^{RN*} \end{array} \right\} \Rightarrow \pi_C^{WA*} < \pi_C^{RN*} < \pi_C^{SA*} \tag{A6}$$

$$\pi_C^{RN*} - \pi_C^{SW*} = \frac{B\beta[c(1+m)-p(1-m)]\{c[4p+3\beta(1+m)]-p[4p+\beta(3+m)]\}}{8p(2p+m\beta+\beta)^2} \tag{A7}$$

$$\pi_C^{SA*} - \pi_C^{SW*} = \frac{\begin{bmatrix} B\beta(p-c+m\beta)\{p[4p^2+p\beta(3+2m)+m\beta^2] \\ -c[4p^2+2m\beta^2(1+m)+3p(\beta+2m\beta)]\} \end{bmatrix}}{2(2p+m\beta)^2(2p+m\beta+\beta)^2} \tag{A8}$$

From Eqs. A5 to A8, we can obtain

$$\begin{cases} \pi_C^{WA*} < \pi_C^{SW*} < \pi_C^{RN*} < \pi_C^{SA*}, \ 0 \le m < m_5 \\ \pi_C^{WA*} < \pi_C^{RN*} < \pi_C^{SW*} < \pi_C^{SA*}, \ m_5 < m < m_6 \\ \pi_C^{WA*} < \pi_C^{RN*} < \pi_C^{SA*} < \pi_C^{SW*}, \ m > m_6 \end{cases}$$

Where m_5 and m_6 satisfy Eqs. A7 and A8, respectively. Additionally, $m_5 = \frac{p-c}{p+c}$ and $m_6 = \frac{2p^2+p\beta-2c(3p+\beta)+\sqrt{p^2(2p+\beta)^2+4cp(2p^2+p\beta-\beta^2)+4c^2(p^2+\beta^2)}}{4c\beta}$.

Genetic programming method for modelling of cup height in deep drawing process

Gusel, L.[a,*], Boskovic, V.[b], Domitner, J.[b], Ficko, M.[a], Brezocnik, M.[a]

[a]University of Maribor, Faculty of Mechanical Engineering, Maribor, Slovenia
[b]Graz University of Technology, Institute of Materials Science, Joining and Forming, Graz, Austria

A B S T R A C T

Genetic programming method for modelling of maximum height of deep drawn high strength sheet materials is proposed in this paper. Genetic programming (GP) is an evolutionary computation approach which uses the principles of Darwin's natural selection to develop effective solutions for different problems. The aim of the research was the modelling of cylindrical cup height in deep drawing process and analysis of the impact of process parameters on material formability. High strength steel sheet materials (DP1180HD and DP780) were formed by deep drawing using different punch speeds and blank holder forces. The heights of specimens before cracks occur were measured. Therefore, four input parameters (yield stress, tensile strength, blank holder force, punch speed) and one output parameter (cup height) were used in the research. The experimental data were the basis for obtaining various accurate prediction models for the cup heights by the genetic programming method. Results showed that proposed genetic modelling method can successfully predict fracture problems in a process of deep drawing.

1. Introduction

Deep-drawing processes are frequently used in the sheet metal forming industry, because of the achievable formability. In deep drawing the sheet blank is deformed by tensile and compressive loads in different directions of action. The sheet thickness should remain as constant as possible. Many parameters influence the deep drawing process and should be carefully selected for effective and economical production. Among them, the influence of punch speed and blank holding force are of great importance, since the shortened process duration leads to an increase in productivity. Therefore, gaining knowledge of the influence of these parameters is of great interest when achieving the greatest possible productivity in producing fault-free parts. Accurate models, which describe the influences of different parameters in deep drawing, can be obtained by many different modelling methods, and serve for optimization and prediction purposes in the process.

In widely used deterministic methods like multi regression method, the form and size of a model is determined in advance. These modelling methods have one common feature: all of them optimize a given model of a problem. That is, each genotype determines a particular combination of values of free variables of the model, and the interpretation of those variables within a model is fixed. When applied to a different problem instances, they often turn out to be infeasible or obtain much worse fitness. Much better results are often achieved by non-deterministic

modelling methods, such as genetic programming method (GP). In GP the model is not known in advance, and only the constraints of the model are given, e.g., instruction set, the maximum size of organisms etc. [1]. Hence GP creates and optimizes entire model of the problem. In other words, GP evolves an executable code that inputs a given problem instance and outputs a solution to the instance. In this sense GP's solutions are active, i.e. adapt to the given problem instance [2].

