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 *Production Engineering Institute* 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

Comparison among four calibrated meta-heuristic algorithms for solving a type-2 fuzzy cell formation problem considering economic and environmental criteria

Arghish, O.[a], Tavakkoli-Moghaddam, R.[b,c,*], Shahandeh-Nookabadi, A.[d], Rezaeian, J.[e]

[a]Department of Industrial Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b]School of Industrial Engineering, College of Engineering, University of Tehran, Tehran, Iran
[c]LCFC, Arts et Métiers Paris Tech, Metz, France
[d]Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran
[e]Department of Industrial Engineering, Mazandaran University of Science and Technology, Babol, Iran

A B S T R A C T

In this paper, a mathematical model is proposed using economic and environmental criteria for a type-2 fuzzy (T2F) cell formation (CF) problem emphasizing the effect of the man-machine relationship aspect. This model aims to show the use of this aspect in CF to minimize the costs of processing, material movement, energy loss, and tooling. For this purpose, a two-stage defuzzification procedure is used to convert the T2F variable into a crisp value. Due to NP-hardness of the model and problem, a genetic algorithm (GA) is used to derive the appropriate solutions. Furthermore, because there is no any existing benchmark to validate the performance of the proposed model, three tuned meta-heuristic algorithms, namely, differential evolution (DE), harmony search (HS) and particle swarm optimization (PSO), are proposed and used. The present research uses the Taguchi method to adjust the parameters in the four proposed algorithms. Furthermore, 15 examples are used to validate the presented model. The results show that PSO is the most appropriate algorithm for solving the model.

1. Introduction

Nowadays, the competitive environment throughout the world has led the involved manufacturing industries to supply the highest quality, affordable products such that recent approaches focus more on the ever-growing manufacturing costs including those associated with location, energy, and transportation system. Group technology (GT) as one of the most efficient approaches tries to group parts and machines in terms of their similarities in production processes, functionalities, and geometries [1]. Cellular manufacturing (CM) as an application of GT in a manufacturing system is utilized to classify similar parts into families assigning different machines to cells [2]. The CMS design involves four principal stages, in which each of them can be considered as an individual problem, namely CF, layout, scheduling and resource assignment (RA) [3]. As the CF problem comes up as the first stage in the CMS design, investigators have attempted to optimally solve the problem.

For instance, Majazi-Delfard [4] introduced a non-linear model for the dynamic cell formation (DCF) in terms of the quantity and length of intra- and inter-cell travels. Deljoo *et al.* [5] provid-

ed a GA-based solution to the DCF problem identifying errors in the models recommended in the literature, declining their helpful perspective and presenting a novel formulation for the DCF problem. Bagheri and Bashiri [6] utilized a LP-metric approach to a proposed model comprising of CF, layout and worker assignment components. Xu *et al.* [7] provided a bat algorithm (BA) to the dual flexible job-shop scheduling problem considering the process sequence and machine selection flexibility. Zupan *et al.* [8] presented a method based on a combination of schmigalla modified triangular method, the schwerdfeger circular process, and a simulation model to the layout optimization of a production cell considering the intensity of the material flow. Nie *et al.* [9] utilized a simulation model for a Token-oriented Petri net-based flexible manufacturing cell.

Mahdavi *et al.* [10] presented a two-stage approach to a CF problem considering interval T2F interactional interests among workers. In the first stage, a multi-depot multiple traveling salesman problem model is used to assign workers to cells. In the second stage, a mathematical model is applied to assign machines to cells. With respect to the historical perspective, sustainability was conceptualized in the late 1960's and early 1970's stressing the environmental effect of industrial projects [11]. Sustainability was represented by Lozano [12] in three dimensions, in which themes in economic, environmental, and social aspects interact in the temporal respective. The literature review also reveals the fact that while the CMS design has received considerable attention, social and environmental aspects have been underexplored with the majority of the criteria investigated in the associated literature devoted to economic rather than social and environmental issues [13].

Niakan *et al.* [14] proposed a mathematical model of the problem comprising of two objectives to investigate the trade-off between the total cost minimization and social issue maximization attempting to solve the problem using a non-dominated sorting genetic algorithm (NSGA-II). In addition, Niakan *et al.* [15] proposed a model for the cost minimization and the machine energy loss minimization incorporating a social constraint developing an NSGA-II solve the problem. Furthermore, Niakan *et al.* [16] introduced a model possessing two objectives for the DCF problem minimizing production and worker costs and total production waste including energy, chemical material, raw material, CO_2 emissions. Proposing the social criteria as constraints, the researchers utilized the NSGA-II and multi-objective simulated annealing (MOSA) algorithms to provide a solution for the bi-objective model.

The literature mostly considers input parameters in the CF to be deterministic. However, in practice, there are a large number of uncertain and imprecise parameters involved. As the quantity of data may not always be sufficient for the uncertain parameters prediction, fuzzy logic is utilized as a robust instrument to gauge this uncertainty through the medium of the personal knowledge [17]. To describe type-2 fuzziness, a T2F variable represents a map extending from the fuzzy possibility space to the real number space [18]. Miller and John [19] assert that further uncertainty degrees supplied through the interval type-2 fuzzy sets (IT-2FS) logic makes it possible to more optimally show the uncertainty and vagueness of resource planning models. Qin *et al.* [18] presented three categories of critical values (CVs) for a regular fuzzy variable (RFV) and three reduction methods for a T2F variable.

Research conducted by Kundu *et al.* [20] examines transportation problems using T2F parameters wherein the first values of the T2F parameters were defuzzified using CV-based reduction methods to type-1 fuzzy variables, and the centroid method was employed for total defuzzification.

The review of the related literature reveals the most commonly used criteria in CF to be the associated costs. Therefore, to bridge this gap, a mathematical model was developed using economic and environmental criteria and a T2F parameter for CF and RA problems simultaneously. The novelty of the present research partly lies in the consideration of worker-machine relationship for worker allocation where the worker can be considered as a significant industrial system component. The present paper consists of the problem statement and mathematical model (Section 2), solution methodologies (Section 3), computational results (Section 4), and conclusion (Section 5).

2. Problem statement and definition of objective function

This section discusses a T2F model for the CF considering economic, environmental and man-machine relationship aspects.

2.1 Worker-machine relationship

Nowadays, numerous machine tools are either totally or partially automatically operated with the worker being usually idle for a portion of the cycle. The potential use of this idle time can enhance worker earnings and the efficiency of a manufacture. The man-machine process chart evidently depicts the respective idle machine time and worker time areas, which normally represent desirable locations to initiate effective improvements. Although this chart is generally implemented to specify the number of machines assigned to a worker, the application of a mathematical model can substantially lessen the time needed to do so. Such ideal situations are generally referred to as synchronous servicing with the number of machines assigned computed as shown in Eq. 1.

$$m = \frac{s + r}{s + w} \tag{1}$$

where m represents the number of machines being operated by each worker, s worker servicing time per machine, r machine working time and w walking time between two machines, the number of machines must be represented by a whole number; otherwise, we have the following equation.

$$ml \leq m < mu \tag{2}$$

If the number of machines does not represent a whole number, the minimum total cost per piece criteria can be used for the optimum operation. The total cost per piece for ml and mu machines are given in Eqs. 3 and 4.

$$TC_{ml} = (C_1.(s + r) + ml.C_2.(s + r))/ml \tag{3}$$

$$TC_{mu} = (s + w)(C_1 + C_2.mu) \tag{4}$$

where C_1 and C_2 are the worker and machine costs, respectively. The number of machines assigned to workers represents the minimum total cost per piece [21].

2.2 Environmental criteria

With the passage of time, humans have damaged the environment through the waste production and uncontrolled use of natural resources to satisfy the ever-increasing unprecedented use level, which was far beyond the nature's capacity to restore and/or regenerate itself. The fear of an uncertain future for the world made it necessary to grasp how individuals, organizations, and governments have cooperated to discover approaches to prevent a global collapse [22]. One of the objective function terms is to minimize the total system energy loss cost. The relationship between man and the machine in the CF model leads to the far better use of both workers and machine time, and more optimal balance in the work cycle. The man and machine relationship tools depict the areas, in which machine and worker idle times take place. Thus, the use of these idle times can enhance worker earnings enhancing the production efficiency [21]. In the researchers' proposed model, all the machines are considered as a multi-functional task. The machines require highly skilled workers thus providing workers with the opportunity to acquire numerous skills expanding their potential.

2.3 T2F set

A T2F set was proposed as an extension of an ordinary fuzzy set. Let Γ be the universe, Pos: $\mathcal{A} \to [0,1]$ be a set function on the ample field $\mathcal{A}$ and *Pos* is a possibility criterion. If $(\Gamma, \mathcal{A}, Pos)$ is a possibility space, then an m-ary regular fuzzy vector $\xi = (\xi_1, \xi_2, ..., \xi_m)$ is a map $\Gamma \to [0,1]^m$ for any $t = (t_1, t_2, ..., t_m) \in [0,1]^m$, one has the following equation:

$$\{\gamma \in \Gamma \mid \xi(\gamma) \leq t\} = \{\gamma \in \Gamma \mid \xi_1(\gamma) \leq t_1, \xi_2(\gamma) \leq t_2, ..., \xi_m(\gamma) \leq t_m\} \in \mathcal{A} \tag{5}$$

as $m=1$, ξ is a RFV. Let $\widetilde{Pos}: \mathcal{A} \to [0,1]$ be a set function on $\mathcal{A}$ such that $\{\widetilde{Pos}\ (A)\mid A\epsilon\ \mathcal{A}\ \}$ is an RFVs and $\widetilde{Pos}$ is a fuzzy possibility criteria. If $\mu_{\widetilde{Pos(\Gamma)}}(1) = 1$, then $\widetilde{Pos}$ is a regular fuzzy possibility criteria. If $(\Gamma, \mathcal{A}, \widetilde{Pos})$ is a fuzzy possibility space (FPS), then an m-ary T2F vector $\tilde{\tilde{\xi}} = (\tilde{\tilde{\xi}}_1, \tilde{\tilde{\xi}}_2, ..., \tilde{\tilde{\xi}}_m)$ is a map $\Gamma \to$ ℜ^m for any $r = (r_1, r_2, \ldots, r_m) \in \mathfrak{R}^m$, as shown in Eq. 6.

$$\{\gamma \in \Gamma \mid \tilde{\tilde{\xi}}(\gamma) \leq r\} = \{\gamma \in \Gamma \mid \tilde{\tilde{\xi}}_1(\gamma) \leq r_1, \tilde{\tilde{\xi}}_2(\gamma) \leq r_2, \ldots, \tilde{\tilde{\xi}}_m(\gamma) \leq r_m\} \in A \tag{6}$$

as $m = 1$, $\tilde{\tilde{\xi}}$ is a T2F variable [18].

Critical value

Let $\xi = (r_1, r_2, r_3, r_4)$ and $\xi = (r_1, r_2, r_3)$ be a trapezoidal and triangular RFV, respectively. Then, we have the following items in Table 1 [18].

Table 1 Component of RFVs

	trapezoidal	triangular
optimistic CV of $\tilde{\tilde{\xi}}$	$r_4/(1 + r_4 - r_3)$	$r_3/(1 + r_3 - r_2)$
pessimistic CV of $\tilde{\tilde{\xi}}$	$r_2/(1 + r_2 - r_1)$	$r_2/(1 + r_2 - r_1)$
The CV of $\tilde{\tilde{\xi}}$	$\begin{cases} \dfrac{2r_2 - r_1}{1 + 2(r_2 - r_1)} & r_2 > \dfrac{1}{2} \\ \dfrac{1}{2} & r_2 \leq \dfrac{1}{2} \leq r_3 \\ \dfrac{r_4}{1 + 2(r_4 - r_3)} & r_3 \leq \dfrac{1}{2} \end{cases}$	$\begin{cases} \dfrac{2r_2 - r_1}{1 + 2(r_2 - r_1)} & r_2 > \dfrac{1}{2} \\ \dfrac{r_3}{1 + 2(r_3 - r_2)} & r_2 \leq \dfrac{1}{2} \end{cases}$

Centroid method

The centroid method is the most commonly used method for transforming the type-1 fuzzy into crisp values. The centroid method can be defined by the Eq. 7 for the discrete case [23].

$$z^* = \frac{\sum_z z.\mu_{\tilde{A}}(z)}{\sum_z \mu_{\tilde{A}}(z)} \tag{7}$$

Defuzzification

In the present research, a two-stage defuzzification procedure is employed to transform the T2F variable to crisp value. Initially, the CV is utilized for RFVs for transforming the T2F into type-1 fuzzy. Then, the centroid method is employed in order to transform the type-1 fuzzy into crisp values.

2.4 Assumptions

For the CF considered in this paper, every single operation on each part classification can be executed on multi-functional and identical machines. Each part type demand, each tool type's tool life, maximum cell number and worker-part-machine-tool-worker combination compatibility are given. The average quantity of energy wasted by each machine type in a unit of time and energy price in a unit of time is also known as is the total servicing time of a worker for each machine.

2.5 Notations and parameters

r_{opmhg}	Machine m working time for performing operation o on part p with tool h by worker g
a_{opmhg}	1, if machine m is employed to operation o for part p with tool h by worker g; and 0, otherwise
α_p^{inter}	Cost related to inter-cell movement for part p in a distance unit
α_p^{intra}	Cost related to intra-cell movement for part p in a distance unit
Q_p	Demand for part p
T_m	Time capacity for machine m

T_g	Time capacity for worker g
U_k	Upper bound of machines allowed in cell k
$d_{kk'}$	Average distance among cells k and k'
s_{opmhg}	Worker servicing time per machine m to perform operation o on part p using tool h by worker g
w_{opmhg}	Walking time between machine m taken to process operation o for part p using tool h by worker g to the next machine
C_{opmhg}	Operating cost on machine m to process operation o on part p using tool h by worker g
$E_{kk'}$	1, if $k \neq k'$; and 0, if $k = k'$
C_{2_m}	Machine (m) cost for a time unit
C_{1_g}	Worker (g) cost for a time unit
γ_h	Cost of tool h
$Batch_p^{inter}$	Inter-cell batch size motion for part p
$Batch_p^{intra}$	Intra-cell batch size motion for part p
$ToolLife_h$	Tool life of tool h
φ_m	Average quantity of energy wasted by each machine m in unit time
$\widetilde{EC}$	Price of energy in unit time

2.6 Decision variables

x_{opmkhg}	1, if machine m is used for operation o of part p using tool h by worker g in cell k; and 0, otherwise
u_{hm}	Number of tool copies for tool h on machine m
y_{opmk}	1, if operation o of part p is performed on machine m in cell k; and 0, otherwise
l_{opm}	1, if operation o of part p is performed on machine m; and 0, otherwise
ll_{mk}	1, if machine m is assigned to cell k; and 0, otherwise
$V1_{mg}$	1, if (mu) machine m is allocated to worker g; and 0, otherwise
$V2_{mg}$	1, if (ml) machine m is allocated to worker g; and 0, otherwise
ml_{mg}	Lower whole quantity of machine m allocated to worker g
mu_{mg}	Upper whole quantity of machine m allocated to worker g
$TC_{ml\ mg}$	Total cost per piece from one machine (ml) m and worker g
$TC_{mu\ mg}$	Total cost per piece from one machine (mu) m and worker g
mm_{mgk}	Number of machine m assigned to worker g in cell k

2.7 Objective function

The objective function of the considered model is to minimize the costs of processing, material movement, energy loss, and tooling.

$$\text{Min} = \sum_{o=1}^{O}\sum_{p=1}^{P}\sum_{m=1}^{M}\sum_{k=1}^{K}\sum_{h=1}^{H}\sum_{g=1}^{G} Q_p C_{opmhg} x_{opmkhg} \tag{8}$$

$$+ \sum_{o=1}^{O-1}\sum_{p=1}^{P}\sum_{k=1}^{K}\sum_{k'=1}^{K} \left\lceil \frac{Q_p}{Batch_p^{inter}} \right\rceil \alpha_p^{inter} E_{kk'} d_{kk'} \left(\sum_{m=1}^{M} y_{opmk}\right)\left(\sum_{m=1}^{M} y_{o+1,pmk'}\right)$$

$$+ \sum_{o=1}^{O-1}\sum_{p=1}^{P}\sum_{k=1}^{K}\sum_{k'=1}^{K} \left\lceil \frac{Q_p}{Batch_p^{intra}} \right\rceil \alpha_p^{intra}(1 - E_{kk'}) d_{kk'} \left(\sum_{m=1}^{M} y_{opmk}\right)\left(\sum_{m=1}^{M} y_{o+1,pmk'}\right)$$

$$+ \sum_{m}^{M}\sum_{g}^{G} \left(ml_{mg}V2_{mg} + mu_{mg}V1_{mg}\right) \times \varphi_m \times \widetilde{EC} + \sum_{h=1}^{H} \gamma_h \sum_{m=1}^{M} u_{hm}$$

$$\text{s.t.}: \sum_{m}^{M} \sum_{k}^{K} \sum_{h}^{H} \sum_{g}^{G} x_{opmkhg}.a_{opmhg} = 1, \qquad \forall o,p \tag{9}$$

$$x_{opmkhg} \leq a_{opmhg}, \qquad \forall o,p,m,k,h,g \tag{10}$$

$$\sum_{m}^{M} \sum_{g}^{G} mm_{mgk} \leq U_k, \qquad \forall k \tag{11}$$

$$y_{opmk} = l_{opm}ll_{mk}, \qquad \forall o,p,m,k \tag{12}$$

$$mm_{mgk}ll_{mk} = ml_{mg}V2_{mg} + mu_{mg}V1_{mg}, \qquad \forall m,g,k \tag{13}$$

$$\sum_{o}^{O} \sum_{p}^{P} \sum_{k}^{K} \sum_{h}^{H} Q_p.x_{opmkhg}(r_{opmhg} + s_{opmhg}) \leq T_g + MV1_{mg}, \quad \forall m,g \tag{14}$$

$$\sum_{o}^{O} \sum_{p}^{P} \sum_{k}^{K} \sum_{h}^{H} Q_p.x_{opmkhg}.mu_{mg}.r_{opmhg} \leq T_g + MV2_{mg}, \quad \forall m,g \tag{15}$$

$$\sum_{o}^{O} \sum_{p}^{P} \sum_{k}^{K} \sum_{h}^{H} Q_p.x_{opmkhg}(r_{opmhg} + s_{opmhg}) \leq T_m + MV1_{mg}, \quad \forall m,g \tag{16}$$

$$\sum_{o}^{O} \sum_{p}^{P} \sum_{k}^{K} \sum_{h}^{H} Q_p.x_{opmkhg}.mu_{mg}.r_{opmhg} \leq T_m + MV2_{mg}, \quad \forall m,g \tag{17}$$

$$\frac{(r_{opmhg} + s_{opmhg}).x_{opmkhg}}{s_{opmhg} + w_{opmhg}} \leq mu_{mg}, \qquad \forall o,p,m,k,h,g \tag{18}$$

$$mu_{mg.} = a_{opmhg}(ml_{mg} + 1), \qquad \forall o,p,m,h,g \tag{19}$$

$$TC_{ml\ mg} = \frac{a_{opmhg}\left((r_{opmhg} + s_{opmhg})\left(C_{1_g} + ml_{mg} \times C_{2_m}\right)\right)}{ml_{mg}}, \forall o,p,m,h,g \tag{20}$$

$$TC_{mu\ mg} = a_{opmhg}\left((s_{opmhg} + w_{opmhg})\left(C_{1_g} + C_{2_m} \times mu_{mg.}\right)\right), \forall o,p,m,h,g \tag{21}$$

$$TC_{ml\ mg} \leq TC_{mu\ mg} + MV1_{mg}, \quad \forall m,g \tag{22}$$

$$TC_{mu\ mg} \leq TC_{ml\ mg} + MV2_{mg}, \quad \forall m,g \tag{23}$$

$$V1_{mg} + V2_{mg} = 1, \qquad \forall m,g \tag{24}$$

$$\sum_{o}^{O} \sum_{p}^{P} \sum_{k}^{K} \sum_{g}^{G} x_{opmkhg}.r_{opmhg} \leq ToolLife_h.u_{hm}, \qquad \forall m,h \tag{25}$$

$$\sum_{h}^{H} \sum_{g}^{G} x_{opmkhg} = y_{opmk}, \qquad \forall o,p,m,k \tag{26}$$

$$x_{opmkhg}, y_{opmk}, V1_{mg}, V2_{mg}, l_{opm}, ll_{mk} \in \{0,1\} \tag{27}$$

$$ml_{mg}, mu_{mg}, mm_{mgk}ll_{mk}, u_{hm} \geq 0, integer \tag{28}$$

The first sentence in the objective function (Eq. 8) represents the total cost of the process. The second and third terms represent the total material transportation cost. The fourth term is the cost of energy loss in the system at a time unit. The fifth term represents the total tool cost. Eq. 9 and Eq. 10 express the operation-part-machine-tool-worker combinations. Eq. 11 limits the cell size. Eq. 12 can be used to define the machine-cell combination. Eq. 13 is used to ensure that the quantity of machine m is assigned to worker g in cell k. Eq. 14 and Eq. 15 express the time capacity of the worker. Eq. 16 and Eq. 17 express the time capacity of the machine. Eq. 18 ensures that the upper whole number of machine m is assigned to worker g. Eq. 19 ensures that a lower whole number of machine m is assigned to worker g. Eq. 20 and Eq. 21 express the cost of a production per cycle from one machine in the lower and upper whole numbers of machine m assigned to worker g. Eq. 22 to Eq. 24 guarantee that the lowest TC is chosen. Eq. 25 represents the tool-machine combinations. Eq. 26 expresses the operation-part-machine-cell combinations. Eq. 27 and Eq. 28 can be used to define the type of variables.

3. Used methods

In the present paper, in order to solve the presented CF model, a GA is implemented and three DE, PSO and HS algorithms are employed to the obtained accredit outcome.

3.1 Genetic algorithm

Holland [24] was the first to develop the GA, which represented coding to a chromosome form. Subsequent to the production of the first random chromosomes, evaluation of performance was undertaken using the fitness function. The remaining chromosomes and offspring produce a generation through the medium of crossover and mutation. In the end, the elitism process is used to produce solutions [25]. The GA used for the CF framework is as follows. Two elements are given in the CF problem. The first element represents the assignment of machines to workers using [Ma_Wo]. This matrix is utilized to define the entirety of the relative constraints. The second element represents an operation-part-machine-cell-tool-worker [OP_Pa_N3], where N3 equals [Ma_Ce_To_Wo]. These matrices are employed to define the entirety of the relative constraints.

3.2 Differential evolution

The DE algorithm represents a recent evolutionary optimization technique for continuous non-linear functions introduced by Noktehdan *et al.* [26] and Storn and Price [27]. The principal stages involved in the DE algorithm are defined as follows. Initially, a random population generation is formed and the objective function is evaluated. For each individual solution in the population, a mutated solution $\widehat{x}_i$ is produced as follows in Eq. 29.

$$\widehat{x}_i = x_{r1} + F(x_{r3} - x_{r2}) \tag{29}$$

where F represents a scalar ($F \in [0, 1]$), and x_{r1}, x_{r3}, x_{r2} represent randomly-selected subjects in the population i ($\widehat{x}_i \neq x_{r1} \neq x_{r3} \neq x_{r2}$). The crossover operation is employed to establish a trial vector through shuffling the information incorporated in the mutated vector and the current solution in Eq. 30.

$$y_i^j = \hat{x}_i^j \text{ for } R_j \leq CR \text{ and } y_i^j = x_i^j \text{ for } R_j > CR \tag{30}$$

where CR represents the crossover rate $\in [0, 1]$, which needs to be specified by the user, and R_j represents a random real number $\in [0,1]$ and j is the j-th parameter. A comparison is made between each trial vector ($\overrightarrow{y_i}$) and its parent ($\overrightarrow{x_i}$), and the more desirable one remains in the population in the selection stage [28].

3.3 Particle swarm optimization

The PSO algorithm represents a population-based stochastic optimization algorithm extended by Kennedy and Eberhart [29] composed of a population (i.e., swarm) of candidate solutions,

referred to as particles, which are in inward motion towards search space with a designated velocity in search of an optimum solution. Each particle maintains a memory assisting it in preserving the path taken by its previous best location. The particle positions are identified as personal best and global best. The principal iterative process to arrive at the solution is executed by:

$$x_i^{k+1} = x_i^k + v_i^{k+1} \qquad (31)$$

$$v_i^{k+1} = wv_i^k + C_1 rand_1 \left(PBest_i - x_i^k\right) + C_2 rand_2 \left(P_{gi} - x_i^k\right) \qquad (32)$$

Eq. 31 is used to calculate the i-th particle movement in the k-th replication, where v_i^k, x_i^k represent the velocity and the current location of the i-th particle in the k-th replication, respectively. Eq. 32 is employed to calculate the latest velocity vector of the i-th particle in the k-th iteration, $PBest_i$ represents the best location of the i-th particle, w represents the inertia factor controlling the magnitude of the old velocity, C_1 and C_2 represent acceleration constants (i.e., cognitive and social) based on the kind of search, local and global P_{gi} are denoted $lBest$ and $gBest$, respectively [30].

3.4 Harmony search

Musical performance can be defined as the quest for the lovely harmony among all harmonies. Geem *et al.* [31] introduced an optimization algorithm on the basis of musical performance, known as HS, searching for the best solution emanating from the objective function. The algorithm starts by playing a new harmony and comparing this harmony with those in harmony memory (HM), which results in the improvement in the harmony quality in a step-by-step manner. Subsequently, the HM updates and verifies the stop criterion. The entirety of the decision variables (notes) in HM together with the values for these notes in the new harmony are determined in the following manner: first, precise choice of the HM domain value. Second, random selection of the full value domain using a selection rate or the harmony memory considering rate (HMCR) between zero and one. Third, selection of ideal identical values for the HM domain with the pitch adjustment rate (PAR) between zero and one and a free distance bandwidth (Bw) [32].

4. Results and discussion

To investigate and evaluate the performance of the four meta-heuristic algorithms on the CF, some randomly-selected numerical examples are produced. To solve the proposed model, MATLAB (R2016b) software is utilized to provide the code the algorithms on a laptop having five Intel Core i5 CPU and 2 GB RAM. The TM is run in MINITAB software version 17.3.1 to calibrate the parameters for a subsequent data analysis.

4.1 Defuzzification of T2F variable

The energy price coefficient in the fourth sentence of the objective function ranges between 4 to 8. The coefficient is represented by the following discrete T2F variable.

$$\widetilde{EC} = \begin{cases} 3 & \text{with } \tilde{\mu}\widetilde{EC}\,(3) = (0.1, 0.4, 0.7) \\ 4 & \text{with } \tilde{\mu}\widetilde{EC}\,(4) = (0.9, 1, 1) \\ 5 & \text{with } \tilde{\mu}\widetilde{EC}\,(5) = (0.1, 0.3, 0.4, 0.6) \end{cases} \qquad \widetilde{EC} = \begin{cases} 7 & \text{with } \tilde{\mu}\widetilde{EC}\,(7) = (0.4, 0.5, 0.7, 0.8) \\ 8 & \text{with } \tilde{\mu}\widetilde{EC}\,(8) = (0.6, 0.8, 0.9) \\ 10 & \text{with } \widetilde{\mu}\widetilde{EC}\,(10) = (0.4, 0.6, 0.7) \end{cases}$$

To solve the CF model under consideration in the initial step, a CV reduction method is employed to convert the energy price T2F variable in unit time to the corresponding type-1 fuzzy variable. During the second step, a centroid method is performed to reduce the type-1 fuzzy variable to the crisp value. The energy price crisp value is obtained using $EC = (3.99, 8.35)$.

4.2 Generating random data

15 random examples are produced in various sizes through the generation of uniformly distributed random points for a number of parameters given. The attributes of 15 designed test examples are shown in Table 2. Also, Table 3 shows the components of the model input parameters required for 15 problem instances.

Table 2 Attribute of test examples

Problem	Part	Operation	Machine	Cell	Tool	Worker	Problem	Part	Operation	Machine	Cell	Tool	Worker	Problem	Part	Operation	Machine	Cell	Tool	Worker
1	2	2	4	2	2	4	6	5	5	3	2	3	3	11	7	2	5	2	3	5
2	2	4	5	2	2	5	7	6	3	4	3	3	4	12	7	5	3	4	3	3
3	4	4	2	3	3	2	8	6	5	3	4	4	3	13	8	2	3	2	3	3
4	4	5	4	3	3	4	9	6	3	5	3	3	5	14	8	5	4	2	2	4
5	5	5	2	3	2	2	10	7	3	4	3	2	4	15	8	5	4	3	3	4

Table 3 Data identifying with random test problems

Parameter	Amount	Parameter	Amount	Parameter	Amount	Parameter	Amount
Q_p	$U(100-500)$	$\widetilde{EC}$	$U(3.99-8.35)$	T_g, T_m	55	C_{opmhg}	$U(9-20)$
$Batch_p^{intra}$	$U(5-10)$	α_p^{intra}	$U(10-30)$	r_{opmhg}	0.02	s_{opmhg}	0.01
$Batch_p^{inter}$	$U(10-15)$	α_p^{inter}	$U(50-75)$	w_{opmhg}	0	C_{1_g}	$U(1-2)$
φ_m	$U(0.5-0.7)$	γ_h	$U(50-80)$	C_{2_m}	$U(2-3)$	$ToolLife_h$	1

4.3 Parameter calibration

The TM is used to calibrate the parameters in the GA, DE, PSO and HS algorithms, as the values of meta-heuristic algorithm parameters influence the solution quality. Nevertheless, the present study, the "smaller is better" response is chosen as S/N should be minimized [25]. To perform the Taguchi procedure, the L^9 design is employed with the values and levels of the GA, DE, PSO and HS algorithm parameters outlined in Table 4 and the values derived after multiple tests on the examples of the classes using the frequent algorithm runs.

Table 4 GA, DE, PSO and HS parameters, and levels

Algorithm	Parameters	(1)	(2)	(3)	Algorithm	Parameters	(1)	(2)	(3)
GA	POP	30	40	50	PSO	NOP	30	40	50
	P_c	0.5	0.6	0.7		C_1	1.5	2	2.5
	P_m	0.01	0.05	0.1		C_2	1.5	2	2.5
	NOG	100	200	300		NOG	100	150	200
DE	NOP	20	30	50	HS	HMS	5	10	20
	NOG	30	50	100		HMCR	0.9	0.95	0.99
	P_c	0.1	0.5	0.9		PAR	0.01	0.1	0.3
	-	-	-	-		BW	0.1	0.5	0.9

4.4 Analysis of results and comparisons

To compare the results emanating from four algorithms and elicit the best methodology to solve the CF having the T2F variable, each of 15 examples, is solved using MATLAB (R2016b) software. In this paper, the GA, PSO, HS and DE algorithms are used, in which the probability of crossover (Pc), the generations number (NOG), the probability of mutation (Pm) and the size of population (POP) are the GA parameters. The generations number (NOG), the acceleration coefficients (C1, C2) and the size of population (NOP) are the PSO parameters. The harmony memory considering rate (HMCR), the Bandwidth (BW), the harmony memory size (HMS) and the pitch adjusting rate (PAR) are the HS parameters. The size of population (NOP), the probability of crossover (Pc) and the generations number (NOG) are the DE parameters.

Tables 5 and 6 incorporates the input parameter and the objective values of four algorithms in each example, in which the optimal values of the parameters are derived using L^9 design and the TM. To compare the performance of the GA, PSO, HS, and DE with reference to the objective

function, a number of approaches are utilized in the present research. Initially, the average and the standard deviation of each of the 15 examples were obtained as shown in the last two rows in Table 6. The results from average and standard deviation of the 15 examples show that PSO has outperformed GA, HS, and DE. The graphical approach depicted in Fig. 1 is also applied twice to compare the algorithm performance in 15 produced examples. Moreover, this figure shows that the PSO appears to represent a better performance than the GA, HS, and DE in the objective function in the entirety of the examples.

Table 5 Input parameters of the PSO, FA and DE algorithms for the generated problems

	HS				PSO				GA				DE		
Prob. No.	HMS	$HMCR$	PAR	BW	NOP	C_1	C_2	NOG	POP	P_c	P_m	NOG	NOP	NOG	P_c
1	20	0.99	0.3	0.9	30	2	2	100	50	0.6	0.01	200	50	100	0.9
2	20	0.99	0.3	0.5	50	25	2	200	50	0.5	0.05	100	30	30	0.9
3	10	0.95	0.01	0.9	50	1.5	2	100	50	0.6	0.05	300	50	100	0.9
4	10	0.95	0.01	0.9	40	1.5	2.5	200	30	0.7	0.1	300	30	30	0.9
5	10	0.99	0.3	0.9	40	1.5	1.5	100	40	0.7	0.05	200	50	50	0.1
6	20	0.95	0.1	0.5	30	2	2	200	30	0.5	0.01	200	30	100	0.1
7	5	0.99	0.01	0.1	30	2	2.5	150	50	0.7	0.1	300	30	100	0.1
8	20	0.95	0.01	0.5	40	1.5	1.5	200	50	0.7	0.1	300	30	50	0.1
9	20	0.9	0.01	0.1	30	1.5	2.5	150	30	0.6	0.01	200	50	100	0.1
10	20	0.99	0.01	0.9	40	2	2.5	200	40	0.6	0.05	200	20	100	0.1
11	20	0.95	0.3	0.9	50	2	1.5	200	30	0.6	0.1	100	30	30	0.1
12	20	0.95	0.01	0.1	50	1.5	2	200	50	0.6	0.01	200	50	100	0.9
13	20	0.95	0.3	0.9	50	2	1.5	200	50	0.7	0.05	300	50	50	0.1
14	10	0.9	0.1	0.5	30	2	2	150	30	0.6	0.05	300	20	30	0.1
15	5	0.99	0.1	0.5	40	1.5	1.5	100	30	0.6	0.1	300	30	100	0.9

Table 6 Objective function of the PSO, FA and DE algorithms for the generated problems

Problem No.	HS Objective function	PSO Objective function	GA Objective function	DE Objective function
1	11579	11449	17874	12530
2	49371	34946	55498	53765
3	61232	59110	64833	63498
4	112440	89216	123982	119870
5	134230	100420	143563	139110
6	114140	101930	129390	128765
7	99366	71208	101550	99875
8	127030	126750	148510	144170
9	106320	105770	110289	107540
10	112430	109290	115410	113980
11	41477	37906	61187	49405
12	215630	214010	265670	231720
13	78911	70402	86832	84695
14	221480	175750	240290	238730
15	221080	220720	247720	225775
Average	113781	101925	127507	120895
St. Dev	64446	61827	73328	68000
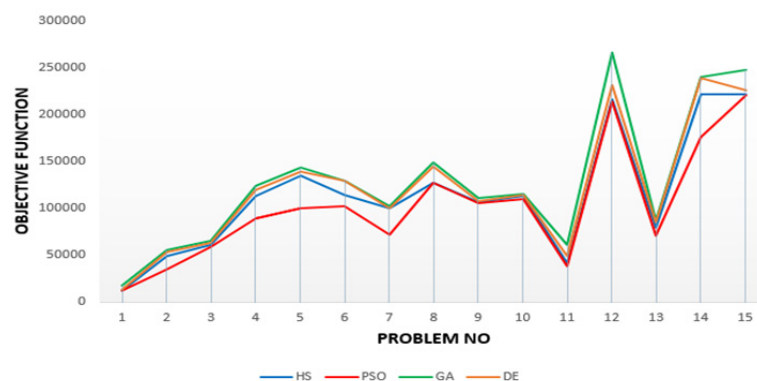
Fig. 1 Trend of the objective function values of the generated problems for the proposed algorithms

Table 7 Analysis of variance results to compare the algorithms in terms of mean objective function value

Source	DF	Adj. SS	Adj. MS	F-Value	P-value
Algorithms	3	5390788752	1796929584	0.40	0.754
Error	56	2.51675E+11	4494189326		
Total	59	2.57065E+11			
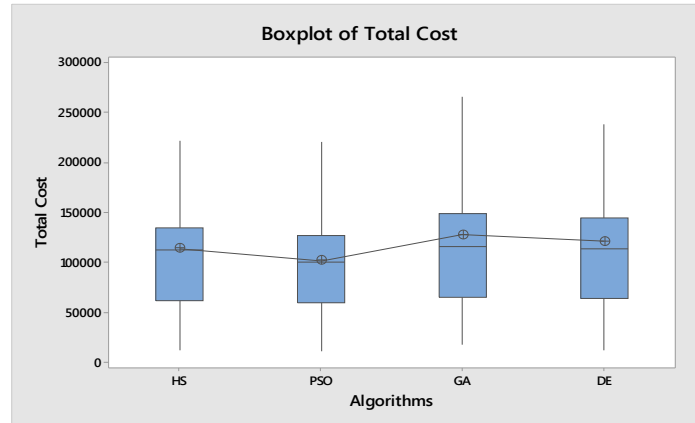
Fig. 2 Boxplot of the total cost values

In the end, the one-way analysis of variance (ANOVA) method is employed to statistically evaluate the performance of the GA, PSO, HS, and DE. This procedure was executed in MINITAB software version 17.3.1. The ANOVA method output outlined in Table 7 demonstrates that at a confidence level of 95 %, four algorithms reveal no significant differences in the mean objective function. Fig. 2 outlines the respective performance of the four algorithms.

5. Conclusion

In this paper, a T2F CF problem was examined with economic and environmental criteria. In the proposed model, the costs associated with processing, material movement, energy loss, and tooling were minimized. The most salient benefits of the mathematical model are as follows: CF using economic and environmental criteria at desirable cost defined by T2F and worker assignment through the man-machine relationship aspect. To solve the presented CF model, a GA was utilized, in which the PSO, HS, and DE algorithms were employed to evaluate the outputs of the proposed algorithm. Another remarkable advantage of the present research is the solution of the 15 random problems produced as the optimal rates of the algorithm parameters using TM for each problem. The results emanating from the algorithms reveal that the PSO algorithm outperforms the GA, DE, and HS algorithms in terms of the objective function on 15 random produced problems. The ANOVA method was also conducted to compare the performance of the GA, PSO, HS and DE algorithms statistically. Moreover, the trend pattern demonstrated that PSO outperformed GA, DE, and HS in the majority of the problems, in which a statistically significant difference was not observed showing that valid results were derived using the PSO. In the end, the following recommendations for further research are made:

- The application of the model can be extended to a stochastic environment.
- The proposed model can be considered with reference to other criteria for sustainability.
- The proposed model may be considered in the multi-period planning horizon.
- The Response Surface Methodology (RSM) may be implemented to set the parameters.
- Future research can concentrate on other meta-heuristic algorithms.
- The model can be extended to other T2F parameters.

References

[1] Rafiei, H., Ghodsi, R. (2013). A bi-objective mathematical model toward dynamic cell formation considering labor utilization, *Applied Mathematical Modelling*, Vol. 37, No. 4, 2308-2316, doi: 10.1016/j.apm.2012.05.015.

[2] Greene, T.J., Sadowski, R.P. (1984). A review of cellular manufacturing assumptions, advantages and design techniques, *Journal of Operations Management*, Vol. 4, No. 2, 85-97, doi: 10.1016/0272-6963(84)90025-1.

[**3**] Mehdizadeh, E., Rahimi, V. (2016). An integrated mathematical model for solving dynamic cell formation problem considering operator assignment and inter/intra cell layouts, *Applied Soft Computing*, Vol. 42, 325-341, doi: 10.1016/j.asoc.2016.01.012.

[4] Majazi Dalfard, V. (2013). New mathematical model for problem of dynamic cell formation based on number and average length of intra and intercellular movements, *Applied Mathematical Modelling*, Vol. 37, No. 4, 1884-1896, doi: 10.1016/j.apm.2012.04.034.

[5] Deljoo, V., Mirzapour Al-e-hashem, S.M.J., Deljoo, F., Aryanezhad, M.B. (2010). Using genetic algorithm to solve dynamic cell formation problem, *Applied Mathematical Modelling*, Vol. 34, No. 4, 1078-1092, doi: 10.1016/j.apm.2009.07.019.

[6] Bagheri, M., Bashiri, M. (2014). A new mathematical model towards the integration of cell formation with operator assignment and inter-cell layout problems in a dynamic environment, *Applied Mathematical Modelling*, Vol. 38, No. 4, 1237-1254, doi:10.1016/j.apm.2013.08.026.

[7] 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.

[8] Zupan, H., Herakovic, N., Zerovnik, J., Berlec, T. (2017). Layout optimization of a production cell, *International Journal of Simulation Modelling*, Vol. 16, No. 4, 603-616, doi: 10.2507/IJSIMM16(4)4.396.

[9] Nie, X.D., Chen, X.D., Chen, X. (2016). Simulation study of flexible manufacturing cell based on token-oriented Petri net model, *International Journal of Simulation Modelling*, Vol. 15, No. 3, 566-576, doi: 10.2507/IJSIMM15(3)CO14.

[10] Mahdavi, I., Mahdavi-Amiri, N., Naderi, F., Paydar, M.M. (2017). Cell formation configuration using interval type-2 fuzzy interactional interests among workers, *International Journal of Operational Research*, Vol. 30, No. 2, doi: 10.1504/IJOR.2017.10007273.

[11] Stocchetti, A. (2012). The sustainable firm: From principles to practice, *International Journal of Business and Management*, Vol. 7, No. 21, 34-47, doi: /10.5539/ijbm.v7n21p34.

[12] Lozano, R. (2008). Envisioning sustainability three-dimensionally, *Journal of Cleaner Production*, Vol. 16, No. 17, 1838-1846, doi: 10.1016/j.jclepro.2008.02.008.

[13] Niakan, F., Baboli, A., Moyaux, T., Botta-Genoulaz, V. (2016). A bi-objective model in sustainable dynamic cell formation problem with skill-based worker assignment, *Journal of Manufacturing Systems*, Vol. 38, 46-62, doi: 10.1016/j.jmsy.2015.11.001.

[14] Niakan, F., Baboli, A., Moyaux, T., Botta-Genoulaz, V.A. (2015). A new multi-objective mathematical model for dynamic cell formation under demand and cost uncertainty considering social criteria, *Applied Mathematical Modelling*, Vol. 40, No. 4, 2674-2691, doi: 10.1016/j.apm.2015.09.047.

[15] Niakan, F., Baboli, A., Moyaux, T., Botta-Genoulaz, V. (2014). A new bi-objective mathematical model for sustainable dynamic cellular manufacturing systems, In: *Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management*, Selangor, Malaysia, 938-942, doi: 10.1109/IEEM.2014.7058776.