Majority of papers dealing with modelling methods in metal forming processes, describing neural network and genetic algorithms (GA) methods while only a few are dealing with GP. In paper [3] authors describe prediction and optimization of sealing cover thinning in deep drawing process by using GA as very accurate and effective method. In [4] GA have been used to develop an optimization strategy for choosing blank holder force and punch force for enable fracture free and wrinkle-free production of deep drawn cups while in [5] optimum blank shape and process parameters were defined and calculated for deep drawing process using Taguchi optimization method and GA. Pareto optimal solution search techniques in GA to reduce excessive thinning and wrinkling occurrence was described in [6] and [7]. An evolutionary structural optimization method for the sheet blank shape optimization in deep drawing process was presented in [8]. Many authors have used combination of neural network and GA approach for modelling and predicting deep drawing parameters and some properties of drawn products [9, 10]. In [11], the effect of hydro-mechanical deep drawing process parameters was investigated by FE simulations and neuro-fuzzy modelling method to predict the maximum thinning of sheet material, while in [12] a fuzzy control algorithm for optimization of loading profiles and drawing ratio was proposed.

An overview of recent applications of evolutionary computing in manufacturing industry was presented in [13]. Very few papers (and especially sheet forming processes) dealing with modelling by the GP modelling method. In papers [14, 15] authors compared genetic algorithm models and GP models for the distribution of effective strain and stress and also some mechanical properties in forward extruded alloy. GP method performed well in developing more accurate genetic models. In [16] authors described the application of bi-objective GP modelling for prediction of the size of austenite grain as a function of heating time. GP method was also used very successfully for modelling of bending capability of titan-zinc sheet in [17]. Different bending parameters and their impact on bending capability were analysed and optimized and a variety of accurate models were developed by GP.

This paper proposed a GP modelling of maximum height of cup deep drawn high strength steel specimens. Maximum height achieved in deep drawing process depends on many parameters and is excellent indicator for formability level of chosen material. Experimental data measured during the deep drawing processes served as an environment for evolution process in genetic programming. Blank holder force, punch speed, tensile and yield stress were used as independent input variables, while maximum height of deep drawn specimens before first cracks occur was a dependent output variable.

2. Materials and methods

2.1 Genetic programming

Genetic programming is a sub brunch of evolutionary computation (EC) emulating the natural evolution of species and is probably the most general approach among EC methods. To imitate the evolutionary process in GP, certain components must be defined, such as mathematical functions, problem variables and genetic operations like reproduction, mutation or crossover [1].

The GP process starts with random generation of initial population of organisms (computer programmes). Terminal genes, such as variables and constants, and function genes (arithmetic operations) compose each programme and must be carefully selected. After that the fitness function must be determined for evaluation of programme adaptation to the environment (e.g., to the experimental data). Adaptation is a main force in natural selection. The change and improvement of programmes fitness during GP process is enabled by genetic operations, such as

reproduction, mutation and crossover. The right selection of genetic operations and their probability is of vital importance for successful GP process (and often vary regarding the problem to be solved), because genetic operations provide an increasing diversity and genetic exchange among computer programmes. The mutation also introduces new code fragments into population and is used as a common workaround for loose of diversity and stagnation, especially in small populations.

The last step of the process is the definition of termination criterion which is usually prescribed number of generations. If the termination criterion is fulfilled the evolution is then terminated. In general, many independent GP runs are needed for successful and accurate problem solutions.

2.2 Experimental details

The main goal of the experiments was determination of the impact of punch speed and blank holder force on the maximum height of deep drawn sheet metal cups before crack occurs. The two chosen parameters are very important for efficient and quality process of deep drawing. Deep drawn cups are cylinders that are closed on one end and open on the other, with or without a flange on the open end. Two different sheet materials were used (DP1180HD and DP780) which are new advanced high strength steels with good cold formability developed for automotive industry. These two materials are also suitable for welding and show excellent characteristics by crash tests.

With the help of tensile tests two important mechanical characteristics were determined: R_m (tensile strength) and yield stress $R_{p0.2}$ for DP1180HD ($R_{p0.2}$ = 1077 N/mm^2, R_m = 1269 N/mm^2) and for DP780 ($R_{p0.2}$ = 490 N/mm^2, R_m = 840 N/mm^2). Material thickness was s = 1.5 mm, diameter of round sheet specimens was D = 215 mm and punch diameter was d = 100 mm. For deep drawing process special experimental tool with a cylinder shaped punch and hydraulic press SHC-400 were used [18], Fig. 1(a).

The process starts with the application of the same amount of Wisura FMO 5020 lubricating oil on both sides of the specimens. Because of cup geometry, anisotropy doesn't have any influence so position of the blank doesn't have to be considered. The measurement of the deep drawn cup's height was executed by laser and acoustic measurement devices. Laser measures drawing height and acoustic sensor device, which was attached in the inner side of the die, detects the moment when crack occurs. With the interaction of these two devices and special software it is possible to detect exact drawing height of the cup just before first crack occurs.

For punch speed the experimental range was set from 50 mm/s to maximum speed of 150 mm/s. The latter is usually the highest punch speed tolerated in praxis for successful and economical deep drawing process of high strength steel. Blank holder force was set to 200 kN and 600 kN respectively.

In order to provide reliable results, three experiments for the same parameters were performed (60 experiments for both materials) and then average value of the height was calculated. The shape of deep drawn cups for both materials after the experiment is shown in Fig. 1(b). Therefore, we obtained a total of 20 combinations of experimental data. The results for measured cup's height are listed in Table 1.


| (a) | (b) |

Fig. 1 (a) Experimental tool for deep drawing, (b) deep drawn cups (v = 125 mm/s, F_b = 200 kN)

Table 1 Experimental results of deep drawing process

Experiment No.	Yield stress $R_{p0.2}$ ($X1$) (N/mm²)	Tensile strength R_m ($X2$) (N/mm²)	Blank holder force F_b ($X3$) (kN)	Punch speed v ($X4$) (mm/s)	Cup height H (mm)
1	490	840	200	50	20.086
2	1077	1269	200	50	15.596
3	490	840	200	75	20.043
4	1077	1269	200	75	15.360
5	490	840	200	100	19.853
6	1077	1269	200	100	15.220
7	490	840	200	125	19.423
8	1077	1269	200	125	14.626
9	490	840	200	150	18.623
10	1077	1269	200	150	13.073
11	490	840	600	50	19.863
12	1077	1269	600	50	15.580
13	490	840	600	75	19.667
14	1077	1269	600	75	15.140
15	490	840	600	100	19.370
16	1077	1269	600	100	14.500
17	490	840	600	125	18.936
18	1077	1269	600	125	14.293
19	490	840	600	150	18.360
20	1077	1269	600	150	13.256

Fig. 2 shows the influence of punch speed v and blank holder force F_b on maximum cup height. It is obvious that an increase of the punch speed leads to a decrease of the maximum drawing height, i.e. to decrease of the formability. This is valid for both examined high strength steels. The highest difference between measured height when blank holder force F_b = 200 kN was used was 7.2 % for DP780. The difference was much higher for DP1180HD, i.e. 16 % decrease of height when punch speed v = 150 mm/s was used compared to the height achieved at v = 50 mm/s.

The increase of blank holder force also leads to a decreasing drawing height but to a smaller extent. A change of the blank holder force does not affect the shape or tendency of the curve. Since increasing punch speed has confirmed the tendency of the formability to decrease, but high speeds are needed to achieve a certain level of productivity, it is important to optimize the process parameters with the goal to improve the efficiency and quality of the deep drawing process.
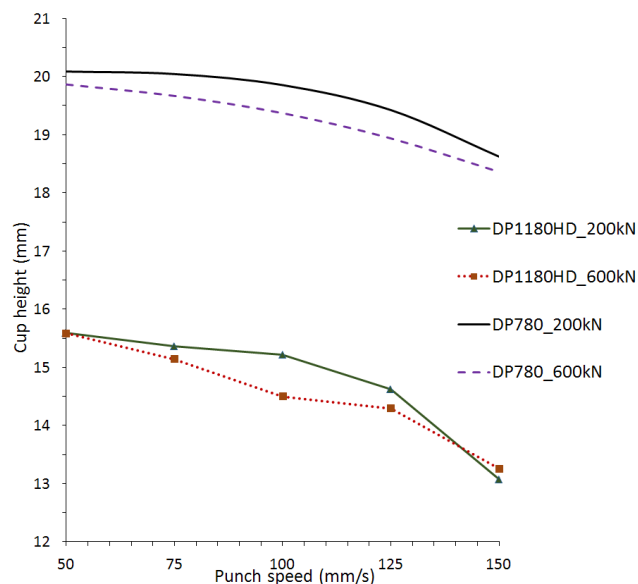
Fig. 2 The influence of punch speed v and blank holder force F_b on maximum cup height