[16] Niakan, F., Baboli, A., Moyaux, T., Botta-Genoulaz, V. (2016). A bi-objective model in sustainable dynamic cell formation problem with skill-based worker assignment, *Journal of Manufacturing Systems*, Vol. 38, 46-62, doi: 10.1016/j.jmsy.2015.11.001.

[17] Safaei, N., Saidi-Mehrabad, M., Tavakkoli-Moghaddam, R., Sassani, F. (2008). A fuzzy programming approach for a cell formation problem with dynamic and uncertain conditions, *Fuzzy Sets and Systems*, Vol. 159, No. 2, 215-236, doi: 10.1016/j.fss.2007.06.014.

[18] Qin, R., Liu, Y.K., Liu, Z.Q. (2011). Methods of critical value reduction for type-2 fuzzy variables and their applications, *Journal of Computational and Applied Mathematics*, Vol. 235, No. 5, 1454-1481, doi: 10.1016/j.cam.2010.08.031.

[19] Miller, S., John, R. (2010). An interval type-2 fuzzy multiple echelon supply chain model, *Knowledge-Based Systems*, Vol. 23, No. 4, 363-368, doi: 10.1016/j.knosys.2009.11.016.

[20] Kundu, P., Kar, S., Maiti, M. (2014). Fixed charge transportation problem with type-2 fuzzy variables, *Information Sciences*, Vol. 255, 170-186, doi: 10.1016/j.ins.2013.08.005.

[21] Niebel, B.W. (1982). *Motion and time study*, Homewood, Irwin, USA.

[22] de Castro Hilsdorf, W., de Mattos, C.A., de Campos Maciel, L.O. (2017). Principles of sustainability and practices in the heavy-duty vehicle industry: A study of multiple cases, *Journal of Cleaner Production*, Vol. 141, 1231-1239, doi: 10.1016/ j.jclepro.2016.09.186.

[23] Wang, L.X. (1996). *A course in fuzzy systems and control*, Prentice-Hall International, Inc., U.S.A.

[24] Holland, J.H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence*, MIT Press Cambridge, MA, USA.

[25] Mousavi, S.M., Hajipour, V., Niaki, S.T.A., Aalikar, N. (2014). A multi-product multi-period inventory control problem under inflation and discount: A parameter-tuned particle swarm optimization algorithm, *The International Journal of Advanced Manufacturing Technology*, Vol. 70, No. 9-12, 1739-1756, doi: 10.1007/s00170-013-5378-y.

[26] Noktehdan, A., Karimi, B., Husseinzadeh Kashan, A. (2010). A differential evolution algorithm for the manufacturing cell formation problem using group-based operators, *Expert Systems with Applications,* Vol. 37, No. 7, 4822-4829, doi: 10.1016/j.eswa.2009.12.033.

[27] Storn, R., Price, K. (1997). Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, Vol. 11, No. 4, 341-359, doi: 10.1023/A:1008202821328.

[28] Sahin, O., Akay, B. (2016). Comparisons of metaheuristic algorithms and fitness functions on software test data generation, *Applied Soft Computing*, Vol. 49, 1202-1214, doi: 10.1016/j.asoc.2016.09.045.

[29] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization, In: *Proceedings of the IEEE International Conference on Neural Networks*, Perth, WA, Australia, 1942-1948.

[30] Pant, M., Thangaraj, R., Abraham, A. (2009). Particle swarm optimization: Performance tuning and empirical analysis, *Foundations of Computational Intelligence*, Vol. 3, 101-128, doi: 10.1007/978-3-642-01085-9_5.

[31] Geem, Z.W., Kim, J.H., Loganathan, G.V. (2001). A new heuristic optimization algorithm: Harmony search, *Simulation*, Vol. 76, No. 2, 60-68, doi: 10.1177/003754970107600201.

[32] Askarzadeh, A., Zebarjadi, M. (2014). Wind power modeling using harmony search with a novel parameter setting approach, *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 135, 70-75, doi: 10.1016/j.jweia.2014.10.012.

Prediction of surface roughness in the ball-end milling process using response surface methodology, genetic algorithms, and grey wolf optimizer algorithm

Sekulic, M.[a,*], Pejic, V.[b], Brezocnik, M.[c], Gostimirović, M.[a], Hadzistevic, M.[a]

[a]University of Novi Sad, Faculty of Technical Sciences, Department of Production Engineering, Novi Sad, Serbia
[b]College of Business and Technical Education, Doboj, Bosnia and Herzegovina
[c]University of Maribor, Faculty of Mechanical Engineering, Production Engineering Institute, Maribor, Slovenia

ABSTRACT

In this research study proposed are a response surface methodology (RSM), genetic algorithm (GA) and a grey wolf optimizer (GWO) algorithm for prediction of surface roughness in ball-end milling of hardened steel. The RSM is a conventional predicting approach, GA is an evolutionary algorithm and GWO is a new swarm intelligence-based algorithm. Spindle speed, feed per tooth, axial depth and radial depth of cut were selected as input parameters. Experiments were performed on a CNC milling center and experimental data were collected based on a four-factor-five-level central composite design (CCD). RSM was applied for establishing the basic relationship between input parameters and surface roughness. After that analysis of variance (ANOVA) was conducted for the evaluation of the proposed mathematical model. A predefined reduced quadratic model was used as a reference model for a build-up of predictive models using GA and GWO algorithm. Predicted values of RSM, GA and GWO models are compared with experimental results. In the comparison of model performance for all the three models it was found that GWO model is the best solution. The model accuracy was found to be at 91.80 % and 89.58 % for training and testing data, respectively, which showed the effectiveness of the GWO algorithm for modeling machining processes.

1. Introduction

The primary objective of machining operations modeling is developing a predictive capability for machining performance in order to facilitate effective planning of machining operations to achieve optimum productivity, quality, and cost [1]. Modeling of output machining performances, fundamental variables (cutting forces, cutting temperature, stress, etc.), and output performances which are relevant for machine industry (tool life, surface roughness, surface integrity, chip form, etc.) is very important due to the fact that conventional machining processes still occupy the dominant part of all production processes [2]. The development of advanced predictive models enables the selection optimal cutting conditions, cooling (kinds of cutting fluids, pressure, flow), cutting tool material (coatings) and its geometry, machine tools and other. Predictive models also help manufacturing engineers to plan and manage machining processes more efficiently and use the sophisticated simulation tools to check the effects of projected process performances far beyond the actual production of any particular product(s).

Since the conventional machine tools era to the present times of CNC multi-tasking machine tools, the prediction of machining performances and optimization of cutting conditions have been interesting research areas. Modeling of output machining performances is mainly a very difficult task due to complexity and stochastic nature of most machining processes. Despite both past and present numerous attempts to analyze metal cutting, the basic relationship between the various variables has still remained unexplained [3]. In the last 60 years, metal cutting researchers have applied many methods or techniques for modeling the output machining performances. Predictive models in practice can be analytical, empirical, numerical, Artificial Intelligence (AI) based, and hybrid, too [2]. Each of these approaches has its advantages and disadvantages, or capabilities and constraints. For example, response surface methodology (RSM) is used for empirical models building. In recent years the research in the area of machining process modeling is oriented towards the use of methods based on Artificial Intelligence. Artificial Intelligence-based models are commonly founded on either a biological, molecular, or neurological phenomenon that resembles the metaphor of natural biological evolution and/or the social behavior of different species of natural organisms [4].

These models have been developed thanks to the rapid advancement of computer technology over the past two decades and are often called nature-inspired algorithms. Nature-inspired algorithms could be classified into three wide groups: physics-based algorithms (PBA), chemistry-based algorithms (CBA) and biology-based algorithms (BBA) [5]. Physics-based algorithms (PBA) actually apply the basic principles of physics, for instance, Newton's laws of gravitation, laws of motion and the like.

Biology-based algorithms (BBA) present a special group of algorithms within algorithms inspired by nature and those can learn and adapt similar to biological organisms [6]. These algorithms can be put into three categories: evolutionary algorithms (EA), bio-inspired algorithms (BIA) and swarm intelligence-based algorithms (SIA) [5]. Biology-based algorithms try to mimic the way in which biological organisms or sub-organisms (e.g., bacteria or neurons) function, so as to achieve a high level of efficiency [6]. Evolutionary algorithms (EA) are based on Darwin's theory of evolution and among them are most famous genetic algorithm (GA) and genetic programming (GP).

Bio-inspired algorithms (BIA) are built on the idea of a commonly observed phenomenon in some animal species and ordered, natural movement of organisms. Flocks of birds and shoal of fish are amazing examples of self-organized coordination. For example, Particle Swarm Optimisation (PSO) simulates the social behavior of birds. In PSO, each solution in search for space is actually analogous to a bird and it is called "a particle". The system is started with a population of random particles (called a swarm) and their searches for optimum value continues by updating new generations. Each particle in the swarm tries to reach a possible solution, whereby the group attempts to meet the collective objective of the group, all that based on the actual feedback from the other members.

Swarm intelligence-based algorithms (SIA) use the fact that collective intelligence of swarm is more than a sum of individual intelligences. Each agent of the swarm may be considered as unintelligent, but it follows some simple rules, so that the whole swarm may show some self-organization behavior and thus can behave like some kind of collective intelligence. These algorithms began from observing the collective behaviors of insects living in colonies such as ants, bees, wasps, termites, respectively and also the collective behaviors of some animal species, such as cuckoos, lions, wolves, etc. The most famous algorithms based on swarm intelligence are Artificial Bee Colony (ABC), Ant Colony Optimization (ACO), Cuckoo search (CS).

The aim of this work is to present the various approaches to predict surface roughness in ball-end milling process and developing a predictive model to obtain surface roughness as a function of machining parameters: spindle speed (n), feed per tooth (f_z), axial depth of cut (a_p) and radial depth of cut (a_e), using response surface methodology (RSM), genetic algorithm (GA) and grey wolf optimizer (GWO) algorithm. RSM is conventional modeling approach, GA is a part of evolutionary algorithms and GWO is new swarm intelligence-based algorithm. GWO is a recently developed algorithm inspired by grey wolves and their life in nature (as for example leadership hierarchy and hunting mechanism) and has been successfully applied for solving different

problems. As the grey wolf optimizer (GWO) is a real novelty, there is no study in the literature about the application of this algorithm for modeling and optimization of machining operations, in other words for prediction of output machining performances.

2. Literature review

One of the most important machining performances is surface roughness, particularly so in finish milling operations [7]. When referencing to the already published papers, obviously there have been a lot of researches regarding the prediction of surface roughness in face, but also end milling process. For this purpose, a number of statistical methods were used such as RSM and soft computing techniques or bio-inspired computing (ANN, GA, GP, ANFIS).

Numerous researchers have studied the influence of input cutting parameters on surface roughness for practical end milling. Most of the research proposal is the multiple regression method to predict surface roughness [8, 9], some research applied ANN, fuzzy logic, ANFIS, GA and grey-fuzzy approaches for surface roughness prediction in end milling process [10-15], everyone proposed GP, namely RSM to predict surface roughness in end milling [16, 15].

Ball-end milling is a type of milling process where ball-end mill is used with the goal to generate 3D free formed sculptured products. In reality, die, mold and aerospace industries mostly apply that type of process. There is a far smaller number of published references dealing with the prediction of the surface roughness for this type of end milling process.

Dhokia *at al*. [17] used GA for predicting of surface roughness in ball-end milling of polypropylene. Establishing a relationship between surface roughness and the process parameters such as feed, speed, and depth of cut for the ball-end milling of polypropylene was the most significant goal for the surface roughness model. The experimental tests were carried out according to the orthogonal array design L_{16}. The mean deviation of R_a obtained by surface roughness model over the validation dataset was obtained as 8.43 %.

In their study Vakondios *at al*. [18] addressed the influence of milling strategy on the surface roughness in ball end milling of the aluminum alloy Al7075-T6. Different strategies were analyzed (vertical, push, pull, oblique, oblique push and oblique pull). A mathematical model of the surface roughness was established for each of them, considering both the down and up milling. Mathematical models for surface roughness were obtained using RSM. For a check up of the validity of models used was the analysis of variance (ANOVA).

Hossain and Ahmad [19] decided to attempt with RSM and adaptive-neuro-fuzzy-inference system (ANFIS) in order to predict surface roughness in ball-end milling. After comparing ANFIS results with the RSM results the superiority of ANFIS results to the RSM results was showed.

Zuperl and Cus [20] developed a simplified model to provide functional relations between surface roughness, cutting chip size, cutting conditions and diameter of the ball-end mill. The ANFIS method was applied for predicting the cutting chip size in ball-end milling. In spite of being simple, the model has reliability and efficiency in predicting the R_a in-process by utilizing the chip size.

Quintana *at al*. [21] used artificial neural networks (ANN) for predicting of surface roughness in the function of spindle speed, feed per tooth, axial depth of cut, radial depth of cut and tool radius. The developed ANN enabled the prediction of surface roughness with a high degree of correlation (R^2 = 0.96296).

3. Materials and methods

3.1 Experimental setup and results

The experimental work explained in this paper referenced the work of Pejic V. [22], and was performed at the Department of Production Engineering, Faculty of Technical Sciences, at University of Novi Sad and at the company "ELMETAL" doo in Senta, Serbia. The experiments were conducted on HAAS VF-3YT vertical three-axis CNC milling machine and on hardened steel X210CR12 with 58 HRC. The cutting tools used were TiAlN-T3 coated two-flutes solid carbide

ball-end milling cutters of diameter 6 mm (Emuge-Franken, type1877A). The dimension of a workpiece 300 mm × 58 mm × 20 mm was used in this study. Based on the tool manufacturer's recommendation, a cold air nozzle, which works on the principle of a vortex tube, was used for cooling the tool. Prior to performing the experiments, the workpiece was divided into 84 fields, with dimensions of 15.33 mm × 3 mm and a height of 2 mm, as shown in Fig. 1. Each field corresponded to one experimental point. This method enables machining in one clamp, and hence the same machining conditions at all experimental points. According to the data taken from the design of experiments (DoE) CNC programming code was set up in Edgecam 2015 for each and every experimental point, as seen in Table 2. The surface roughness value of the machined workpiece was measured using the MarSurf PS1 portable unit.

Applying the rotatable central composite design (RCCD), the Design of experiment (DOE) was obtained. Using different combinations of the input parameters levels performed was a total of 30 experiments. The machining parameters chosen to analyze their effect on surface roughness were spindle speed, feed per tooth, axial depth of cut and radial depth of cut. Values of cutting conditions have been determined for a 4-factor design of experiments according to tool producer recommendations and the workpiece material, Table 1.
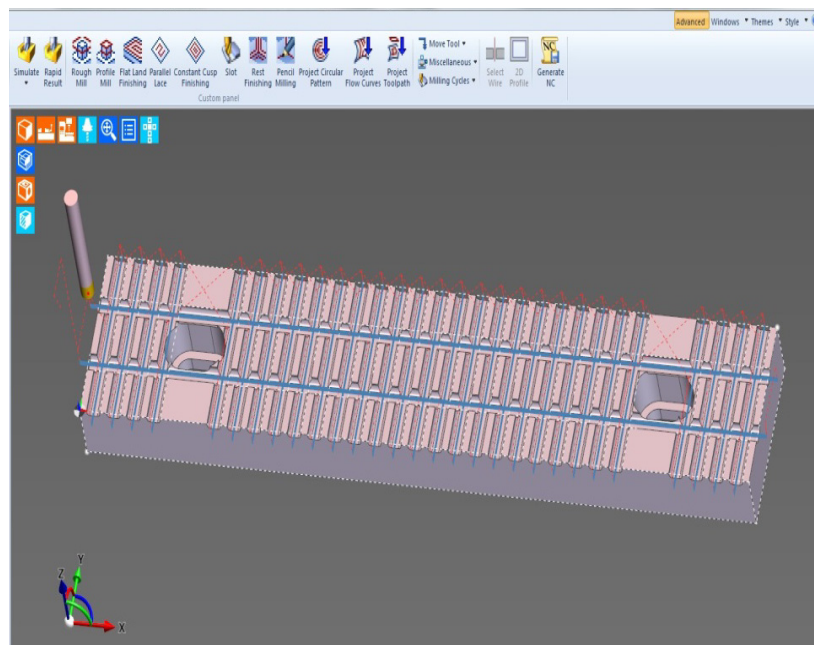
Fig. 1 Test part designed in Edgecam 2015

Table 1 Machining parameters and their levels

Parameters	Levels				
	-2	-1	0	1	1
Spindle speed, n (min^{-1})	3981	4777	5573	6369	7169
Feed per tooth, f_z (mm/tooth)	0.018	0.024	0.030	0.036	0.042
Axial depth of cut, a_p (mm)	0.04	0.08	0.12	0.16	0.20
Radial depth of cut, a_e (mm)	0.20	0.40	0.60	0.80	1.00

Spindle speed is calculated by the equation below:

$$n = \frac{v}{2 \cdot \pi \cdot \sqrt{a_p \cdot (d_1 - a_p)}} \tag{1}$$

where v is the cutting speed, a_p is the axial depth of cut and d_1 is a diameter of the tool.

Measured results of surface roughness are presented in Table 2.

Table 2 Experimental results for surface roughness R_a

Trial No.	Code					Parameters				Measured value
	x_0	x_1	x_2	x_3	x_4	n (min^{-1})	f_z (mm/z)	a_p (mm)	a_e (mm)	R_a (µm)
1	1	-1	-1	-1	-1	4777	0.024	0.08	0.40	0.745
2	1	1	-1	-1	-1	6369	0.024	0.08	0.40	0.305
3	1	-1	1	-1	-1	4777	0.036	0.08	0.40	0.643
4	1	1	1	-1	-1	6369	0.036	0.08	0.40	0.497
5	1	-1	-1	1	-1	4777	0.024	0.16	0.40	0.662
6	1	1	-1	1	-1	6369	0.024	0.16	0.40	0.569
7	1	-1	1	1	-1	4777	0.036	0.16	0.40	0.850
8	1	1	1	1	-1	6369	0.036	0.16	0.40	0.425
9	1	-1	-1	-1	1	4777	0.024	0.08	0.80	3.370
10	1	1	-1	-1	1	6369	0.024	0.08	0.80	3.040
11	1	-1	1	-1	1	4777	0.036	0.08	0.80	3.302
12	1	1	1	-1	1	6369	0.036	0.08	0.80	3.149
13	1	-1	-1	1	1	4777	0.024	0.16	0.80	3.261
14	1	1	-1	1	1	6369	0.024	0.16	0.80	3.116
15	1	-1	1	1	1	4777	0.036	0.16	0.80	3.379
16	1	1	1	1	1	6369	0.036	0.16	0.80	3.113
17	1	0	0	0	0	5573	0.030	0.12	0.60	1.677
18	1	0	0	0	0	5573	0.030	0.12	0.60	1.518
19	1	0	0	0	0	5573	0.030	0.12	0.60	1.571
20	1	0	0	0	0	5573	0.030	0.12	0.60	1.296
21	1	-2	0	0	0	3981	0.030	0.12	0.60	1.926
22	1	2	0	0	0	7166	0.030	0.12	0.60	1.159
23	1	0	-2	0	0	5573	0.018	0.12	0.60	1.334
24	1	0	2	0	0	5573	0.042	0.12	0.60	1.299
25	1	0	0	-2	0	5573	0.030	0.04	0.60	1.324
26	1	0	0	2	0	5573	0.030	0.20	0.60	1.285
27	1	0	0	0	-2	5573	0.030	0.12	0.20	0.245
28	1	0	0	0	2	5573	0.030	0.12	1.00	4.258
29	1	0	0	0	0	5573	0.030	0.12	0.60	1.470
30	1	0	0	0	0	5573	0.030	0.12	0.60	1.471

3.2 Response surface methodology (RSM) background

A collection of statistical and mathematical methods is named response surface methodology (RSM) and it can be used in modeling and optimization of different machining processes [23-25]. In the actual development of conventional predictive modeling this methodology is commonly present. The input variables are here referred to as independent variables, and they are subjected to the control of an engineer or a scientist, with the purpose of completing a test or an experiment.

Measurable output performance of the process is called response and it is a dependent variable being tested. RSM quantifies the relationship between the controllable input parameters and the obtained response, in other words, attempts to analyze the influence of independent parameters on a specific dependent response. For modeling and optimization of machining processes data are needed which are collected through experimental work. This methodology represents the empirical statistical technique, which is applied for the regression analysis of the data obtained through the experiment in order to obtain the equation which represents the response function (depending on the variable size being examined). This function can be graphically displayed as a response surface, whereby which this methodology has been named. A few advantages that Response surface methodology (RSM) offers when compared to the classical experimental or optimization methods where only one variable at a time technique is used are as follows [4]: implies a large amount of information from not so many experiments and a possibil-

ity to notice easily the interaction effect of the independent parameters on the response. To obtain information about the process useful was the empirical model as it linked the response to the independent variables and it could also be a practical tool for the optimization of machining processes.

According to RSM, the measurability of all the input process parameters is assumed and the corresponding equation expresses the process:

$$y = f(x_1, x_2, \ldots, x_k) + \varepsilon \tag{2}$$

where y is the response, f is the unknown function of response, x_1, x_2,…, x_k refer to the independent parameters or variables, k is the independent variables number and in the end ε is the statistical error that denotes other sources of variability not accounted for by f.

In RSM two models are commonly used. The first-degree model:

$$y = b_0 + \sum_{i=1}^{k} b_i \cdot x_i + \varepsilon \tag{3}$$

and the second-degree model:

$$y = b_0 + \sum_{i=1}^{k} b_i \cdot x_i + \sum_{i=1}^{k} b_{ii} \cdot x_i^2 + \sum_{j>1}^{k} b_{ij} \cdot x_i \cdot x_j + \varepsilon \tag{4}$$

where b_0 is coefficient of the free term, coefficients b_i are the linear terms, coefficients b_{ii} are quadratic terms and coefficients b_{ij} are interaction terms.

Value of coefficients b_i, b_{ii}, b_{ij} is determined using the method of least squares (MLS).

3.3 Genetic algorithm (GA) background

The GA is a search algorithm for optimization, based on a Darwinian theory of evolution and on the concept of "survival of the fittest". As in nature, the strong species remain intact, while the sleazy species is eliminated. The two most significant advantages of the GA approach are its simplicity of operation and computational efficiency. GA deals with chromosome populations. Actually, string representations of solutions to a particular problem are called chromosomes. Using the real analogy with biology, the chromosome is presented as the genotype, whereas the solution it describes is called the phenotype.

The simplest form of GA involves three types of operators: selection, crossover, and mutation. Fig. 2 illustrates the flowchart of GA for optimization [26]. For using this algorithm, a problem solution is defined in terms of the fitness function. A fitness function is used to evaluate each of the solutions in the population, represented by the chromosomes. Defining this function for the given problem is one of the most difficult tasks in creating a good genetic algorithm.

To allow the entire range of possible solutions (the search space) the original chromosome population is created randomly. Several hundreds or thousands of possible solutions is a typical content of the population size. The fitness function measures the quality of the represented solution and it depends on the nature of a problem. Selection is the process of choosing two or more parents from the population for crossing. The mixing of genetic material from two selected parent chromosomes to produce one or two child chromosomes is called the crossover operator. That crossover operation can be presented in many alternatives, and mutation is one of last GA operators. This is a background operator which produces spontaneous random changes in various chromosomes. The purpose of mutation in GA is preserving and introducing diversity. Population size, number of generations, crossover rate, and mutation rate are the dependable variables for the GA performance.

The GA can actually be stopped under some strict general criteria. Looking at the number of generations the stopping criteria are prescribed. In case the maximum number of the generations exceeds the number of generations, the GA process is terminated and optimised results are provided [17].
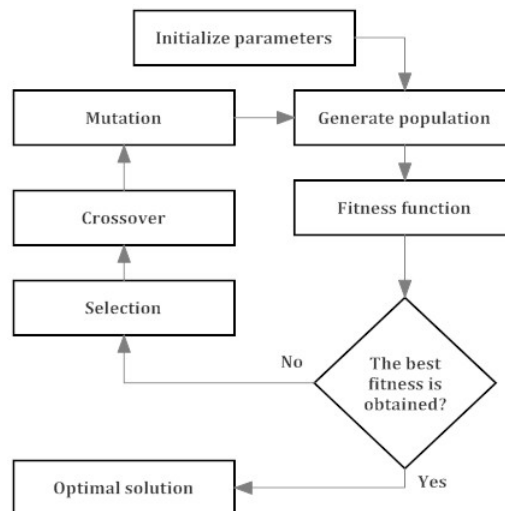
Fig. 2 The flow of GA for optimization

The pseudo code of the GA is presented in Fig. 3 [17]:

```
begin
        k = 0;
        initialize (k);
        evaluate P(t);
while (gen ≤ Max_gen)
        begin
        k = k + 1
        evaluate P(k –1)
        select P(k) from P(k – 1)
        crossover on P(k)
        mutation on P(k)
        evaluate P(t)
        end
end
```
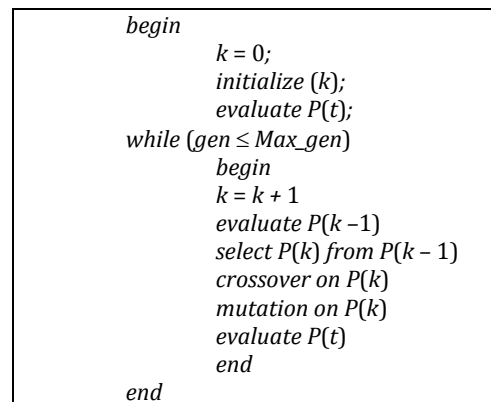
Fig. 3 Pseudo code of the GA

3.4 Gray wolf optimizer (GWO) algorithm background

The leadership hierarchy and hunting mechanism of grey wolves in nature is imitated in the grey wolf optimizer (GWO) algorithm. In 2014 S. Mirjalilli *at al*. developed the algorithm [27]. Grey wolves mainly have a preference for living in a pack. To simulate the leadership hierarchy observed were four types of grey wolves such as alpha beta, delta, and omega. The alpha is actually the first level in the hierarchy and is referred to as the leader of the pack. In the hierarchy of grey wolves beta is the second, next level. The betas are subordinated to alphas but can be of assistance in decision-making or other pack activities. Omega is the lowest ranking of grey wolves. Even delta wolves are dominating over omega, but of course they have to submit to alphas and betas.

The main inspiration of this algorithm apart from their social leadership was the hunting technique of grey wolves. Alpha wolves always lead the way, but grey wolves hunt in groups and coordinate with each other very well. Multiple steps are taken when hunting a prey. That hunting can be divided into a few main stages: tracking, chasing, and approaching the prey; pursuing, encircling and harassing the prey until it stops moving; attack towards the prey [27]. Mathematical modeling hunting technique and social behavior of grey wolves is observed in this GWO algorithm. The prey is located by alpha, beta and delta through its body stance, uncoordinated movements or the smell of wounds and then the chasing starts. The dominant wolves are followed by omegas. The prey is first approached as near as possible, encircled and then harassed until it stops moving. The next step by the pack is jumping and attacking the prey. The flowchart of the GWO algorithm is shown in Fig. 4 [28].
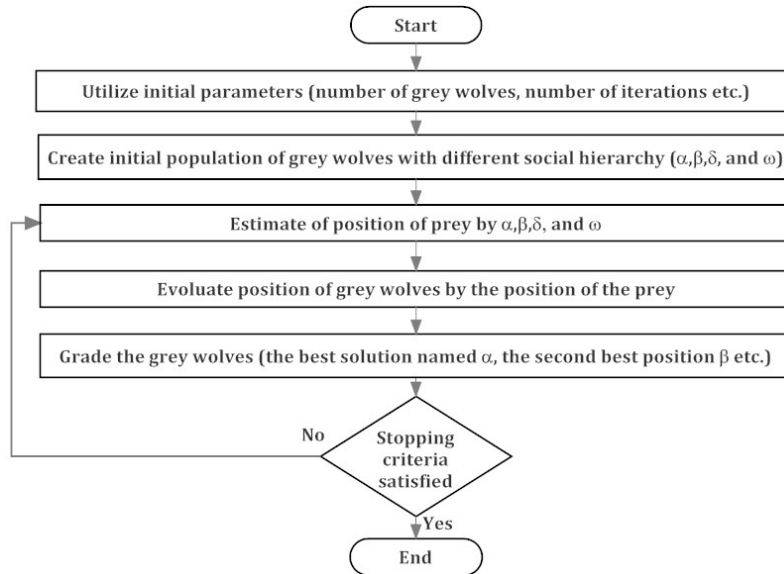
Fig. 4 Flowchart of the GWO algorithm

The most appropriate solution is α and this solution is followed by β and δ, respectively, and the rest of the solutions belong to the ω, in respect to the GWO. The first three fittest wolves that are closest to the prey are α, β, and δ who guide ω to search prey in promising search areas. Omega updates its location based on the location of alpha, beta, and delta in a 2D search space. In the GWO algorithm the search process is started by forming a random population of grey wolves (candidate solutions). The estimation of probable position of the prey is conducted by alpha, beta, and delta wolves using multiple iterations. The distance from the prey is updated by each candidate solution.

With the goal to stimulate the main phases of grey wolf hunting proposed are more vector equations [27], which indicate the position of the prey and grey wolfs during encircling prey, hunting and attacking prey. The pseudo code of the GWO algorithm is presented in Fig. 5 [27] where t indicates the current iteration, a is a component which linearly decreased from 2 to 0 over the course of iterations, A and C are coefficient vectors, $X_{\alpha\beta\delta}$ are the position vectors of grey wolfs.

```
Initialize the grey wolf population Xi (i = 1,2,...,n)
Initialize a, A and C
Calculate the fitness of each search agent
Xα = the best search agent
Xβ = the second best search agent
Xδ = the third best search agent
while (t < Max number of iterations)
    for each search agent
            Update the position of the current search agent
    end for
Update a, A and C
Calculate the fitness of all search agent
Update Xα, Xβ and Xδ
t = t + 1
end while
return Xα
```

Fig. 5 Pseudo code of the GWO algorithm

4. Results and discussion

4.1 Modeling surface roughness by RSM

Applying Design Expert software created was RSM model with the intention to model and analyze surface roughness. Setting up a relationship between surface roughness and the process parameters such as spindle speed, feed per tooth, axial depth of cut and radial depth of cut a for the ball-end milling of hardened steel is the most important objective of the surface roughness model. The mathematical model for surface roughness as a function of machining parameters was developed by using a reduced second-order polynomial response surface mathematical equation. Analysis of variance (ANOVA) for response surface reduced quadratic model is the basis for this kind of equation, as shown in Table 3.

The developed mathematical model to predict surface roughness R_a is:

$$R_{a(RSM)} = 0.95 - 1.85 \cdot 10^{-4} \cdot n + 1.53 \cdot f_z + 0.26 \cdot a_p - 0.85 \cdot a_e + 5.76 \cdot a_e^2 \qquad (5)$$

Table 3 Choice of model type based on ANOVA

Response	R_a						
ANOVA for response surface							
Analysis of variance table (Partial sum of squares – Type III)							
Source	Sum of squares	df	Mean square	F Value	p-value Prob > F		PC (%)
Model	37.24	5	7.45	140.10	< 0.0001	significant	
A-n	0.52	1	0.52	9.78	0.0046		1.35
B-f_z	2.017E-03	1	2.017E-03	0.038	0.8472		0.01
C-a_p	2.521E-03	1	2.521E-03	0.047	0.8294		0.01
D-a_e	35.19	1	35.19	661.90	< 0.0001		91.36
D^2	1.53	1	1.53	28.72	< 0.0001		3.96
Residual	1.28	24	0.053				3.31
Lack of fit	1.20	19	0.063	3.93	0.0677	not significant	3.10
Pure error	0.080	5	0.016				0.21
Corrected total	38.51	29					100
	R² = 0.9669; Adj R² = 0.9599						

With the aim to justify the validity of the model the ANOVA was conducted. The p-value is lower than 0.05 which proves that the model is considered adequate at the 95 % confidence level. The validity of the model is confirmed, by the determination coefficient $R^2 = 0.9669$. If R^2 convergences unity the response model gives better results and there exists less difference between predicted and measured data. The variability measure of the observed output is the adjusted correlation coefficient (Adj $R^2 = 0.96$). The approximate value of the Adj R^2 with R^2 determines the fitness of the model The p-value is separately calculated for all the parameters of the proposed model and it can be concluded the radial depth of cut a_e (p < 0.0001) is the most significant parameter on surface roughness.

4.2 Modeling surface roughness by GA

The RSM was used to develop basically the reduced second-order polynomial response surface mathematical model for prediction of surface roughness in ball-end milling and GA was used for fine-tuning of the constants in Eq. 5, which obtained from RSM. The fine-tuning of the constants in GA is performed in order to find the minimum value of the fitness function.

The fitness function is defined as:

$$\Delta = \frac{1}{n} \sum_{i=1}^{n} \frac{|E_i - G_i|}{E_i} \cdot 100\% \qquad (6)$$

where n is the size of sample data, E_i the measured R_a and G_i the predicted R_a calculated by GA.

The lower the values of Eq. 6, the better agreement of the model is to the experimental data. For implementing GA GATool was used in MATLAB. The GA predictive model is developed using 25 datasets selected based on experimental results, without 6 datasets on the average level (center points), Table 2. Six datasets on the average level were used as one average value. The best result was obtained with population size of 1500. The developed mathematical model to predict surface roughness R_a using GA is:

$$R_{a(GA)} = 1.48 - 1.85 \cdot 10^{-4} \cdot n + 4.75 \cdot f_z + 0.79 \cdot a_p - 3.94 \cdot a_e + 8.8 \cdot a_e^2 \tag{7}$$

4.3 Modeling surface roughness by GWO algorithm

The fitness/objective function in GWO algorithm is calculated identically as in GA, using Eq. 6 and same datasets of experimental results. Solving of the optimization problem is finding the minimum value of fitness function. For implementing GWO was used GWO toolbox in MATLAB, developed by S. Mirjalilli *at al.* [27]. This toolbox is very simple and can be used without high programming skills because its user-friendly graphical interface. In the toolbox the parameters of the GWO algorithm parameters can easily be calculated. Cost Function is the objective function default name for this toolbox.

The developed mathematical model to predict surface roughness R_a using GWO algorithm is:

$$R_{a(GWO)} = 1.47 - 1.81 \cdot 10^{-4} \cdot n + 4.30 \cdot f_z + 0.86 \cdot a_p - 4.0 \cdot a_e + 8.88 \cdot a_e^2 \tag{8}$$

4.4 Comparison of RSM, GA and GWO model performance

Predicted values for surface roughness as obtained in the RSM, GA and GWO are compared with the experimental values. In Table 4 presented are the compared values of the predictive performance for all the three models (RSM model, GA model, and GWO model) with the measured value. The prediction accuracy PA of each datasets was calculated using Eq. 9 [29]. The model accuracy of the developed RSM, GA and GWO models to predict the surface roughness, was evaluated using Eq. 10.

$$PA = \left[1 - \frac{|Expt\_value_i - Model\_pred_i|}{Expt\_value_i} \right] \cdot 100 \tag{9}$$

$$MA = \frac{1}{n} \sum_{i=1}^{n} PA_i \tag{10}$$

The model accuracy of the RSM, GA and GWO models are 86.79 %, 91.78 %, and 91.80 %, respectively. It can be concluded that the model accuracy of the GWO and GA models almost the same although they have been developed using the various approaches to predict surface roughness in ball-end milling process and different pseudo codes. Both models showed good agreement with the experimental data, but GWO model was more accurate than the GA model. It should be noted that the GWO algorithm is easier for application than the GA and the time required to calculate the minimum value of the fitness function is shorter than when using the GA.

Confirmation for the models developed has also been conducted using ten additional experiments, which were randomly selected from the set of experiments performed as according to Taguchi orthogonal array L_{25} (5^6). Those confirmation tests have also proved good accuracy of all the models obtained, having in mind the model based on GWO enabled the best predictability of surface roughness for the given case. In Table 5 presented are the values of the input parameters as well as the results of the roughness measured for all the ten additional measurements.

Table 4 Comparison of RSM, GA, and GWO predictive models

Trial No.	Measured value R_a (µm)	RSM Predicted value R_a (µm)	RSM Prediction accuracy (%)	GA Predicted value R_a (µm)	GA Prediction accuracy (%)	GWO Predicted value R_a (µm)	GWO Prediction accuracy (%)	
1	0.745	0.705	94.63	0.600	80.56	0.593	79.66	
2	0.305	0.411	65.25	0.306	99.79	0.305	99.96	
3	0.643	0.724	87.40	0.657	97.80	0.645	99.68	
4	0.497	0.429	86.32	0.363	72.97	0.356	71.73	
5	0.662	0.726	90.33	0.663	99.82	0.662	99.99	
6	0.569	0.432	75.92	0.369	64.79	0.373	65.64	
7	0.850	0.745	87.65	0.720	84.73	0.714	83.96	
8	0.425	0.450	94.12	0.426	99.85	0.425	99.98	
9	3.370	3.130	92.88	3.246	96.31	3.254	96.55	
10	3.040	2.836	93.29	2.951	97.08	2.965	97.53	
11	3.302	3.149	95.37	3.303	99.98	3.305	99.90	
12	3.149	2.854	90.63	3.008	95.53	3.017	95.80	
13	3.261	3.151	96.63	3.309	98.54	3.322	98.12	
14	3.116	2.856	91.66	3.014	96.73	3.034	97.35	Training data
15	3.379	3.169	93.79	3.366	99.61	3.374	99.85	
16	3.113	2.875	92.35	3.071	98.66	3.085	99.11	
17	1.677	1.560	93.02	1.484	88.48	1.484	88.50	
18	1.518	1.560	97.23	1.484	97.75	1.484	97.77	
19	1.571	1.560	99.30	1.484	94.45	1.484	94.47	
20	1.296	1.560	79.63	1.484	85.51	1.484	85.49	
21	1.926	1.854	96.26	1.778	92.33	1.773	92.04	
22	1.159	1.265	90.85	1.189	97.40	1.195	96.87	
23	1.334	1.541	84.48	1.427	93.04	1.432	92.62	
24	1.299	1.578	78.52	1.541	81.38	1.536	81.78	
25	1.324	1.539	83.76	1.421	92.69	1.415	93.09	
26	1.285	1.580	77.04	1.547	79.62	1.553	79.17	
27	0.245	0.056	22.86	0.246	99.75	0.245	99.99	
28	4.258	4.906	84.78	5.537	69.97	5.565	69.30	
29	1.470	1.560	93.88	1.484	99.06	1.484	99.04	
30	1.471	1.560	93.95	1.484	99.13	1.484	99.11	
Model accuracy			**86.79**		**91.78**		**91.80**	
1	1.587	1.854	83.16	1.778	87.94	1.773	88.30	
2	1.402	1.678	80.30	1.543	89.97	1.542	89.98	
3	3.235	3.141	97.08	3.277	98.70	3.288	98.37	
4	5.259	5.064	96.29	5.715	91.32	5.744	90.78	
5	0.523	0.598	85.57	0.576	89.88	0.578	89.51	Testing data
6	1.328	1.548	83.43	1.449	90.86	1.441	91.47	
7	3.640	3.010	82.70	3.184	87.47	3.187	87.55	
8	1.602	1.424	88.89	1.371	85.59	1383	86.30	
9	4.851	4.758	98.07	5.386	88.96	5.412	88.43	
10	3.405	2.710	79.60	2.870	84.28	2.898	85.10	
Model accuracy			**87.51**		**89.50**		**89.58**	

Table 5 Experimental results of confirmation tests

Trial No.	Parameters n (min⁻¹)	f_z (mm/z)	a_p (mm)	a_e (mm)	Measured value R_a (µm)
1	3981	0.030	0.12	0.60	1.587
2	4777	0.018	0.08	0.60	1.402
3	4777	0.024	0.12	0.80	3.235
4	4777	0.030	0.16	1.00	5.259
5	5573	0.030	0.20	0.40	0.523
6	5573	0.036	0.04	0.60	1.328
7	5573	0.042	0.08	0.80	3.640
8	6369	0.024	0.20	0.60	1.602
9	6369	0.036	0.08	1.00	4.851
10	7166	0.018	0.20	0.80	3.405

5. Conclusion

This paper presents the predictive models for surface roughness R_a during ball-end milling process which were developed using RSM, GA and GWO algorithm. In the first step of the research basic mathematical model was developed by the use of RSM. The developed model is adequate. The validity of the model was confirmed using ANOVA. This reduced quadratic model was used in next steps as basic shape for a build-up of predictive models using GA and GWO algorithm. In the second step, the predictive models were developed applying GA and GWO algorithm. The fitness function was same for both algorithms and was calculated identically, using Eq. 6. For implementing GA and GWO algorithm were used toolboxes in MATLAB. The predictive capability developed models were compared. Experimental results were compared with predicted values for all three the models. The predictive model developed using GWO algorithm providing the best prediction accuracy. The model accuracy for surface roughness was 91.8 % and 89.58 % for training and testing data, respectively. On comparison RSM, GA and GWO models were found that nature-inspired algorithms show the good ability for prediction of surface roughness in ball-end milling process. Results have confirmed that new swarm intelligence-based algorithm, called GWO as useful for modeling machining processes.

Acknowledgement

References

[1] Van Luttervelt, C.A., Childs, T.H.C., Jawahir, I.S., Klocke, F., Venuvinod, P.K., Altintas, Y., Armarego, E., Dornfeld, D., Grabec, I., Leopold, J., Lindstrom, B., Lucca, D.,Obikawa, T., Shirakashi, Sato, H. (1998). Present situation and future trends in modelling of machining operations progress report of the CIRP working group 'Modelling of machining operations', *CIRP Annals*, Vol. 47., No. 2, 587-626, doi: 10.1016/S0007-8506(07)63244-2.

[2] Arrazola, P.J., Özel, T., Umbrello, D., Davies, M., Jawahir, I.S. (2013). Recent advances in modelling of metal machining processes, *CIRP Annals*, Vol. 62., No. 2, 695-718, doi: 10.1016/j.cirp.2013.05.006.

[3] Finnie, I. (1956). Review of the metal cutting analysis of the past hundred years, *Mechanical Engineering*, Vol. 78, No. 8, 715-721.

[4] Venkata Rao, R. (2011). *Advanced modelling and optimization of manufacturing processes*, Springer-Verlag, London ,United Kingdom, doi: 10.1007/978-0-85729-015-1.