3. Results and discussion

Evolutionary parameters that were used for GP modelling processes were: population size was 400 for all GP runs, maximum number of generations varied from 500 to 1000, reproduction probability was set to 0.1, probability of crossover varied from 0.1 to 0.3, while probability of mutation varied from 0.6 to 0.8. Such a high probability of mutation has been chosen because of the relatively small allowable maximal depth of organisms. In this way, sufficient diversity of organisms in the population is preserved. Maximum allowed depth for organisms created in the initial generation was 6 and maximum depth after crossover was 10 (in some cases only 8). Experimental results from Table 1 were used as a training data set for GP process. Additional experiments were also performed for testing data purpose and were not used in training data set.

Three different function sets were applied for GP modelling. Each function set contained function genes. First set contained three operations (function genes): addition, subtraction and multiplication, i.e. $F = \{+, -, *\}$, while division operation was added to the second function set, i.e. $F = \{+, -, *, /\}$. In third function set, natural exponential function was added to the second function set $\{+, -, *, /, ZEXP\}$. All function genes were protected against extreme values that can be occurred during simulated evolution process [1].

Terminal set comprised of terminal genes. In our case, the terminal consisted of four input variables, and random real numbers, i.e. $T = \{X1, X2, X3, X4, \mathbb{R}\}$. $X1$ is yield strength ($R_{p0.2}$), $X2$ is tensile strength (R_m), $X3$ is blank holder force (F_b), $X4$ is punch speed (v), and $\mathbb{R}$ are random generated real constants from the interval -1 to 1.

By applying the first function set and terminal set, the most accurate genetically developed model for a cup height H developed by genetic programming was:

(+ (* -8.71014 -2.13249) (* (+ (* (+ (- (+ (- (* -9.70734 *X1*) (* *X1 X4*)) (* (- *X2 X1*) (+ *X2* 9.80336))) (+ (* (+ 9.80336 *X4*) (+ *X1 X1*)) (* *X1* (+ *X4 X1*)))) (+ (- (- (* -9.70734 *X1*) (* *X4 X3*)) (+ (* *X3 X4*) (* *X4 X4*))) (- (- (* -9.70734 *X1*) (* *X4 X4*)) (* (+ 9.80336 *X4*) (+ *X1 X3*)))))) (- -2.07369 -2.07544)) (- (+ (+ (- (- *X2* 1.62782) (* -0.0546093 *X4*)) (+ 9.76417 1.167)) (- (+ (- (- *X2* 1.62782) (* -0.0546093 *X4*)) (+ 9.76417 1.167)) *X1*)) (- (- (- (+ *X1* (* -0.0546093 *X4*)) 9.80336) 9.80336) (- (- (- *X2* 1.62782) (* -0.0546093 *X4*)) (+ *X1* (+ (* 3.08974 -6.87342) (- -7.08458 4.51225)))))))))) (- -2.07369 -2.07544))))

The above model is presented in prefix notation of programming language LISP. After simplifying, the above model for a cup height H can be written as the mathematical expression (please note that in Eq. 1 instead of variable names $X1$, $X2$, $X3$, $X4$ original symbols are used):

$$18.6958 - 3.0625 * 10^{-6}R_p{}^2 + 0.00528002\,R_m + 3.0625 * 10^{-6}R_m{}^2 - 0.0000300228\,F_b$$
$$+R_p(-0.00545928 - 3.0625 * 10^{-6}\,R_m - 0.0000153125\,v) + 0.000382265\,v \tag{1}$$
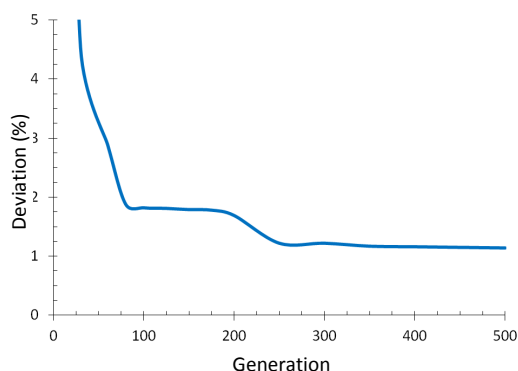$$-9.1875 * 10^{-6}F_b v - 6.125 * 10^{-6}\,v^2$$

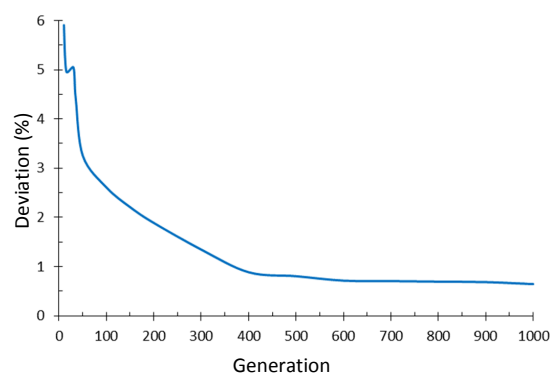**Fig. 3** Deviation of best GP model using first function set {+, -, *} and training data

Fig. 4 Deviation of best GP model using second function set {+, -, *, /} and training data

The model in Eq. 1 was generated in generation 500. It has a total of 127 genes, 63 functional genes and the depth of this model is 10. Probability of reproduction was 0.1, probability of mutation 0.6, and probability of crossover 0.3. Maximal allowed depth after crossover was 10. The average difference between experimental results and the results predicted by the genetic developed model in Eq. 1 was $\delta = 1.14$ % and $\delta = 1.18$ % for training data set and testing data set, respectively.

Fig. 3 presents the average deviation (error) of best GP models when applying first function set. In first few generations there is a large deviation between the prediction results of the GP developed model and experimental data but after several generations natural selection randomly added new genes. Thus, diversity get bigger and deviation becomes better, i.e. lower. After generation 250, the deviation of best models improves very slowly and does not change much until final generation.

The most accurate GP model of all experiments for modelling of cup height was developed by applying the second function set (+, -, *, /). In the prefix LISP form it can be expressed as:

(+ (/ (+ (- (/ (- (* 9.40782 *X2*) (* *X2* -5.06327)) (+ (- *X3* 9.69337) (* -0.346372 *X3*))) (+ (+ (-3.49304 5.8424) -8.94781) (+ (- *X1* 0.583014) -8.27698)))) (- (/ (- (- *X2* -5.60735) (- *X3* 9.69337)) 4.48845) (+ (/ (* *X3* 3.73772) (- *X2* *X1*)) (+ (- -9.68001 3.38485) (- *X4* -0.0676581)))))) (+ (+ (/ (+ (* *X1* 3.13824) (* *X1* 3.13824)) (- (+ 7.6819 5.58638) (- *X3* *X4*))) (+ (/ (+ *X2* *X3*) (- *X1* *X2*)) (+ -1.18787 *X2*))) (/ (+ (- (* *X1* -8.97815) (* *X1* *X4*)) (* (- *X4* -8.75927) (* *X1* -2.32527))) (- (* (- *X2* 8.5169) 0.560745) (+ (/ *X1* *X4*) (/ *X1* *X4*)))))) (+ (/ 5.35335 (/ (+ (- (+ *X1* -8.27698) (/ *X2* *X4*)) (/ (+ *X2* 5.21486) -9.66085)) (+ *X2* (/ (* *X1* 8.08171) (+ -6.90552 *X3*)))))) 8.66904)))