[5] Siddique, N., Adeli, H. (2015). Nature inspired computing: An overview and some future directions, *Cognitive Computation,* Vol. 7, No. 6, 706-714, doi: 10.1007/s12559-015-9370-8.

[6] Kar, A.K. (2016). Bio inspired computing – A review of algorithms and scope of applications, *Expert Systems with Applications*, Vol. 59, 20-32, doi: 10.1016/j.eswa.2016.04.018.

[7] Simunovic, G., Simunovic, K., Saric, T. (2013). Modelling and simulation of surface roughness in face milling. *International Journal of Simulation Modelling*, Vol. 12, No. 3, 141-153, doi: 10.2507/IJSIMM12(3)1.219.

[8] Lou, S.-J. (1997). *Development of four in-process surface recognition systems to predict surface roughness in end milling,* Ph.D. Thesis, Iowa State University, Iowa, USA.

[9] Lou, M.S., Chen, J.C., Li, C.M. (1998). Surface roughness prediction technique for CNC end-milling, *Journal of Industrial Technology*, Vol. 15, No. 1, 2-6.

[10] Chen, J.C., Lou, M.S. (2000). Fuzzy-nets based approach to using an accelerometer for in-process surface roughness prediction system in milling operations, *International Journal of Computer Integrated Manufacturing*, Vol. 13, No. 4, 358-368, doi: 10.1080/095119200407714.

[11] Ali, Y.M., Zhang, L.C. (1999). Surface roughness prediction of ground components using a fuzzy logic approach, *Journal of Material Processing Technology*, Vol. 89-90, 561-568, doi: 10.1016/S0924-0136(99)00022-9.

[12] Chen, J. (2000). Neural networks and neural-fuzzy approaches in an in-process surface roughness recognition system for end milling, In: Kusiak, A., Wang, J. (eds.), *Computational intelligence in manufacturing handbook*, CRC Press, Boca Raton, USA, doi: 10.1201/9781420041934.ch16.

[13] Suresh, P.V.S., Venkateswara Rao, P., Deshmukh, S.G. (2002) A genetic algorithmic approach for optimization of surface roughness prediction model, *International Journal of Machine Tools and Manufacture,* Vol. 42, No. 6, 675-680, doi: 10.1016/S0890-6955(02)00005-6.

[14] Tamiloli, N., Venkatesan, J., Vijaya Ramnath, B. (2016). A grey-fuzzy modeling for evaluating surface roughness and material removal rate of coated end milling insert, *Measurement*, Vol. 84, 68-82, doi: 10.1016/ j.measurement. 2016.02.008.

[15] Karkalos, N.E., Galanis, N.I., Markopoulos, A.P. (2016). Surface roughness prediction for the milling of Ti–6Al–4V ELI alloy with the use of statistical and soft computing techniques, *Measurement,* Vol. 90, 25-35, doi: 10.1016/j.measurement.2016.04.039.

[16] Brezocnik, M., Kovacic, M., Ficko, M. (2004). Prediction of surface roughness with genetic programming, *Journal of Materials Processing Technology*, Vol. 157-158, 28-36, doi: 10.1016/j.jmatprotec.2004.09.004.

[17] Dhokia, V.G., Kumar, S., Vichare, P., Newman, S.T. (2008). An intelligent approach for the prediction of surface roughness in ball-end machining of polypropylene, *Robotics and Computer-Integrated Manufacturing,* Vol. 24, No. 6, 835-842, doi: 10.1016/j.rcim.2008.03.019.

[18] Vakondios, D., Kyratsis, P., Yaldiz, S., Antoniadis, A. (2012). Influence of milling strategy on the surface roughness in ball end milling of the aluminum alloy Al7075-T6, *Measurement*, Vol. 45, No. 6, 1480-1488, doi: 10.1016/j.measurement.2012.03.001.

[19] Hossain, S.J., Ahmad, N. (2012). Surface roughness prediction modeling for AISI 4340 after ball end mill operation using artificial intelligence, *International Journal of Scientific & Engineering Research*, Vol. 3, No. 5, 1-10, doi: 10.3968/j.mse.1913035X20120602.2933.

[20] Zuperl, U., Cus, F. (2015). Simulation and visual control of chip size for constant surface roughness, *International Journal of Simulation Modelling*, Vol. 14, No. 3, 392-403, doi: 10.2507/IJSIMM14(3)2.282.

[21] Quintana, G., Garcia-Romeu, M.L., Ciurana, J. (2011). Surface roughness monitoring application based on artificial neural networks for ball-end milling operations, *Journal of Intelligent Manufacturing*, Vol. 22, No. 4, 607-617, doi: 10.1007/s10845-009-0323-5.

[22] Pejic, V. (2016). *Modeling and optimization in the ball end milling process* (original Serbian title*: Modelovanje i optimizacija procesa glodanja vretenastim glodalima*), Ph.D. Thesis, University of Novi Sad, Serbia.

[23] Myers, R.H., Montgomery, D.C., Anderson-Cook, C.M. (2009). *Response surface methodology: Process and product optimisation using designed experiments,* 3rd edition, John Wiley & Sons, New Jersey, USA.

[24] Sung, A.N., Loh, W.P., Ratnam, M.M. (2016). Simulation approach for surface roughness interval prediction in finish turning, *International Journal of Simulation Modelling*, Vol. 15, No. 1, 42-55, doi: 10.2507/IJSIMM15(1)4.320.

[25] Boyacı, A.İ., Hatipoğlu, T., Balcı, E. (2017). Drilling process optimization by using fuzzy-based multi-response surface methodology, *Advances in Production Engineering & Management*, Vol. 12, No. 2, 163-172, doi: 10.14743/apem2017.2.248.

[26] Zain, A.M., Haron, H., Sharif, S. (2010). Application of GA to optimize cutting conditions for minimizing surface roughness in end milling machining process, *Expert Systems with Applications*, Vol. 37, No. 6, 4650-4659, doi: 10.1016/j.eswa.2009.12.043.

[27] Mirjalili, S., Mirjalili, S.M., Lewis, A. (2014). Grey wolf optimizer, *Advances in Engineering Software*, Vol. 69, 46-61, doi: 10.1016/j.advengsoft.2013.12.007.

[28] Rezaei, H., Bozorg-Haddad, O., Chu, X. (2018). Grey wolf optimization (GWO) algorithm, In: Bozorg-Haddad, O. (ed.), *Advanced optimization by nature-inspired algorithms*, Springer Nature, Singapore, 81-92, doi: 10.1007/978-981-10-5221-7.

[29] Tamang, S.K., Chandrasekaran, M. (2015). Modeling and optimization of parameters for minimizing surface roughness and tool wear in turning Al/SiCp MMC, using conventional and soft computing techniques, *Advances in Production Engineering & Management*, Vol. 10, No. 2, 59-72, doi: 10.14743/apem2015.2.192.

Optimal production planning with capacity reservation and convex capacity costs

Huang, D.[a,b], Lin, Z.K.[a,b,*], Wei, W.[a,b]

[a]School of Economics and Management, Beijing Jiaotong University, Beijing, P.R. China
[b]Beijing Key Laboratory of Logistics Management and Technology, Beijing Jiaotong University, Beijing, P.R. China

A B S T R A C T

In this paper, we develop an analytical model for a multi-period production planning problem with dual supply sources of production capacity, where the supply price of one source (i.e., the spot market) is random and the supply capacity of the other source (i.e., the contract supplier) is limited. The purchasing cost of reserved capacity is assumed to be a convex function, rather than a linear function. We solve this problem by first characterizing the structure of the optimal production policy by employing a stochastic dynamic programming approach, and then determining the optimal capacity reservation level by applying a single-variable optimization method. For any given level of capacity reservation, the optimal periodic production policy is a quantity-dependent base stock policy with a threshold of production quantity that increases with the spot price. With this structure of the optimal production policy, the expected total discounted cost function is shown to be convex in the capacity reservation level. These results are also extended to the infinite-horizon case. A numerical study is conducted to examine the impacts of spot market characteristics on the optimal capacity reservation level and the corresponding optimal total cost.

1. Introduction

Supply chain disruptions caused by natural disasters or man-made interruptions have drawn significant attention in the past decade [1]. To manage the risk of disruptions and to mitigate their impacts on the whole supply chain, many companies have adopted dual sourcing (backup supply) as an easy-to-implement but quite effective operational mitigation strategy. In the presence of a spot trading market, the manufacturer inclines to build a long-term relationship with the primary supplier through supply contracts while utilizing spot purchase as a backup resource. A typical example is the Chinese steel-makers' procurement practice for iron ore [2]. They receive their supply of iron ore from long-term contracts with major iron miners in the world. Besides the long-term contracts, they also purchase iron ore cargoes from the spot market with a volatile trading price.

There are numerous forms of supply contracts, among which quantity-flexible or options contracts such as capacity reservation are more preferred in industrial procurement. In a capacity reservation contract, the manufacturer first pays a reservation fee to set a capacity reservation quantity. After some uncertainties of the spot market price and consumer demand are resolved, the manufacturer has the right to use any desired portion of the reserved capacity according to the contract agreement. The manufacturer then pays an additional capacity cost for only the

used capacity. This provides the manufacturer with flexibility in responding to changes in market demands and spot prices. In practice, such supply contracts are observed in apparel, power generation and high-tech industries such as semiconductors, electronics, and telecommunications equipment. For example, Apple's major chip suppliers were reported to reserve capacity for the second and third quarters of 2016 for iPhone 7 production [3].

There is a rich body of literature in operations research/management science that addresses various procurement problems with capacity reservation contracts. Early work mainly focused on deriving the optimal contract parameters to minimize the total cost of the supply chain system. Silver and Jain [4] and later Jain and Sliver [5] considered the case where the buyer pays a premium to the supplier for reserving a certain level of dedicated capacity in response to capacity supply uncertainty. They developed computational methods for the buyer to determine the capacity reservation level and the periodic replenishment quantity. Costa and Silver [6] extended the work of [4] to the case with discretely distributed demand and capacity supply. Brown and Lee [7] considered a pay-to-delay capacity reservation contract which is commonly seen in the semiconductor manufacturing. Peleg *et al.* [8] studied a two-period inventory model where the manufacturer (buyer) can procure from a long-term contract supplier with a fixed price and from the suppliers on the Internet with a random price. They discussed three alternative procurement strategies and identify the conditions under which each strategy is optimal. A comprehensive review of earlier literature on option contracts is provided by Kleindorfer and Wu [9].

Some research papers are devoted to the design of reservation contracts to achieve channel coordination. Barnes-Schuster *et al.* [10] studied channel coordination with option contracts in a two-period buyer-supplier model with correlated demand. They found that channel coordination can be achieved in case of piecewise linear option exercise prices. Moreover, they also showed that the backup agreement (Eppen and Iyer [11]), the quantity flexibility contract (Tsay and Lovejoy [12]), and the pay-to-delay capacity reservation contract are special cases of a general option contract they discussed. Erkoc and Wu [13] analysed a single-period supplier-buyer model with capacity reservation in the context of high-tech manufacturing. They proposed two reservation contracts under which channel coordination can be achieved and discussed the implications of them under different situations on contract compliance and demand uncertainty resolution. In a similar model setting, Jin and Wu [14] compared the deductible reservation contract with the take-or-pay contract in terms of channel coordination. They also extended their model to the multi-buyer case where one buyer with unsatisfied demand may use the unused capacity reserved by other buyers. In the presence of spot market trading, Pei *et al.* [15] studied the design of option contracts including exercise price and reservation price in a two-period seller-buyer model. They showed that the optimal contract can be either a sales contract or an options contract depending on the value of the optimal exercise price, while the optimal reservation price can be either fixed or volume-dependent (i.e., volume discounts or volume premia).

The research on capacity reservation in a multi-period setting arises in recent years. Serel *et al.* [16] first studied a multi-period buyer-supplier model with capacity reservation. They showed that the buyer's optimal sourcing decision is a base-stock type policy with two order-up-to levels corresponding to different supply strategies. Serel [17] extended the work of [16] to the case of considering supply uncertainty in the spot market. Inderfurth and Kelle [18] incorporated spot price uncertainty into the model discussed in [16], and provided a base stock policy for the dual sourcing problem which is suboptimal but easy to implement. For the same model, Inderfurth *et al.* [19] derived the optimal policy which contains three critical parameters: a capacity reservation level, a base stock level for sourcing from the long-term supplier, and a spot-price-dependent base stock level for sourcing from the spot market.

In addition to these above mentioned papers, many researchers have been working on the issues of production planning optimization and simulation [20-23], production coordination [24], cell production [25] and spot market price volatility [26-30]. In these related papers it is usually assumed that the reserved capacity is charged at a fixed unit capacity cost. However, as stated in [13], "capacity expansion demonstrates diseconomy of scale in this context (high-tech manufacturing), we (should) assume convex capacity costs". It is recognized that the reserved capacity price should not be linear for some industries such as high-tech industries, the apparel industry

and the power generation industry. The presence of nonlinear purchasing cost of reserved capacity creates new challenges for manufacturers in making their production planning decisions. Previous research work mainly focus on the production planning problem with capacity reservation in the context of a linear capacity cost (see, e.g., [16-19]), very limited literature has been devoted to the production planning problem with nonlinear capacity costs.

In this paper, we consider a manufacturer who procures capacity to produce finished products to satisfy random market demands. During the planning horizon, the manufacturer pays a unit premium per period to the contract supplier for reserving a certain level of production capacity. In each period, the manufacturer can either procure any amount up to the capacity reservation level from the contract supplier or procure from the spot market. The purchasing cost for the reserved capacity is increasing and convex, which reflects the diseconomy of scale when the contract supplier builds capacity. The spot market price follows a Markov process, whose realization is revealed to the manufacturer before the periodic procurement decisions are made. If the reserved capacity is not fully utilized in a period, the unused capacity cannot be sold back to the spot market and has no value to the manufacturer. The manufacturer's objective is to make optimal capacity reservation level and production decisions to minimize the expected total discounted cost over the planning horizon.

This problem can be modelled as a cost minimization problem in which the long-term capacity reservation level and the periodic production decisions are determined sequentially. For any given capacity reservation level, we formulate the problem as a stochastic dynamic program and characterize the structure of the optimal production policy. Our analysis reveals that the optimal periodic production policy is a quantity-dependent base stock policy, characterized by a threshold increasing with the spot price. That is, if the production quantity is lower than the threshold, the manufacturer produces exclusively from the reserved capacity; otherwise he produces from both the reserved capacity and the purchased capacity. For the former case the optimal produce-up-to level is increasing with the initial inventory level, whereas for the latter case the optimal produce-up-to level is decreasing with the spot price. With this structure of the production policy, the expected total discounted cost function is shown to be convex in the capacity reservation level. These results are also extended to the infinite-horizon case. Numerical experiments are conducted to examine the impacts of spot market characteristics on the optimal capacity reservation decisions.

The main contributions of this paper are threefold. First, in the present of convex capacity costs we find that the optimal order-up-to level for purchasing from the long-term supplier is an increasing function of the initial inventory level, rather than a fixed value in the linear capacity cost case. Second, we show that with the increase of spot price variance, the optimal capacity reservation level may decrease while the corresponding optimal total cost increases, which implies that the capacity reservation contract might not be an effective operational hedging tool against spot market price volatility. Third, the optimal capacity reservation level is shown to be decreasing with the initial inventory level at the beginning of the planning horizon, which demonstrates the substitution effect in mitigating supply chain risks between the carrying inventory strategy and the sourcing strategy [31].

The remainder of this paper is organized as follows. In Section 2, we introduce the model and its mathematical formulation in detail. In Section 3, we study the structural properties of the model and analyse the manufacturer's optimal capacity reservation and production strategies. A numerical study is conducted in Section 4. We conclude the paper in Section 5.

2. The model

We consider a manufacturer who procures production capacity from upstream suppliers and supplies finished products to end-consumers. The planning horizon is finite and indexed by $t = 1, \dots, T$. The random demands D_t for the finished products in each period are independently and identically distributed. Without loss of generality, we assume that the production lead time is negligible compared to the period length and one unit of capacity is required for producing

one unit of finished product. Thus, D_t can also be regarded as the manufacturer's demand for production capacity in each period.

In a capacity reservation contract, the long-term supplier allows the manufacturer to pay a unit premium c_r per period for keeping a certain level of reserved capacity K throughout the planning horizon. In each period, the manufacturer has the right, but not the obligation, to produce Q ($0 \leq Q \leq K$) units of products by the reserved capacity. The capacity cost charged by the long-term supplier, $R(Q)$, is a convex and increasing function of Q, where $R'(0) = 0$.

The manufacturer can also purchase capacity from the spot market – an outside market which is established by a group of homogeneous short-term suppliers. Let C_t be the spot price in period t and c_t be its realization. We assume that the spot prices $C_1,\dots, C_T$ follow a Markov process. That is, the distribution of c_{t+1} is only determined by c_t. In addition, D_t and C_t are assumed to be independent since the spot market size is sufficiently large.

The unit production cost c_m is identical for both the reserved capacity and the purchased capacity. At the end of each period, excess inventory of finished products is carried over to the next period and unsatisfied demand is backlogged. Denote $L_t(y)$ as the expected holding/backlogging cost in period t if the inventory level after production is y, where

$$L_t(y) = hE[(y - D_t)^+] + bE[(D_t - y)^+] \tag{1}$$

where h and b are respectively the unit inventory holding cost and unit demand backlogging cost in one period.

The sequence of events is as follows. At the beginning of the planning horizon, the manufacturer determines the long-term capacity reservation level K. At the beginning of each period t, the manufacturer first observes his on-hand inventory level x_t and the realized spot price c_t. Then he decides on how many to produce from each capacity. We denote by y_{1t} and y_{2t} the inventory level after production from the reserved capacity and the inventory level after production from all the capacity, respectively. Hence, $y_{1t} - x_t$ is the production quantity from the reserved capacity, and $y_{2t} - y_{1t}$ is production quantity from the capacity purchased from the spot market. Finally, demand is realized and satisfied by the available inventory. All costs are incurred and calculated at the end of the period. The objective of the manufacturer is to make optimal capacity reservation and production decisions to minimize the expected total discounted cost over the planning horizon.

Let $V_t(x_t, c_t, K)$ be the optimal expected total discounted cost from period t to T, given the initial inventory level x_t, the realized spot price c_t, and the capacity reservation level K. Then the manufacturer's production planning problem can be formulated as the following dynamic program: for $t = 1,\dots,T$,

$$V_t(x_t, c_t, K) = \min_{x_t + K \geq y_{1t} \geq x_t, y_{2t} \geq y_{1t}} \{c_r K + R(y_{1t} - x_t) - c_t y_{1t} - c_m x_t + G_t(y_{2t})\} \tag{2}$$

where

$$G_t(y_{2t}) = (c_t + c_m)y_{2t} + L_t(y_{2t}) + \alpha E_t[V_{t+1}(y_{2t} - D_t, C_{t+1}, K)] \tag{3}$$

The boundary condition is $V_{T+1}(x_{T+1}, c_{T+1}, K) = 0$. Here we use $E_t[\cdot]$ to denote $E_{D_t}\left[E_{C_{t+1}}[\cdot \mid C_t = c_t]\right]$, where the first expectation, E_{D_t}, is taken with respect to the random demand D_t, and the second expectation, $E_{C_{t+1}}[\cdot \mid C_t = c_t]$, is taken with respect to the random spot price in period $t + 1$, C_{t+1}, given the realized spot price in period t, c_t.

Our model includes the models in Henig $et\ al.$ [32] and Inderfurth $et\ al.$ [19] as special cases. Henig $et\ al.$ [32] studied a multi-period inventory problem embedded in a transportation contract assuming that the unit transportation cost for the delivery amount beyond the contract volume is larger. In the presence of a spot market, we assume that the spot price is random and thus may be either smaller or larger than the contract price. Inderfurth $et\ al.$ [19] studied the multi-period inventory problem with a linear reserved capacity cost. Here we generalize the model in Inderfurth $et\ al.$ [19] to the case with convex reserved capacity costs.

3. Optimal policies

In this section, we first characterize the structure of the optimal production policies for a given capacity reservation level, and then determine the optimal capacity reservation level.

For notational convenience, we define

$$H_t(y) = c_m y + L_t(y) + \alpha E_t[V_{t+1}(y - D_t, C_{t+1}, K)] \tag{4}$$

Then, the optimality equation (2) with respect to initial inventory level x can be rewritten as

$$V_t(x) = \min_{x+K \geq y_{1t} \geq x} \left\{ R(y_{1t} - x) - c_t y_{1t} + \min_{y_{2t} \geq y_{1t}} [c_t y_{2t} + H_t(y_{2t})] \right\} + c_r K - c_m x \tag{5}$$

The convexity of the cost-to-go function is established in Lemma 1.

Lemma 1: *In period $t = 1, \ldots, T$, for any given K and c_t, $V_t(x)$ is convex in x.*

Proof: We prove by induction that, when $V_{t+1}(x)$ is convex in x, then $V_t(x)$ is convex. Since $V_{T+1}(x) = 0$ is clearly convex, we assume that the convexity of $V_{t+1}(x)$ holds for period $t + 1$ where $t \leq T$. Then it is easy to verify that $H_t(y)$ is convex in y. Thus, $c_t y_{2t} + H_t(y_{2t})$ is a convex function of y_{2t}. Note that the feasible region $y_{2t} \geq y_{1t}$ is a convex set of (y_{1t}, y_{2t}). Then it follows from Proposition B-4 in Heyman and Sobel [33] that, $\min_{y_{2t} \geq y_{1t}} [c_t y_{2t} + H_t(y_{2t})]$ is convex in y_{1t}. Given c_t, $R(y_{1t} - x) - c_t y_{1t} + \min_{y_{2t} \geq y_{1t}} [c_t y_{2t} + H_t(y_{2t})]$ is jointly convex in x and y_{1t}. For a given K, the feasible region $x + K \geq y_{1t} \geq x$ is a convex set of (x, y_{1t}). Then by [33] again we obtain that $V_t(x)$ is convex in x. ∎

Define

$$S_H = \arg \min_y H_t(y) \tag{6}$$

$$S_R(x) = \arg \min_y R(y - x) + H_t(y) \tag{7}$$

$$S_F(c_t) = \arg \min_y c_t y + H_t(y) \tag{8}$$

$$S_K = \{y | R'(K) + H_t'(y) = 0\} \tag{9}$$

$$M_t = R'^{-1}(c_t) \tag{10}$$

and

$$c_K = R'(K) \tag{11}$$

where S_H is the minimizer of $H_t(y)$, $S_R(x)$ is the minimizer of $R(y - x) + H_t(y)$ for any given x, $S_F(c_t)$ is the minimizer of $c_t y + H_t(y)$ for any spot price realization c_t, S_K is the minimizer of $c_K y + H_t(y)$ for any given K, M_t is a threshold capacity level for any c_t, and c_K is a threshold price for any K. Some order relations and monotonic properties of these critical points are given in Lemma 2.

Lemma 2: *For any period $t = 1, \ldots, T$, we have the following results.*

(a) $S_K \leq S_F(c_t) < S_H$ if $c_t \leq c_K$; otherwise $S_F(c_t) < S_K < S_H$
(b) $S_R(S_H) = S_H$, $S_R(S_F(c_t) - M_t) = S_F(c_t)$, $S_F(0) = S_H$, $S_F(c_K) = S_K$
(c) $S_R(x)$ is increasing in x
(d) $S_R(x) - x$ is decreasing in x
(e) $S_F(c_t)$ is decreasing in c_t
(f) S_K is decreasing in K
(g) M_t is increasing in c_t
(h) If $c_t \leq c_K$, then $K \geq M_t$; otherwise $K < M_t$

Proof: Please see the Appendix A. ∎

Based on these critical points, the optimal production policy is characterized as follows.

Theorem 1: *In period* t $(t = 1, \ldots, T)$, *for a given state* (x, c_t, K) *the optimal production policy,* (y_1^*, y_2^*), *is:*

$$\text{if } c_t \leq c_K, (y_1^*, y_2^*) = \begin{cases} (x + M_t, S_F(c_t)), & \text{if } x < S_F(c_t) - M_t \\ (S_R(x), S_R(x)), & \text{if } S_F(c_t) - M_t \leq x < S_H \\ (x, x), & \text{if } S_H \leq x \end{cases}$$

$$\text{otherwise, } (y_1^*, y_2^*) = \begin{cases} (x + K, S_F(c_t)), & \text{if } x < S_F(c_t) - K \\ (x + K, x + K), & \text{if } S_F(c_t) - K \leq x < S_K - K \\ (S_R(x), S_R(x)), & \text{if } S_K - K \leq x < S_H \\ (x, x), & \text{if } S_H \leq x \end{cases}$$

Proof: It follows from the convexity of $R(x)$ that, the marginal reserved capacity cost $R'(x) \leq c_t$ if $x \leq M_t$; and $R'(x) > c_t$ otherwise. Thus, when the production quantity is smaller than M_t, it is optimal to produce exclusively from the reserved capacity (the exclusive case); otherwise it is optimal to produce from both types of capacity in which only M_t reserved capacity is used (the non-exclusive case).

When $c_t \leq c_K$, the capacity reservation level K is larger than M_t. For the exclusive case, from the convexity of $R(y - x) + H_t(y)$, the optimal production policy is to make the inventory level after production from the reserved capacity, y_{1t}, as close to $S_R(x)$ as possible. If $S_F(c_t) - M_t \leq x < S_H$, it is optimal to set $y_{2t} = y_{1t} = S_R(x)$ since the optimal production quantity satisfies that $0 < S_R(x) - x \leq M_t$. Next, if $x \geq S_H$, it is optimal to set $y_{2t} = y_{1t} = x$ since for any $y \geq x \geq S_H$, $R(y - x) + H_t(y)$ is monotonically increasing in y. For the non-exclusive case, the optimal production policy is to make the inventory level after production from both capacity, y_{2t}, as close to $S_F(c_t)$ as possible. Then, if $x < S_F(c_t) - M_t$, it is optimal to set $y_{1t} = x + M_t$ and $y_{2t} = S_F(c_t)$.

When $c_t > c_K$, the reserved capacity level K is smaller than M_t. For the non-exclusive case, the optimal policy is modified by replacing M_t with K since only K reserved capacity can be used. For the exclusive case, following a similar proof to that of Theorem 1(a), it can be shown that it is optimal to set $y_{2t} = y_{1t} = S_R(x)$ if $S_K - K \leq x < S_H$ and $y_{2t} = y_{1t} = x$ if $x \geq S_H$. Finally, if $S_F(c_t) - K \leq x < S_K - K$, it is optimal to set $y_{1t} = x + K$ to make y_{1t} as close to $S_R(x)$ as possible, and set $y_{2t} = y_{1t}$ since the optimal inventory level after production from the reserved capacity, $x + K$, is larger than $S_F(c_t)$. ∎

Fig. 1 illustrates the optimal policy in Theorem 1. For a given capacity reservation level K, the state space (x, c_t) is divided into different decision regions. If the state (x, c_t) falls into region I, it is optimal to raise the inventory level to $S_F(c_t)$ by producing an amount of $\min(M_t, K)$ from the reserved capacity and the additional amount over $\min(M_t, K)$ from the capacity purchased in the spot market. For (x, c_t) in region II, it is optimal to produce up to $S_R(x)$ exclusively from the reserved capacity. However, in region II′, the desired inventory level $S_R(x)$ cannot be reached and it is optimal to exploit all the reserved capacity to raise the inventory level as close to $S_R(x)$ as possible. Finally, for (x, c_t) in region III, it is optimal for the manufacturer to do nothing.
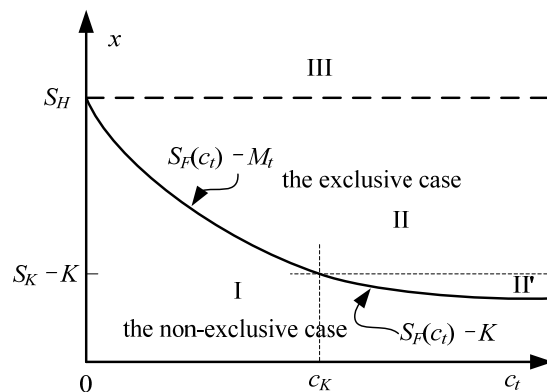

Fig. 1 The optimal policy for different x and c_t

The optimal production policy in Theorem 1 can be interpreted as a quantity-dependent base stock policy with dual production capacity supply sources. If the production quantity is lower than a threshold level, the manufacturer produces exclusively from the reserved capacity; otherwise he produces an amount equal to the threshold level from the reserved capacity and the additional amount over the threshold level from the capacity purchased from the spot market. The threshold level of production quantity, $\min(M_t, K)$, is non-decreasing in the spot price c_t. When the capacity reservation level is sufficient, the optimal produce-up-to level for the exclusive case is $S_R(x)$; otherwise, this desired inventory level cannot be reached and the optimal produce-up-to level is $x + K$. For the non-exclusive case the optimal produce-up-to level is $S_F(c_t)$, which is decreasing in c_t.

Corollary 1: *The threshold of x that triggers spot purchase is decreasing in c_t.*

Proof: It is clearly seen from Theorem 1 that spot purchase is used only when the initial inventory level x is low (i.e., $x < S_F(c_t) - \min(M_t, K)$). Since $S_F(c_t)$ is decreasing in c_t and $\min(M_t, K)$, is non-decreasing in c_t, the threshold that triggers spot purchase, $S_F(c_t) - \min(M_t, K)$, is decreasing in c_t. ∎

Corollary 1 implies that with a higher spot market price, spot purchase will be used by the manufacturer at a lower initial inventory level.

It is shown in Theorem 1 that if the capacity reservation level K is insufficient, the manufacturer cannot extract the benefit of lower capacity cost from capacity reservation. However, if K is too large, some amount of the reserved capacity will never be used. In Lemma 3, given the initial inventory level at the beginning of the planning horizon, x_1, we characterize the convexity of the expected total discounted cost $V_1(x_1, K)$ in Eq. 2 with respect to the capacity reservation level.

Lemma 3: *Given the initial inventory level x_1, $V_1(x_1, K)$ is convex in K.*

Proof: The convexity of $V_1(x_1, K)$ can be shown by induction. The proof is similar with that of Lemma 1. ∎

Lemma 3 guarantees that for a given x_1 the optimal capacity reservation level can be obtained by minimizing the expected total discounted cost $V_1(x_1, K)$ with respect to K.

Theorem 2: *When demands are stationary, all cost parameters are time homogeneous, and the discount factor $\alpha \in (0,1)$, the optimal policy for the infinite-horizon case (i.e., $T \to +\infty$) takes the same structure as that of the finite-horizon case.*

Proof: We prove this theorem by verifying that conditions (a)-(d) and (f) in Theorem 8-15 in Heyman and Sobel [33] is hold here. Consider a (non-optimal) production policy which selects $y_1 = x$, $y_2 = S_F(c_t)$ for all t, and let $v^\lambda(x)$ denote its total expected (infinite horizon) discounted cost when initial inventory level is x. Since the discount factor $\alpha \in (0,1)$, it is easy to prove that $v^\lambda(x)$ is bounded. From Theorem 8-13 in Heyman and Sobel [33], condition (a) holds. Since all costs are non-negative, the single-period cost function is increasing and condition (b) holds. Conditions (c) and (d) are immediate. Furthermore, for a given x, the total inventory level after production, y_2, is bounded by $\max(S_H, x)$ and thus Condition (f) is valid. □

Theorem 2 implies that for a given capacity reservation level the optimal policy for the infinite-horizon problem inherits the structure of the T-period problem. Moreover, the convexity of the expected total discounted cost function in Lemma 3 is still hold in the infinite-horizon case as the limit of the convex function $V_1(x_1, K)$ is convex as well.

4. A numerical study: Results and discussion

In this section, we conduct numerical experiments to examine the impacts of spot market characteristics on the optimal capacity reservation level and the corresponding optimal total cost. We consider a five-period example with the following cost parameters: $c_m = 10$, $c_r = 5$, $h = 8$, $b = 50$ and $\alpha = 0.95$. The reserved capacity cost is assumed to be $R(x) = 0.2x^2$. We assume that demand in each period follows a discrete uniform distribution on $[A, B]$, where $A = 1$ and

$B = 20$. We use a simple Markov chain with three possible states: $(c_1, c_2, c_3) = (12 - \Delta, 12, 12 + \Delta)$ to model the spot price process, where a larger value of Δ implies a higher level of spot price variance. The one-step transition probabilities $p_{ij} = pr(c_{t+1} = c_j | c_t = c_i)$ for $i, j = 1,2,3$ are given in the following matrix:

$$P = \begin{bmatrix} 0.80 & 0.15 & 0.05 \\ 0.15 & 0.70 & 0.15 \\ 0.15 & 0.25 & 0.60 \end{bmatrix}$$

Figs 2 and 3 illustrate the optimal capacity reservation level and the corresponding optimal total cost with respect to the initial inventory level at different level of Δ. It is shown that the optimal capacity reservation level is decreasing in the initial inventory level. The explanation is that with a larger initial inventory level, a fewer production quantity is needed and the manufacturer would like to keep a lower capacity reservation level. The optimal total cost is observed to be convex in the initial inventory level, which is in line with the preceding analytical results.
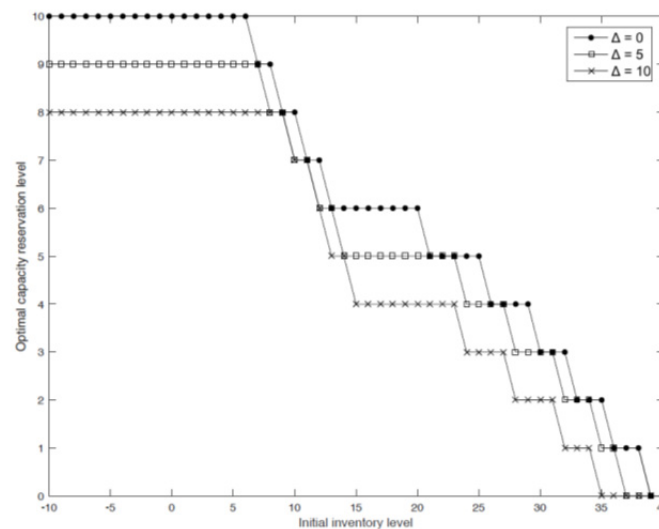

Fig. 2 Optimal capacity reservation level vs. initial inventory level at different level of Δ
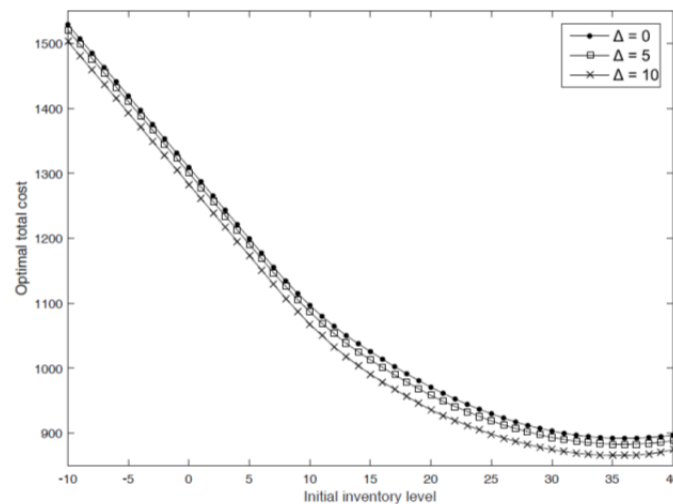
Fig. 3 Optimal total cost vs. initial inventory level at different level of Δ

Another observation is that both the optimal capacity reservation level and the corresponding optimal total cost are decreasing in Δ. This is because that as Δ increases, the likelihood that the manufacturer can purchase capacity at a lower price also increases, which directly benefits the manufacturer and makes the capacity reservation contracts less valuable. This observation implies that the more volatile the spot market price is, the less production capacity the manufacturer would like to reserve in advance.

Next, we examine the impacts of the expected spot price. We set the possible spot price realizations as $(c_1, c_2, c_3) = (S - 2, S, S + 2)$. Here, a larger value of S implies a higher level of expected spot price. Figs 4 and 5 show the optimal capacity reservation level and the corresponding optimal total cost with respect to the initial inventory level at different level of S. We find that both the optimal capacity reservation level and the corresponding optimal total cost are increasing in S. This is in line with the intuition that as the expected spot price increases the manufacturer becomes more dependent on supply contracts with the long-term supplier. However, when the expected spot price is at a lower level, the optimal capacity reservation level may drop to 0 more quickly as the initial inventory level increases. Under such spot market conditions, the manufacturer will refuse to sign the long-term reservation contract with the supplier and fully depends on the spot market. This is what the Chinese steel makers did in their iron ore procurement in 2014!

Fig. 4 Optimal capacity reservation level vs. initial inventory level at different level of S

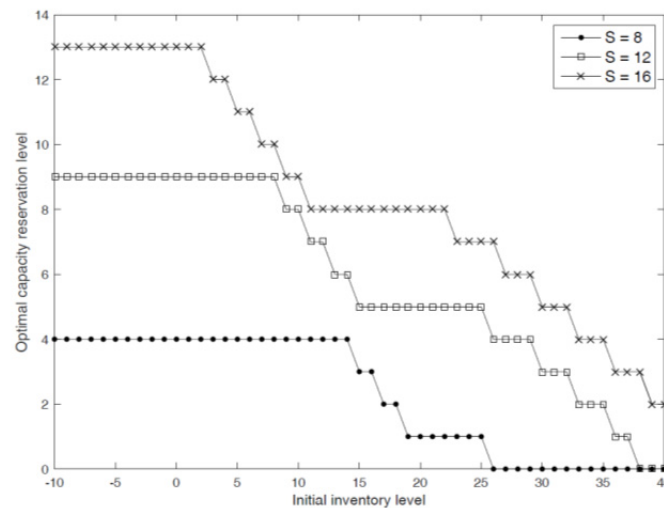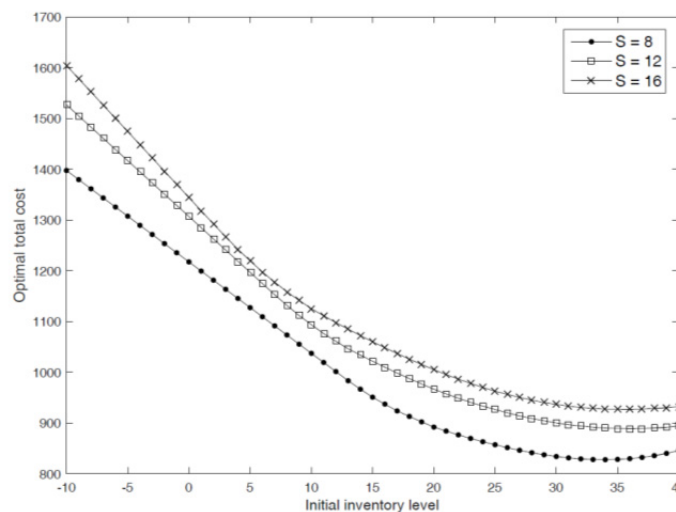**Fig. 5** Optimal total cost vs. initial inventory level at different level of S

For the final analysis, we investigate the impacts of spot price distribution. Denote the steady-state distribution of the spot prices as $p_i = pr(c_t = c_i)$ for $i = 1, 2, 3$. We set $(p_1, p_2, p_3) = (\beta, 0, 1 - \beta)$, where the value of $\beta \in [0, 1]$ affects the spot price distribution. It can be shown that the expected spot price, $12 + (1 - 2\beta)\Delta$, is decreasing in β and the spot price variance, $4\beta(1 - \beta)\Delta^2$, is convex in β.

The observation that both the optimal capacity reservation level and the corresponding optimal total cost are decreasing in Δ is still valid for the case $\beta > 0.5$. The interpretation is as follows. In this case, a larger Δ will lead to a lower expected spot price and a higher level of spot price variance, both of which result in a decreasing trend in the optimal capacity reservation level and the corresponding optimal total cost.

However, for the case $\beta < 0.5$, these observations on the impacts of Δ may be quite different. For example, as shown in Fig. 6, the optimal capacity reservation level is increasing with Δ when $\beta = 0.2$. The explanation is that when the spot price is more possible to take the high realization, a larger price difference Δ will increase the manufacturer's risk of purchasing at a higher spot price and thus makes the capacity reservation contract more favorable.
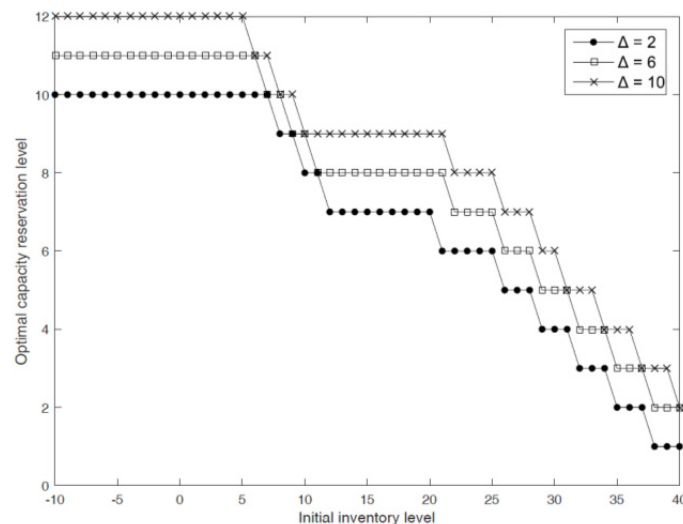
Fig. 6 The optimal capacity reservation level vs. initial inventory level at different level of Δ for the case $\beta = 0.2$

5. Conclusion

In this paper, we have studied how a manufacturer jointly uses a long-term supplier and the spot market to procure production capacity and produce finished products to satisfy random market demands. A multi-period production planning problem with a convex purchasing cost of reserved capacity is discussed. The manufacturer's optimal capacity reservation and production decisions that minimize the expected total discounted cost over the planning horizon are provided.

For a given capacity reservation level, the optimal production policy is a quantity-dependent base stock policy, characterized by a threshold of production quantity that increases with the spot price. That is, if the production quantity is lower than the threshold, the manufacturer produces exclusively from the reserved capacity; otherwise he produces an amount equal to the threshold level from the reserved capacity and the additional amount over the threshold level from the capacity purchased from the spot market. For the former case the optimal produce-up-to level is increasing with the initial inventory level, whereas for the latter case the optimal produce-up-to level is decreasing with the spot price. Compared with the linear reserved capacity purchasing cost case, we find that the optimal order-up-to level for purchasing from the long-term supplier is an increasing function of the initial inventory level, rather than a fixed value.