The above model can be written in the infix mathematical expression as:

$$
8.66904 + \frac{5.35335 \left(R_{\mathrm{m}} + \frac{8.08171\, R_{\mathrm{p}}}{-6.90553 + F_{\mathrm{b}}} \right)}{-8.81677 + R_{\mathrm{p}} - 0.103511\, R_{\mathrm{m}} - \frac{R_{\mathrm{p}}}{v}}
$$
$$
+ \frac{\left(36.5633 - R_{\mathrm{p}} + R_{\mathrm{p}} \left(0.222794 - \frac{14.4711}{9.69337 - 0.653628\, F_{\mathrm{b}}} \right) - 0.222794\, F_{\mathrm{b}} + \frac{3.73772\, F_{\mathrm{b}}}{R_{\mathrm{p}} - R_{\mathrm{m}}} - v \right)}{-1.18787 + R_{\mathrm{m}} + \frac{R_{\mathrm{p}} + F_{\mathrm{b}}}{R_{\mathrm{p}} - R_{\mathrm{m}}} + \frac{6.27648\, R_{\mathrm{p}}}{13.2683 - F_{\mathrm{b}} + v} + \frac{R_{\mathrm{p}}(-29.3458 - 3.32527\, v)\, v}{-2 R_{\mathrm{p}} - 4.77581\, v + 0.560745\, R_{\mathrm{m}}\, v}} \tag{2}
$$

In this process we increased the highest allowed number of generations from 500 to 1000. Also the values of probability for some genetic operations were changed. The obtained model (Eq. 2) has an average error of $\delta = 0.65$ % ($\delta = 0.68$ % for testing set) and was generated in generation 1000, has a total of 139 genes, 69 functional genes and the depth of this model is 8 which was also the highest allowed depth after crossover. In this case the values of main evolutionary parameters were: probability of reproduction was 0.1, probability of crossover 0.1 while mutation probability was set to 0.8. This model is more complicated compared to GP model in Eq. 1 but it is also much more accurate.

Fig. 4 presents the average deviation (error) of the best GP model when second function data set was applied for GP process. It can be seen that after generation 50 all calculated average errors of best GP models are smaller than 3 % and after generation 360 they are all under 1 %. From that point and up to final generation the percentage deviation decreases very slowly. If more generations would be used in genetic programming the accuracy of the models would not increase significantly, but processing time would be much longer.

Some very simple genetic models were also obtained by GP process. One of the best simple model with average error of 1.31 % was developed with first function set (+, -, *) and is presented in Eq. 3.

$$
-0.30418 - 0.0469128\, R_{\mathrm{p}} + 0.0532937\, R_{\mathrm{m}} - 0.000797616\, F_{\mathrm{b}} - 0.0170231\, v \tag{3}
$$

The above model is the evidence that GP process is capable of developing not only complex genetic models, but also very simple models with satisfactory accuracy. Simple models are easy to use, but when accuracy of the model is priority, such as in our research, the more complex and more accurate models should be used for prediction and simulation purposes.

We have also performed a few runs of GP modelling with third function set by adding natural exponential function (ZEXP) to four basic math functions. The best obtained genetic models were complex but not so accurate. It was obvious that adding ZEXP function does not results in getting better genetic models, in some cases natural selection in GP even eliminates exponential function and so the best developed models were without ZEXP function. The reason for this happening could be in relatively simplicity of the studied problem.

4. Conclusion

In the paper a GP modelling method for maximum height of cup deep drawn high strength steel specimens, which is an indicator for cold formability level of the material, was presented. The research showed that an increase of the punch speed and blank holder force leads to a decrease of the maximum drawn height. By GP modelling it was possible to develop many different and very accurate genetic models. By using and combining different values for evolution parameters the optimal and most suitable GP models were developed. In our case, a high probability of mutation was vital for good convergence of the algorithms, because it preserves high diversity of a population. With lower value of probability of mutation, the convergence becomes either very slow due to low diversity of organisms in a population or organisms do not even converge to a sufficiently good solution. In the paper only three of the best developed genetic models were presented. The most accurate GP model was also very complex, but this is not an obstacle because modern production processes are all supported by high performance computers, so it is easy to use even the most complex models for accurate prediction of chosen parameters.

In mass production processes, such as deep drawing, sometimes even the tiny improvement in optimization of process parameters can lead to massive reduction of production costs and consequently highly accurate models are desired. In our future work we intend to perform experiments with much more different materials and also additional parameters with the goal to optimize the input parameters for achieving better formability in deep drawing process.