With this structure of periodic production policy, the expected total discounted cost function is shown to be convex in the capacity reservation level. These results are also extended to the

infinite-horizon case. We have also examined the impacts of initial inventory level and spot characteristics on the capacity reservation decisions, and found that: (1) when the spot price is more possible to take the lower (higher) realization, the capacity reservation level may decrease (increase) with the price difference between different spot price realizations; (2) the capacity reservation level increases with the expected spot price; and (3) the capacity reservation level decreases with the initial inventory level at the beginning of the planning horizon.

For future research, there are several important and possible extensions to this study. First, we assume that the demand and spot price distributions are independent as most related models did in the literature. However, in some cases, the spot price of production capacity may influence the selling price of the finished product, which in turn affects the demand for the finished product as well as the demand for the production capacity. Hence, developing production planning models with considering the correlation between demand and spot price is an interesting direction for future research. Second, we assume that the spot market is reliable with an infinite supply capacity in this paper. Considering various supply conditions in the spot market such as supply capacity limits, supply uncertainty and capacity-dependent spot price will be another interesting direction. Finally, incorporating endogenous sales price of the finished products into our model and investigating the periodic pricing decisions in the production planning problem is also an interesting direction for future research.

Acknowledgement

References

[1] Snyder, L.V., Atan, Z., Peng, P., Rong, Y., Schmitt, A.J., Sinsoysal, B. (2016). OR/MS models for supply chain disruptions: A review, *IIE Transactions*, Vol. 48, No. 2, 89-109, doi: 10.1080/0740817X.2015.1067735.

[2] Chen, Y., Xue, W., Yang, J. (2013). Technical note – Optimal inventory policy in the presence of a long-term supplier and a spot market, *Operations Research*, Vol. 61, No. 1, 88-97, doi: 10.1287/opre.1120.1114.

[3] Appleinsider. Apple chip suppliers gearing up capacity for 'iPhone 7' production – report, from *https://forums. appleinsider.com/discussion/191903*, accessed June 17, 2017.

[4] Silver, E.A., Jain, K. (1994). Some ideas regarding reserving supplier capacity and selecting replenishment quantities in a project context, *International Journal of Production Economics*, Vol. 35, No. 1-3, 177-182, doi: 10.1016/0925-5273(94)90079-5.

[5] Jain, K., Silver, E.A. (1995). The single period procurement problem where dedicated supplier capacity can be reserved, *Naval Research Logistics*, Vol. 42, No. 6, 915-934, doi: 10.1002/1520-6750(199509)42:6<915::AID-NAV3220420605>3.0.CO;2-M.

[6] Costa, D., Silver, E.A. (1996). Exact and approximate algorithms for the multi-period procurement problem where dedicated supplier capacity can be reserved, *OR Spectrum*, Vol. 18, No. 4, 197-207, doi: 10.1007/BF01540156.

[7] Brown, A.O., Lee, H.L. (1997). Optimal *pay-to-delay capacity reservation with application to the semiconductor industry,* Working paper, Department of Industrial Engineering and Engineering Management, Stanford University, USA.

[8] Peleg, B., Lee, H.L., Hausman, W.H. (2002). Short-term e-procurement strategies versus long-term contracts, *Production and Operations Management*, Vol. 11, No. 4, 458-479, doi: 10.1111/j.1937-5956.2002.tb00472.x.

[9] Kleindorfer, P.R., Wu, J.D. (2003). Integrating long- and short-term contracting via business-to-business exchanges for capital-intensive industries, *Management Science*, Vol. 49, No. 11, 1597-1615, doi: 10.1287/mnsc.49.11.1597.20583.

[10] Barnes-Schuster, D., Bassok, Y., Anupindi, R. (2002). Coordination and flexibility in supply contracts with options, *Manufacturing & Service Operations Management*, Vol. 4, No. 3, 171-207, doi: 10.1287/msom.4.3.171.7754.

[11] Eppen, G.D., Iyer, A.V. (1997). Backup agreements in fashion buying – The value of upstream flexibility, *Management Science*, Vol. 43, No. 11, 1469-1484, doi: 10.1287/mnsc.43.11.1469.

[12] Tsay, A.A., Lovejoy, W.S. (1999). Quantity flexibility contracts and supply chain performance, *Manufacturing & Service Operations Management*, Vol. 1, No. 2, 89-111, doi: 10.1287/msom.1.2.89.

[13] Erkoc, M., Wu, S.D. (2005). Managing high-tech capacity expansion via reservation contracts, *Production and Operations Management*, Vol. 14, No. 2, 232-251, doi: 10.1111/j.1937-5956.2005.tb00021.x.

[14] Jin, M., Wu, S.D. (2007). Capacity reservation contracts for high-tech industry, *European Journal of Operational Research*, Vol. 176, No. 3, 1659-1677, doi: 10.1016/j.ejor.2005.11.008.

[15] Pei, P.P.-E., Simchi-Levi, D., Tunca, T.I. (2011). Sourcing flexibility, spot trading, and procurement contract structure, *Operations Research*, Vol. 59, No. 3, 578-601, doi: 10.1287/opre.1100.0905.

[16] Serel, D.A., Dada, M., Moskowitz, H. (2001). Sourcing decisions with capacity reservation contracts, *European Journal of Operational Research*, Vol. 131, No. 3, 635-648, doi: 10.1016/S0377-2217(00)00106-5.

[17] Serel, D.A. (2007). Capacity reservation under supply uncertainty, *Computers & Operations Research*, Vol. 34, No. 4, 1192-1220, doi: 10.1016/j.cor.2005.06.018.

[18] Inderfurth, K., Kelle, P., Kleber, R. (2013). Dual sourcing using capacity reservation and spot market: Optimal procurement policy and heuristic parameter determination, *European Journal of Operational Research*, Vol. 225, No. 2, 298-309, doi: 10.1016/j.ejor.2012.08.025.

[19] Inderfurth, K., Kelle, P. (2011). Capacity reservation under spot market price uncertainty, *International Journal of Production Economics*, Vol. 133, No. 1, 272-279, doi: 10.1016/j.ijpe.2010.04.022.

[20] Wang, C., Liu, X.-B., Zhao, G.-Z., Chin, K.O. (2014). Multi-objective integrated production planning model and simulation constrained doubly by resources and materials, *International Journal of Simulation Modelling*, Vol. 13, No. 2, 243-254, doi: 10.2507/IJSIMM13(2)CO10.

[21] Chen, Y.X. (2016). Integrated optimization model for production planning and scheduling with logistics constraints, *International Journal of Simulation Modelling*, Vol. 15, No. 4, 711-720, doi: 10.2507/IJSIMM15(4)CO16.

[22] Gordić, B. (2017). Flexible optimization in the process of planning and production control, *Tehnički Vjesnik – Technical Gazette*, Vol. 24, No. 4, 1087-1094, doi: 10.17559/TV-20160112132024.

[23] Tang, M., Gong, D., Liu, S., Lu, X. (2017). Finding key factors affecting the locations of electric vehicle charging stations: A simulation and ANOVA approach, *International Journal of Simulation Modelling*, Vol. 16, No. 3, 541-554, doi: 10.2507/IJSIMM16(3)CO15.

[24] Tang, M., Qi, Y., Zhang, M. (2017). Impact of product modularity on mass customization capability: An exploratory study of contextual factors, *International Journal of Information Technology & Decision Making*, Vol. 16, No. 4, 939-959, doi: 10.1142/S0219622017410012.

[25] Gong, D., Tang, M., Liu, S., Li, Q. (2017). Reconsidering production coordination: A principal-agent theory-based analysis, *Advances in Production Engineering & Management*, Vol. 12, No. 1, 51-61, doi: 10.14743/apem2017.1.239.

[26] Wen, F., He, Z., Dai, Z., Yang, X. (2014). Characteristics of investors' risk preference for stock markets, *Economic Computation and Economic Cybernetics Studies and Research*, Vol. 48, No. 3, 235-254.

[27] Wen, F., Gong, X., Cai, S. (2016). Forecasting the volatility of crude oil futures using HAR-type models with structural breaks, *Energy Economics*, Vol. 59, 400-413, doi: 10.1016/j.eneco.2016.07.014.

[28] Wen, F., Xiao, J., Huang, C., Xia, X. (2018). Interaction between oil and US dollar exchange rate: Nonlinear causality, time-varying influence and structural breaks in volatility, *Applied Economics*, Vol. 50, No. 3, 319-334, doi: 10.1080/00036846.2017.1321838.

[29] Zhou, M., Liu, X., Pan, B., Yang, X., Wen, F., Xia, X. (2017). Effect of tourism building investments on tourist revenues in China: A spatial panel econometric analysis, *Emerging Markets Finance and Trade*, Vol. 53, No. 9, 1973-1987, doi: 10.1080/1540496X.2016.1237353.

[30] Hu, C., Liu, X., Pan, B., Chen, B., Xia, X. (2017). Asymmetric impact of oil price shock on stock market in China: A combination analysis based on SVAR model and NARDL model, *Emerging Markets Finance and Trade*, Vol. 53, No. 9, 1973-1987, doi: 10.1080/1540496X.2017.1412303.

[31] Tomlin, B. (2006). On the value of mitigation and contingency strategies for managing supply chain disruption risks, *Management Science*, Vol. 52, No. 5, 639-657, doi: 10.1287/mnsc.1060.0515.

[32] Henig, M., Gerchak, Y., Ernst, R., Pyke, D.F. (1997). An inventory model embedded in designing a supply contract, *Management Science*, Vol. 43, No. 2, 184-189, doi: 10.1287/mnsc.43.2.184.

[33] Heyman, D.P., Sobel, M.J. (1984). *Stochastic models in operations research* (*Vol. II*), McGraw-Hill, New York, USA.

Appendix A

Proof of Lemma 2: By Lemma 1, it is straightforward to verify that $H_t(y)$, $R(y - x) + H_t(y)$ and $c_t y + H_t(y)$ are both convex in y. Then S_H, $S_R(x)$, $S_F(c_t)$ and S_K are given by the first order condition, where

$$H_t'(S_H) = 0 \tag{A.1}$$

$$R'(S_R(x) - x) + H_t'(S_R(x)) = 0 \tag{A.2}$$

$$H_t'(S_F(c_t)) + c_t = 0 \tag{A.3}$$

and

$$R'(K) + H_t'(S_K) = 0 \tag{A.4}$$

It follows from Eq. A.1, Eq. A.3 and Eq. A.4 that

$$H_t'(S_K) = -R'(K) \le H_t'\big(S_F(c_t)\big) = -c_t < H_t'(S_H) = 0, \text{ when } c_t \le c_K, \tag{A.5}$$

and

$$H_t'\big(S_F(c_t)\big) = -c_t < H_t'(S_K) = -R'(K) < H_t'(S_H) = 0, \text{ when } c_t > c_K. \tag{A.6}$$

Then, the result of Lemma 2(a) is obtained by the fact that $H_t'(y)$ is increasing in y.

From Eq. A.2, it is easy to show that:

- if $S_R(x) = S_H$, then x is given by $R'(S_H - x) + H_t'(S_H) = R'(S_H - x) = 0$;
- if $S_R(x) = S_F(c_t)$, then x is given by $R'(S_F(c_t) - x) + H_t'(S_F(c_t)) = R'(S_F(c_t) - x) - c_t = 0$.

From Eq. A.3, it is easy to show that

- if $c_t = 0$, then $S_F(0)$ is given by $H_t'\big(S_F(0)\big) = 0$.
- if $c_t = c_K$, then $S_F(c_K)$ is given by $H_t'\big(S_F(c_K)\big) = -c_K = -R'(K)$.

Then, the results of Lemma 2(b) are obtained by the convexity of $R(\cdot)$ and the definition of S_H, S_K and M_t.

For Lemma 2(c) and 2(d), taking the first derivative of $S_R(x)$ with respect to x, we have

$$\frac{dS_R(x)}{dx} = \frac{R''(S_R(x) - x)}{R''(S_R(x) - x) + H_t''(S_R(x))} > 0 \tag{A.7}$$

and

$$\frac{dS_R(x) - x}{dx} = \frac{dS_R(x)}{dx} - 1 \tag{A.8}$$

For Lemma 2(e), taking the first derivative of $S_F(c_t)$ with respect to c_t, we have

$$\frac{dS_F(c_t)}{dc_t} = -\frac{1}{H_t''\big(S_F(c_t)\big)} < 0 \tag{A.9}$$

For Lemma 2(f), taking the first derivative of S_K with respect to K, we have

$$\frac{dS_K}{dK} = -\frac{R''(K)}{H_t''(S_K)} < 0 \tag{A.10}$$

For Lemma 2(g), by the definition of M_t, we have

$$\frac{dM_t}{dc_t} = \frac{1}{R''(M_t)} > 0 \tag{A.11}$$

It follows from the convexity of $R(\cdot)$ and the definition of M_t and c_K that,

- if $c_t = R'(M_t) \le c_K = R'(K)$, then $K \ge M_t$;
- if $c_t = R'(M_t) > c_K = R'(K)$, then $K < M_t$.

This completes the proof of Lemma 2.

Solving fuzzy flexible job shop scheduling problem based on fuzzy satisfaction rate and differential evolution

Ma, D.Y.[a,b], **He, C.H.**[c], **Wang, S.Q.**[d], **Han, X.M.**[b], **Shi, X.H.**[d,e,*]

[a]Business School, Jilin University, Changchun, P.R. China
[b]School of Computer Science and Engineering, Changchun University of Technology, Changchun, P.R. China
[c]School of Electronics Engineering and Computer Science, Peking University, Beijing, P.R. China
[d]Key Laboratory of Symbol Computation and Knowledge Engineering, Ministry of Education, Changchun, P.R. China
[e]College of Computer Science and Technology, Jilin University, Changchun, P.R. China

A B S T R A C T

Focused on a variety of JSSP considered flexibility and fuzziness, namely the fuzzy flexibility JSSP (FfJSSP), a novel method based on fuzzy satisfaction rate and differential evolution (DE) algorithm is proposed in this paper. In the method, the fuzzy membership functions' parameters are determined according to normal distribution for maximum satisfaction rate calculation. Then a DE algorithm is proposed by well designing the coding for the problem and extending the related operators on the coding. A local exploring search for operation and machine parts of the coding is also introduced to improve the performance of the method. Experimental results show that our proposed method is effective compared with other five popular existed methods. Comparisons between different mutation and crossover strategies are also performed. Numerical results show that the proposed method could be applied to real FfJSSP problems.

1. Introduction

Job shop scheduling problem (JSSP) is one of the best known combinatorial optimization problems, and has received tremendous attention in different fields. Lots of real problems could be modelled into JSSP, such as cargo port scheduling [1], health care activity scheduling [2], electric vehicle charging station arrangement [3], and so on. In conventional JSSP, each operation could access to only one machine. But in flexible JSSP (fJSSP), it could access to multiple machines, which makes fJSSP more complex than JSSP. In recent years, fJSSP is attracting more and more attentions in the related fields. The existing methods for fJSSP could be divided into two categories: evolution algorithm and non-evolution algorithm. For non-evolution algorithms, Wu and Weng proposed a multi-agent scheduling method with earliness and tardiness objectives in flexible job shops [4]. Sahin *et al.* also developed a multi-agent based system to simultaneous schedule flexible machine groups and material handling system working under a manufacturing dynamic environment [5]. Zhang *et al.* developed a dynamic game theory based two-layer scheduling method to reduce makespan, the total workload of machines and energy consumption to achieve real-time multi-objective flexible job shop scheduling [6]. While the evolution algorithm based methods were a lot more than non-evolution algorithms. For example, Li and Pan presented a novel discrete artificial bee colony algorithm for solving the multi-objective flexible job

shop scheduling problem with maintenance activities in 2014 [7]. Gao proposed a hybrid genetic and variable neighborhood descent algorithm [8]. Li and Gao [9] and Zhang *et al.* [10] hybridized the genetic algorithm (GA) and tabu search (TS) and both developed a GA and TS based algorithm genetic algorithm for flexible job shop. Focused on sequence flexibility, Huang *et al.* proposed an improved GA method to minimize the makespan [11]. By defining the objective function as the weighted combination of the maximum of the completion time (makespan) and the mean of earliness and tardiness, Gao *et al.* proposed a discrete harmony search (DHS) algorithm to solve FJSP [12]. Li and Pan adopted an effective discrete chemical-reaction optimization to solve the multi-objective fJSSP [13], and Ozgüven *et al.* proposed a concept of removing invalid nodes before modeling, and considering a new problem with sequence dependent setup times [14]. Latest developments is that Yuan and Xu proposed a hybrid differential evolution (HDE) algorithms to minimize the makespan [15], and Xu *et al.* proposed an effective teaching-learning based optimization algorithm to solve the fJSSP with fuzzy processing time [16]. Focused on the machine part manufacturing, Supsomboon and Vajasuvimon presented a simulation model which is beneficial to assist in performance improving [17].

Fuzzy job shop scheduling problem (FJSSP) is also an extension of conventional JSSP in which the uncertain parameters of practical manufacturing are considered. For the fuzzy description of problem, many researchers represented fuzzy processing time by fuzzy triples, for example, Mehrabad *et al.* described the fuzzy delivery time with fuzzy number [18], Lei [19] and Kilic [20] constructed the objective function according to the intersected area of membership function and customer's requirements. Usually, FJSSP is solved by fuzzy theory combined with computational intelligence methods rather than traditional methods. For instance, Lei developed an achieve Pareto-optimality with particle swarm algorithm [19], Niu *et al.* proposed a fuzzy particle swarm algorithm [21], Kilic adopted the artificial ant colony algorithm to solve this problem [20], Zhang *et al.* put forward a generalized IFS to process the additive operation and to compare the operation between two IFSs [22], and Gao *et al.* proposed a discrete harmony search (DHS) algorithm for FJSSP [23]. By using multiobjective evolutionary algorithm combined with a novel dominance-based tabu search method to optimize both the expected makespan and the predictive robustness of the fuzzy schedule, José Palacios *et al.* developed a robust multiobjective optimization algorithm for fuzzy job shop scheduling problem [24].

Fuzzy flexible job shop scheduling problem (FfJSSP) is the combination of FJSSP and fJSSP. Comparing fJSSP and FJSSP, the studies of FfJSSP are quite few. Kacem *et al.* proposed a pareto-optimality approach for FfJSSP with hybridization of evolutionary algorithms and fuzzy logic [25]. Zheng *et al.* solved the multi-objective FfJSSP with swarm-based neighborhood search algorithm [26]. Lei solved the FfJSSP with the objective of minimizing fuzzy scheduling time with co-evolutionary algorithm [27]. Gao *et al.* developed an artificial bee colony algorithm or flexible job-shop scheduling problem with fuzzy processing time [28]. In the existing methods, the membership functions are mostly determined by experience in FfJSSP or even in FJSSP. In order to minimize the influence of human factors, Azadeh *et al.* adopted random probability normal distribution, got the membership function with historical data, and solved multi-product assembly shop problems with fuzzy simulation [29].

In this paper, focused on FfJSSP, a differential evolution algorithm is proposed based on fuzzy satisfaction rate. Firstly, according to the experience data we construct the fuzzy triple membership functions of machine operation following the normal probability distribution. Then the objective function is defined by fuzzified machine processing time and costumer's expected time. At last, a novel Differential evolution (DE) algorithm is developed for scheduling process by defining a new representation of scheduling and corresponding operations. We also introduce a local search process to improve the performance of the method. To test the effectiveness of our proposed method, it is applied to two experiments for comparison. Numerical results show that our method is effective and practical.

2. Fuzzy flexible job shop scheduling problem

Compared with JSSP, one operation in FfJSSP problem could be processed on multiple machines rather than only on one machine. On the other hand, the processing time of an operation is fuzzy but not crisp, as well as the costumer's requirements. Therefore, FfJSSP is more complex than JSSP. The fuzziness of the problem is always described by fuzzy theory, for example, using fuzzy 3-tuples to represent processing times and fuzzy 4-tuples to represent costumer's requirements.

Supposed we have n jobs, m machines, denote J_i as Job i ($I \in [1, n]$), M_i as Machine i ($I \in [1, m]$), Num_i as the number of operations of J_i, $Q_{i,j}$ as the j-th operation of J_i, $S_{i,j}$ as the beginning time of $Q_{i,j}$, $C_{i,j}$ as the completing time of $Q_{i,j}$, $T_{i,j,k}$ as the processing time of $Q_{i,j}$ processed on machine k, $Q_{i,j,k}$ as the average $T_{i,j,k}$, $\sigma_{i,j,k}$ as the standard deviation of $T_{i,j,k}$. Then FfJSSP should satisfy some constraints as follows:

- Each operation of any jobs must be processed once and only once

$$\sum_{i=1}^{n} \sum_{j=1}^{Num} \sum_{k=1}^{m} X_{i,j,k} = \sum_{i=1}^{n} Num_i \tag{1}$$

Here $X_{i,j,k}$ is the indicator to represent whether $Q_{i,j}$ is processed on machine k, namely that

$$X_{i,j,k} = \begin{cases} 1 & \text{if } O_{i,j} \text{ is processed on machine } k \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- One operation must be processed after its precedent operations. Namely, for two operations of J_i, $O_{i,j}$, and $O_{i,k}$ ($j < k$), Eq. 3 must be satisfied:

$$S_{i,k} > C_{i,j} \ (j < k) \tag{3}$$

- All operations are non-preemptive. It means that once a machine starts processing an operation, the next operation to be processed on this machine must wait until the present operation completes, namely,

$$S_{l,m,k} \geq C_{i,j,k} (S_{l,m,k} \geq S_{i,j,k}) \tag{4}$$

- The processing time of every operation being processed on every machine agrees with normal distribution:

$$T_{i,j,k} \sim N(Q_{i,j,k}, \sigma_{i,j,k}) \tag{5}$$

Generally, the objective function to be optimized for FfJSSP could be divided into two classes, namely, the mono-objective function and multi-objective function:

- Mono-objective
  - Minimizing the makespan [27]

$$f = \min(C_{max}) \tag{6}$$

  C_{max} is the makespan.
  - Fuzzy scheduling based on E/T penalty [30]
  - Scheduling based on satisfaction rate [26], which is adopted in this paper and discussed in detail in Section 3.

- Multi-objective [25]
  - Integrating multiple objectives using weights

$$f = \sum_{i=1}^{l} (w_i \times f_i) \tag{7}$$

  where f_i is a certain objective and w_i is the corresponding weights.
  - Integrating objectives using Pareto optimality.

3. Fuzzy satisfaction rate based on normal distribution

In practice, the manufacturing system is fuzzy because of its changing environment conditions. For example, a breakdown of machine might lead to an uncertain delivery time. At the same

time, the costumer's requirements are often fuzzified, e.g., the costumers are more likely to provide a fuzzified time interval for production delivery but not a crisp time point. Hence, more and more researchers focused on the study of so call fuzzy flexible job shop scheduling problem. There are two kinds of objective functions for FfJSSP, one is mono-objective function and the other is multi-objective function. Zheng *et al.* defined the concept of satisfaction rate to measure the effectiveness of the scheduling [26]. However, in [26] and most other studies before, the fuzzy numbers used to describe the final solution are generated by trial and error, which are heavily relied on the experience and human factor. Azadeh *et al.* solved multi-product assembly shop problems by applying normal distribution of historical data to the membership function [29]. Similarly in this paper, we exploit the historical data, i.e. processing time and assuming it obeys normal distribution to determine the fuzzy numbers. In this way, the influence of human factor on evaluation of effectiveness of manufacturing is decreased. And then, we combined it with the satisfaction rate to evaluate the effectiveness of scheduling. To describe our method clearly, we firstly introduce the satisfaction rate concept in Section 3.1 as stated bellow.

3.1 Satisfaction rate

The membership function of fuzzy set can be described by fuzzy numbers. In this problem the triangle fuzzy number and trapezoidal fuzzy number are applied to represent processing time and costumer requirement, respectively. Triangle fuzzy number can be described by a triple (L, M, R) and the trapezoidal one can be represented by a 4-tuple (d_1, d_2, d_3, d_4). Denote the membership as μ, these two membership functions are described as Eqs. 8 and 9 [26]:

$$\mu = \begin{cases} 0 & x < L \\ 2(x - L)/(R - L) & L \leq x < (R + L)/2 \\ 2(R - x)/(R - L) & (R + L)/2 \leq x \leq R \\ 0 & x > R \end{cases} \tag{8}$$

$$\mu = \begin{cases} 0 & x < d_1 \\ (x - d_1)/(d_2 - d_1) & d_1 \leq x < d_2 \\ 1 & d_2 \leq x < d_3 \\ (d_4 - x)/(d_4 - d_3) & d_3 \leq x \leq d_4 \\ 0 & x > d_4 \end{cases} \tag{9}$$

where L and R are the lower and upper bounds of confidence interval of machine processing time, d_1 is the most optimistic time, d_2 and d_3 are the lower and upper bounds of the acceptable time interval, and d_4 is the most pessimistic time, respectively.

Because the makespan of a batch of jobs is fuzzy resulting from the fuzzy processing time, the membership functions of the makespan of job and the costumer requirement can be put in the same coordinate system, Fig. 1. Denote S_T as the area modelled by the triangular membership function, S_D the area modelled by the triangular membership function and S_E the intersected area of the two membership functions, respectively.
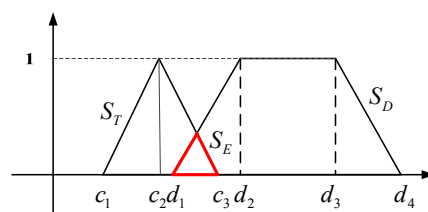

Fig. 1 Membership functions of the job makespan and the costumer requirement

Then, the satisfaction rate is defined as

$$fitness = \frac{S_E}{S_T} \tag{10}$$

3.2 Parameter estimation based on normal distribution

In the existing methods, the parameters of confidence interval of machine processing time, namely the lower and upper bounds L and R, are generated by experience. To avoid the influence of human factor, a normal distribution based method is developed in this section. Because there are many factors affecting the exactly machine processing time, we assume it obeys the normal distribution. Then we have

$$f(T_{i,j,k}) = Normal(Q_{i,j,k}, \sigma_{i,j,k}) = \frac{1}{\sigma_{i,j,k}\sqrt{2\pi}} exp\left(-\frac{(T_{i,j,k} - Q_{i,j,k})}{2\sigma_{i,j,k}^2}\right) \tag{11}$$

Therefore we could determine the parameters from the observation sample data according to statistical theory which is described step by step as follows.

1) Computation of average processing time. For machine M_i, the average processing time $Q_{i,j,k}$ of a specific operation and a specific job can be calculated by

$$Q_{i,j,k} = \frac{1}{L_{i,j,k}}\sum_{l=1}^{L_{i,j,k}} P_{i,j,k,l}, \ (1 \leq i \leq n, \ 1 \leq j \leq Num_i, 1 \leq k \leq m) \tag{12}$$

Here, $L_{i,j,k}$ represents the record number of processing time of the *j*-th operation of job *i* processed on machine *k*, $P_{i,j,k,l}$ processing time of the *j*-th operation of job *i* processed on machine *k* of the *l*-th record, respectively.

2) Computation of the standard deviation. For the standard deviation of $T_{i,j,k}$, it could be calculated according to

$$\sigma_{i,j,k} = \sqrt{\frac{1}{L_{i,j,k}}\sum_{l=1}^{L_{i,j,k}}\left(P_{i,j,k,l} - Q_{i,j,k}\right)^2} , \ (1 \leq i \leq n, \ 1 \leq j \leq Num_i, 1 \leq k \leq m) \tag{13}$$

3) Determination of the confidence interval. For a confidence value α, the confidence interval of $T_{i,j,k}$ could be calculated by

$$[\theta_1(\alpha), \theta_2(\alpha)] = \left[Q_{i,j,k} - \frac{Z_{\alpha/2}\sigma_{i,j,k}}{\sqrt{L_{i,j,k}}}, Q_{i,j,k} + \frac{Z_{\alpha/2}\sigma_{i,j,k}}{\sqrt{L_{i,j,k}}}\right] \tag{14}$$

where $\theta_1(\alpha)$ and $\theta_1(\alpha)$ are the lower bound and upper bound of the confidence interval of $T_{i,j,k}$ on confidence level of $1-\alpha$, and $-Z_{\alpha/2}$ and $Z_{\alpha/2}$ are the lower bound and upper bound of the confidence interval of $N(0, 1)$, which are easy to obtain by checking from table.

Compared Eq. 8 and Eq. 14, the lower and upper bounds of confidence interval of machine processing time *L* and *R* are set as

$$L = Q_{i,j,k} - \frac{Z_{\alpha/2}\sigma_{i,j,k}}{\sqrt{L_{i,j,k}}} \tag{15}$$

$$R = Q_{i,j,k} + \frac{Z_{\alpha/2}\sigma_{i,j,k}}{\sqrt{L_{i,j,k}}} \tag{16}$$

4. Scheduling method based on differential evolution algorithm

Differential evolution algorithm (DE) is a novel evolution algorithm, which is firstly proposed by Rainer, S and Price K. to solve continuous space optimization problem [31]. It is a swarm-based global parallel evolutionary algorithm of which individuals in the population can evolve simultaneously. DE can generate new individual variation with straightforward operations, and then generate the new population with greedy comparison strategies. Compared with other evolution algorithms, DE has obvious advantages, reproducing the population based global search ability, real-numbered coding, and straightforward operations. In addition, it has a special memory ability, which makes DE able to track the current evolution situation and adjust the searching strategies. In spite of the intensive research, mostly, DE is used to solve non-flexible job shop scheduling problem. In this paper, a novel scheduling method is developed to apply DE to flexible job shop scheduling problem by well defining operation code and machine code, which are both coded with real numbers. The basic differential evolution algorithm will be described in Section 4.1.

4.1 Differential evolution algorithm

Similar as genetic algorithm (GA), differential evolution (DE) algorithm also performs selection, crossover, and mutation operations. While in DE, differential idea is introduced into evolution operations and it is coded in real number. Denote the i-th individual in the G-th generation as

$$X_i^G = \left[X_{i1}^G, X_{i2}^G, \cdots, X_{iD}^G\right], (i = 1,2,\cdots, NP) \tag{17}$$

where $x_{ij}{}^G \in [lower, upper]$ ($j = 1,2,...,D$), which is called gene, *lower* and *upper* are the predefined lower bound and upper bound of the real code, D is the dimension of the problem, and NP is the size of population, respectively.

In each generation, the new population is generated by mutation and crossover operations first, and then selection operation is adopted to optimize the population.

Mutation

Mutation is used to maintain the diversity of population. There are many mutation strategies in the existing methods. Table 1 shows 5 most popular strategies. Taking the first one as an example, the mutation operation is performed according to

$$V_i^{G+1} = x_{r1}^G + \gamma \times \left(x_{r2}^G - x_{r3}^G\right) \tag{18}$$

where $V_i{}^{G+1}$ is the new individual after mutation, $r1$, $r2$, and $r3$ are three different randomly selected number within the set $\{1,...,i-1, i+1,..., NP\}$, and γ is the scaling factor. If any components of $V_i{}^{G+1}$ exceed the boundary, they should be revised by

$$V_{i,k}^{G+1} = \begin{cases} lower + \dfrac{V_{i,k}^{G+1} - upper}{2} & if\ (V_{i,k}^{G+1} > upper) \\ upper - \dfrac{lower - V_{i,k}^{G+1}}{2} & if\ (V_{i,k}^{G+1} < lower) \end{cases} \tag{19}$$

where $V_{i,k}{}^{G+1}$ represents the k-th component (gene) of $V_i{}^{G+1}$.

Table 1 Evolution strategies of DE

Strategies	Formula
rand/1	$V_i^{G+1} = x_{r1}^G + \gamma \times (x_{r2}^G - x_{r3}^G)$
best/1	$V_i^{G+1} = x_{best}^G + \gamma \times (x_{r2}^G - x_{r3}^G)$
rand-to-best/1	$V_i^{G+1} = x_i^G + \lambda \times \left(x_{best}^G - x_i^G\right) + \gamma \times (x_{r2}^G - x_{r3}^G)$
best/2	$V_i^{G+1} = x_{best}^G + \gamma \times (x_{r1}^G + x_{r2}^G - x_{r3}^G - x_{r4}^G)$
rand/2	$V_i^{G+1} = x_{r5}^G + \gamma \times (x_{r1}^G + x_{r2}^G - x_{r3}^G - x_{r4}^G)$

Crossover

Crossover is performed based on the results of mutation. It can be classified into binomial crossover and exponential crossover types, which are described by Eqs. 20 and 21, respectively.

$$U_{i,j}^{G+1} = \begin{cases} V_{i,j}^{G+1} & if\ (rand \leq CR\ or\ j = rand(1,D)) \\ X_{i,j}^G & otherwise \end{cases} \tag{20}$$

$$U_{i,j}^{G+1} = \begin{cases} V_{i,j}^{G+1} & if\ (rand \leq CR\ and\ j < D) \\ X_{i,j}^G & otherwise \end{cases} \tag{21}$$

where $U_{i,j}{}^{G+1}$ is the j-th component of the new individual $U_i{}^{G+1}$ after crossover, *rand* is a random real number within [0,1], and CR is the crossover rate. It could be found that in binomial crossover at least one gene in $V_i{}^{G+1}$ will be included, so the diversity of the population is improved.

Selection

Selection is defined by

$$X_i^{G+1} = \begin{cases} U_i^{G+1} & if\ (f\left(U_i^{G+1}\right) > f(X_i^G)) \\ X_i^G & otherwise \end{cases} \tag{22}$$

where f (*x*) is the objective function, X_i^{G+1} represents the newly generated individual in generation $G + 1$.

4.2 Differential evolution algorithm

Because DE uses real number for coding, it could not be applied to the scheduling problem directly. Qian *et al.* adopted a random key based Largest-Order-Value (L-O-V) rule to convert the continuous values of individuals in DE to job permutations [32, 33]. To the best of our knowledge, DE based methods are all applied to solve the JSSP problem, there are no related works for FfJSSP, and even for fJSSP. In this paper, by well defining the machine code and its operations to handle the flexibility, we extend Qian's method to the FfJSSP problems. Also a local search process is introduced in our proposed method.

Coding and decoding methods

Because in FfJSSP an operation could be processed on different machines, we design the code of a potential solution with two parts, namely the operation part and machine part (Fig. 2). For a FfJSSP problem with *n* jobs, denote m_i as the operation number of *i*-th job, then the total number of operations is $D = \sum_{i=1}^{n} m_i$, so the length of operation part and that of machine part both are *D*. Denote CO_i = ($CO_{i,1}$, $CO_{i,2}$,..., $CO_{i,D}$) as the operation part of the *i*-th individual and CM_i = ($CM_{i,1}$, $CM_{i,2}$,..., $CM_{i,D}$) as the machine part, each element of which is coded by a real number within [*lower*, *upper*]. Next the decoding methods of these two parts will be described.

Operation based coding						Machine based coding					
$CO_{i,1}$	$CO_{i,2}$	$CO_{i,3}$	•••	$CO_{i,D-1}$	$CO_{i,D}$	$CM_{i,1}$	$CM_{i,2}$	$CM_{i,3}$	•••	$CM_{i,D-1}$	$CM_{i,D}$
						Job 1		•••		Job n	
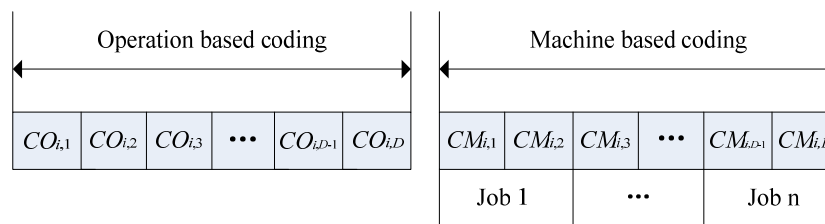
Fig. 2 Coding schedule

Decoding of operation part

Sort the elements of CO_i in a descending order, which is denoted as SO_i = ($SO_{i,1}$, $SO_{i,2}$,..., $SO_{i,D}$), where $SO_{i,p} \geq SO_{i,q}$ (*p<q*). Then we define an index vector for CO_i, denoting as $IndexO_i$ = ($indexo_{i,1}$, $indexo_{i,2}$,..., $indexo_{i,D}$), where $indexo_{i,k}$ = arg($SO_{i,j} = CO_{i,k}$). Define $m_0 = 0$, then the index number will be converted into job number according to

$$\pi_{i,j} = k, \text{ if } \sum_{s=1}^{k-1} m_s < \mu_{i,j} \leq \sum_{s=1}^{k} m_s \tag{23}$$

Here $\pi_{i,j}$ represents the corresponding job number of the operation $CO_{i,j}$. After obtained the index sequence of CO_i, it is easy to get the final job operation scheduling sequence: if $\pi_{i,j}$ is the *q*-th integer number *p*, it means that the *j*-th operation in the scheduling sequence should be the *q*-th operation of job *p*.

For clearly understand, a simple scheduling sequence is taken as an example to describe the method. The scheduling sequence has 3 jobs with 2, 2, and 3 operations, respectively. The initial individual is CO_i = (3.2, 1.4, 6.5, 3.7, 4.9, 7.9, 6.3). Then the sorted sequence CO_i, index sequence $IndexO_i$ and job sequence $\pi_{i,j}$ are listed in Table. 2. The result of job sequence is $\pi_{i,j}$ = (3, 3, 1, 3, 2, 1, 2), so the final operation schedule should be: (O_{31}, O_{32}, O_{11}, O_{33}, O_{21}, O_{12}, O_{22}).

Table 2 An example of decoding operation part

Dimension *j*	1	2	3	4	5	6	7
CO_{ij}	3.2	1.4	6.5	3.7	4.9	7.9	6.3
O_{ij}	7.9	6.5	6.3	4.9	3.7	3.2	1.4
Index O_{ij}	6	7	2	5	4	1	3
π_{ij}	3	3	1	3	2	1	2

Decoding of machine part

An element in machine part decoding represents which machine the corresponding operation is processed on. Different from the operation part, each element of machine part is corresponding to a FIXED operation, namely that the q-th operation of p-th job $O_{p,q}$ is represented by the k-th element of machine part, where

$$k = \sum_{i=1}^{p-1} m_i + q \tag{24}$$

To decode it, we firstly construct the available machine set for each operation process $O_{p,q}$ as $S_{p,q}$ ={machines that can process $O_{p,q}$}, and denote $|S_{p,q}|$ as the size of $S_{p,q}$, namely that the available machine number of operation $O_{p,q}$. Then the decoding method is shown as follows:

a) Calculate the size of interval [*lower*, *upper*]: $\Phi = upper - lower$.
b) For the position corresponding to the operation $O_{p,q}$ in the machine part, we divide the interval [*lower*, *upper*] into $|S_{p,q}|$ subintervals: $\varphi_k = \Phi / |S_{p,q}|$, where k is defined by Eq. 24.
c) For each individual code of machine part, the machine numbers of operations are determined by

$$\omega_{i,k} = \left\lceil \frac{CM_{i,k}}{\varphi_k} \right\rceil, (i = 1,2,\cdots,NP; \; k = 1,2,\cdots,D) \tag{24}$$

where i is the index number of individual in the population, k is the position in the code of machine part, $\omega_{i,k}$ represents the serial number of the dispatched machine in the corresponding set for $O_{p,q}$, and k, p, q satisfy Eq. 24. Then it is easy to find the dispatched machine of $O_{p,q}$ in the set of $S_{p,q}$. For example, there are 2 jobs with 2 and 4 operations, respectively, and there are 6 machines. So, $D = 2 + 4 = 6$. Besides, it is set that *lower* = 0, and *upper* = 10. The corresponding optional machine sets are $S_{1,1}$ = {1,3,5}, $S_{1,2}$ = {1,2,4,5}, $S_{2,1}$ = {2,4}, $S_{2,2}$ = {1,2,3,4,6}, $S_{2,3}$ = {4,6}, and $S_{2,3}$ = {2,3,4,5}. For individual CM_i = (4.4, 1.6, 1.2, 8.9, 0.3, 7.9), the calculation steps are shown in Table 3. Table 3 shows that the dispatched machines for the 6 operations are (3, 1, 2, 6, 4, 5). It means that the 1st and 2nd operations of job 1 are processed on machine 3 and 1 respectively; the 1st, 2nd, 3rd, and 4th operations of job 2 are processed on machine 2, 6, 4, and 5, respectively.

Table 3 An example of decoding machine part

Dimension j	1(O_{11})	2(O_{12})	3(O_{21})	4(O_{22})	5(O_{23})	6(O_{24})
CM_{ij}	4.4	1.6	1.2	8.9	0.3	7.9
φ_j	3.33	2.5	5	2	5	2.5
ω_{ij}	2	1	1	5	1	4
Corresponding machine	3	1	2	6	4	5

Local search option

In order to improve the search ability, local search process is usually performed within the general DE framework. For example, Zhang and Wu proposed a tree-based local exploration to solve JSSP problem [34]. In this paper, we use the idea in [34] to do the local search for our operation part of the code, and develop a novel local exploration which is applied to the machine part of the code. Next, they will be described concretely in this section.

Local search exploration for operation part

Defining a function swap (a, b) which means swapping a and b, then the pseudocode of this algorithm could be described by Fig. 3.

```
Select an operation individual α
For d = 1 to D
      Set P = {ϕ}
      For k = 1 to RandNum1
       Randomly select a position pair of α for swapping, gener-
           ating a new individual pₖ.
        Add pₖ into P.
      End For
      Denote the best one as η.
      For k = 1 to RandNum1
            For j = 1 to RandNum2
             Randomly select a position pair of pₖ for swapping,
             generating a new individual p_{RandNum1×k +j}.
                    Add p_{RandNum1×k +j} into P.
            End For
              Calculate the fitness of individuals in P. Keep the
              top RandNum1 individuals in the queue and discard
              the others.
              Update η.
      End For
End For
Output η.
```

Fig. 3 Pseudocode of local search exploration for operation part

Local search exploration for machine population

Corresponding to our designed machine code, a novel local exploration process is proposed in the paper. The pseudocode of the method is described by Fig. 4.

```
Set Φ = upper – lower
Select a machine individual β={b₁, b₂, …, b_D}
For j = 1 to D
      Calculate p and q according to Eq. 24
      φₖ = Φ / |S_{p,q}|
      For k = 1 to |S_{p,q}|
            cₖ = bⱼ + k×φₖ
            Update cₖ according to Eq. 20
      End For
      Use the best cₖ replacing bⱼ.
End For
Output updated β.
```

Fig. 4 Pseudocode of local search exploration for machine part

4.3 The framework of the algorithm

To sum up, the framework of the algorithm is described as follows:

1) Initialize the scaling factor F, the crossover rate CR, the size of the population NP, and the total iteration time N; set the current iteration number $G = 0$; input the fuzzy number of the processing time.
2) Initialize the population (Including operation part CO, and machine part CM), and calculate the best fitness BF.
3) Perform the mutation operation on CO and get population V; perform the crossover operation on V and get population U; perform the selection operation on V and U to get the new population CO.
4) Apply the local exploration search operation to CO and update population.
5) Perform the mutation operation on CM and get population MV; perform the crossover operation on MV and get population MU; perform the selection operation on MV and MU to get the new population CM.
6) Apply the local exploration search operation to CM and update population.
7) Set $G = G + 1$, if $G > N$ output the best individual, otherwise go to Step 2.

5. Results and discussion

To examine the effectiveness of the algorithm, it is better to implement it on some benchmark problems and compare the results with other existing methods. On the best of our knowledge, there is still no benchmark dataset for Fuzzy Flexible Job Shop Scheduling Problems (FfJSSP) at all. Therefore a Flexible Job Shop Scheduling Problem (fJSSP) is taken from Ref. [25] firstly, by which we could compare the results of our proposed method with other existing algorithms and examine its basic performance on fJSSP. To further evaluate our method on FfJSSP, we construct 30 problems for our second experiment, which could also be used as benchmark dataset for other researchers to study FfJSSP.

5.1 Case 1

The data set of first experiment is selected from Ref. [25]. There are 4 jobs in the problem, and the numbers of operation of jobs are 3, 3, 4 and 2, respectively. All operations are processed on 5 machines. The processing time is shown in Table 4. The result of our method is compared with five often used existing methods [35]. The result is listed in Table 5. In the table, C* = 11 is the optimal result of the problem. It can be seen that our proposed method could obtain the optimal result together with other 4 existing methods, while AL+CGA method only reaches 16, which are much more than others. Fig. 5 shows the Gantt chart of one feasible solution obtained by our method.

5.2 Case 2

Because there is no suitable data set for FfJSSP, we construct 30 FfJSSP problems to examine our proposed DE method. From the historical data, namely the historical processing time of different operations on different machines, the fuzzy numbers of membership function parameters are deduced according to normal distribution.

Table 4 Processing time

Jobs	Operations	M_1	M_2	M_3	M_4	M_5
J_1	$O_{1,1}$	2	5	4	1	2
	$O_{1,2}$	5	4	5	7	5
	$O_{1,3}$	4	5	5	4	5
J_2	$O_{2,1}$	2	5	4	7	8
	$O_{2,2}$	5	6	9	8	5
	$O_{2,3}$	4	5	4	54	5
J_3	$O_{3,1}$	9	8	6	7	9
	$O_{3,2}$	6	1	2	5	4
	$O_{3,3}$	2	5	4	2	4
	$O_{3,4}$	4	5	2	1	5
J_4	$O_{4,1}$	1	5	2	4	12
	$O_{4,2}$	5	1	2	1	2

Table 5 Comparisons of different algorithms

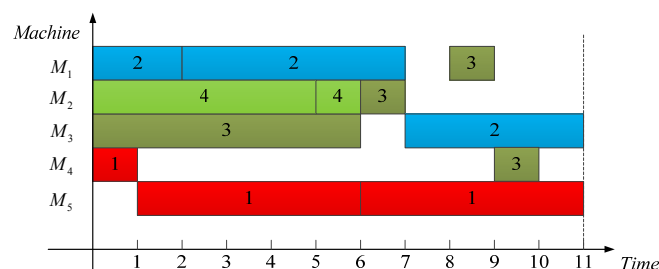| $n \times m$	C*	AL+CGA	GEN ACE	KBACO	TSPCB	ABC	Proposed DE
4×5	11	16	11	11	11	11	11

Fig. 5 Gantt chart of one feasible solution

The fuzzy numbers are needed to calculate the objective function in DE scheduling method. The 30 FfJSSP dataset together with their fuzzy numbers could be available on the following link _https://github.com/BrotherQi/APEM_DATA_, the scale of which ranges from 5×8 to 10×10, being more complicated than that of Case 1. Performing our proposed DE based scheduling method, 10 different strategies are executed for comparison, namely that 5 mutation strategies multiplying 2 crossover strategies described in Section 4.1. The results are shown in Table 6. In Table 6, the second column d_1, d_2, d_3, d_4 represent the fuzzy numbers of costumer's requirements, from the third to the 13th columns are the max satisfied rates of the ten strategies, in the order of rand/1 mutation + binomial crossover, best/1 mutation + binomial crossover, rand-to-best/1 mutation + binomial crossover, best/2 mutation + binomial crossover, rand/2 mutation + binomial crossover, rand/1 mutation + exponential crossover, best/1 mutation + exponential crossover, rand-to-best/1 mutation + exponential crossover, best/2 mutation + exponential crossover, and rand/2 mutation + exponential crossover. The results demonstrate our proposed method is applicable for FfJSSP. Comparing the average results of different strategies, it can be seen that those results are similar, the best one is 0.708 of the rand/2 mutation + binomial crossover strategy, and the worst one is 0.643 of rand-to-best/1 mutation + exponential crossover strategy. The average result of five binomial crossover strategies is 0.678, a little bigger than that of five exponential crossover strategies 0.662.

Table 6 Results of Case 2

Data set	d_1,d_2,d_3,d_4	Strategies									
		1	2	3	4	5	6	7	8	9	10
1	300,350,390,440	0.620	0.572	0.570	0.680	**0.692**	0.521	0.541	0.680	0.632	0.434
2	330,370,400,430	0.845	0.934	0.940	0.904	**0.987**	0.934	0.958	0.706	0.894	0.935
3	300,330,380,410	0.653	0.660	**0.758**	0.662	0.599	0.448	0.729	0.685	0.497	0.486
4	70,77,85,93	0.546	0.590	0.556	0.596	0.649	0.653	0.550	0.688	**0.694**	0.629
5	70,75,83,93	0.460	0.524	0.395	0.400	0.551	**0.622**	0.585	0.460	0.487	0.504
6	78,83,89,98	0.431	0.425	0.595	0.612	0.624	0.486	0.856	0.504	**0.882**	0.622
7	110,120,125,145	0.719	0.816	0.755	0.710	0.763	0.690	**0.875**	0.717	0.591	0.585
8	100,110,120,145	0.483	0.484	0.528	0.340	0.575	0.468	0.493	0.484	0.491	**0.577**
9	90,100,105,125	0.645	0.707	0.669	0.669	0.739	0.733	**0.742**	0.646	0.711	0.715
10	55,60,65,75	0.880	0.890	0.911	0.925	0.854	0.888	0.856	0.898	0.813	**0.933**
11	60,65,70,85	0.918	0.925	0.916	0.907	0.905	0.880	0.909	**0.971**	0.932	0.929
12	70,75,80,100	0.669	0.726	0.654	0.699	0.718	0.681	0.713	0.682	0.666	**0.778**
13	95,100,105,130	0.639	0.615	**0.683**	0.579	0.656	0.631	0.662	0.646	0.576	0.668
14	110,115,118,135	0.496	0.506	0.446	0.437	**0.627**	0.511	0.544	0.476	0.335	0.309
15	105,110,120,135	0.695	0.637	**0.798**	0.679	0.659	0.698	0.629	0.647	0.712	0.605
16	60,65,70,85	0.789	0.784	0.781	0.767	0.844	0.736	0.733	0.824	**0.847**	0.801
17	85,90,95,110	0.772	0.708	0.815	0.834	0.877	**0.893**	0.796	0.749	0.712	0.734
18	70,75,80,95	**0.684**	0.569	0.591	0.538	0.539	0.620	0.675	0.515	0.591	0.486
19	120,125,130,150	0.492	0.415	0.535	0.528	0.594	0.558	**0.716**	0.490	0.504	0.340
20	130,135,140,160	0.898	0.942	0.887	0.918	0.929	0.924	0.871	0.866	0.943	**0.949**
21	120,125,130,150	0.476	0.255	0.467	0.281	0.505	0.472	0.568	**0.645**	0.358	0.511
22	140,145,150,170	0.675	0.620	0.749	0.744	0.740	0.713	0.715	**0.758**	0.575	0.719
23	135,140,150,165	0.772	0.708	0.815	0.834	0.877	**0.893**	0.796	0.749	0.712	0.734
24	125,130,135,155	0.676	**0.787**	0.669	0.723	0.708	0.703	0.693	0.672	0.640	0.694
25	220,230,240,255	0.582	0.739	0.702	0.738	0.701	**0.791**	0.577	0.569	0.674	0.651
26	240,245,250,265	0.661	0.726	0.787	0.787	0.669	0.510	0.544	0.416	**0.839**	0.776
27	225,230,235,255	**0.784**	0.667	0.647	0.648	0.661	0.615	0.715	0.615	0.525	0.484
28	145,150,155,170	0.608	0.648	0.687	0.668	0.607	**0.743**	0.617	0.459	0.637	0.643
29	165,170,175,200	0.621	**0.668**	0.650	0.626	0.635	0.602	0.538	0.528	0.536	0.498
30	175,180,185,210	0.712	0.568	0.610	0.708	**0.762**	0.467	0.719	0.550	0.605	0.662
Ave	-	0.663	0.661	0.686	0.671	**0.708**	0.669	0.697	0.643	0.654	0.646

5.3 Discussion

The experiment of Case 1 is simple, the result of which shows that our proposed method is no less than the existing methods, and obviously better than AL+CGA method. In Case 2, we construct 30 more complicated datasets, and apply our method to them with different combinations of mutation and crossover strategies. The results are really effective and show that our method could be applied to real fuzzy flexible JSSP problems. It could be found that there are no significant difference among different strategy combinations, especially between two different crossover strategies.

6. Conclusion

Focused on FfJSSP, this paper develops a novel scheduling method based on fuzzy satisfaction rate and differential evolution algorithm. In the method, fuzzy parameters are deduced based on normal distribution for the calculation for satisfaction rate firstly. And then, a differential evolution is proposed for scheduling. Firstly, the coding method is well designed for FfJSSP, and then mutation, crossover and selection operations are proposed corresponding to the coding method. Moreover, local exploration search process is introduced in the method to improve the performance. At last, our proposed method is applied to two experiments to test the performance. Numerical results show that the developed method is practical and effective.

Acknowledgement

References

[1] Tang, M., Gong, D., Liu, S., Zhang, H. (2016). Applying multi-phase particle swarm optimization to solve bulk cargo port scheduling problem, *Advances in Production Engineering & Management*, Vol. 11, No. 4, 299-310, doi: 10.14743/apem2016.4.228.

[2] Burdett, R.L., Kozan, E. (2018). An integrated approach for scheduling health care activities in a hospital, *European Journal of Operational Research*, Vol. 264, No. 2, 756-773, doi: 10.1016/j.ejor.2017.06.051.

[3] Tang, M., Gong, D., Liu, S., Lu, X. (2017). Finding key factors for electric vehicle charging station location: a simulation and ANOVA approach, *International Journal of Simulation Modelling*, Vol. 16, No. 3, 541-554, doi: 10.2507/IJSIMM16(3)CO15.

[4] Wu, Z.B., Weng, M.X. (2005). Multiagent scheduling method with earliness and tardiness objectives in flexible job shops, *IEEE Transactions on Man, and Cybernetics, Part B: Cybernetics*, Vol. 35, No. 2, 293-301, doi: 10.1109/TSMCB.2004.842412.

[5] Sahin, C., Demirtas, M., Erol, R., Baykasoğlu, A., Kaplanoğlu, V. (2017). A multi-agent based approach to dynamic scheduling with flexible processing capabilities, *Journal of Intelligent Manufacturing*, Vol. 28, No. 8, 1827-1845, doi: 10.1007/s10845-015-1069-x.

[6] Zhang, Y.F., Wang, J., Liu, Y. (2017). Game theory based real-time multi-objective flexible job shop scheduling considering environmental impact, *Journal of Cleaner Production*, Vol. 167, 665-679, doi: 10.1016/j.jclepro.2017.08.068.

[7] Li, J.Q., Pan, Q.K., Tasgetiren, M.F. (2014). A discrete artificial bee colony algorithm for the multi-objective flexible job-shop scheduling problem with maintenance activities, *Applied Mathematical Modelling*, Vol. 38, No. 3, 1111-1132, doi: 10.1016/j.apm.2013.07.038.

[8] Gao, J., Sun, L.Y., Gen, M. (2008). A hybrid genetic and variable neighborhood descent algorithm for flexible job shop scheduling problems, *Computers & Operations Research*, Vol. 35, No. 9, 2892-2907, doi: 10.1016/j.cor.2007.01.001.

[9] Li, X.Y., Gao, L. (2016). An effective hybrid genetic algorithm and tabu search for flexible job shop scheduling problem, *International Journal of Production Economics*, Vol. 174, 93-110, doi: 10.1016/j.ijpe.2016.01.016.

[10] Zhang, Q., Manier, H., Manier, M.-A. (2012). A genetic algorithm with tabu search procedure for flexible job shop scheduling with transportation constraints and bounded processing times, *Computers & Operations Research*, Vol. 39, No. 7, 1713-1723, doi: 10.1016/j.cor.2011.10.007.

[11] 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.

[12] Gao, K.Z., Suganthan, P.N., Pan, Q.K., Chua, T.J., Cai, T.X., Chong, C.S. (2016). Discrete harmony search algorithm for flexible job shop scheduling problem with multiple objectives, *Journal of Intelligent Manufacturing*, Vol. 27, No. 2, 363-374, doi: 10.1007/s10845-014-0869-8.

[13] Li, J.Q., Pan, Q.K. (2012). Chemical-reaction optimization for flexible job-shop scheduling problems with maintenance activity, *Applied Soft Computing*, Vol. 12, No. 9, 2896-2912, doi: 10.1016/j.asoc.2012.04.012.

[14] Özgüven, C., Yavuz, Y., Özbakır, L. (2012). Mixed integer goal programming models for the flexible job-shop scheduling problems with separable and non-separable sequence dependent setup times, *Applied Mathematical Modelling*, Vol. 36, No. 2, 846-858, doi: 10.1016/j.apm.2011.07.037.

[15] Yuan, Y., Xu, H. (2013). Flexible job shop scheduling using hybrid differential evolution algorithms, *Computers & Industrial engineering*, Vol. 65, No. 2, 246-260, doi: 10.1016/j.cie.2013.02.022.

[16] Xu, Y., Wang, L., Wang, S.Y., Liu, M. (2015). An effective teaching–learning-based optimization algorithm for the flexible job-shop scheduling problem with fuzzy processing time, *Neurocomputing*, Vol. 148, 260-268, doi: 10.1016/j.neucom.2013.10.042.

[17] Supsomboon, S., Vajasuvimon, A. (2016). Simulation model for job shop production process improvement in machine parts manufacturing, *International Journal of Simulation Modelling*, Vol. 15, No. 4, 611-622, doi: 10.2507/IJSIMM15(4)3.352.

[18] Mehrabad, M.S., Pahlavani, A. (2009). A fuzzy multi-objective programming for scheduling of weighted jobs on a single machine, *The International Journal of Advanced Manufacturing Technology*, Vol. 45, No. 1-2, 122-139, doi: 10.1007/s00170-009-1947-5.

[19] Lei, D.M. (2008). Pareto archive particle swarm optimization for multi-objective fuzzy job shop scheduling problems, *The International Journal of Advanced Manufacturing Technology*, Vol. 37, No. 1-2, 157-165, doi: 10.1007/s00170-007-0945-8.

[20] Kilic, S. (2007). Scheduling a fuzzy flowshop problem with flexible due dates using ant colony optimization, In: Giacobini, M. *et al.* (eds.), *Applications of Evolutionary Computing, EvoWorkshops 2007,* Vol. 4448, Springer-Verlag, Berlin, Heidelberg, 742-751, doi: 10.1007/978-3-540-71805-5_80.

[21] Niu, Q., Jiao, B., Gu, X.S. (2008). Particle swarm optimization combined with genetic operators for job shop scheduling problem with fuzzy processing time, *Applied Mathematics and Computation*, Vol. 205, No. 1, 148-158, doi: 10.1016/j.amc.2008.05.086.

[22] Zhang, X.G., Deng, Y., Chan, F.T.S, Xu, P.D., Mahadevan, S., Hu, Y. (2013). IFSJSP: A novel methodology for the job-shop scheduling problem based on intuitionistic fuzzy sets. *International Journal of Production Research*, Vol. 51, No. 17, 5100-5119, doi: 10.1080/00207543.2013.793425.

[23] Gao, K.Z., Suganthan, P.N., Pan, Q.K., Tasgetiren, M.F. (2015). An effective discrete harmony search algorithm for flexible job shop scheduling problem with fuzzy processing time, *International Journal of Production Research*, Vol. 53, No.19, 5896-5911, doi: 10.1080/00207543.2015.1020174.

[24] Palacios, J.J., González-Rodríguez, I., Vela, C.R., Puente, J. (2017). Robust multiobjective optimisation for fuzzy job shop problems, *Applied Soft Computing*, Vol. 56, 604-616, doi: 10.1016/j.asoc.2016.07.004.

[25] 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.

[26] Zheng, Y.L., Li ,Y.X., Lei, D.M. (2012). Multi-objective swarm-based neighborhood search for fuzzy flexible job shop scheduling, *The International Journal of Advanced Manufacturing Technology*, Vol. 60, No. 9-12, 1063-1069, doi: 10.1007/s00170-011-3646-2.

[27] Lei, D.M. (2012). Co-evolutionary genetic algorithm for fuzzy flexible job shop scheduling, *Applied Soft Computing*, Vol. 12, No. 8, 2237-2245, doi: 10.1016/j.asoc.2012.03.025.

[28] Gao, K.Z., Suganthan, P.N., Pan, Q.K., Chua, T.J., Chong, C.S., Cai, T.X. (2016). An improved artificial bee colony algorithm for flexible job-shop scheduling problem with fuzzy processing time, *Expert Systems with Applications*, Vol. 65, 52-67, doi: 10.1016/j.eswa.2016.07.046.

[29] Azadeh, A., Hatefi, S.M., Kor, H. (2012). Performance improvement of a multi product assembly shop by integrated fuzzy simulation approach, *Journal of Intelligent Manufacturing*, Vol. 23, No. 5, 1861-1883, doi: 10.1007/s10845-011-0501-0.

[30] Shadrokh, S., Kianfar, F., (2007). A genetic algorithm for resource investment project scheduling problem, tardiness permitted with penalty, *European Journal of Operational Research*, Vol. 181, No. 1, 86-101, doi: 10.1016/j.ejor.2006.03.056.

[31] Storn, R., Price, K. (1997). Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization*, Vol. 11, No. 4, 341-359, doi: 10.1023/A:1008202821328.

[32] Bean, J.C. (1994). Genetic algorithms and random keys for sequencing and optimization, *ORSA journal on computing*, Vol. 6, No. 2, 154-160, doi: 10.1287/ijoc.6.2.154.

[33] Qian, B., Wang, L., Huang, D.X., Wang, W.L., Wang, X. (2009). An effective hybrid DE-based algorithm for multi-objective flow shop scheduling with limited buffers, *Computers & Operations Research*, Vol. 36, No. 1, 209-233, doi: 10.1016/j.cor.2007.08.007.

[34] Zhang, R., Wu, C. (2011). A hybrid differential evolution and tree search algorithm for the job shop scheduling problem, *Mathematical Problems in Engineering*, Vol. 2011, Article ID: 390593, doi: 10.1155/2011/390593.

[35] 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.

An experimental and modelling approach for improving utilization rate of the cold roll forming production line

Jurkovic, M.[a], **Jurkovic, Z.**[a,*], **Buljan, S.**[b], **Obad, M.**[c]

[a]Faculty of Engineering, University of Rijeka, Rijeka, Croatia
[b]Federal Ministry of Energy, Mining and Industry, Mostar, Bosnia and Herzegovina
[c]Faculty of Mechanical Engineering, Computing and Electrical Engineering, University of Mostar, Mostar, Bosnia and Herzegovina

ABSTRACT

This paper presents theoretical approach and complex experimental research which was conducted within the real production conditions of cold roll forming channel sections. The experimental investigation was focused on forming forces measuring on the rolls and the deflections of roll stands due to the forming loads. The comparison and analysis of the obtained experimental results was performed for the majority roll stands. Based on the experimental results mathematical modelling of the forming a force-roll load was performed by response surface methodology for different values of the input parameters of the process: material properties, sheet thickness, and sheet width. The defined force model and experimental research show insufficient energetic and technological utilization of the existing production line. After the conducted research in the production process a sheet thickness of up to 1.40 mm is used instead of 0.70 mm, and the utilization of the installed energy has increased from 20 % to 75 %. This is confirmed by the measured deformations of the roll stands and the energy consumption of the powered electric motor. Through realized modernization of the cold roll forming production line, 30 % higher productivity is achieved, which is a result of optimal number planning of roll forming stations and approximately the same load of all roll stands, as well as the higher flow rate of the profile sheet.

1. Introduction

1.1 Cold roll forming process

Profiles can be produced, from strip or sheet metal in three possible ways of bending: press brake forming (e.g. die bending) – by linear tool motion (Fig. 1), using rotating rollers-roll forming (Fig. 2a), and draw roll forming (Fig. 2b) [1]. The cold roll forming, high volume production process, plays an important role in sheet metal forming processes used to create wide variety of product. It creates segments like automotive, aircraft, buildings, construction, agricultural, office furniture, consumer appliances and other sheet metal products according to customer's requirements for high quality and quantity, at the right time and for the right price [2]. The roll forming is a continuous bending operation in which sheet or strip metal is formed between successive pairs of rolls (rotating tools) that progressively bend and form until the desired cross-sectional configuration, known as profiles (C- and U-channel, L-, T-, V- and Z-sections, trapezoidal profiles, and many others), is obtained (Fig. 3) [3]. The advantage of roll forming in compari-
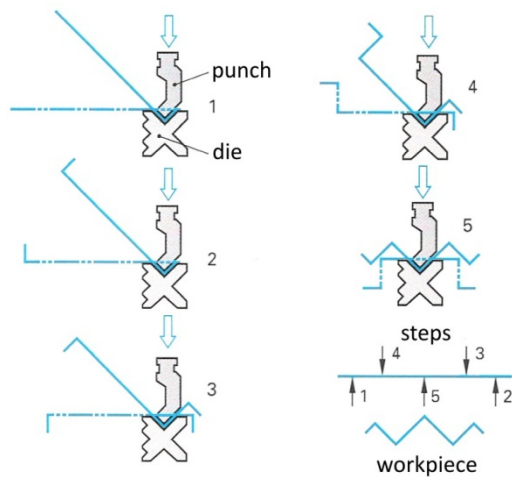
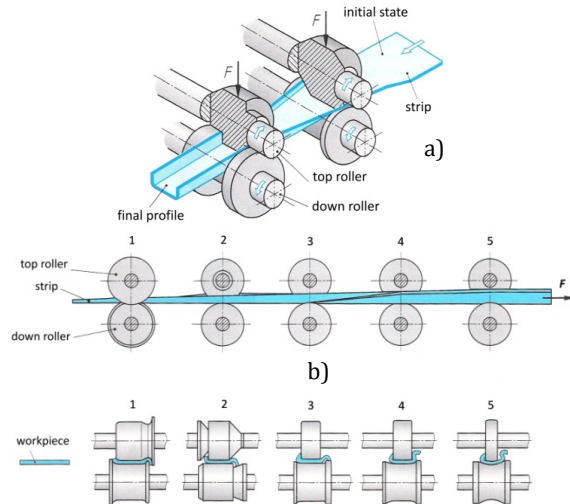**Fig. 1** The profile forming by die bending process

Fig. 2 The profile forming by, a) roll forming process, b) draw roll forming process

son with the other bending processes is that strips-sheets of any length (simple or complex cross sections) can be formed at a relatively high speed [4, 5].

1.2 Literature review

In general, forming process is characterized by various process parameters including geometrical data of tool and workpiece, material properties, contact conditions and different technological parameters. Therefore, the determination of the optimal forming parameters by using optimization techniques is a continuous engineering task with the main aim to reduce the production cost and achieve desired product quality [6]. Therefore, several authors with different aspects of roll forming process were studied from theoretical point of view to application through experimental research and mathematical modelling to numerical simulation. This was always with only one aim-to upgrade the process efficiency and for controlling of the process to be more predictable. Bhattacharyya *et al.* developed semi-empirical method for calculation of roll load in the cold roll forming process and obtained theoretical results which were verified with the experimental ones [7]. Paralikas *et al.* presented an optimization approach of the roll forming process parameters, for U-channel, based on semi-empirical method [8]. The optimal values of the process parameters for each roll station resulted in with reducing longitudinal and shear strains, and lastly with better product quality. Also, introduced by Paralikas *et al.* was used modelling for prediction deformation, as a function of the main roll forming process parameters and its impact on the V-section profile quality [9]. The study of Traub *et al.* presented a numerical modelling approach combining global and submodels enabling a high resolution of the strain distribution in the bending zone at acceptable computational costs [10]. The simulation concept was proofed by experimental results. Numerical simulation of the cold roll forming process of the U-channel was done by Bui *et al.* [11] and it was also studied by Lindgren [12, 13] with emphasis on change in the longitudinal peak membrane strain and deformation length of strip when yield strength increase. This was used to determine the number of forming steps and a distance between the roll stand. The proposed FE-method was shown to be an appropriate approach to research cold roll forming process. Paralikas *et al.* introduced energy efficiency indicator and presented on application a U-channel profile and analyzed influence of process parameters on energy efficiency indicator [14]. Ferreira [15] studied both experimentally and numerically, a deflection of roll stands during the forming process with a main aim to better predict defects. In the paper by Jurkovic *et al.* [16] an experimental investigation was presented of force and torque during the sheet roll forming in order to achieve improvement of the process. The application on response surface method and FE simulation in paper Zeng *et al.* [17] was studied on the optimization design of the cold roll forming process of a channel section. Liu *et al.* [18] developed mathematical model to analyze the deformation in roll forming a channel section.
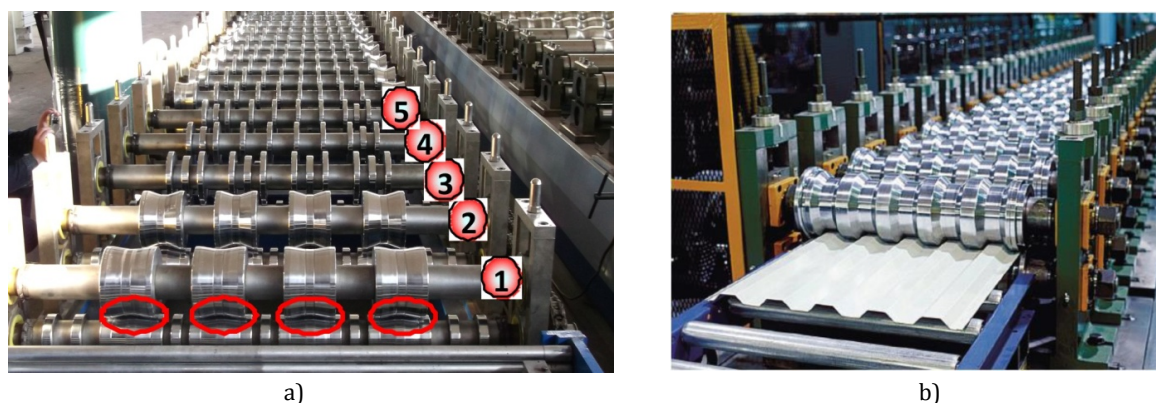
Fig. 3 The cold roll forming production line, a) beginning, b) ending of line

In paper [19] Wang *et al*. investigate the fracture mechanism on roll formed part. This paper has been presented experimental work of measuring elastic deformation of the roll forming stands under forming forces (also measured) occurring when strip is being bent in desired trapezoidal profile. The experimental research was carried out on the real production system in which the computerized, measuring and sensory equipment was installed. Based on the obtained experimental results mathematical modelling of the roll forming force was performed for different values of the input parameters of the process: material properties (R_m), sheet thickness (*s*), and sheet width (*b*). That is of a high importance to bring the roll forming force under control to achieve closer tolerances of profiles, which was crucial in the leading process in stable ones.

2. Materials and methods

The experiment was carried out in real production conditions on the roll forming production line. The production line was specially prepared with the equipment for forming forces measuring on the rolls and the roll stands deformation on forming stations. It was also prepared with other necessary resources for implementation of the design of experiment.

2.1 Roll forming production line – basic facility data

The production line has 20 forming stations (length 10.20 m and width 2.0 m). The experimental research was carried out on it in the real production process (Fig. 4). Each station had two rollers: upper and lower one. The electric motor power is 11 kW, the output speed of the gearbox is 41 min⁻¹, the transmission drives from electric motors to the shafts is a chain, the maximum sheet width is 1250 mm and the sheet thickness is between 0.4-1.0 mm. Further, the production system consists of decoiler with a coil weight of 10000 kg, in feed unit, straightening machine, roll forming machine, and shear cutting of the profiled sheet at a certain length [20, 21].

2.2 Materials used in experiments

In this experimental research, different type of materials were used [16]:

Fig. 4 Roll forming production line used in experiments
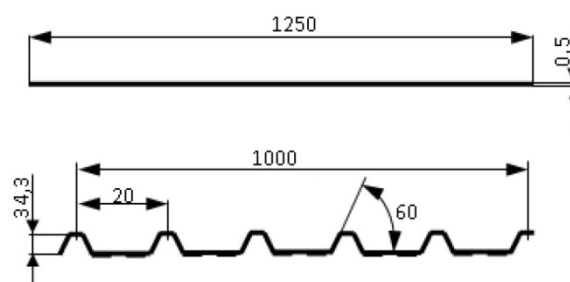
Fig. 5 Geometry of trapezoidal profiled sheet

a)   b)
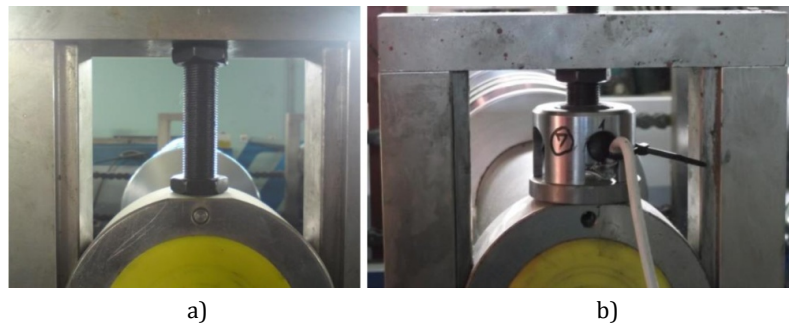
Fig. 6 Force transducer location on roll stand, a) before, b) after setup

- steel sheet DX 51D (DIN 17162-1, EN 10327), mechanical properties: R_m = 383 MPa, R_e = 278 MPa, percentage elongation after fracture A_{10} = 31.3 %,
- steel sheet DX 53D (DIN 17162-1, EN 10327), mechanical properties: R_m = 270 MPa, R_e = 140 MPa and A_{10} = 30 %,
- aluminium sheet Al 99.5 (EN 1050), mechanical properties: R_m = 130 MPa, A_{10} = 5 %.

Sample sheets used in the experiment were: three types of sheet materials (130 MPa, 270 MPa, 383 MPa), three widths of the sheets (950 mm, 1100 mm, 1250 mm), and three thicknesses (0.50 mm, 0.60 mm, 0.70 mm) and a length of 2500 mm. The profiled geometry of sheet metal (Fig. 5) was obtained by continuous forming of the sheet between the rollers, through twenty of the forming stations. At each forming station, the bending occurred according to flower pattern and a geometrical shape of rollers.

2.3 Measurement devices setup

Measurement devices used in this experiment can be classified into two main groups [16]:

1) Measurement device used for measuring rollers load:

- force transducer with strain gauges, Fig. 6.
- multi-channel measuring amplifier HBM-QuantumX (MX840A), Fig. 7.
- multi-channel measuring amplifier HBM-SPIDER8 (8xSR55), Fig. 7.
- data acquisition software, Fig. 7.

Force transducer for measuring force-roller load during roll forming of sheet is located between the top shaft bearing housings and the top crossbar of roll stand (Fig. 6b).
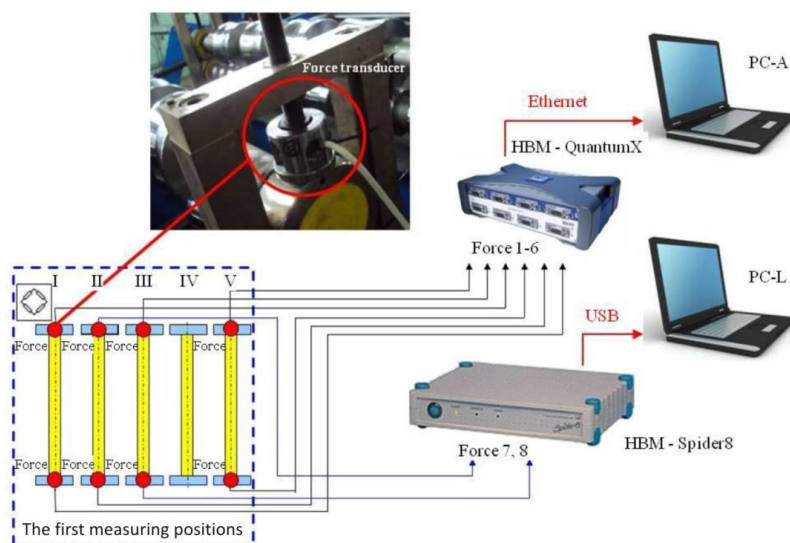
Fig. 7 Roll load – measurement device setup
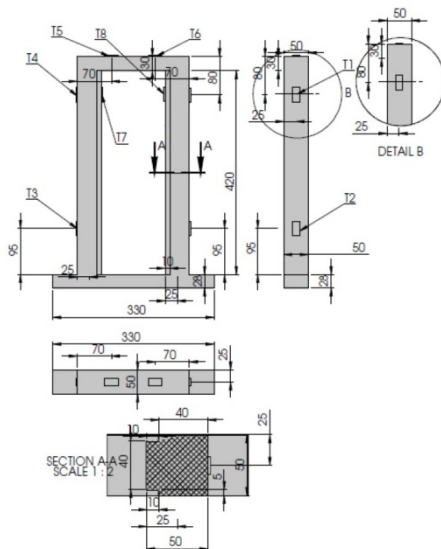
Fig. 8 The scheme of measurement points from T1 to T8   **Fig. 9** Deflection measuring on roll stand – setup
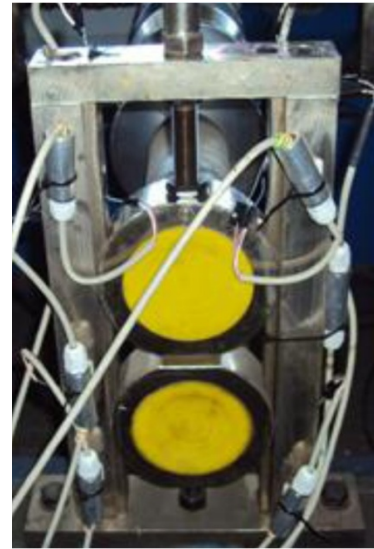
2) Measurement device used for measuring the roll stand deflection:

For measuring of the roll stand deflection strain gauges were used with the following character-istics: resistance of 120 Ω ± 0.35 %, k = 2.05 ± 1.0 %, type 6/120LY4, temperature compensation α = 10.8 (10^{-6}/° C). The Fig. 8 present the roll stand with locations of strain gauges (T1-T8) [21, 22]. The roll stand prepared for the strain measuring is shown on Fig. 9.

In recording of the experimental data, three amplifiers were used (Fig. 10): the first one – DMCplus with eight channels for strain data acquisition from the eight position (T1-T8) of strain gauges placed on the roll stands. The second one QUANTUM X has also eight channels to forces data collect (F1-F6), while the two other channels collect torque data (M1-M2) from the rollers with the telemetry box. The last amplifier SPIDER 8 is connected by two channels for forces data collect (F7 and F8). Measurement setup was defined in a way to measured force, torque, and strain values in real time from the eighteen positions.

2.4 Design of experiment and experimental results

The measurement devices used in this experimental research enabled data acquisitions. Table 1 presents the measured values for the roll forces (Fig. 6b) and strains of the roll stand (Fig. 9) for the second forming station.
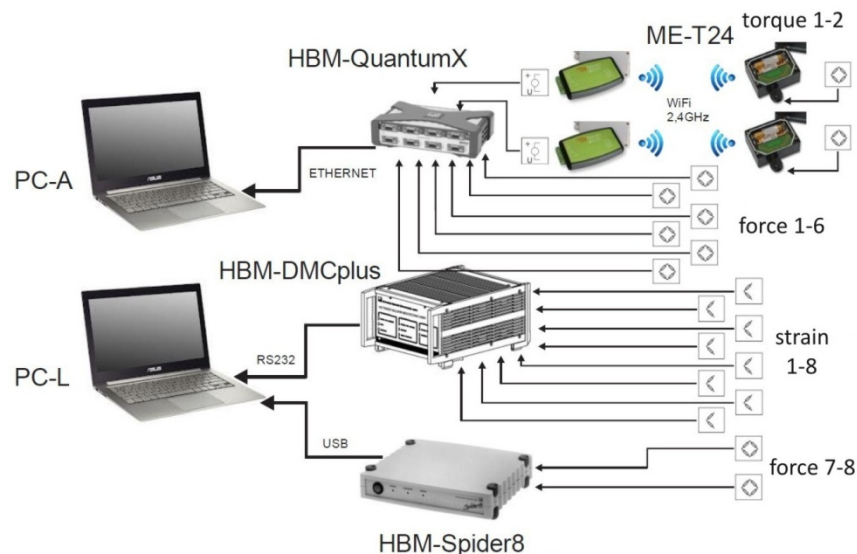
Fig. 10 Connecting scheme of amplifiers with measuring positions
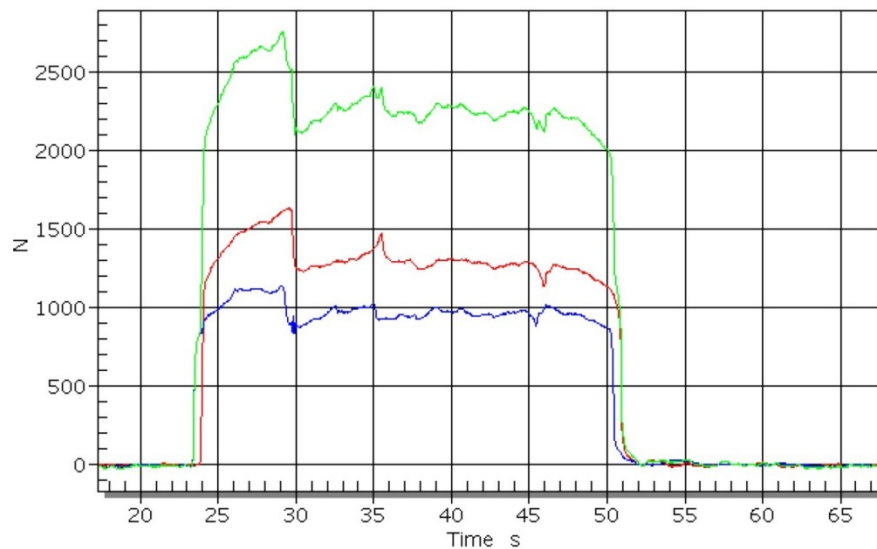
Fig. 11 Force diagram for the 3rd forming station, green curve – total force, red & blue forces on both side of a roller

Table 1 Design of experiment with experimental results – second forming station

Trial No.	Input process parameters			Coded process parameters			Experimental results	
							Force	Strain $T_1/T_2/T_3/T_4$,
	R_m, MPa	s, mm	b, mm	X_1	X_2	X_3	F_R^E, N	Δl_i, μm/m
1	130	0.5	950	-1	-1	-1	372.3	1.019/1.609/-1.637/-0.418
2	383	0.5	950	1	-1	-1	908.0	1.867/2.333/-1.120/0.588
3	130	0.7	950	-1	1	-1	654.2	3.732/1.741/1.799/-1.877
4	383	0.7	950	1	1	-1	1543.7	2.085/2.646/-1.644/0.516
5	130	0.5	1250	-1	-1	1	425.3	3.822/1.386/0.843/-1.785
6	383	0.5	1250	1	-1	1	1458.6	4.507/2.499/2.536/-2.414
7	130	0.7	1250	-1	1	1	1373.6	4.190/2.798/2.159/-1.727
8	383	0.7	1250	1	1	1	2558.3	2.269/3.306/-1.724/-0.160
9	270	0.6	1100	0	0	0	1173.5	3.664/2.159/1.650/-1.855
10	270	0.6	1100	0	0	0	1100.1	2.470/3.592/1.944/1.431
11	270	0.6	1100	0	0	0	1011.4	2.066/1.438/-2.377/-0.584
12	270	0.6	1100	0	0	0	1142.8	2.024/1.547/-2.456/-0.649

The experiment was performed according to design of experiments with three input parameters of the roll forming process: material properties (R_m), sheet thickness (s), and sheet width (b), resulting in with $N = 2^3 + 4 = 12$ experiments [23]. The force diagram for the third forming station (experiment No. 4) is shown in Fig. 11. According to Fig. 11, diagrams for other forming stations and for all the experiments also presented the change of forming force intensity (roll loads). The obtained data are necessary for a comprehensive analysis of the roll forming process, which consists of roll loads, optimal schedule of sheet strain through each station, the optimal flower pattern of sheet roll forming, optimal utilization rate of the energy consumption and the proposal for possible equally load on each forming stations of roll forming production line. The software for data acquisitions automatically processing the measurement results in real time.