Acknowledgement

References

[1]   Koza, J. (1992). *Genetic programming*, The MIT Press, Massachusetts, USA.
[2]   Pawlak, T.P. (2015). *Competent algorithms for geometric semantic genetic programming*, PhD thesis, University of Technology, Poznan, Poland, doi: 10.13140/RG.2.1.1240.7760.
[3]   Kakandikar, G.M., Nandedkar, V.M. (2016). Prediction and optimization of thinning in automotive sealing cover using genetic algorithm, *Journal of Computational Design and Engineering*, Vol. 3, No. 1, 63-70, doi: 10.1016/j.jcde.2015.08.001.
[4]   Gharib, H., Wifi, A.S., Younan, M., Nassef, A. (2006). Optimization of the blank holder force in cup drawing, *Journal of Achievements in Materials Manufacturing Engineering*, Vol. 18, No. 1-2, 291-294.
[5]   Sener, B., Kurtaran, H. (2016). Optimization of process parameters for rectangular cup deep drawing by the Taguchi method and genetic algorithm, *Materials Testing*, Vol. 58, No. 3, 238-245, doi: 10.3139/120.110840.
[6]   Wei, L., Yuying, Y. (2008). Multi-objective optimization of sheet metal forming process using Pareto-based genetic algorithm, *Journal of Materials Processing Technology*, Vol. 208, No. 1-3, 499-506, doi: 10.1016/j.jmatprotec.2008.01.014.
[7]   Di Lorenzo, R., Ingarao, G., Marretta, L., Micari, F. (2009). Deep drawing process design: A multi objective optimization approach, *Key Engineering Materials*, Vol. 410-411, 601-608, doi: 10.4028/www.scientific.net/KEM.410-411.601.

[8] Naceur, H., Guo, Y.Q., Batoz, J.L. (2004). Blank optimization in sheet metal forming using an evolutionary algorithm, *Journal of Materials Processing Technology*, Vol. 151, No. 1-3, 183-191, doi: 10.1016/j.jmatprotec.2004.04.036.

[9] Singh, D., Yousefi, R., Boroushaki, M. (2011). Identification of optimum parameters of deep drawing of a cylindrical workpiece using neural network and genetic algorithm, *International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering,* Vol. 5, No. 6, 987-993, doi: 10.1999/1307-6892/2043.

[10] Zhao, J., Wang, F. (2005). Parameter identification by neural network for intelligent deep drawing of axisymmetric workpieces, *Journal of Materials Processing Technology*, Vol. 166, No. 3, 387-391, doi: 10.1016/j.jmatprotec.2004.08.020.

[11] Tinkir, M., Dilmeç, M., Türköz, M., Halkaci, H.S. (2015). Investigation of the effect of hydromechanical deep drawing process parameters on formability of AA5754 sheets metals by using neuro-fuzzy forecasting approach, *Journal of Intelligent & Fuzzy Systems,* Vol. 28, No. 2, 647-659, doi: 10.3233/IFS-141346.

[12] Öztürk, E., Türköz, M., Halkacı, H.S., Koç, M. (2017). Determination of optimal loading profiles in hydro–mechanical deep drawing process using integrated adaptive finite element analysis and fuzzy control approach, *The International Journal of Advanced Manufacturing Technology*, Vol. 88, No. 9-12, 2443-2459, doi: 10.1007/s00170-016-8912-x.

[13] Oduguwa, V., Tiwari, A., Roy, R. (2005). Evolutionary computing in manufacturing industry: An overview of recent applications, *Applied Soft Computing,* Vol. 5, No. 3, 281-299, doi:10.1016/j.asoc.2004.08.003.

[14] Brezocnik, M., Kovacic, M., Gusel, L. (2005). Comparison between genetic algorithm and genetic programming approach for modeling the stress distribution, *Materials and Manufacturing Processes*, Vol. 20, No. 3, 497-508, doi:10.1081/AMP-200053541.

[15] Gusel, L., Rudolf, R., Brezocnik, M. (2015). Genetic based approach to predicting the elongation of drawn alloy, *International Journal of Simulation Modelling*, Vol. 14, No. 1, 39-47, doi: 10.2507/IJSIMM14(1)4.277.

[16] Halder, C., Madej, L., Pietrzyk, M., Chakraborti, N. (2015). Optimization of cellular automata model for the heating of dual-phase steel by genetic algorithm and genetic programming, *Materials and Manufacturing Processes*, Vol. 30, No. 4, 552-562, doi: 10.1080/10426914.2014.994765.