2.5 Response surface methodology

Development and application of mathematical models in manufacturing processes is based on the application of knowledge, which is a prerequisite for the transformation of conventional and less productive processes in modern ones [23]. Therefore, mathematical modelling is based on the qualitative and quantitative relation between the input variables (X_i) and the output effects of the process (Y_j) in the form $Y_j = Y(X_i)$, which is the result of technology and technological process, workpiece material, and manufacturing system. In general, the theoretical model can be presented in the form of:

$$y = f(x_1, x_2, \ldots, x_k, \beta_1, \beta_2, \ldots, \beta_k)\tag{1}$$

where are:
- x_i – variable parameters of process i = 0, 1, 2,...,k,
- β_i – theoretical coefficients of regression model.

When mechanism of process is unknown, mathematical model can be shown in polynomial form:

$$Y = \sum_{i=0}^{k} \beta_i X_i + \sum_{1 \le i < m}^{k} \beta_{im} X_i X_m + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{1 \le i < m \le k} \beta_{imk} X_i X_m X_k\tag{2}$$

which insures optimal approximation of change of value x by means of function $f(x)$.

Based on the conducted experimental and regression analysis, the statistical model is determined by real regression coefficients b_i, b_{ii}, b_{im}, b_{imk} so that the mathematical model can get form:

$$Y = \sum_{i=0}^{k} b_i X_i + \sum_{1 \le i < m}^{k} b_{im} X_i X_m + \sum_{i=1}^{k} b_{ii} X_i^2 + \sum_{1 \le i < m \le k} b_{imk} X_i X_m X_k\tag{3}$$

3. Results and discussion

3.1 Modelling

Through application of the experiment, it is possible to define the accurate mathematical model with a minimal number of experimental data. In order to do so, the input variables of the process need to be determined first. In the experiment, constant parameters of the sheet roll forming are: forming machine, rollers geometry, lubrication, radius of rollers, bending angle, and stiffness of the rolls (Fig. 12).

Force model

Based on the experimental results by applying the response surface methodology, a force model in coded form was obtained (4) for the force profiling sheet metal by rollers:

$$y = 1143.483 + 455.4 \cdot X_1 + 370.7 \cdot X_2 + 292.2 \cdot X_3 + 63.15 \cdot X_1 \cdot X_2 + 99.1 \cdot X_1 \cdot X_3 \\ + 141.3 \cdot X_2 \cdot X_3 - 25.3 \cdot X_1 \cdot X_2 \cdot X_3\tag{4}$$

After decoding model (4), the mathematical model in physical form was:

$$F = 6560.701 - 13.969 \cdot R_m - 11697.472 \cdot s - 7.095 \cdot b + 19.658 \cdot R_m \cdot s + 0.013 \cdot R_m \cdot b \\ + 12.84 \cdot s \cdot b - 0.013 \cdot R_m \cdot s \cdot b\tag{5}$$
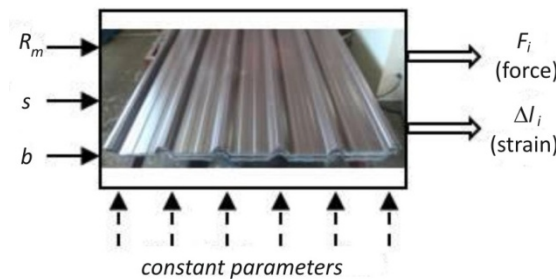
Fig. 12 Black box – input, output parameters of the roll forming process

The obtained force model (5) adequately describes the force dependence of the roll forming parameters $F = F (R_m, s, b)$. Multiple regression coefficient is $R = 0.99$, which confirms the obtained mathematical model´s adequacy.

3.2 The roll load analysis at forming stations

In Fig. 13, measured intensities of the roll forming force on a roller of forming stations are shown: the first measuring positions (1, 2, 3 and 5), the second positions (9, 10, 11 and 13) and the third positions (16, 17 and 18) corresponding to a linear stations layout in the production line. The forming stations: 4, 6, 7, 8, 12, 14, 15, 19 and 20 don't have force transducer and there are not measurements. This is not a lack of research because the force was measured on 11 stations out of 20 stations. The results of measurements were shown on Fig. 13:

- load increase from first to third position of force transducer, so the station (1) – for a first position force has a value F_1 = 1504.9 N, station (9) – second position F_9 = 3137.7 N and the station (16) – third position F_{16} = 6058.2 N, four times compared to the first station or two times to the ninth station,
- the roll load increase by stations is not a constant although there is a general trend of upward load, as shown by the fact form 1 > 2 <3 > 5 < 9 > 10 < 11 > 13 < 16 < 17 > 18, forming stations,
- optimal technological process should provide approximately the same roll load at each forming station,
- the maximum roll profile force is on the station 17 (F_{17} = 8377.8 N),
- in the experiment No. 6 sheets had the same dimensions and mechanical properties in all forming stations.

On the basis of experimental research, it can be concluded that the roll load increases with the number of stations (Fig. 13). Fig. 14 shows that the force intensity changes for forming stations 1, 2, 3 and 5, wherein the change of force is shown as a function of the number of experiments from 1 to 12. Thus, the force F_1 refers to the first station, the force F_2 to the second station, and so on. Also, Fig. 14 shows the maximum value of the roll profile force (F_i) for the four stations (1, 2, 3 and 5). The analysis by Fig. 14 shows the maximum force (F_3) to be on the third station. Also, this station has a maximum value of force in experiments No. 2, 4, 6 and 8, which was expected because the strength of the sheet was the largest, R_m = 383 MPa.

forming stations	1, 9 & 16	2, 10 & 17	3, 11 & 18	5 & 13
No. 6, first measuring positions	1504,9	1458,6	2999,8	1315,1
No. 6, second positions	3137,7	2863,6	5254,2	1588,3
No. 6, third positions	6058,2	8377,8	7882,9	
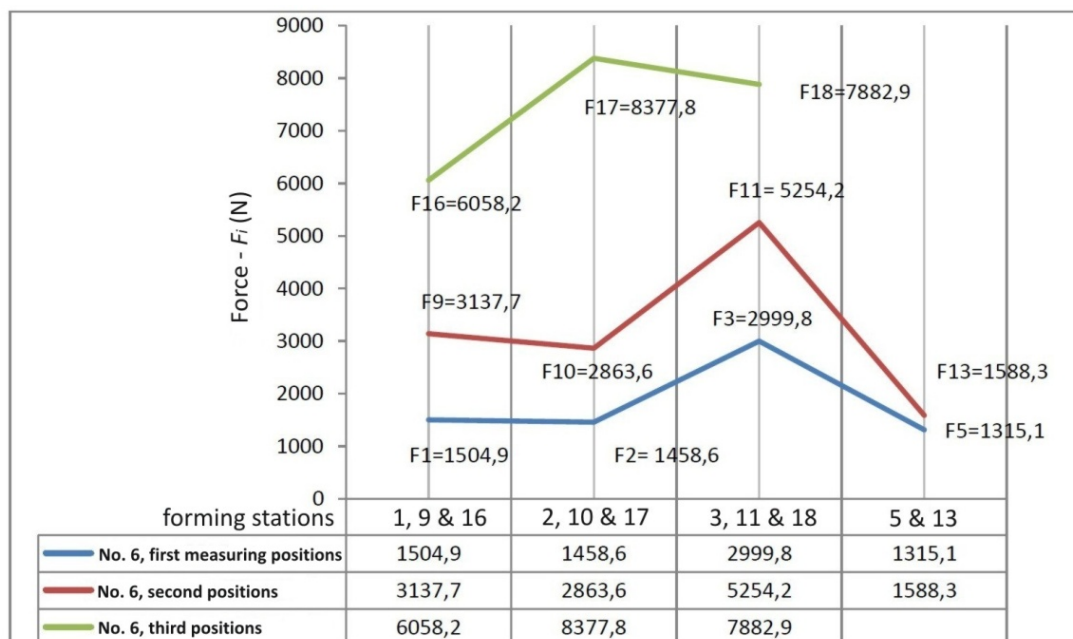
Fig. 13 The forces obtained in the experiment No. 6 on the forming stations: 1, 2, 3 & 5; 9, 10, 11 & 13; 16, 17 & 18

3.3 Roll stand deflection analysis

The results of the strain analysis obtained for measuring points (Table 1 and Fig. 9) 1, 2, 3 and 4, show that their absolute values, the largest values (Δl_1) at the measuring point 1 (T1), and the least values (Δl_4) at the measuring point 4 (T4). On the basis of that, it can be concluded:

- a strain is not classified by any logical sequence nor is the absolute value, nor is the sign of strain (+, -),
- on intensity and the sign of strain the profile tension between forming stations influences (due to a contact friction between the sheet and the roller). This is because there is no symmetry between the strain on the measuring points T1 and T4, T2 and T3, T1 and T2, and T3 and T4, although the profile velocity is equal between the stations, i.e. constant,
- the differences in the strength of the sheet material can be affected on the load change of roll stand,
- a contact between the sheet and the rollers with respect to friction is not constant, although the roll profile processes with a constant cross-sectional area because there is no change in the thickness of the sheet,
- on the measuring points T1 and T2 there is an obtained strain have + sign, while on the measuring points T3 and T4 have + and – sign, without being able to carry out a specific conclusion because there are more influential variables in such a result,
- the maximum strain generated in the experiment (8) and (6) wherein is the biggest strength and width of the sheet, while the least is expected in the experiment (1) and (2),
- the intensity of the deflection of the roll stand is very small and it meets the rigidity and accuracy of the roll profile. However, with the use of a thicker sheet, deflection will be increasingly significant.

3.4 The directions of modernization of the roll forming production line

The presented experimental research and modelling were carried out due to the modernization of the process and the roll forming production line for the profile sheet in order to achieve the optimal values of technological, energy consumption, and economical parameters. The roll load of the forming stations depends on: the material strength of the sheet (R_m), the thickness (s), and the sheet width (b). The model (2) obtained by modelling approach is in function of the mentioned parameters.

	1	2	3	4	5	6	7	8	9	10	11	12
F1	1104	1180	1830	2083	1436	1505	2212	2332	2421	2577	2597	2684
F2	372,3	908	654,2	1544	425,3	1459	1374	2558	1174	1100	1011	1143
F3	716,2	2041	1543	2758	828,7	3000	1575	3574	2390	2355	2102	2279
F5	283,5	823,7	749,4	1489	356,7	1315	937,7	1902	1165	1021	1241	976,7
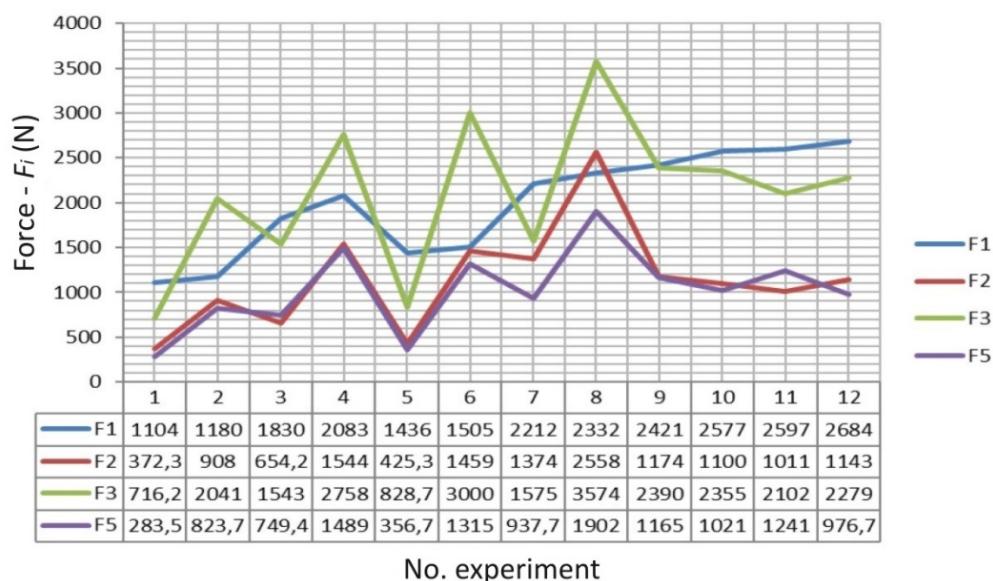
No. experiment

Fig. 14 The roll profile forces for experiments 1 to 12 and the forming stations 1, 2, 3 and 5

The basic parameters of modernization are from technological, energy and economical point of view:

- technological parameters should have the optimal values that are determined by minimizing force and taking into account the economic criterion-minimum costs,
- energy criterion is based on the minimum energy consumption, which implies: an optimized flower pattern of the sheet metal processing through the forming stations, less energy idle utilization, which requires optimum construction of the drive and transmission mechanism from the drive motor to the roller,
- also, an important indicator of the optimization of the process and the production system is the degree of energy utilization i.e.: $\eta_e = f(n, N, M_p, M_{ph})$, where: n – electric motor spindle rotation; N – number of forming stations; M_p – torque and M_{ph} – torque of idle.

Achievements and benefits of modernization

Modernization of the roll forming production line is carried out in order to obtain the highest achievements that will bring competitive advantage. Hence, to take advantage of that it was necessary to do the following steps:

- to design an optimal technological process of the sheet roll forming: a) approximately equalization of the roll load on each forming station, b) the increased production efficiency, and c) the increased part of the useful work of sheet roll forming in the total energy consumed,
- the redesign of flower pattern ensures: a) the optimal technological process (optimal flower pattern), b) greater bending angle per forming station (greater strain of bending sheet), c) larger sheet thicknesses with respect to the installed energy of the existing production system, and d) optimum flow rate of the production line depending on the complexity of the profile, type of material and product quality,
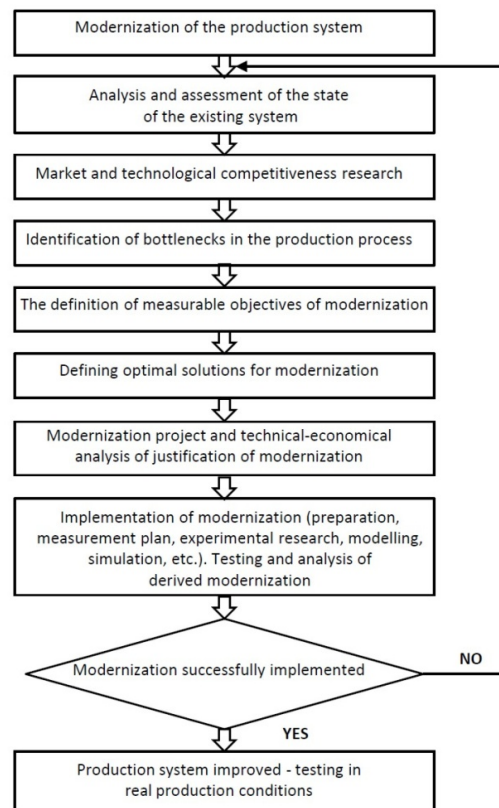
Fig. 15 Flow chart of complete modernization of the production system

- 30 % higher productivity of the production line after modernization as a result of the optimal flower pattern, higher utilization rate of the forming stations capacity, and the uniform roll loading on each roll stand,
- achieved high alignment of the production process and system by technological and energy point of view.

Flow chart modernization of the production system

The modernization of the production system beginning with the identification and analysis of the current state of the production system for that modernization is needed and the definition of the aims of modernization. The project of modernization and techno-economical analysis should show the validity of the modernization implementation (Fig. 15). Naturally, implementation of modernization needs to have measurable aims, experimental research, and relevant results to prove realized modernization. Previously mentioned things are fundamental of successful process and system modernization.

4. Conclusion

In this paper, the experimental research and modelling were carried out in order to test the roll forming process and system in real production conditions, with the aim of optimal process parameters of the sheet roll forming is increasing productivity of the production line. The obtained force model and deformation of the roll stand showed an insufficient utilization rate of the forming stations and consequently the production line. Therefore, in the process of the work, after modernization, the maximum sheet thickness of 1.40 mm was used instead of the thickness of up to 0.70 mm. The measured deformations of the roll stand were confirmed through this approach. The modernization of roll forming production line for sheet profile has been achieved with the following main objectives:

- technological process of the roll forming was optimized – forming stations are approximately equalization of a roll load, what resulted in the production line with a maximum efficiency and higher product quality,
- technological process of the sheet roll forming has been redesigned: the bending deformation for each forming station is greater, defining the optimal flower pattern, and an increased flow rate of the profiled sheet through the forming stations of the production line,
- and finally, productivity of modernized roll forming production line is higher for 30 %, what is the result of higher utilization rate of forming stations capacity and the balanced roll load on the forming stations.

Acknowledgement

References

[1] Scheipers, P. (2002). *Handbuch der Metallbearbeitung*, Verlag Europa-Lehrmittel, Haan-Gruiten, Germany.
[2] Halmos, G.T. (2006). *Roll forming handbook*, CRC Press, Boca Raton, USA.
[3] Brandegger, R. (2004). Predicting stress and strain in roll formed profiles, *The Fabricator*, Vol. 34, No. 5, 56-57.
[4] Lange, K. (1994). *Handbook of metal forming*, SME, Dearborn, USA.
[5] Schuler GmbH (1998). *Metal forming handbook,* Springer-Verlag, Berlin Heidelberg, Germany, doi: 10.1007/978-3-642-58857-0.
[6] Kuzman, K. (2001). Problems of accuracy control in cold forming, *Journal of Materials Processing Technology*, Vol. 113, No. 1-3, 10-15, doi: 10.1016/S0924-0136(01)00688-4.
[7] Bhattacharyya, D., Smith, P.D., Thadakamalla, S.K., Collins, I.F. (1987). The prediction of roll load in cold roll-forming, *Journal of Mechanical Working Technology*, Vol.14, No.3, 363-379, doi:10.1016/0378-3804(87)90019-2.

[8] Paralikas, J., Salonitis, K., Chryssolouris, G. (2010). Optimization of the roll forming process parameters – a semi-empirical approach, *The International Journal of Advanced Manufacturing Technology*, Vol. 47, No. 9-12, 1041-1052, doi: 10.1007/s00170-009-2252-z.

[9] Paralikas, J., Salonitis, K., Chryssolouris, G. (2009). Investigation of the effects of main roll-forming process parameters on quality for a V-section profile from AHSS, *The International Journal of Advanced Manufacturing Technology*, Vol. 44, No. 3-4, 223-237, doi: 10.1007/s00170-008-1822-9.

[10] Traub, T., Chen, X., Groche, P. (2017). Experimental and numerical investigation of the bending zone in roll forming, *International Journal of Mechanical Sciences*, Vol. 131-132, 956-970, doi: 10.1016/j.ijmecsci.2017.07.056.

[11] Viet, B.Q., Papeleux, L., Boman, R., Ponthot, J.-P., Wouters, P., Kergen, R., Daolio, G. (2005). Numerical simulation of cold roll forming process, In: *Proceedings of the 8th ESAFORM Conference on material forming,* Cluj-Napoca, Romania, 141-144.

[12] Lindgren, M. (2007). Cold roll forming of a U-channel made of high strength steel, *Journal of Materials Processing Technology*, Vol. 186, No. 1-3, 77-81, doi: 10.1016/j.jmatprotec.2006.12.017.

[13] Lindgren, M. (2005). *Modelling and simulation of the roll forming process*, Licentiate thesis, Luleå University of Technology, Sweden.

[14] Paralikas, J., Salonitis, K., Chryssolouris, G. (2013). Energy efficiency of cold roll forming process, *The International Journal of Advanced Manufacturing Technology*, Vol. 66, No. 9-12, 1271-1284, doi: 10.1007/s00170-012-4405-8.

[15] Ferreira, P.B.F. (2016). *Roll forming – A study on machine deflection by means of experimental analysis and numerical developments*, Master of science thesis, Faculty of Engineering, University of Porto, Portugal.

[16] Jurkovic, M., Jurkovic, Z., Obad, M., Buljan, S., Mustafic, E. (2015). An investigation of the force and torque at profile sheet metal rolling – Input data for the production system reengineering, *Tehnički Vjesnik – Technical Gazette*, Vol. 22, No. 4, 1029-1034, doi: 10.17559/TV-20150310092726.

[17] Zeng, G., Li, S.H., Yu, Z.Q., Lai, X.M. (2009). Optimization design of roll profiles for cold roll forming based on response surface method, *Materials & Design*, Vol. 30, No. 6, 1930-1938, doi: 10.1016/j.matdes.2008.09.018.

[18] Liu, C.-F., Zhou, W.-L., Fu, X.-S., Chen, G.-Q. (2015). A new mathematical model for determining the longitudinal strain in cold roll forming process, *The International Journal of Advanced Manufacturing Technology*, Vol. 79, No. 5-8, 1055-1061, doi: 10.1007/s00170-015-6845-4.

[19] Wang, H., Yan, Y., Jia, F., Han, F. (2016). Investigations of fracture on DP980 steel sheet in roll forming process, *Journal of Manufacturing Processes*, Vol. 22, 177-184, doi: 10.1016/j.jmapro.2016.03.008.

[20] Hasanagić, N. (2013). *Theoretical and experimental research of the force of sheet metal profiling by rollers*, (original title: *Teoretsko i eksperimentalno istraživanje sile profilisanja lima pomoću valjaka*), Master of science thesis, Bihać, Bosnia and Herzegovina.

[21] Jušić, A. (2013). *Theoretical research and experimental diagnostics load production line for the forming of open profile of the sheet metal by rollers*, (original title: *Teorijsko istraživanje i eksperimentalna dijagnostika opterećenja proizvodne linije za oblikovanje otvorenih profila iz lima pomoću valjaka*), Master of science thesis, Bihać, Bosnia and Herzegovina.

[22] Jurković, M., Jušić, A., Hasanagić, N. (2013). Experimental diagnostics line profiling sheet metal using rollers, In: *Proceedings of the 9th International Scientific Conference on Production Engineering*, Budva, Montenegro, 69-74.

[23] Jurković, M. (1999). *Mathematical modelling and optimization of machining processes*, (original title: *Matematičko modeliranje i optimizacija obradnih procesa*), Faculty of Engineering University of Rijeka, Rijeka, Croatia.

Risk management in automotive manufacturing process based on FMEA and grey relational analysis: A case study

Baynal, K.[a,*], Sarı, T.[b,*], Akpınar, B.[c]

[a]Kocaeli University, Department of Industrial Engineering, Kocaeli, Turkey
[b]Konya Food and Agriculture University, Department of International Trade and Business, Konya, Turkey
[c]General Electric, İstanbul, Turkey

A B S T R A C T

Risk management is an important issue in manufacturing companies in to-day's competitive market. Failure modes and effects analysis (FMEA) method is a risk management tool to stabilize production and enhance market competitiveness by using risk priority numbers (RPN). Although the traditional FMEA approach is an effectively and commonly used method, it has some shortcomings such as assumption of equal importance of the factors, severity, occurrence and detectability, and not following the ordered weighted rule. Thus, in order to improve RPN, an integrated method combining grey relational analysis (GRA) with FMEA is used in this study. The purpose of this paper is to contribute to risk management activities by proposing solutions to assembly line problems in an automotive manufacturing company by using combined GRA and FMEA method. In the proposed method, the priorities of production failures were determined by GRA approach and these failures were minimized by using FMEA technique. The study results indicated the actions that lead to enhancement in the product. The implementation of corrective/preventive activities resulted in 96 % improvement in door seal cuts problem caused by the door step assembly. Door seal cuts problem caused by instrument panel assembly and the noisy door window problem are solved completely.

1. Introduction

Before a new product is introduced to a market, the manufacturing companies probably face many problems in the stages of design, plan, production and delivery. It is very critical to detect and solve these problems, before the product reaches a customer. Some failures are easy to detect while some of them remain hidden. The whole process should be evaluated carefully and the appropriate quality control techniques should be used in order to find out these hidden failures. One of the most effective methods to determine the failures in any process is Failure Mode and Effects Analysis (FMEA).

FMEA can be expressed as a specific methodology in order to evaluate a process, system, service or design for possible ways in which failure can occur [1]. Risks, problems, concerns or errors are different type of failures. Failure mode can be described as a product failing to perform its desired function, described by the expectations of the customers. Failure emerges from the deviation from standards in the conditions of machine, method, material and workforce, affecting the quality of a product or a process. The FMEA analysis follows a well-defined sequence of

steps that includes (1) failure mode, (2) failure effects, (3) causes, (4) detectability, (5) corrective or preventive actions and (6) rationale for acceptance [2]. Today, with the increasing competition in the market, any deficiency and deviation in product performance result in market share loss. Although the traditional FMEA employing risk priority numbers (RPN) is an efficient and effective tool to stabilize production and enhance market competitiveness, it has been criticized for the following shortcomings: its (1) high duplication rate, (2) not following the ordered weighted rule, (3) assumption of equal importance of severity (S), occurrence (O), and detectability (D) and (4) failure to consider the direct and indirect relationships between the modes and the causes of failure [3].

In this study, an integrated method combining FMEA and Grey Relational Analysis (GRA) is used in order to overcome the shortcomings of traditional FMEA method. GRA is used to determine the priorities of production failures in an automotive manufacturing company. The two failures, door seal cut and noisy window problems, have been minimized by using FMEA methodology.

2. Literature review

FMEA methodology is widely used to manage risk in industries such as manufacturing, automotive, and aerospace. Vinodh and Santhosh [4] reported an application of design FMEA to an automotive leaf spring manufacturing organization in India. Implementation of fuzzy developed FMEA method to aircraft landing system, which is one of the important potential failure mode in aerospace industry, has shown the strength of the method in managing risk [5]. Chang [6] combined generalized multi-attribute FMEA and multi-attribute FMEA to improve LCD manufacturing process in a company in Taiwan. Segismundo and Miguel [7] proposed a methodological approach to effective risk management in new product development in a Brazilian automaker company by using FMEA technique. Banduka *et al.* [8] integrated lean approach with process FMEA in automotive industry. Liu *et al.* introduced a risk priority model by combining hesitant 2-tuple linguistic term sets and an extended QUALIFLEX method and FMEA methodology for handling a health care risk analysis problem [9, 10]. Barkovic *et al.* used FMEA method in improvement of newspaper production system quality [11].

GRA has been used by managers to make decisions under uncertainty in many different areas since 1982. Feng and Wang [12] measured the financial performance of airway companies with the help of GRA. Hsu and Wen [13] proposed a design to deal with the traffic and flight frequency in airways using GRA. In the study of Lin and Lin, one of the techniques used for optimization of wire erosion system was grey relation analysis method [14]. Wang *et al.* [15] proposed a hybrid methodology using grey relational analysis and experimental design to solve several multi-criteria decision making problems such as, a flexible manufacturing system, a rapid prototyping process and an automated inspection system. Palanikumar et.al. [16] optimized the results of polymer material process with grey relation analysis method. Rajeswari and Amirthagadeswaran [17] used grey relational approach to improve machinability properties of end milling process. Wang [18] developed a model for measuring the performance of logistic companies via grey relational analysis method. Ramesh *et al.* [19] proposed an effective model to investigate turning of magnesium alloy by using grey relational analysis method.

A combination of FMEA and GRA techniques are used by authors in order to eliminate the shortcomings of FMEA. Pillay and Wang [2] used an integrated method combining FMEA and grey theory to investigate the system failures in fuzzy environment for an ocean-going fishing vessel in their study. Baghery *et al.* [20] implemented process FMEA method combining with DEA (data envelopment analysis) and GRA in an automotive company producing auto parts for Samand, Peugeot 405 and Peugeot 206.

3. Materials and methods

In this study a combined methodology of failure modes and effect analysis and grey relational analysis is used. The priorities of production failures were determined by GRA approach and failures were minimized by using FMEA technique.

3.1 FMEA method

FMEA was first developed as an assessment tool to improve the evaluation of the reliability of military systems and weapons in the US army in the late 1940s. This method was also used for Apollo space missions in the 1960s by the National Aeronautics and Space Administration (NASA) [3]. In the late 1970s FMEA was used by Ford Motor Company in automotive production processes [6]. Because these applications resulted in satisfactory improvements in the Ford Company, the method has been widely used in automotive industry as a risk assessment tool. Today FMEA is applied successfully to industries such as aircraft, automotive, medicine, semiconductors and food industry. In the FMEA approach, for each of the failures identified (whether known or potential), an estimate is made of its occurrence (O), severity (S) and detection (D) [1]. Occurrence is the probability of occurrence of the failure and its cause. Detection is an evaluation process to find potential failures in the product. Severity is an expression of importance and emergency of potential system default mode. FMEA technique evaluates the risk of failure by using RPNs. The RPN value is found by taking the product of S, O, and D on a scale from 1 to 10. Higher RPN value indicates a higher priority.

3.2 Grey relational analysis

Grey relational analysis is a multi-criteria decision making method used by decision makers to take the right decision under circumstances with limited and uncertain data [21]. GRA approach explores system behavior using relational analysis and model constructions [2]. The grey system provides solutions to problems where the information is incomplete, limited or characterized by random uncertainty. The grey theory has become a popular technique providing multidisciplinary approaches in recent twenty years. The grey relational analysis was first developed by Julong Deng in 1982 [22]. The model includes three types of information points: white, grey or black. The main goal is to transfer black points in the system to the grey points. Grey relation analysis consists of six basic steps. These steps are explained below [19, 23, 24]:

Step 1: Construct a norm matrix X. It is assumed that there are n data sequences including m criteria:

$$X = \begin{bmatrix} x_1(1) & x_1(2) & \dots & x_1(m) \\ x_2(1) & x_2(2) & \dots & x_2(m) \\ \dots & \dots & \dots & \dots \\ x_n(1) & x_n(2) & \dots & x_n(m) \end{bmatrix} \quad (1)$$

where $x_i(j)$ is the entity in the *i*-th data sequence corresponding to the *j*-th criterion.

Step 2: Since multi-criteria decision making (MCDM) problems may contain a variation of different criteria, the solution needs normalization. Normalization process based on properties of three types of criteria, larger the better, smaller the better, and nominal the best:

$$x'_i(j) = \frac{x_i(j) - min_{j=1}^n[x_i(j)]}{max_{i=1}^n[x_i(j)] - min_{j=1}^n[x_i(j)]} ; \text{ larger the better} \quad (2)$$

$$x'_i(j) = \frac{max_{i=1}^n[x_i(j)] - x_i(j)}{max_{i=1}^n[x_i(j)] - min_{j=1}^n[x_i(j)]} ; \text{ smaller the better} \quad (3)$$

$$x'_i(j) = 1 - \frac{|x_i(j) - x_{obj}(j)|}{max\{max_{i=1}^n[x_i(j)] - x_{obj}(j), x_{obj}(j) - min_{j=1}^n[x_i(j)]\}} ; \text{ nominal the best} \quad (4)$$

$x_{obj}(j)$ is the target value for the criterion *j*, and $min_{i=1}^n[x_i(j)] \leq x_{obj}(j) \leq max_{i=1}^n[x_i(j)]$.

Step 3: Normalize the data set and generate a reference sequence based on Eq. 2 to Eq. 4. Normalized matrix is expressed as X':

$$X' = \begin{bmatrix} x_1'(1) & x_1'(2) & ... & x_1'(m) \\ x_2'(1) & x_2'(2) & ... & x_2'(m) \\ ... & ... & ... & ... \\ x_n'(1) & x_n'(2) & ... & x_n'(m) \end{bmatrix} \tag{5}$$

Step 4: Calculate absolute value table. The difference between a normalized entity and its reference value is calculated. The difference is shown as $\Delta_{0i}(j)$.

$$\Delta_{0i}(j) = |x_0'(j) - x_i'(j)| \tag{6}$$

$$\Delta = \begin{bmatrix} \Delta_{01}(1) & \Delta_{01}(2) & ... & \Delta_{01}(m) \\ \Delta_{02}(1) & \Delta_{02}(2) & ... & \Delta_{02}(m) \\ ... & ... & ... & ... \\ \Delta_{0n}(1) & \Delta_{0n}(2) & ... & \Delta_{0n}(m) \end{bmatrix} \tag{7}$$

Step 5: Compute grey relational coefficient $\gamma_{0i}(j)$, applying following grey relational equation:

$$\gamma_{0i}(j) = \frac{\Delta_{min} + \zeta\Delta_{max}}{\Delta_{0i}(j) + \zeta\Delta_{max}} \tag{8}$$

where $\Delta_{max} = max_{i=1}^{n} max_{j=1}^{m} \Delta_{0i}(j)$, $\Delta_{min} = min_{i=1}^{n} min_{j=1}^{m} \Delta_{0i}(j)$, and $\Delta_{0i}(j)$ and $\zeta \in [0,1]$. ζ is the distinguishing index and in most cases it takes the value of 0.5 offering moderate distinguishing effect.

Step 6: Compute the grey relational degree. Grey relational degree which indicates the magnitude of correlation or similarity. The overall grey relational degree (Γ_{0i}) is calculated by taking average value of grey relational coefficients by using the following equation:

$$\Gamma_{0i}(j) = \sum_{j=1}^{m} \gamma_{0i}(j)w(j) \tag{9}$$

where $w(j)$ refers to the weight of the *j*-th criterion. The sum of the weights of all criteria must equal to 1.

3.3 Integration of grey theory and FMEA method

The traditional FMEA method cannot assign the possibility of occurrence of failure, its detectability and severity comply with the real world. The integration of grey theory to FMEA allows engineers and decision makers to assign relative weights depending on research and production strategies. In decision making problems, the factor series with the highest grey relation degree gives the best alternative. The greater the relation degree means the smaller effect of failure source in FMEA application. For this reason, the increasing relative degree shows the decrease in risk priority of potential sources which have to be improved.

4. Case study

The case study was held in a car manufacturing company in the Turkish automotive industry. The aim of the study was to solve the assembly line problems. The company mainly faced two types of problems. The first one was the car door seal problem and the second one was the noisy car window problem.

4.1 Formulation of problems and causes

The car door seal problem can be explained as a tear or cut in the seal of the car doors. The door seal serves as a barrier protecting the inner car against dust and water from the outside. If the

seal is damaged and if this damage cannot be detected through quality control processes, it may cause severe customer complaints. The noisy car window problem is the annoying noise and shaking problem when the window glass moves up and down.

The factory has used Pareto analysis to find out the failures in manufacturing and their occurrence probability. The numbers about the occurrence probability, the causes of failures and the failure detection points are given in Table 1. The line point in the table represents the problems detected through the control points of the assembly line itself. Final point is the point of a detailed control of the car just before it leaves the assembly line. Pre-quality point is a control point for repaired cars before quality control. After quality control of the products, five of them are very carefully controlled in detail at a quality control point. The company's expert team has determined two important production problems and the most probable causes by using FMAE technique. The resulting causes are listed below:

Causes for car door seal cut or tear problem are summarized below:

Step: Tear or cut in the car door seal during the door step assembly
IP: Tear or cut in the car door seal during instrument panel assembly
Door lock: Tear or cut in the car door seal during door lock assembly
Seat: Tear or cut in the car door seal during car seat assembly
Operator: Damaging of the car door seal by operator during placement of the seal

Causes for noisy window problem are as follows:

Rivet position: The effect of the position of the rivet of window mechanism
Fixing equipment: The incompatibility between window mechanism and fixing equipment
Hole position: The effect of the position of the rivet hole

Table 1 Problems and causes

Failure	Cause	Detection Point					
		Line	Final	Pre-quality	Quality	Detailed quality	Customer
Cut/tear in door seal	Step	0	4	0	5	0	2
	IP	0	5	0	2	0	0
	Door lock	0	2	0	2	0	0
	Seat	0	0	0	2	0	0
	Operator	0	0	0	2	0	0
		Line	Final	Pre-quality	Quality	Detailed quality	Customer
Noisy window	Rivet position	0	88	0	9	0	0
	Fixing equipment	0	108	0	20	0	0
	Hole position	0	190	0	20	0	0

4.2 Probability of occurrence, detectability, severity

Occurrence (O):

Occurrence is the probability of failure occurrence and its cause. The precautions for detecting the failures are not taken into consideration in this step. Only the methods determined for preventing failure are considered. If the process is under statistical process control, the evolution depends on the statistical data. Otherwise, intangible data from judgments are used for evolution. In the factory, the occurrence and the detection points of failures are expressed by Pareto analysis and then transferred to a "1-10" scale. In the table, the O column describes the occurrence of probability between the values 1 and 10.

Table 2 Calculation of O values based on Pareto analysis

Failure	Cause	Detection Point						
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	O
Cut/tear in door seal	Step	0	54	0	55	0	12	10
	IP	0	25	0	32	0	0	7
	Door lock	0	12	0	0	0	0	3
	Seat	0	0	0	12	0	0	3
	Operator	0	0	0	12	0	0	3
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	O
Noisy window	Rivet position	0	88	0	9	0	0	7
	Fixing equipment	0	190	0	20	0	0	10
	Hole position	0	109	0	20	0	0	9

Detectability (D):

Detection is an evaluation process to find the potential failures in a product, before it leaves the assembly line. The failure should be accepted as it has occurred and the criteria for failure detection should be detected before the product has been introduced to a consumer. In the factory, according to results from Pareto analysis, the failures are scored for their detection points to calculate D values. The scale used in the calculation of D value is below:

Line point: 1-2
Final point: 3-4
Pre-quality: 5-6
Quality: 7-8
Detailed quality: 9
Customer: 10

Since all the failures are detected more than once in different points, D values are calculated by the weighed matrix (Table 3) and based on QLS Pareto analysis (Table 4).

Table 3 Weighted D values

Failure	Cause	Detection Point						
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	D
Cut/tear in door seal	Step	0	54X4	0	55X8	0	12X10	776
	IP	0	25X4	0	32X8	0	0	356
	Door lock	0	12X4	0	0	0	0	48
	Seat	0	0	0	12X8	0	0	96
	Operator	0	0	0	12X8	0	0	96
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	D
Noisy window	Rivet position	0	88X4	0	9X8	0	0	424
	Fixing equipment	0	190X4	0	20X8	0	0	920
	Hole position	0	109X4	0	20X8	0	0	596

Table 4 Calculation of D values based on QLS Pareto analysis

Failure	Cause	Detection Point						
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	D
Cut/tear in door seal	Step	0	54X4	0	55X8	0	12X10	6
	IP	0	25X4	0	32X8	0	0	6
	Door lock	0	12X4	0	0	0	0	4
	Seat	0	0	0	12X8	0	0	8
	Operator	0	0	0	12X8	0	0	86
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	D
Noisy window	Rivet position	0	88X4	0	9X8	0	0	4
	Fixing equipment	0	190X4	0	20X8	0	0	4
	Hole position	0	109X4	0	20X8	0	0	5

Severity (S):

Severity is an expression of importance and urgency of a potential system default mode. The only evaluation criterion for severity is the effect of a failure. The severity degrees are defined according to the degree of the effects on product, system, customer and legal obligations. The failure scale matrix of the factory's quality system is used directly in this study (Table 5). The severity values are calculated with the help of this table. Table 6 gives the S values determined by the company's experts using quality leadership system (QLS) Pareto analysis.

Table 5 Failure increase matrix for severity calculations

Failure Severity (Quality Standards)	Pick Level	Increase Level				
		Line team leader	Team leader	Area manager	Quality assurance manager	Factory manager
Blitz (10-9) Failure effects auto/driver control, customer safety and legal conditions.	Cause or Quality					
	1	X	X	X		
	3				X	X
	Sampling					
	1	X	X	X	X	
	3					X
A (8-7) Failure is very annoying and customer files a complaint to vendor/service.	Cause or Quality					
	1	X	X	X		
	3				X	
	Sampling					
	1	X	X	X	X	
	3					X
B (6-5) Failure is annoying, causing customer unsatisfaction and complaints of guarantee.	Cause or Quality					
	5	X	X			
	8			X		
	10				X	X
	Sampling					
	2	X	X	X		
	4				X	X
C (4-3) Failure is detected by educated/critical customers and it needs long time improvements.	Cause or Quality					
	10	X	X			
	15			X		
	Sampling					
	4	X	X	X		
	6				X	

Table 6 Severity calculation based on QLS Pareto values

Failure	Cause	Detection Point						
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	S
Cut/tear in door seal	Step	0	54	0	55	0	12	7
	IP	0	25	0	32	0	0	7
	Door lock	0	12	0	0	0	0	7
	Seat	0	0	0	12	0	0	7
	Operator	0	0	0	12	0	0	7
		Line	Final	Pre-quality	Quality	Detailed quality	Customer	S
Noisy window	Rivet position	0	88	0	9	0	0	6
	Fixing equipment	0	190	0	20	0	0	6
	Hole position	0	109	0	20	0	0	6

4.3 Calculation of risk priority number (RPN)

RPN is calculated by multiplication of O, D and S values. RPN shows the relative importance of failure causes. The resulting rank of RPN values help the decision makers to decide which cause should be improved first. The highest the RPN value means the first rate. The ranking according to RPN is shown in Table 7.

Table 7 Risk priority numbers

Failure	Cause	O	D	S	RPN	Rank
Cut/tear in door seal	Step	10	6	7	420	1
	IP	7	6	7	294	2
	Door lock	3	4	7	84	6
	Seat	3	8	7	168	5
	Operator	3	8	7	168	5
Noisy window	Rivet position	7	4	6	168	5
	Fixing equipment	10	4	6	240	4
	Hole position	9	5	6	270	3

4.4 Calculation of grey relational coefficient

The RPN in Table 7 are transferred to grey RPN values and reordered by using grey relational analysis. Then the difference matrix is constructed by using Eq. 8:

$$\begin{bmatrix} x_1(1) & x_1(2) & x_1(3) \\ x_2(1) & x_2(2) & x_2(3) \\ x_3(1) & x_3(2) & x_3(3) \\ x_4(1) & x_4(2) & x_4(3) \\ x_5(1) & x_5(2) & x_5(3) \\ x_6(1) & x_6(2) & x_6(3) \\ x_7(1) & x_7(2) & x_7(3) \\ x_8(1) & x_8(2) & x_8(3) \end{bmatrix} = \begin{bmatrix} 10 & 6 & 7 \\ 7 & 6 & 7 \\ 3 & 4 & 7 \\ 3 & 8 & 7 \\ 3 & 8 & 7 \\ 7 & 4 & 6 \\ 10 & 4 & 6 \\ 9 & 5 & 6 \end{bmatrix}$$

$$\begin{bmatrix} \Delta_1(1) & \Delta_1(2) & \Delta_1(3) \\ \Delta_2(1) & \Delta_2(2) & \Delta_2(3) \\ \Delta_3(1) & \Delta_3(2) & \Delta_3(3) \\ \Delta_4(1) & \Delta_4(2) & \Delta_4(3) \\ \Delta_5(1) & \Delta_5(2) & \Delta_5(3) \\ \Delta_6(1) & \Delta_6(2) & \Delta_6(3) \\ \Delta_7(1) & \Delta_7(2) & \Delta_7(3) \\ \Delta_8(1) & \Delta_8(2) & \Delta_8(3) \end{bmatrix} = \begin{bmatrix} 9 & 5 & 6 \\ 6 & 5 & 6 \\ 2 & 3 & 6 \\ 2 & 7 & 6 \\ 2 & 7 & 6 \\ 6 & 3 & 5 \\ 9 & 3 & 5 \\ 8 & 4 & 5 \end{bmatrix}$$

According to difference matrix, $\Delta_{min} = 2$, $\Delta_{max} = 9$ and ζ is assumed as 0.5 value. After calculation of difference matrix, grey relational coefficients are calculated by using Eq. 8.

$$\gamma_{0i}(j) = \frac{\Delta_{min} + \zeta\Delta_{max}}{\Delta_{0i}(j) + \zeta\Delta_{max}} = \frac{2 + 0.5 \cdot 9}{9 + 0.5 \cdot 9} = 0.481$$

The following matrix is constructed by using grey relational coefficients:

$$
\begin{bmatrix}
\gamma_1(1) & \gamma_1(2) & \gamma_1(3) \\
\gamma_2(1) & \gamma_2(2) & \gamma_2(3) \\
\gamma_3(1) & \gamma_3(2) & \gamma_3(3) \\
\gamma_4(1) & \gamma_4(2) & \gamma_4(3) \\
\gamma_5(1) & \gamma_5(2) & \gamma_5(3) \\
\gamma_6(1) & \gamma_6(2) & \gamma_6(3) \\
\gamma_7(1) & \gamma_7(2) & \gamma_7(3) \\
\gamma_8(1) & \gamma_8(2) & \gamma_8(3)
\end{bmatrix}
=
\begin{bmatrix}
0.481 & 0.684 & 0.619 \\
0.619 & 0.684 & 0.619 \\
1.000 & 0.867 & 0.619 \\
1.000 & 0.565 & 0.619 \\
1.000 & 0.565 & 0.619 \\
0.714 & 1.000 & 0.789 \\
0.556 & 1.000 & 0.789 \\
0.600 & 0.882 & 0.789
\end{bmatrix}
$$

The last step is to calculate the Grey RPN to determine the priorities. Table 8 shows the weights of cost based priorities.

Table 8 Cost based weights

	W_O	W_D	W_S
Cost	2.6 €	1.3 €	2.6 €
Weight	0.4	0.2	0.4

Grey relational degrees are found by the formula in Eq. 8. The grey relational degree of the first failure level is calculated as 0.577 by using Eq. 9.

$$\Gamma_{0i}(j) = \sum_{j=1}^{m} \gamma_{0i}(j)w(j) = 0.481 \cdot 0.4 + 0.684 \cdot 0.2 + 0.619 \cdot 0.4 = 0.577$$

The weighted grey RPN values are found as follows:

$$
\text{Weighted Grey RPN} =
\begin{bmatrix}
0.577 \\
0.632 \\
0.821 \\
0.761 \\
0.761 \\
0.823 \\
0.759 \\
0.742
\end{bmatrix}
$$

The RPN and Grey RPN values are listed comparatively in Table 9. As shown in the table, the weight of rivet position differs from one method to the other. According to the ranking in Table 9, the most important problem is the door seal cuts caused by step assembly. The second important problem is the seal cuts caused by instrument panel assembly. The third one is the noise problem of the window caused by the position of the hole. The least important problem is defined as noisy window problem caused by rivet hole position. The priority of decision makers is to initiate improvement on these most urgent problems.

Table 9 The comparison of FMEA RPN and grey RPN

Failure	Cause	O	D	S	FMEA RPN	Rank	Grey RPN	Rank
Cut/tear in door seal	Step	10	6	7	420	1	0.577	1
	IP	7	6	7	294	2	0.632	2
	Door lock	3	4	7	84	6	0.821	6
	Seat	3	8	7	168	5	0.761	5
	Operator	3	8	7	168	5	0.761	5
Noisy window	Rivet position	7	4	6	168	5	0.823	7
	Fixing equipment	10	4	6	240	4	0.759	4
	Hole position	9	5	6	270	3	0.742	3

4.5 Results and discussion

The solutions are developed according to the resulting grey RPN ranks. After improvements better results in the quality metrics are obtained.

- The solution for door seal cuts problem causing from door step assembly:

  Since a door seal prevents water and dust leakages, a tear or cut in the door seal causes customer complaints. There are five basic causes for the door seal problems. Based on Table 9, the most important reason for this problem is tear or cut in door seal during step assembly process. When the door step assembly process was inspected in detail, it was determined that the sharp corners of the step caused cuts in the seal during assembly of the step by an operator. As a part of corrective or preventive activities, all the areas of seal to where the step corners hit were detected. Magnetic protectors were made. The operators have begun to use these protectors in relevant areas during assembly. One month later, quality records indicated an important decrease in seal cuts by the rate of 96 %.

- The solution for door seal cuts problem causing from instrument panel assembly:

  Cut or tear in door seal during assembly of instrument panel gets the second rank in priority. The sharp-edged frame of the instrument panels was identified as the cause for this problem. The corrective and preventive activities were developed as effective solutions to the problem. As a result of corrective or preventive activities, the potential tangible areas of seal were determined. Magnetic protectors were designed to protect the surfaces which are likely to be damaged. The operators have begun to use these protectors in relevant areas during assembly. One month later, quality records indicated that cuts and tears in door seal caused by instrument panel assembly were prevented by the ratio of 100 %.

- The solution for noisy window glass problem causing from hole position:

  Noisy window glass is a problem which causes a disturbing noise and jolt in the vehicle, while the window is moving up and down. According to Pareto analysis, the most important reason with the third lowest degree in priority level is the rivet hole position. Riveting process in window installation were inspected in detail and it was detected that the distance between mechanism and the rivet hole was too small (2 mm). This short distance caused the mechanism parts to hit the rivet which resulted in a disturbing noise and jolt in windows. As a result of corrective/preventive activities, the position of rivet hole was moved to a 3 mm lower position. Therefore the distance became 5 mm which was sufficient for preventing the hitting of window mechanism parts. The preventive activities have resulted in 100 % improvement in the noise problem in one month.

  The quality reports indicated that 2 operators have spent 48 working hours in a month to deal with quality problems before improvements. After implementation of grey FMEA technique, this time was reduced to 2 hours which means saving cost by 2300 Euro in a month and 27600 Euro in a year.

5. Conclusion

FMEA is widely used as an efficient decision-making tool to control the stability of the manufacturing process and to improve product and system performance by decreasing failure rate. Although the traditional FMEA, employing risk priority numbers, stabilize production and increase the market competitiveness, it has some limitations such as failing to evaluate the relative relationship of each weight of those parameters. In this study the limitations of FMEA are overcome by using an integrated method of grey theory and FMEA. First, the possible causes of failure and their detection points are determined by FMEA. Second, the priorities of the factors (causes) are determined by using grey RPN values. According to the results of case application in an automotive manufacturing factory, a 96 % improvement was achieved for a door seal cuts problem caused by the door step assembly. A further door seal cuts problem caused by the instrument

panel assembly was solved completely. As a third improvement, the noisy door window problem, caused by riveting hole position, is prevented by 100 %.

The main advantage of the integrated GRA and FMEA method in this study is the flexibility of assigning weight to each factor in FMEA, providing an effective and consistent methodology to identify weak parts in the component studied. This integrated approach is convenient to deal with risk assessment problems under circumstances where the information is incomplete or uncertain. The processing of linguistic information based on expert knowledge and experience enables a realistic, practical and flexible way to express judgments.

References

[1] Stamatis, D.H. (2003). *Failure mode effect analysis: FMEA from theory to execution,* 2nd edition, ASQ Quality Press, Wisconsin, USA.

[2] Pillay, A., Wang, J. (2003). Modified failure mode and effects analysis using approximate reasoning, *Reliability Engineering & System Safety*, Vol. 79, No. 1, 69-85, doi: 10.1016/S0951-8320(02)00179-5.

[3] Chang, K.H., Chang, Y.C., Tsai, I.T. (2013). Enhancing FMEA assessment by integrating grey relational analysis and the decision making trial and evaluation laboratory approach, *Engineering Failure Analysis*, Vol. 31, 211-224, doi: 10.1016/j.engfailanal.2013.02.020.

[4] Vinodh, S., Santhosh, D. (2011). Application of FMEA to an automotive leaf spring manufacturing organization, *The TQM Journal,* Vol. 24, No. 3, 260-274, doi: 10.1108/17542731211226772.

[5] Yazdi, M., Daneshvar, S., Setareh, H. (2017). An extension to fuzzy developed failure mode and effects analysis (FDMEA) application for aircraft landing system. *Safety Science,* Vol. 98, 113-123, doi: 10.1016/j.ssci.2017.06.009.

[6] Chang, K.H. (2016). Generalized multi-attribute failure mode analysis, *Neurocomputing*, Vol. 175, Part A, 90-100, doi: 10.1016/j.neucom.2015.10.039.

[7] Segismundo, A., Miguel, P.A.C. (2008). Failure mode and effects analysis (FMEA) in the context of risk management in new product development: A case study in an automotive company, *International Journal of Quality & Reliability Management*, Vol. 25, No. 9, 899-912, doi: 10.1108/02656710810908061.

[8] Banduka, N., Veža, I., Bilić, B. (2016). An integrated lean approach to process failure mode and effect analysis (PFMEA): A case study from automotive industry, *Advances in Production Engineering & Management,* Vol. 11, No. 4, 355-365, doi:10.14743/apem2016.4.233.

[9] Liu, H.C., Li, P., You, J.X., Chen, Y.Z. (2015). A novel approach for FMEA: Combination of interval 2-tuple linguistic variables and gray relational analysis, *Quality and Reliability Engineering International,* Vol. 31, No. 5, 761-772, doi: 10.1002/gre.1633.

[10] Liu, H.C., You, J.X., Li, P., Su, Q. (2016). Failure mode and effect analysis under uncertainty: An integrated multiple criteria decision making approach, *IEEE Trnasactions on Reliability,* Vol. 65, No. 3, 1380-1392, doi: 10.1109/TR.2016.2570567.

[11] Borković, J., Milčić, D., Donevski, D. (2017). Implementation of differentiated quality management system and FMEA method in the newspaper production, *Tehnički Vjesnik – Technical Gazette*, Vol. 24, No. 4, 1203-1211, doi: 10.17559/TV-20160222082713.

[12] Feng, C.M., Wang, R.T. (2000). Performance evaluation for airlines including the consideration of financial ratios, *Journal of Air Transport Management,* Vol. 6, No. 3, 133-142, doi: 10.1016/S0969-6997(00)00003-X.

[13] Hsu, C.I., Wen, Y.H. (2000). Application of grey theory and multi objective programming towards airline network design, *European Journal of Operational Research*, Vol. 27, No. 1, 44-68, doi: 10.1016/S0377-2217(99)00320-3.

[14] Lin, J.L., Lin, C.L. (2002). The use of the orthogonal array with grey relational analysis to optimize the electrical discharge machining process with multiple performance characteristics, *International Journal of Machine Tools and Manufacture*, Vol. 42, No. 2, 237-244, doi: 10.1016/S0890-6955(01)00107-9.

[15] Wang, P., Meng, P., Zhai, J.Y., Zhu, Z.Q. (2013). A hybrid method using experiment design and grey relational analysis for multiple criteria decision making problems, *Knowledge-Based Systems,* Vol. 53, 100-107, doi: 10.1016/j.knosys.2013.08.025.

[16] Palanikumar, K., Karunamoorthy, L., Karthikeyan, R. (2006). Multiple performance optimization of machining parameters on the machining of GFRP composites using carbide (K10) tool, *Materials and Manufacturing Processes*, Vol. 21, No. 8, 846-852, doi: 10.1080/03602550600728166.

[17] Rajeswari, B., Amirthagadeswaran, K.S. (2017). Experimental investigation of machinability characteristics and multi-response optimization of end milling in aluminium composites using RSM based grey relational analysis, *Measurement,* Vol. 105, 78-86, doi: 10.1016/j.measurement.2017.04.014.

[18] Wang, Y.J. (2009). Combining grey relation analysis with FMCGDM to evaluate financial performance of Taiwan container, *Expert Systems with Applications*, Vol. 36, No. 2, Part 1, 2424-2432, doi: 10.1016/j.eswa.2007.12.027.

[19] Ramesh, S., Viswanathan, R., Ambika, S. (2016). Measurement and optimization of surface roughness and tool wear via grey relational analysis, TOPSIS and RSA techniques, *Measurement,* Vol. 78, 63-72, doi: 10.1016/j.measurement.2015.09.036.

[20] Baghery, M., Yousefi, S., Rezaee, M.J. (2016). Risk measurement and prioritization of auto parts manufacturing processes based on process failure analysis, interval data envelopment analysis and grey relational analysis, *Journal of Intelligent Manufacturing*, Vol. 1, 1-23, doi: 10.1007/s10845-016-1214-1.

[21] Chan, J.W.K., Tong, T.K.L. (2007). Multi-criteria material selections and end-of-life product strategy: Grey relational analysis approach, *Materials & Design,* Vol. 28, No. 5, 1539-1546, doi: 10.1016/j.matdes.2006.02.016.

[22] Deng, J.L. (1989). Introduction to grey system theory, *The Journal of Grey System*, Vol. 1, No. 1, 1-24.

[23] Zhai, L.Y., Khoo, L.P., Zhong, Z.W. (2009). Design concept evaluation in product development using rough sets and grey relation analysis, *Expert System with Applications,* Vol. 36, No. 3, Part 2, 7072-7079, doi: 10.1016/j.eswa.2008.08.068.

[24] Wu, H.H. (2002). A comparative study of using grey relational analysis in multiple attribute decision making problems, *Quality Engineering,* Vol.15, No. 2, 209-2017, doi: 10.1081/QEN-120015853.

Revenue sharing or profit sharing? An internet production perspective

Gong, D.[a], **Liu, S.**[a], **Tang, M.**[a], **Ren, L.**[b,*], **Liu, J.**[c], **Liu, X.**[d]

[a]School of Economics and Management, Beijing Jiaotong University, Beijing, P.R. China
[b]Research Center for Contemporary Management, Key Research Institute of Humanities and Social Sciences at Universities, School of Economics and Management, Tsinghua University, Beijing, P.R. China
[c]Institute of Economics, School of Social Science, Tsinghua University, Beijing, P.R. China
[d]School of Business and Law, Foshan University, Guangdong, P.R. China

A B S T R A C T

The booming of Internet economics brings new opportunities for small- and medium-sized product and service providers in e-platforms. Usually, the Internet platform (thereafter "platform") and Internet providers (thereafter "providers") operate under a consignment revenue sharing production model. Another production model is profit sharing, under which the platform undertakes part of the providers' operational costs. Intuitively, common sense conjectures that the platform prefers the revenue sharing model while the providers may prefer the profit sharing scheme. However, this is not the case. In this paper, we compare these two forms of emerging production models with a theoretical framework and investigate both market participants' performance under different schemes. Starting from the single provider case, we find that the provider has less incentive to operate under the revenue sharing contract when compared with the profit sharing contract. Counter intuitively, we identify the threshold of the cutting ratio above which it is more beneficial for the platform to choose the profit sharing mode. Our results are proved to be robust when the number of the providers increases. A numerical study is provided to illustrate this effect.

1. Introduction

With the ascending of Internet economics, more product (service) providers are cooperating with platforms such as Airbnb, Amazon, Uber, and eBay. According to a survey by JP Morgan & Chase, 4.3 % of adults earned income from the Internet production economy by 2016 [1]. An increasing number of workforces are seeking job opportunities through the platform production eco-system. Homeowners coordinate through the online rental platform, Airbnb, to provide cheaper and more convenient houses [2]. On the other side, tenants are able to find more households through the Internet with less transaction and searching costs than ever before. Another example comes with Internet producers such as Amazon and Alibaba, through which the sellers can sell to more potential end customers. Due to the convenience and cost saving of FBA (fulfillment by Amazon) service, more small business sellers are doing business through this powerful Internet production. Apart from that, thousands of active Amazon prime members are another important impetus to sell on the platform [3]. The sellers on the platform retain the authority to decide what kind of product to sell, how much to stock, and even the selling price [4].

In platform production eco-systems, the product (service) providers and the e-platform constitute a complex social system. Usually, the platform and the product (service) providers cooperate under a consignment revenue sharing (thereafter "R-S") arrangement [5], under which the sales revenue is split between the providers and the platform for each successful sale. Under such a production model, the platform bears no operational costs. However, the product (service) provider undertakes much of the over-stock or under-stock losses, which in turn reduces the provider's operational incentive. In comparison with the prevailing consignment R-S mode, profit sharing (thereafter "P-S") mechanisms can coordinate the interest between the platform and the providers at the cost of the platform taking on more risk. We realize that some core enterprises in a supply chain system have started to build up a P-S mechanism by holding the shares of their main suppliers in recent years, such as the Li & Fung Group, an HK based company [6]. The supply chain has achieved very good performance by taking the upstream suppliers' interest into account.

Based on this Internet production phenomenon, some interesting research questions arise: 1) How does the production model type impact both market participants' operational decisions? 2) Which model is more profitable for the platform, the R-S or the P-S mode? 3) How will the system performance vary with different production models? 4) Will the competition landscape alter the results above? To answer these research questions, we construct a theoretical model by considering a powerful platform hosting several products (service) providers. We start from the single provider case, which means there is only one provider on the digital platform. We show that the R-S production model lessens the provider's incentive to operate. The system performance decreases in cutting ratio. Intuitively, the platform prefers the R-S mode while the providers may prefer the P-S from a risk-sharing perspective. However, we do not identify a threshold of the cutting ratio above which the platform is better off by selecting the P-S mode. We then extend the basic model to cases where multiple providers exist. The results derived from the single provider model prove to be robust.

The remainder of the paper proceeds as follows. In Section 2, we review the literature that has the closest relationship with this paper. Section 3 follows with the model preliminaries. We introduce the time sequence and the notation conventions. We start with the single provider case in Section 4, which means there is only one provider selling through the platform. Our discussion is divided into two parts depending on the production model variations studied in this paper, namely, R-S and P-S. Section 5 extends the basic model with a single provider to multiple providers. We summarize our main findings and conclusions in the last part.

2. Literature review

Two streams of literature relate to our research; these streams are R-S and P-S in the supply chain and horizontal competitions in the supply chain.

The first stream lies in R-S and P-S in a supply chain. Dana and Spier [7] make an early attempt to investigate the revenue-sharing contract and vertical control in a video rental market. Wang *et al.* [5] study the channel performance under a consignment revenue-sharing contract by taking the e-retailing guru, Amazon, as a prototype, and finding that both the overall channel performance and the performance of individual firm depend on demand price elasticity as well as the retailer's share of the channel cost. Gong *et al.* [8] examine the production coordination problem from the perspective of asymmetric information: how a manufacturer coordinates the relationships with its subsidiary firm(s). Giannoccaro and Pontrandolfo [9] propose a model of a supply chain contract aimed at coordinating a three-stage supply chain, which is based on the revenue sharing mechanism. Gerchak and Wang [10] compare two distinct types of arrangements between an assembler/retailer and its suppliers. One scheme is a vendor-managed inventory with revenue sharing, and the other is a wholesale-price driven contract. Cachon and Lariviere [11] demonstrate that revenue sharing coordinates a supply chain with a single retailer and arbitrarily allocates the total supply chain profit. They compare this contract with other contract forms such as buy-back contracts, price-discount contracts, and quantity-flexibility contracts. For more application of revenue sharing contracts in operations management literature, please

see the in-depth review by Cachon [12]. Chen and Gupta [13] study a problem where the budget constraint supplier and retailer operates under a consignment revenue sharing contract in the presence of external financing. On the other hand, some literature highlights the advantages of P-S within a supply chain coordination framework. Çanakoğlu and Bilgic [14] analyze a two-stage telecommunication supply chain consisting of one operator and one vendor under a multiple period setting. They suggest a P-S contract where firms share both the revenue and operating costs. Wei and Choi [15] provide an industry practice of P-S in the apparel industry. Then, they explore the use of a wholesale pricing and profit sharing scheme for supply chain coordination under the mean-variance criteria. We differ significantly from the existing literature because we compare two different emerging production models, the R-S and P-S, within the framework of an online platform.

The second stream of research that relates to our research is on horizontal supply chain competition. Please see Singh and Vives [16], and Kreps and Scheinkman [17] for further comparison between quantity and pricing competitions. Gong *et al.* [18] investigate resources sharing's impact on the supply chain revenue. Van Meighem and Dada [19] summarize the value of strategic postponement for two firms competing in capacity and responsive pricing. Goyal and Netessine [20] evaluate the strategic value of manufacturing flexibility in an uncertain environment to understand whether the value of flexibility increases or decreases under competition. Anupindi and Jiang [21] consider duopoly models where firms make decisions on capacity, production, and price under demand uncertainty. In their model, flexible firms can postpone production decisions until the actual demand curve is observed, but inflexible firms cannot. Under general demand structures and cost functions, they characterize the equilibrium for symmetric duopoly and establish the strategic equivalence of price and quantity competitions when firms are flexible. In addition, Wen *et al.* [22, 23] also propose related methods for volatility of energy market, which provides support for this study. We apply the horizontal quantity competition in our model. We combine the production model selection with horizontal competition in an e-retailing supply chain framework.

3. Model preliminaries

Consider a powerful platform (thereafter "she"), hosting at least one product (service) provider (thereafter "he"), i.e., small- and medium-sized sellers selling through an e-retailing platform such as Amazon or Taobao, which is a Chinese based e-commerce platform owned by Alibaba. The platform and providers reach an agreement on the profit structure beforehand. In this paper, we consider two types of modes, namely, R-S and P-S. The former means that for each successful sale, the platform takes a certain percentage of the total sales revenue, which is known as a "referral fee" according to industrial practice. Usually, referral fees are different across different product categories. Note that under such a production model, the platform undertakes almost no operational risk since she refuses to share the production costs with the providers. However, the providers take on much of the inventory costs. Apart from the R-S mode, there is another production model, called P-S. In such a case, for each profit made, the platform takes a cut of the total profit. Therefore, the time sequence is as follows: the powerful digital platform negotiates with the providers on the production model type (R-S or P-S). Then, providers compete in quantities simultaneously. The market clearance price is determined then. The platform takes the revenue (profit) cut according to the agreement.

For notation conventions, we denote $q^i, i = 1, 2, \ldots, n$ as the quantity provided by the provider i, where the subscript states the number of providers. We apply a linear demand system where the market clearance price is determined by $p = a - \sum_{i=1}^{n} q^i$. Intercept a denotes the basic market demand. For simplicity without any loss of generality, there is no uncertainty considered in our model. γ $(0 \leq \gamma \leq 1)$ is the revenue (profit) cutting ratio, which is negotiated between the platform and the providers ex ante. Therefore, we assume the ratio is exogenously given according to industry practice. Let $\pi^i, i = 1, 2, \ldots, n$ stand for the providers' profit and Π denote the platform's profit.

4. Single provider case

We start from a case where only one provider is selling through the platform, which can be regarded as a benchmark for the competing scenario. We divide the discussion into two parts depending on the production model chosen. Then, we compare both agents' performance under these two production models.

Before we investigate the details of these two production models, we show the "first-best" solution when $\gamma = 0$ for both cases. In such case, the provider's profit is:

$$\pi(q;\gamma) = (a - q)q - cq \tag{1}$$

Therefore, the corresponding first-best quantity decision is the same in the monopoly case as $q_{FB} = \frac{a-c}{2}$. Then, we start from the R-S model.

4.1 R-S model

Under the R-S model, the provider's profit is:

$$\pi(q;\gamma) = (1 - \gamma)(a - q)q - cq \tag{2}$$

We can solve the optimal quantity decision of the provider as $q^*_{RS}(\gamma) = \frac{a}{2} - \frac{c}{2(1-\gamma)} \le q_{FB} = \frac{a-c}{2}$. $q^*_{RS}(\gamma)$ is determined when the marginal revenue is equal to the marginal cost. Accordingly, by substituting $q^*_{RS}(\gamma)$ into the provider's profit. We have $\pi_{RS}\big(q^*_{RS}(\gamma)\big) = \frac{(a(1-\gamma)-c)^2}{4(1-\gamma)}$.

Proposition 1: *With only one product (service) provider, the provider's quantity decision and profit decreases in γ under R-S.*

Proposition 1 shows that the revenue cut ratio reduces the provider's incentive to operate. As γ increases, the provider produces less and receives less profit. That is, because the platform takes a larger part of the total sales revenue as γ increases. On the other side, the provider's quantity decision deviates from the first-best solution due to a decrease in operational incentive.

In such a case, the platform's profit is:

$$\Pi_{RS}\big(q^*_{RS}(\gamma)\big) = \frac{(a(1-\gamma)-c)\gamma(a(1-\gamma)+c)}{4(1-\gamma)^2} = \frac{\gamma(a^2(1-\gamma)^2 - c^2)}{4(1-\gamma)^2} \tag{3}$$

Proposition 2: *With only one product (service) provider, the platform's profit is concave in γ. The optimal referral fee γ^* can be derived with $\frac{\partial \Pi_{RS}(q^*_{RS}(\gamma))}{\partial \gamma} = 0$ under R-S.*

When there is only one product (service) provider on the platform, the platform's profit is not always increasing in γ for the following reasons. Even though the total percentage of sales revenue increases as γ increases, the provider also reduces the output amount as γ increases. Consider an extreme case such that when the referral fee is extremely large, providers may shut down production, which leaves both market participants with 0 profit.

Denote the system profit as:

$$\begin{aligned} W_{RS}\big(q^*_{RS}(\gamma)\big) &= \pi_{RS}\big(q^*_{RS}(\gamma)\big) + \Pi_{RS}\big(q^*_{RS}(\gamma)\big) \\ &= c^2(1 - 2\gamma) + a^2(1-\gamma)^2 - 2ac(1-\gamma)^2 4(1-\gamma)^2 \end{aligned} \tag{4}$$

By checking the properties of $W_{RS}\big(q^*_{RS}(\gamma)\big)$, we have Proposition 3 below:

Proposition 3: *With only one product (service) provider, the system profit decreases in γ under R-S.*

Proposition 3 shows that the sum of the platform and providers' profit decreases in γ. When $\gamma = 0$, the revenue sharing production model degenerates to the first best model. As γ increases, the interest conflict between the platform and provider deepens because the platform is not burdened by the production cost carried by the provider. To eliminate the incentive mismatch,

they can operate under a P-S model, as noted by Wei and Choi [14]. Then, we analyze the case when the P-S model is present.

4.2 P-S model

Under a P-S model, the platform will take a cut of the total profit gained. Thus, the provider's profit is:

$$\pi_{PS}(q;\gamma) = (1-\gamma)(a-c-q)q \tag{5}$$

where γ is the cut ratio under a P-S model determined ex ante. We can find from the expression that the platform helps to bear the production cost under such a case.

We can solve the optimal quantity decision for the provider as $q^*{}_{PS}(\gamma) = \frac{a-c}{2} = q^*{}_{RS}(0) = q_{FB} = \frac{a-c}{2}$. Additionally, $q^*{}_{PS}(\gamma)$ is determined when the marginal revenue is equal to the marginal cost. Observe that the optimal quantity is independent of γ under P-S, which is the same as the first-best solution. The provider's profit is $\pi_{PS}\left(q^*{}_{PS}(\gamma)\right) = \frac{(1-\gamma)(a-c)^2}{4}$, which is decreasing in γ.

Proposition 4: *With only one product (service) provider, the provider's profit decreases in γ under P-S. The optimal quantity decision is independent of γ and achieves the first-best solution.*

Proposition 4 shows that the provider achieves the first-best solution under P-S because at this time the platform takes on the operational risks together with the provider. Unlike the revenue sharing scenario, the quantity decision is independent of γ.

In such a case, the platform's profit is:

$$\Pi_{PS}(q^*{}_{PS}(\gamma)) = \frac{\gamma(a-c)^2}{4} \tag{6}$$

which is increasing in γ. As γ increases, the platform takes a larger part of the total profit. The system profit is:

$$W_{PS}(q^*{}_{PS}(\gamma)) = \pi_{PS}(q^*{}_{PS}(\gamma)) + \Pi_{PS}(q^*{}_{PS}(\gamma)) = \frac{(a-c)^2}{4} \tag{7}$$

As we can find from the system profit, it equals the first-best profit. Under the P-S model, the interest conflicts between both market participants have been eliminated. We then need to compare their performance under different production models.

4.3 A comparison between R-S and P-S

We summarize both parties' profit in Table 1, where "R-S" is the revenue sharing and "P-S" is the profit sharing.

We then compare both players' profit under R-S and P-S for a given γ. We start from the provider's side.

Denote $\Delta\pi_{PS}(\gamma) = \pi_{PS}\left(q^*{}_{PS}(\gamma)\right) - \pi_{RS}\left(q^*{}_{RS}(\gamma)\right)$. We have

$$\Delta\pi_{PS}(\gamma) = \frac{(a(1-\gamma)-c)^2}{4(1-\gamma)} - \frac{(1-\gamma)(a-c)^2}{4} = \frac{c(c(-2+\gamma)-2a(-1+\gamma))\gamma}{4(-1+\gamma)} \tag{8}$$

$\Delta\pi_{PS}(0) = 0$ holds.

$$\frac{\partial\Delta\pi_{PS}(\gamma)}{\partial\gamma} = \frac{c(-2a(-1+\gamma)^2 + c(2-2\gamma+\gamma^2))}{4(-1+\gamma)^2} \tag{9}$$

Table 1 Profits with a single provider

	Provider	Platform	System (Provider + Platform)
R-S	$\dfrac{(a(1-\gamma)-c)^2}{4(1-\gamma)}$	$\dfrac{\gamma(a^2(1-\gamma)^2-c^2)}{4(1-\gamma)^2}$	$\dfrac{c^2(1-2\gamma)+a^2(1-\gamma)^2-2ac(1-\gamma)^2}{4(1-\gamma)^2}$
P-S	$\dfrac{(1-\gamma)(a-c)^2}{4}$	$\dfrac{\gamma(a-c)^2}{4}$	$\dfrac{(a-c)^2}{4}$

$$\frac{\partial^2 \Delta \pi_{PS}(\gamma)}{\partial \gamma^2} = \frac{c^2}{2(1-\gamma)^3} > 0 \tag{10}$$

By solving $\Delta \pi_{PS}(\gamma) = 0$, we have

$$\gamma^\pi = \frac{2(a-c)}{2a-c} = 1 - \frac{c}{2a-c}, \tag{11}$$

which we call the "R-S or P-S" threshold for the provider. Recall that to avoid triviality, we must have $(1-\gamma)a > c$, which is equivalent to $\gamma < 1 - \frac{c}{a} < \frac{c}{2a-c}$. Therefore, $\Delta \pi_{PS}(\gamma)$ is decreasing in the interval. In other words, $\pi_{PS}\left(q^*{}_{PS}(\gamma)\right) \le \pi_{PS}\left(q^*{}_{RS}(\gamma)\right)$ for $\gamma \in [0, 1 - \frac{c}{a}]$.

We then compare the platform's profit under R-S and P-S.

Denote $\Delta \Pi_{PS}(\gamma) = \Pi_{PS}\left(q^*{}_{PS}(\gamma)\right) - \Pi_{RS}\left(q^*{}_{RS}(\gamma)\right)$. We have

$$\Delta \Pi_{PS}(\gamma) = \frac{\gamma(a^2(1-\gamma)^2 - c^2)}{4(1-\gamma)^2} - \frac{\gamma(a-c)^2}{4} = \frac{c\gamma\left(2a(1-\gamma)^2 - c(1+(1-\gamma)^2)\right)}{4(1-\gamma)^2} \tag{12}$$

$\Delta \Pi_{PS}(0) = 0$ holds.

$$\frac{\partial \Delta \Pi_{PS}(\gamma)}{\partial \gamma} = \frac{c(2a(-1+\gamma)^3 - c(-2+2\gamma-3\gamma^2+\gamma^3))}{4(-1+\gamma)^3} \tag{13}$$

$$\frac{\partial^2 \Delta \Pi_{PS}(\gamma)}{\partial \gamma^2} = -\frac{c^2(2+\gamma)}{2(1-\gamma)^4} < 0 \tag{14}$$

Therefore, there is another root γ^Π to make $\Delta \Pi_{PS}(\gamma^\Pi) = 0$ hold. $\gamma^\Pi = 1 - \frac{\sqrt{2ac-c^2}}{2a-c}$, which we call the "R-S or P-S" threshold for the platform. Then, we compare $\frac{\sqrt{2ac-c^2}}{2a-c}$ with $\frac{c}{a}$. When $a > c$, we have $\frac{\sqrt{2ac-c^2}}{2a-c} > \frac{c}{a}$. Therefore, $1 - \frac{\sqrt{2ac-c^2}}{2a-c} < 1 - \frac{c}{a}$.

In other words, $\Pi_{PS}\left(q^*{}_{PS}(\gamma)\right) \ge \Pi_{RS}\left(q^*{}_{RS}(\gamma)\right)$ for $\gamma \in [0, 1 - \frac{\sqrt{2ac-c^2}}{2a-c}]$ and $\Pi_{PS}\left(q^*{}_{PS}(\gamma)\right) < \Pi_{RS}\left(q^*{}_{RS}(\gamma)\right)$ for $\gamma \in (1 - \frac{\sqrt{2ac-c^2}}{2a-c}, 1 - \frac{c}{a}]$. However, from the platform's side, this does not always hold for $\Pi_{PS}\left(q^*{}_{PS}(\gamma)\right) > \Pi_{RS}\left(q^*{}_{RS}(\gamma)\right)$. Theorem 1 below summarizes the above results:

Theorem 1: *When $a(1-\gamma) > c$, the powerful platform selects a revenue sharing contract for $\gamma \in [0, 1 - \frac{\sqrt{2ac-c^2}}{2a-c}]$, and selects a profit sharing contract for $\gamma \in (1 - \frac{\sqrt{2ac-c^2}}{2a-c}, 1 - \frac{c}{a}]$. The platform and the product (service) provider achieve interest alignment when $\gamma \in [1 - \frac{\sqrt{2ac-c^2}}{2a-c}, 1 - \frac{c}{a}]$.*

Theorem 1 seems slightly counterintuitive since common sense conjectures that the platform will always prefer the R-S model since she will not bear any operational costs (take no additional operational risks). The provider seems to always prefer the P-S model as long as the platform is willing to share risks with him. However, we identify some conditions under which the platform may be better off by providing the P-S model, which means she must have no hesitation with undertaking more operational risks when γ is relatively high. If not, the provider loses the desire to operate. We then try to figure out how the system profit varies under different production models. We have Proposition 5 below.

Proposition 5: *The system profit under R-S is no more than the profit with P-S. The system achieves first-best performance with P-S.*

The platform and product (service) provider's interest conflicts under the R-S because the platform is unwilling to undertake more operational costs. However, with a P-S alignment, the system achieves the *first-best* solution. It seems that P-S is the best mechanism from a systematic

view. However, in reality, the powerful platform may control the game and selects the best scheme for herself, which in turn harms the social welfare.

How will the results change when more than one provider is present in the market? Will the competition landscape alter our main results? We extend our basic model to a case where multiple providers exist.

5. Competing providers case

Usually, the product (service) providers on the platform always face cutthroat competition. The competition landscape may have an impact on the performance of both players. We consider that n providers are selling through the platform. As we have introduced in the model preliminaries section, the market clearance price is determined by $p = a - \sum_{i=1}^{n} q_i$. Similar to the single provider case, we start from the R-S model.

5.1 R-S model

The provider i's profit function is

$$\pi^i{}_{RS}\big(q^i; q^{-i}\big) = (1 - \gamma)\left(a - q^i - \sum_{j \neq i} q^j\right)q^i - cq^i \tag{15}$$

By solving the Nash equilibrium (NE) with the responsive functions, the unique NE is derived as $q^i{}_{RS}(\gamma; n) = \frac{2}{n+1}\left(\frac{a}{2} - \frac{c}{2(1-\gamma)}\right) = \frac{2}{n+1} q^*{}_{RS}(\gamma)$. We can find that with competition each provider's willingness to operate decreases as the number of providers increases. Additionally, the equilibrium quantity is decreasing in γ. The total providers' profit is

$$\pi_{PS}\big(q^i{}_{RS}(\gamma; n), n\big) = \frac{n(c + a(-1 + \gamma))^2}{(1 + n)^2(1 - \gamma)} \tag{16}$$

Proposition 6: *With n product (service) provider, the provider's profit is decreases in n and γ under R-S.*

From Proposition 6, we know that as competition becomes more cutting-throat, the provider's profit drops due to a diminishing marginal sales revenue.

We can derive the platform's profit

$$\Pi_{RS}\big(q^i{}_{RS}(\gamma; n), n\big) = \frac{n(a(1 - \gamma) - c)\gamma(a + cn - a\gamma)}{(1 + n)^2(1 - \gamma)^2} \tag{17}$$

Proposition 7: *With n product (service) provider, the provider's profit decreases in n under P-S.*

As the product market becomes more competitive, the total output in the market increases dramatically, which leads to a decrease in the market clearance price. Therefore, the total sales revenue decreases. Accordingly, the platform gets less from the revenue cut.

Denote the system profit as:

$$\begin{aligned} W_{RS}\big(q^i{}_{RS}(\gamma; n), n\big) &= \pi_{RS}\big(q^i(\gamma; n), n\big) + \Pi_{RS}\big(q^i(\gamma; n), n\big) \\ &= \frac{n(a(1 - \gamma) - c)(a - a\gamma + c(-1 + \gamma + n\gamma))}{(1 + n)^2(1 - \gamma)^2} \end{aligned} \tag{18}$$

We then analyze the P-S model with n symmetric providers.

5.2 P-S model

We then consider the P-S model, which means the platform will take a profit cut from each provider. In such a case, the provider i's profit function is:

$$\pi^i{}_{PS}(q^i; q^{-i}) = (1 - \gamma)\left(a - c - q^i - \sum_{j \neq i} q^j\right)q^i \qquad (19)$$

By solving the problem above, we can derive the equilibrium quantity as $q^i{}_{PS}(\gamma; n) = \frac{a-c}{n+1} >$ $\frac{2}{n+1}\left(\frac{a}{2} - \frac{c}{2(1-\gamma)}\right) = q^i{}_{RS}(\gamma; n)$. It shows that the equilibrium quantity is independent of the revenue cut parameter γ. The provider under the revenue sharing contract has less incentive to produce. We take the equilibrium quantity into the provider's profit function:

$$\pi^i{}_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) = \frac{(1 - \gamma)n(a - c)^2}{(n + 1)^2} \qquad (20)$$

The platform's profit is

$$\Pi_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) = \frac{\gamma n(a - c)^2}{(n + 1)^2} \qquad (21)$$

Similar to Section 4, we summarize the market players' profit function in Table 2.
We then compare both market agents' profit under R-S and P-S for a given γ.
Denote $\Delta\pi_{PS}(\gamma, n) = \pi_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) - \pi_{RS}\big(q^i{}_{RS}(\gamma; n), n\big)$, we have

$$\Delta\pi_{PS}(\gamma, n) = \frac{cn(c(-2 + \gamma) - 2a(-1 + \gamma))\gamma}{(1 + n)^2(-1 + \gamma)} \qquad (22)$$

We can check that the SOC of $\Delta\pi_{PS}(\gamma, n)$ w.r.t. γ is greater than 0. To make $\Delta\pi_{PS}(\gamma, n) = 0$, we have $\gamma^\pi(n) = 0, \gamma^\pi(n) = \frac{2(a-c)}{2a-c} = 1 - \frac{c}{2a-c}$. Recall that we call $\gamma^\pi(n)$ as the "R-S or P-S" threshold for the provider with n symmetric providers. To avoid triviality, we must have $(1 - \gamma)a > c$. Therefore, $\gamma < 1 - \frac{c}{a} < \frac{c}{2a-c}$. Thus, $\pi_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) \leq \pi_{RS}\big(q^i{}_{RS}(\gamma; n), n\big)$ for $\gamma \in [0, 1 - \frac{c}{a}]$. Note that the "R-S or P-S" threshold has nothing to do with the number of providers. From the provider's perspective, it always holds for $\pi_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) < \pi_{RS}\big(q^i{}_{RS}(\gamma; n), n\big)$ when $\gamma \in [0, 1 - \frac{c}{a}]$.

We then check the R-S or P-S threshold for the platform. Denote

$$\Delta\Pi_{PS}(\gamma, n) = \Pi_{PS}\big(q^i{}_{PS}(\gamma; n), n\big) - \Pi_{RS}\big(q^i{}_{RS}(\gamma; n), n\big).$$

Therefore,

$$\Delta\Pi_{PS}(\gamma, n) = -\frac{cn\big(c(n + (-1 + \gamma)^2) + a(1 + n - 2\gamma)(-1 + \gamma)\big)\gamma}{(1 + n)^2(-1 + \gamma)^2} \qquad (23)$$

$$\frac{\partial^2\Delta\Pi_{PS}(\gamma, n)}{\partial\gamma^2} = -\frac{2cn\big(a(-1 + n)(-1 + \gamma) + cn(2 + \gamma)\big)}{(1 + n)^2(-1 + \gamma)^4} < 0 \qquad (24)$$

$\Delta\Pi_{PS}(\gamma, n)$ is concave in γ. To make $\Delta\pi_{PS}(\gamma, n) = 0$, we have

$\gamma = 0$,

$\gamma^\Pi(n) = \frac{3a - 2c + an - \sqrt{a^2 - 2a^2n + 8acn - 4c^2n + a^2n^2}}{2(2a-c)} = 1 - \frac{\sqrt{a^2 - 2a^2n + 8acn - 4c^2n + a^2n^2} - a(n-1)}{2(2a-c)}$,

$\gamma(n) = 0$.