[17] Kovacic, M., Uratnik, P., Brezocnik, M., Turk, R. (2007). Prediction of the bending capability of rolled metal sheet by genetic programming, *Materials and Manufacturing Processes*, Vol. 22, No. 5, 634-640, doi: 10.1080/10426910701323326.

[18] Radakovics, S.M. (2017). *Einfluss der Tiefziehgeschwindigkeit auf die Kaltumformbarkeit von HSS und AHSS Stählen*, Bachelor's thesis, Fakultät für Maschinenbau und Wirtschaftswissenschaften, TU Graz, Austria (in German).

Calendar of events

- Industrial Engineering and Manufacturing Technologies, Phuket, Thailand, September 21-23, 2018.

- 26th International Conference on Materials and Technology, Portorož, Slovenia, October 3-5, 2018.

- International Conference on Additive Technologies, Maribor, Slovenia, October 10-11, 2018.

- European Simulation and Modelling Conference (ESM 2018), Ghent, Belgium, October 24-26, 2018.

- 5th International Conference on Materials, Mechatronics, Manufacturing and Mechanical, Kuching, Malaysia, November 2-4, 2018.

- International Conference on Industrial Engineering and Engineering Management (IEEM), Bangkok, Thailand, December 16-19, 2018.

- 3rd International Conference on 3D Printing Technology and Innovations, Rome, Italy, March 25-26, 2019.

- 9th IFAC Conference on Manufacturing Modeling, Management, and Control, Berlin, Germany, August 28-30, 2019.

Notes for contributors

General

Articles submitted to the *APEM journal* should be original and unpublished contributions and should not be under consideration for any other publication at the same time. Manuscript should be written in English. Responsibility for the contents of the paper rests upon the authors and not upon the editors or the publisher. Authors of submitted papers automatically accept a copyright transfer to *Chair of Production Engineering*, *University of Maribor*. For most up-to-date information on publishing procedure please see the *APEM journal* homepage *apem-journal.org*.

Submission of papers

A submission must include the corresponding author's complete name, affiliation, address, phone and fax numbers, and e-mail address. All papers for consideration by *Advances in Production Engineering & Management* should be submitted by e-mail to the journal Editor-in-Chief:

Miran Brezocnik, Editor-in-Chief
UNIVERSITY OF MARIBOR
Faculty of Mechanical Engineering
Chair of Production Engineering
Smetanova ulica 17, SI – 2000 Maribor
Slovenia, European Union
E-mail: editor@apem-journal.org

Manuscript preparation

Manuscript should be prepared in *Microsoft Word 2007* (or higher version) word processor. *Word .docx* format is required. Papers on A4 format, single-spaced, typed in one column, using body text font size of 11 pt, should not exceed 12 pages, including abstract, keywords, body text, figures, tables, acknowledgements (if any), references, and appendices (if any). The title of the paper, authors' names, affiliations and headings of the body text should be in *Calibri* font. Body text, figures and tables captions have to be written in *Cambria* font. Mathematical equations and expressions must be set in *Microsoft Word Equation Editor* and written in *Cambria Math* font. For detail instructions on manuscript preparation please see instruction for authors in the *APEM journal* homepage *apem-journal.org*.

The review process

Every manuscript submitted for possible publication in the *APEM journal* is first briefly reviewed by the editor for general suitability for the journal. Notification of successful submission is sent. After initial screening, and checking by a special plagiarism detection tool, the manuscript is passed on to at least two referees. A double-blind peer review process ensures the content's validity and relevance. Optionally, authors are invited to suggest up to three well-respected experts in the field discussed in the article who might act as reviewers. The review process can take up to eight weeks on average. Based on the comments of the referees, the editor will take a decision about the paper. The following decisions can be made: accepting the paper, reconsidering the paper after changes, or rejecting the paper. Accepted papers may not be offered elsewhere for publication. The editor may, in some circumstances, vary this process at his discretion.

Proofs

Proofs will be sent to the corresponding author and should be returned within 3 days of receipt. Corrections should be restricted to typesetting errors and minor changes.

Offprints

An e-offprint, i.e., a PDF version of the published article, will be sent by e-mail to the corresponding author. Additionally, one complete copy of the journal will be sent free of charge to the corresponding author of the published article.

APEM
journal

Contents

9 771854 625008