Table 2 Profits with n symmetric provider

	Provider	Platform	System (Provider + Platform)
R-S	$\dfrac{n(c + a(-1 + \gamma))^2}{(1 + n)^2(1 - \gamma)}$	$\dfrac{n(a(1 - \gamma) - c)\gamma(a + cn - a\gamma)}{(1 + n)^2(1 - \gamma)^2}$	$\dfrac{n(a(1 - \gamma) - c)(a - a\gamma + c(-1 + \gamma + n\gamma))}{(1 + n)^2(1 - \gamma)^2}$
P-S	$\dfrac{(1 - \gamma)n(a - c)^2}{(n + 1)^2}$	$\dfrac{\gamma n(a - c)^2}{(n + 1)^2}$	$\dfrac{n(a - c)^2}{(n + 1)^2}$

$\gamma^{\Pi}(n)$ is called the platform's "R-S or P-S" threshold with n symmetric providers.

$$\frac{\partial^2 \gamma^{\Pi}(n)}{\partial n^2} = -\frac{2(a-c)^2 c}{(a^2(-1+n)^2 + 8acn - 4c^2 n)^{3/2}} < 0 \qquad (25)$$

Proposition 8: *The provider's "R-S or P-S" threshold is independent of the number of product (service) providers in the market n. However, the platform's "R-S or P-S" threshold is concave in n.*

Up until now, we have shown that most of the results remain robust when the number of providers increases. Several numerical studies are illustrated to show the effects.

6. Numerical study

From Fig. 1, the horizontal axis shows the cutting ratio γ, while the vertical axis denotes the profit. The blue line is the provider's profit with P-S, which is a linearly decreasing function in γ. The red line is the provider's profit with R-S, which is convex decreasing in γ. The spread, denoted in the purple line, is a uni-modal function in γ. It starts from the zero point, then decreases, which means the P-S is always better than the R-S for the provider. We then illustrate the platform's profit under different production models.
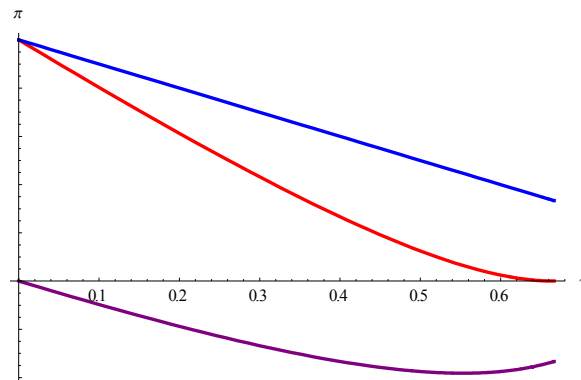

Fig. 1 Provider's profit

Fig. 2 shows the platform's profit as γ varies. The blue line denotes the profit under P-S, which is an increasing function in γ. The red line is the platform's profit under R-S, which is a concave function in γ. The purple line illustrates the spread between the R-S and the P-S. It starts from the zero point, then increases and finally decreases below 0. When γ is larger than a threshold, itis more profitable for the platform to select the P-S model.
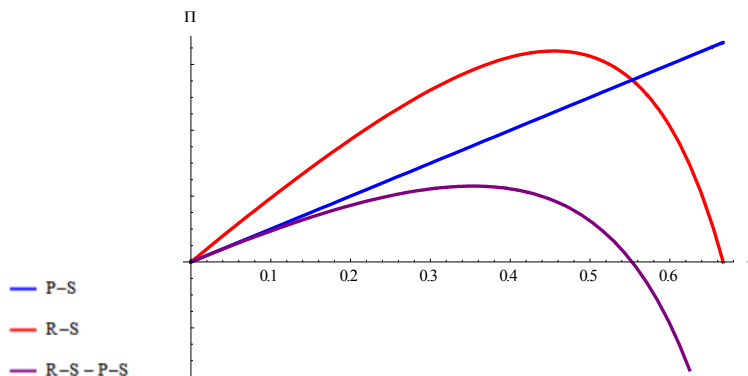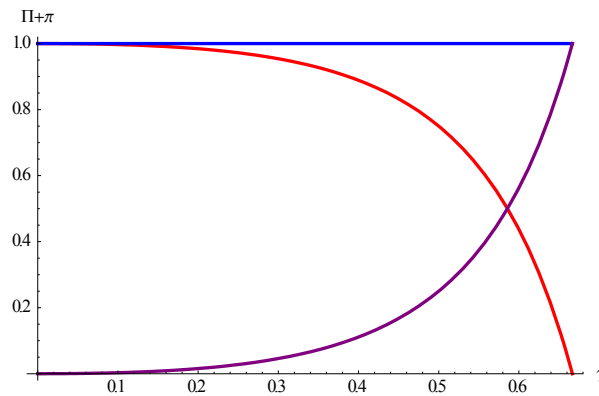

Fig. 2 Platform's profit

Fig. 3 System performance

Fig. 3 shows the system performance as γ varies. The blue line denotes the system performance under P-S, which remains stable as γ increases. The red line is the total profit under R-S, which is a decreasing function in γ. As γ increases, the mismatch between the provider and platform deepens. Therefore, the system performance under R-S is always below that under P-S. The purple line illustrates the spread between the P-S and the R-S, which is increasing as γ increases.

7. Discussion and conclusion

With the emergence of Internet economics, more product (service) providers are selling through e-platforms. In this paper, we compare two distinguishing production models, namely, the revenue sharing (R-S) model and the profit sharing (P-S) model. We find that the providers have less incentive to operate under the R-S model. The system performance decreases as the cutting ratio increases with R-S. Astonishingly, we show that it is not always beneficial for the platform to select the revenue sharing model, under which she seems to take no inventory risks by just taking a sales revenue cut. Additionally, the providers may be better off when the cutting ratio is larger than a threshold. We extend the basic model to a case when multiple providers exist. We find that the providers' "R-S or P-S" threshold remains stable as the number of players increases. However, the platform's "R-S or P-S" threshold is concave in the number of players in the market. Most of the results are proved to be robust.

There are several promising extensions along this theoretical framework. We ignore the randomness in our model. When the market is uncertain, the platform is exposed to more operational risks when she takes the P-S model. How will the results change in such a case? We assume perfect substitution in this paper, does it matter when these providers are selling partial substitute (complimentary) products or service via the same platform? Expanding the basic model into a dynamic setting would be another valuable study that could provide managerial insights for managers.

Acknowledgement

References

[1] Farrell, D., Greig, F. (2017). *The online platform economy: Has growth peaked?,* JP Morgan Chase Institute, New York, USA, doi: 10.2139/ssrn.2911194.

[2] Zervas, G., Proserpio, D., Byers, J.W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry, *Journal of Marketing Research*, Vol. 54, No. 5, 687-705, doi: 10.1509/jmr.15.0204.

[3] Nunes, P.F., Bellin, J., Lee, I., Schunck, O. (2013). Converting the nonstop customer into a loyal customer, *Strategy & Leadership*, Vol. 41, No. 5, 48-53, doi: 10.1108/sl-05-2013-0035.

[4] Dong, L., Ren, L., Zhang, D. (2017) *Financing small and medium-sized sellers via E-commerce platform,* Working paper*, Washington University, St. Louis.

[5] Wang, Y., Jiang, L., Shen, Z.-J. (2004). Channel performance under consignment contract with revenue sharing, *Management Science*, Vol. 50, No. 1, 34-47, doi: 10.1287/mnsc.1030.0168.

[6] Christopher, M., Towill, D. (2001). An integrated model for the design of agile supply chains, *International Journal of Physical Distribution & Logistics Management*, Vol. 31, No. 4, 235-246, doi: 10.1108/09600030110394914.

[7] Dana Jr., J.D., Spier, K.E. (2001). Revenue sharing and vertical control in the video rental industry, *The Journal of Industrial Economics*, Vol. 49, No. 3, 223-245, doi: 10.1111/1467-6451.00147.

[8] Gong, D., Tang, M., Liu, S., Li, Q. (2017). Reconsidering production coordination: A principal-agent theory-based analysis, *Advances in Production Engineering & Management*, Vol. 12, No. 1, 51-61, doi: 10.14743/apem2017.1.239.

[9] Giannoccaro, I., Pontrandolfo, P. (2004). Supply chain coordination by revenue sharing contracts, *International Journal of Production Economics*, Vol. 89, No. 2, 131-139, doi: 10.1016/s0925-5273(03)00047-1.

[10] Gerchak, Y., Wang, Y. (2004). Revenue-sharing vs. wholesale-price contracts in assembly systems with random demand, *Production and Operations Management*, Vol. 13, No. 1, 23-33, doi: 10.1111/j.1937-5956.2004.tb00142.x.

[11] Cachon, G.P., Lariviere, M.A. (2005). Supply chain coordination with revenue-sharing contracts: Strengths and limitations, *Management Science*, Vol. 51, No. 1, 30-44, doi: 10.1287/mnsc.1040.0215.

[12] Cachon, G.P. (2003). Supply chain coordination with contracts, *Handbooks in Operations Research and Management Science*, Vol. 11, No. 6, 227-339, doi: 10.1016/S0927-0507(03)11006-7.

[13] Chen, Y., Gupta, D. (2014). *Trade-finance contracts for small-production suppliers and the effect of third-party financing,* Working paper, University of Minnesota.

[14] Çanakoğlu, E., Bilgiç, T. (2007). Analysis of a two-stage telecommunication supply chain with technology dependent demand, *European Journal of Operational Research*, Vol. 177, No. 2, 995-1012, doi: 10.1016/j.ejor.2006.01.006.

[15] Wei, Y., Choi, T.-M. (2010). Mean-variance analysis of supply chains under wholesale pricing and profit sharing schemes, *European Journal of Operational Research*, Vol. 204, No. 2, 255-262, doi: 10.1016/j.ejor.2009.10.016.

[16] Singh, N., Vives, X. (1984). Price and quantity competition in a differentiated duopoly, *The RAND Journal of Economics*, Vol. 15, No. 4, 546-554, from *http://www.jstor.org/stable/2555525*, accessed February 19, 2018.

[17] Kreps, D.M., Scheinkman, J.A. (1983). Quantity precommitment and bertrand competition yield cournot outcomes, *The Bell Journal of Economics*, Vol. 14, No. 2, 326-337, doi: 10.2307/3003636.

[18] Gong, D., Liu, S., Lu, X. (2015). Modelling the impacts of resource sharing on supply chain efficiency, *International Journal of Simulation Modelling*, Vol. 14, No. 4, 744-755, doi: 10.2507/IJSIMM14(4)CO20.

[19] Van Mieghem, J.A., Dada, M. (1999). Price versus production postponement: Capacity and competition, *Management Science*, Vol. 45, No. 12, 1639-1649, doi: 10.1287/mnsc.45.12.1631.

[20] Goyal, M., Netessine, S. (2007). Strategic technology choice and capacity investment under demand uncertainty, *Management Science*, Vol. 53, No. 2, 192-207, doi: 10.1287/mnsc.1060.0611.

[21] Anupindi, R., Jiang, L. (2008). Capacity investment under postponement strategies, market competition, and demand uncertainty, *Management Science*, Vol. 54, No. 11, 1876-1890, doi: 10.1287/mnsc.1080.0940.

[22] Wen, F., Xiao, J., Huang, C., Xia, X. (2018). Interaction between oil and US dollar exchange rate: Nonlinear causality, time-varying influence and structural breaks in volatility, *Applied Economics*, Vol. 50, No. 3, 319-334, doi: 10.1080/00036846.2017.1321838.

[23] Wen, F., Gong, X., Cai, S. (2016). Forecasting the volatility of crude oil futures using HAR-type models with structural breaks, *Energy Economics*, Vol. 59, 400-413, doi: 10.1016/j.eneco.2016.07.014.

Appendix A

Proof of Proposition 1:

$$\frac{\partial \pi_{RS}(\gamma)}{\partial \gamma} = -\frac{(a(1-\gamma)-c)(a+c-a\gamma)}{4(1-\gamma)^2} \leq 0$$

Proof of Proposition 2:

$$\frac{\partial \Pi_{RS}(q^*_{RS}(\gamma))}{\partial \gamma} = \frac{a^2(-1+\gamma)^3 + c^2(1+\gamma)}{4(-1+\gamma)^3} = \frac{a^2}{4} - \frac{c^2(1+\gamma)}{4(1-\gamma)^3}, \frac{\partial^2 \Pi_{RS}(q^*_{RS}(\gamma))}{\partial \gamma^2} = -\frac{c^2(2+\gamma)}{2(1-\gamma)^4} < 0. \text{ By solving the FOC,}$$

we have $\gamma^* = 1 - \dfrac{c^2}{3^{1/3}\left(-9a^4c^2 + \sqrt{3}\sqrt{27a^8c^4 + a^6c^6}\right)^{1/3}} + \dfrac{\left(-9a^4c^2 + \sqrt{3}\sqrt{27a^8c^4 + a^6c^6}\right)^{1/3}}{3^{2/3}a^2}$

Proof of Proposition 3:

$$\frac{\partial W_{RS}(q^*_{RS}(\gamma))}{\partial \gamma} = \frac{c^2\gamma}{2(1-\gamma)^3} < 0$$

Proof of Proposition 5:

$$\frac{c^2(1-2\gamma)+a^2(1-\gamma)^2-2ac(1-\gamma)^2}{4(1-\gamma)^2} - \frac{(a-c)^2}{4} = -\frac{c^2\gamma^2}{4(1-\gamma)^2} < 0$$

Proof of Proposition 6:

$$\frac{\partial \pi_{PS}\left(q^i{}_{RS}(\gamma;n),n\right)}{\partial n} = -\frac{(-1+n)(c+a(-1+\gamma))^2}{(1+n)^3(1-\gamma)} < 0$$

$$\frac{\partial \pi_{PS}\left(q^i{}_{RS}(\gamma;n),n\right)}{\partial \gamma} = \frac{n(c+a(-1+\gamma))(a+c-a\gamma)}{(1+n)^2(1-\gamma)^2}$$

$$\frac{\partial^2 \pi_{PS}\left(q^i{}_{RS}(\gamma;n),n\right)}{\partial \gamma^2} = \frac{2c^2n}{(1+n)^2(1-\gamma)^3} > 0$$

Proof of Proposition 7:

$$\frac{\partial \Pi_{RS}\left(q^i{}_{RS}(\gamma;n),n\right)}{\partial n} = \frac{(a(1-\gamma)-c)(2cn+a(-1+n)(-1+\gamma))\gamma}{(1+n)^3(-1+\gamma)^2} < 0$$

A bi-objective environmental-economic optimisation of hot-rolled steel coils supply chain: A case study in Thailand

Somboonwiwat, T.[a,*], Khompatraporn, C.[a], Miengarrom, T.[a], Lerdluechachai, K.[a]

[a]Department of Production Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi (KMUTT), Bangkok, Thailand

A B S T R A C T

Steel production is an energy intensive industry, emitting a considerable amount of CO_2 which contributes to global warming. Many sources of energy may be used in steel production, incurring different costs. This research studies the effects of different decisions in the supply chain network for the production of hot-rolled steel coils (HRSC) in Thailand. The objectives are to minimise the total cost of HRSC production as well as to minimise the total CO_2 emissions in order to reduce environmental impact. Towards meeting these two objectives, a mathematical model is proposed to simultaneously determine the choices of energy, raw materials, and transportation modes with regard to production, network, and business constraints. Via examination of the price differences for available raw materials and energy sources several scenarios are investigated to evaluate their impact on both environmental and economic requirements of the supply chain. The analysis shows that the optimal solutions are greatly affected by changes in the prices of slabs and scrap, and the cost of electricity, whereas fuel oil and natural gas prices only affect the choice of fuel for the pre-heating process of the slabs. Strategies to operate under different scenarios are also discussed.

1. Introduction

Climate change induced by global warming and carbon emissions has been a growing concern in recent decades. Steel manufacturing is one of the most energy intensive industries, and is responsible for 25 % of industrial carbon emissions, or about 9 % of the carbon emissions from energy and processes [1]. The World Steel Association [2] collected crude steel production data from 66 countries and recorded that the annual production in 2015 reached 1,621 million metric tonnes, accounting for approximately 98 % of the total world crude steel production. In order to tackle climate change and reduce carbon emissions caused by the iron and steel industry, the International Iron and Steel Institute (IISI) submitted proposals to governments in many countries to seek cooperation in order to develop and create new methods of production [3]. New technologies alone, however, are not sufficient to effectively reduce carbon emissions, and there is a need to develop suitable management decision tools to facilitate the carbon reduction mission.

There are several studies addressing energy consumption, environmental impacts, and cost reduction in the steel industry in many countries across the globe. For instance, Geilen and Moriguchi [4] developed a mathematical model in the context of the Japanese iron and steel industry, and investigated the impact of CO_2 taxes on technology selection and other factors. Ruth and Amato [5] investigated the implications of changes in the cost of carbon on energy usage and

carbon emission. They indicated that in the case of the US iron and steel industry, emissions could be reduced by accelerating the shift to electric arc furnaces or by increases in the penalty costs of carbon emissions by perhaps technology-led policies. Zhang and Wang [6] identified that energy in iron and steel manufacturing could be saved by adopting certain production techniques such as pulverised coal injection and continuous casting, and observed that large Chinese steel manufacturers tended to be more energy efficient than the smaller ones. Soheili [7] estimated that by improving production technology the oil consumption in the Iran iron and steel industry could be reduced by about 45.8 million barrels. Tongpool *et al.* [8] studied environmental impacts of various steel product forms such as slab, hot-rolled, cold-rolled, and galvanised steel in Thailand. They recommended that energy consumption could be reduced by the use of more steel scrap. Johansson and Soderstrom [9] investigated energy efficiency measures and fuel conversion options in order to reduce CO_2 emissions in the Swedish iron and steel industry. It must be noted that the above studies tend to be based on iron and steel production processes in individual plants in particular countries, rather than on considerations of the supply chain as a whole, and as such the studies have not taken the effects of transportation into account.

Modern management forces firms to integrate transportation planning in their management decisions in order to reduce costs and provide improved service to customers. Simultaneous integration of production and transportation in a supply chain has gained considerable interest in various models. Lee and Kim [10] extended the concept of Byrne and Bakir [11] by combining analytical and simulation models to solve production-distribution planning problems that involved multi-products and multi-periods in supply chains. Zamarripa *et al.* [12] considered two different supply chains, each consisting of three stages, to simultaneously minimise system costs and accumulated delivery times from different production echelons to the storage centres. They used a mixed integer linear programming model to optimise the supply chain planning problem. Dehghanbaghi and Sajadieh [13] studied complementary products and jointly optimised their production, inventory, and transportation by assuming certain convexity properties. Gholamian and Heydari [14] integrated transportation operations and routing with location decision and inventory control to minimise the overall cost. Koc *et al.* [15] coordinated inbound and outbound transportation with production schedules. Their study assumed that the vehicles for both inbound and outbound transportation could be jointly utilised.

Even though steel is being increasingly substituted by lighter materials in automobiles, it still accounts for about half of the vehicle weight. Steel sheets produced from hot-rolled coils (HRSC) are perhaps the highest value added item among steel products and are extensively used in the automotive industry [16]. Since the whole steel supply chain encompasses a spectrum of steel manufacturers and products, this research is focused on the HRSC portion of the supply chain, by using the case of Thailand. Thailand is the second largest economy in the Southeast Asia region (based on the 10 member states of ASEAN), and a leading car manufacturing country in the region. In 2016, the country was ranked 12th in the world in total automotive production with specialisation in light trucks and passenger cars [17]. It is also the third largest steel producer in the ASEAN region with annual production of 10 million metric tonnes [18].

The organisation of the paper is as follows. The next section gives the background of the HRSC supply chain problem in more detail. Section 3 describes the mathematical model of the problem by which the energy cost and CO_2 emission are to be minimised. The energy cost in the model considers both the production and transportation costs. Section 4 contains numerical data of the HRSC supply chain case study in Thailand. The results and operational strategies are then discussed in Section 5. Finally, Section 6 concludes the observations of the study.

2. Background of hot-rolled steel coils supply chain in Thailand

The steel supply chain starts from iron ore smelting factories as the upstream part of the supply chain. The products from the smelting factories are called pig iron, crude iron or crude steel (depending upon the shape and chemical composition). Refined liquid steel is cast into billets, blooms, and slabs to suit the production in the next part of the chain. The midstream steel mills transform these primary steel products into beams, rods, or coils, known as semi-finished prod-

ucts, through various methods such as hot rolling and cold rolling. The semi-finished products are later processed into the final products in various industries in the downstream of the supply chain [19].

Thailand is the second largest economy among the 10 member states of ASEAN, behind only Indonesia. The two countries together account for over half of the GDP of the ASEAN region, a contribution of over 1.25 trillion USD annually [20]. They are also leading automotive manufacturers of the region, and are major users of steel products. Nonetheless, the steel supply chain in these countries differ as there are iron ore smelting factories in Indonesia but none in Thailand. Hot-rolled steel coils (HRSC) make up over 30 % of the semi-finished steel products in Thailand as they are a versatile material that can be used in a wide–range of economically important industries [21].

For the HRSC supply chain in Thailand, there are two steel mills that can produce suitable steel slabs. The slabs are processed by five HRSC factories, two of which are also slab producers. The study also considers 13 customers who use the HRSC as the raw material for their production routes. These customers manufacture steel products such as cold-rolled steel pipes and cold formed structural steel sections. The pig iron used in primary steel production for HRSC can be imported through four ports. The mode of transportation from the ports is by road, with additional marine transportation to a customer from a HRSC producer.

The two steel mills can also use steel scrap as raw material for primary steel production cast as slabs and can then send slabs directly through conveyers for integrated HRSC production. Therefore, they require only internal transportation between the slab and HRSC production processes. Imported steel scrap arrives at sea-ports. Domestic scrap is collected and the suppliers also often transport the scrap by marine transportation to save cost. Thus, it is assumed that steel scrap is obtained through ports. Some HRSC are directly imported from overseas and sold through the HRSC factories. Fig. 1 summarises the HRSC supply chain network in Thailand.
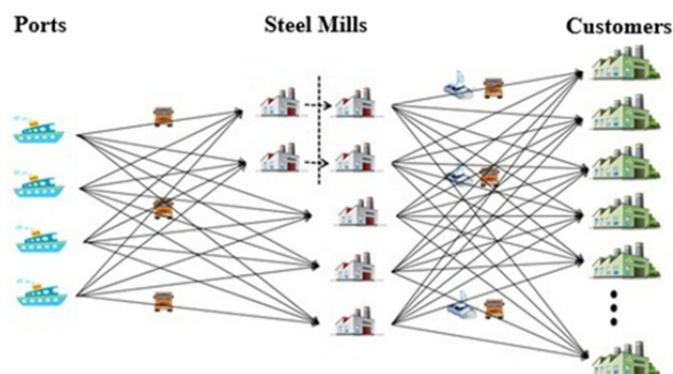
Fig. 1 The structure of the HRSC supply chain network in Thailand

To minimise the total costs in the HRSC supply chain, there are three sets of decision variables involved. The first set indicates the amounts of pig iron, steel scrap, and slabs to be acquired through the ports, and the proportions to be sent to the different HRSC mills. The second set determines the suitable production at each of the HRSC mills. The third set represents the amounts of HRSC to be transported to the customers via the two transportation modes. The values of these decision variables influence the total cost as well as the total CO_2 emissions in the supply chain. The objectives are to minimise both the total costs and the CO_2 emissions.

HRSC production may use fuel oil or natural gas as the energy source in the pre-heating process. Some of the mills in the supply chain do not locate along the national natural gas pipeline, so these mills do not have access to natural gas and can only use fuel oil in their production. The HRSC supply chain situation can be formulated as a bi-objective optimisation problem as shown in the next section.

3. Mathematical formulation

The HRSC supply chain optimisation problem with total costs and CO_2 emission minimisation objectives can be formulated as follows.

3.1 Notation

The notation of the mathematical programme is below:

Indices

i	fuel input i = 1,2,3 (1 = electricity, 2 = fuel oil, 3 = natural gas)
j	product j = 1,2,3,4 (1= pig iron, 2 = steel scrap, 3 = steel slabs, 4 = HRSC)
k	steel mill k = 1,2,…,5
l	customer l = 1,2,…,13
m	mode of transportation m = 1,2 (1 = road, 2 = marine)
n	source of raw material n = 1,2 (1 = domestic, 2 = imported)
p	port p = 1,2,3,4
q	% load

Decision variables

X_{ijk}	production volume by fuel input i to produce product j at steel mill k
Y_{jknp}	purchased volume of product j for steel mill k from source of raw material n and from port p
HRC_{klm}	volume of HRSC from steel mill k to customer l by transportation mode m

Costs

PC_{ijk}	unit production cost by fuel input i to produce product j at steel mill k
MC_{jknp}	unit material cost of product j for steel mill k from source of raw material n from port p
MTC_{jknp}	unit material transportation cost of product j for steel mill k from source of raw material n from port p
$FGTC_{klm}$	unit HRSC transportation cost from steel mill k to customer l by transportation mode m

Demand and capacities

D_l	demand volume of customer l
$CAPF_{jk}$	capacity of production of product j by steel mill k
$CAPRM_{jn}$	maximum amount of product j that can be purchased from source n
$CAPT_m$	capacity of transportation mode m

Emission factors

PE_{ijk}	emission factor of production process by fuel input i to produce product j at steel mill k
MTE_{jknpq}	emission factor of material transportation of product j to steel mill k from source of raw material n and from port p with % load q
$FGTE_{klmq}$	emission factor of HRSC transportation from steel mill k to customer l with % load q

Loss and raw materials

α	proportion of pig iron used to produce slabs
β	loss percentage of raw material in the production process
RM	amount of raw material

3.2 Constraints

Demand and supply

The total production and imported HRSC must meet the total demand of the customers.

$$\sum_{k=1}^{5} \sum_{i=1}^{3} X_{i4k} \geq \sum_{l=1}^{13} D_l \tag{1}$$

The HRSC delivered from the steel mills must meet the demand of each customer.

$$\sum_{m=1}^{2} \sum_{k=1}^{5} HRC_{klm} \geq D_l, \qquad \forall l \tag{2}$$

The total amount of HRSC produced and imported must at least equal to the total amount of HRSC delivered to the customers.

$$\sum_{k=1}^{5} \sum_{i=1}^{3} X_{i4k} \geq \sum_{m=1}^{2} \sum_{l=1}^{13} \sum_{k=1}^{5} HRC_{klm} \tag{3}$$

The amount of HRSC produced and imported by each HRSC mill must at least equal to the total amount sent to the customers.

$$\sum_{i=1}^{3} X_{i4k} + \sum_{p=1}^{4} Y_{4k2p} \geq \sum_{m=1}^{2} \sum_{l=1}^{13} HRC_{klm}, \qquad \forall k \tag{4}$$

There is no HRSC received from domestic sources.

$$Y_{4k1p} = 0, \qquad \forall k, p \tag{5}$$

The total amount of slabs produced and/or purchased as the raw material by each steel mill must be at least equal to the total amount of HRSC.

$$\sum_{i=1}^{3} X_{i3k} + \sum_{p=1}^{4} Y_{3k2p} \geq \sum_{i=1}^{3} X_{14k}, \qquad \forall k \tag{6}$$

For HRSC mills 3, 4 and 5, there is no HRSC received from domestic sources.

$$Y_{4k1p} = 0, \qquad \text{for } k = 3,4,5 \text{ and } \forall p \tag{7}$$

Capacity and utilization

The total production of each product at each HRSC mill must not exceed its capacity.

$$\sum_{i=1}^{3} X_{ijk} \leq CPAF_{jk}, \qquad \forall j, k \tag{8}$$

The HRSC mills 3, 4, and 5 cannot produce slabs.

$$\sum_{k=3}^{5} \sum_{i=1}^{3} X_{i3k} = 0 \tag{9}$$

Every HRSC mill cannot produce pig iron or scrap.

$$\sum_{j=1}^{2} \sum_{i=1}^{3} X_{ijk} = 0, \qquad \forall k \tag{10}$$

Material balance

The production of slabs in the mills 1 and 2 must equal the production of HRSC.

$$\sum_{i=1}^{3} X_{i3k} = \sum_{i=1}^{3} X_{i4k}, \qquad \text{for } k = 1,2 \tag{11}$$

A portion of the raw material for slabs is from pig iron.

$$\sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} Y_{1knp} = \alpha RM \tag{12}$$

The rest of the raw material for slabs is from scrap.

$$\sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} Y_{2knp} = (1-\alpha)RM \tag{13}$$

The amount of raw material must accommodate the loss in the slab production process.

$$RM = \beta \sum_{k=1}^{5} \sum_{i=1}^{3} X_{i3k} \tag{14}$$

The scrap purchased must not exceed the amount of available scrap.

$$\sum_{p=1}^{4} \sum_{k=1}^{5} Y_{jknp} \leq CAPRM_{jn} , \qquad \forall j,n \tag{15}$$

Business structure

The following constraints are based on the business structure of the HRSC supply chain network in Thailand in which there are certain subsidiary companies within the chain. As a result, certain mills prefer to serve some customers more than the others.

The total transportation from HRSC mills 1 and 2 must meet the demand of customer 6.

$$\sum_{m=1}^{2} \sum_{k=1}^{2} HRC_{k6m} + \sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{2} Y_{4knp} \geq D_{6} \tag{16}$$

The total HRSC transported from mill 3 must meet the demand of customer 1.

$$\sum_{m=1}^{2} HRC_{31m} + \sum_{p=1}^{4} \sum_{n=1}^{2} Y_{43np} \geq D_{1} \tag{17}$$

The total HRSC transported from mill 4 must satisfy the demand of customer 12.

$$\sum_{m=1}^{2} HRC_{412m} + \sum_{p=1}^{4} \sum_{n=1}^{2} Y_{412np} \geq D_{12} \tag{18}$$

The total HRSC transported from mill 5 must meet the demand of customer 11.

$$\sum_{m=1}^{2} HRC_{511m} + \sum_{p=1}^{4} \sum_{n=1}^{2} Y_{411np} \geq D_{11} \tag{19}$$

The HRSC mills 1, 2, 4, and 5 cannot access marine transportation.

$$HRC_{kl2} = 0 , \qquad \text{for } k = 1,2,4,5 \text{ and } \forall l \tag{20}$$

Decision variables

All of the decision variables are non-negative.

$$X_{ijk} \geq 0 , \qquad \forall i,j,k \tag{21}$$

$$Y_{jknp} \geq 0 , \qquad \forall j,k,n,p \tag{22}$$

$$HRC_{klm} \geq 0 , \qquad \forall k,l,m \tag{23}$$

3.3 Objectives

Two objectives as in this formulation.

z_1 : Cost Optimisation (CO). This objective is to minimise the total costs which consist of production cost (TPC), raw material cost (TMC), and transportation cost (TTC).

Min
$$z_1 = TPC + TMC + TTC \tag{24}$$

z_2 : Emission Optimisation (EO). This objective is to minimise the total CO_2 emission which includes emission from production process (TPE) and transportation (TTE).

Min
$$z_2 = TPE + TTE \tag{25}$$

Total costs

The total production cost is determined by the unit production cost multiplied by the production volume.

$$TPC = \sum_{k=1}^{5} \sum_{j=1}^{5} \sum_{i=1}^{3} C_{ijk} X_{ijk} \tag{26}$$

The material cost is calculated from the unit material cost multiplied by the total amount of material acquired.

$$TMC = \sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} \sum_{j=1}^{5} MC_{jknp} Y_{jknp} \tag{27}$$

The transportation cost includes the raw material transportation cost from the ports to the mills, and the HRSC transportation cost from the mills to the customers. The raw material transportation cost is determined by the unit material transportation cost multiplied by the number of trips from the ports. Similarly, the HRSC transportation cost is given by the unit HRSC transportation cost multiplied by the number of trips that the HRSC is transported. The notation [•] represents rounding up of the number inside the symbol.

$$TTC = \sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} \sum_{j=1}^{5} MTC_{jknp} \left\lceil \frac{Y_{jknp}}{CAPT_m} \right\rceil + \sum_{m=1}^{2} \sum_{l=1}^{13} \sum_{k=1}^{5} FGTC_{klm} \left\lceil \frac{HRC_{klm}}{CAPT_m} \right\rceil \tag{28}$$

Total CO_2 emission

The total emission is determined by the unit emission per tonne of production multiplied by the production volume.

$$TPE = \sum_{k=1}^{5} \sum_{j=1}^{5} \sum_{i=1}^{3} PE_{ijk} X_{ijk} \tag{29}$$

The emission from transportation consists of the emissions from raw material and HRSC transportation. The raw material transportation emission comprises of the emissions of full-load and no-load of the trucks. Similarly, the HRSC transportation emission includes the full-load transportation emission calculated from the unit emission of full-load HRSC transportation per tonne multiplied by the total HRSC volume, and no-load transportation emission determined by the unit emission of no-load HRSC transportation multiplied by the total number of no-load trips.

$$
\begin{aligned}
TTE = & \left(\sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} \sum_{j=1}^{4} MTE_{jknp1} Y_{jknp} + \sum_{p=1}^{4} \sum_{n=1}^{2} \sum_{k=1}^{5} \sum_{j=1}^{4} MTE_{jknp2} \left\lceil \frac{Y_{jknp}}{CAPT_m} \right\rceil \right) \\
& + \left(\sum_{m=1}^{2} \sum_{l=1}^{13} \sum_{k=1}^{5} HRC_{klm} FGTE_{klm1} + \sum_{m=1}^{2} \sum_{l=1}^{13} \sum_{k=1}^{5} FGTE_{klm2} \left\lceil \frac{HRC_{klm}}{CAPT_m} \right\rceil \right)
\end{aligned}
\tag{30}
$$

4. A case study

The steel mills are indicated as P1-P5. Their production capacities are shown in Table 1. The demands of the customers, D1-D13, are given in Table 2. The unit production costs based on the energy source in different steel mills are displayed in Table 3 (THB refers to Thai Baht which is Thailand's currency). The unit production costs of slabs in the steel mills 3, 4, and 5 are not

shown in the table since these mills cannot produce slabs. Table 4 presents the unit material costs and availability of different materials.

Table 5 presents the unit transportation cost from the ports to the mills by trucks, each with 40 tonnes capacity. The unit transportation costs by road and marine of HRSC are shown in Table 6. Trucks of the same 40 tonnes capacity are also used in HRSC transportation. Only P3 has access to the marine transportation mode with 8,000 tonnes capacity per trip by barges. However, the barge transportation cannot be applied to customer 1.

Table 1 Production capacities of the steel mills (in 1,000 tonnes)

Product	Plant				
	P1	P2	P3	P4	P5
Slab	3,000	1,800	0	0	0
HRSC	3,000	1,800	4,000	1,000	500

Table 2 HRSC demands of the customers (in 1,000 tonnes)

Customer	D1	D2	D3	D4	D5	D6	D7
HRSC	720	240	600	150	180	336	60
Customer	D8	D9	D10	D11	D12	D13	
HRSC	180	176.4	104.4	3,900	54	180	

Table 3 Unit production costs by energy source and product (THB/tonne)

Energy	Plant				
	P1	P2	P3	P4	P5
Electricity – Slab	1,200	1,314	-	-	-
Electricity – HRC	1,104	930	408	540	474
Fuel oil – HRC	464	487	876	914	895
Natural gas – HRC	320	336	2,183	1,108	1,330

Table 4 Unit material costs and availability of raw materials

Cost/Availability	Product				
	Imported pig iron	Domestic scrap	Imported scrap	Imported slab	Imported HRC
Price (THB/tonne)	16,591	12,276	14,731	18,228	35,000
Capacity (tonnes)	560,063	2,478,273	12,599,807	452,9038	23,534,220

Table 5 The unit material transportation cost from ports to steel mills (THB/trip)

Port	Plant				
	P1	P2	P3	P4	P5
Port1	4,011	1,995	37,170	7,980	11,970
Port2	1,799	4,543	33,460	4,669	8,330
Port3	34,020	37,170	469	29,750	25,410
Port4	8,540	11,760	26,110	4,333	1,225

Table 6 The unit HRSC transportation cost by truck and barge (THB/trip)

Customer	Plant					
	P1	P2	P3	P3 (Barge)	P4	P5
D1	34,090	37,310	147	-	29,890	25,550
D2	3,948	1,806	36,750	17,000	7,770	11,830
D3	3,941	1,925	37,100	17,000	7,910	11,900
D4	1,316	2,611	34,930	14,500	5,719	9,730
D5	2,303	1,561	35,490	17,000	6,258	10,290
D6	10,010	13,230	24,780	13,500	5,950	1,547
D7	8,400	11,620	26,110	13,500	4,193	903
D8	10,360	13,510	26,950	13,500	6,111	1,540
D9	9,800	13,020	26,390	13,500	5,586	1,015
D10	8,890	12,110	25,970	13,500	4,669	791
D11	9,240	12,390	25,830	13,500	4,998	427
D12	8,890	12,110	25,340	13,500	4,746	427
D13	1,316	2,611	34,720	14,500	5,747	9,800

The unit production emission factors are presented in Table 7. The CO_2 emission in the production process is calculated from the energy used in the melting and pre-heating processes. Tables 8 to 12 show unit material and HRSC transportation emission factors, for both full-load and no-load of the two transportation modes.

Table 7 Unit production emission factors ($kgCO_2$/tonne)

Energy	Plant				
	P1	P2	P3	P4	P5
Elec – Slab	224.40	245.72	0	0	0
Elec – HRC	206.45	173.91	76.30	100.98	88.64
Fuel – HRC	65.37	68.62	123.32	128.74	126.03
NG – HRC	47.40	49.76	89.42	93.35	91.39

Table 8 Unit material transportation emission factors by full-load truck ($kgCO_2$/tonne)

Port	Plant				
	P1	P2	P3	P4	P5
Port1	2.30	1.14	21.29	4.57	6.86
Port2	1.03	2.60	19.17	2.67	4.77
Port3	19.49	21.29	0.27	17.04	14.56
Port4	4.89	6.74	14.96	2.48	0.70

Table 9 Unit material transportation emission factors by no-load truck ($kgCO_2$/trip)

Port	Plant				
	P1	P2	P3	P4	P5
Port1	44.72	22.24	414.45	88.98	133.47
Port2	20.06	50.65	373.08	52.06	92.88
Port3	379.32	414.45	5.23	331.71	283.32
Port4	95.22	131.12	291.13	48.31	13.66

Table 10 Unit HRSC transportation emission factors by full-load truck ($kgCO_2$/tonne)

Customer	Plant				
	P1	P2	P3	P4	P5
D1	19.53	21.37	0.08	17.12	14.64
D2	2.26	1.03	21.05	4.45	6.78
D3	2.26	1.10	21.25	4.53	6.82
D4	0.75	1.50	20.01	3.28	5.57
D5	1.32	0.89	20.33	3.58	5.89
D6	5.73	7.58	14.20	3.41	0.89
D7	4.81	6.66	14.96	2.40	0.52
D8	5.93	7.74	15.44	3.50	0.88
D9	5.61	7.46	15.12	3.20	0.58
D10	5.09	6.94	14.88	2.67	0.45
D11	5.29	7.10	14.8	2.86	0.24
D12	5.09	6.94	14.52	2.72	0.24
D13	0.75	1.50	19.89	3.29	5.61

Table 11 Unit HRSC transportation emission factors by no-load truck ($kgCO_2$/trip)

Customer	Plant				
	P1	P2	P3	P4	P5
D1	380.10	416.01	1.64	333.27	284.88
D2	44.02	20.14	409.76	86.64	131.90
D3	43.94	21.46	413.67	88.20	132.69
D4	14.67	29.11	389.47	63.77	108.49
D5	25.68	17.41	395.71	69.78	114.73
D6	111.61	147.51	276.30	66.34	17.25
D7	93.66	129.56	291.13	46.75	10.07
D8	115.51	150.64	300.49	68.14	17.17
D9	109.27	145.17	294.25	62.28	11.32
D10	99.12	135.03	289.57	52.06	8.82
D11	103.03	138.15	288.00	55.73	4.76
D12	99.12	135.03	282.54	52.92	4.76
D13	14.67	29.11	387.13	64.08	109.27

Table 12 Unit HRSC transportation emission factors of marine transportation of P3 to customers (kgCO$_2$/trip)

Customer	D1	D2	D3	D4	D5	D6	D7
Barge (full-load)	-	0.43	0.43	0.38	0.43	0.27	0.27
Barge (no-load)	-	0.43	0.43	0.38	0.43	0.27	0.27
Customer	D8	D9	D10	D11	D12	D13	
Barge (full-load)	0.27	0.27	0.27	0.27	0.27	0.38	
Barge (no-load)	0.27	0.27	0.27	0.27	0.27	0.38	

5. Results, discussion and operational strategies

5.1 Numerical results

Using the data in Section 4 and Excel Premium Solver to obtain the optimal solutions, the results show that the total cost obtained based on the cost optimisation (CO) model is lower than for that based on emission optimisation (EO) by approximately THB 9.56 million or 16.08 %. Table 13 shows these results in more detail. On the other hand, the EO model emphasises reduction of the overall CO$_2$ emission of the supply chain. Therefore, the EO model yields lower total CO$_2$ emission than for that based on the CO model by about 620.63 thousand tonnes of CO$_2$ or 47.82 %. It is important to note that the solution from the EO model yields higher slab usage for which (not shown here) a large part must be imported. The EO model also suggests substantially lower scrap usage to reduce the energy (which leads to CO$_2$ emission) required for slab production.

The cost saved by the CO model is mainly from processing scrap to produce slabs because the price of imported slabs is higher than that of domestic slabs produced by P1 and P2. However, the cost saving incurs higher energy usage of electricity and fuel oil in comparison to the EO model. Electricity is consumed mainly in P1 and P2 to produce slabs. Moreover, among all the production plants P3 is the most active one using the highest amount of imported slabs. It should be noted that P3 utilises mainly fuel oil, and this leads to a sharp increase (46 %) in fuel oil usage in the solution to the EO model.

These results imply that there is a price to pay to obtain lower CO$_2$ emission. In comparing the results based on the CO and EO models, the reduction of about 620.63 thousand tonnes of CO$_2$ emission comes with the disadvantage of increased total cost of over 9.56 billion THB, or approximately 15.40 THB/kgCO$_2$. This cost per kgCO$_2$ may be used as a benchmark for the government when introducing policies to encourage steel mills to reduce their emissions. For example, CO$_2$ emission may be reduced by simply using cleaner energy sources such as natural gas.

The government may choose to subsidise the cost of natural gas to lower its price and encourage steel mills to use more natural gas for production. On the other hand, the results from the EO model involve greater use of imported slabs. To prevent excessive use of imported slabs and the weakening of the competitiveness of domestic slab manufacturers, the government may temporarily impose a higher import tax rate on slabs sourced overseas as and when required.

The production of HRSC involves costs and CO$_2$ emissions and their distribution (in percentage) is as shown in Table 14. From the table, the majority of the cost is located at P1 and P2 where scrap is processed to make slabs and then HRSC. In addition, these two mills also have access to the national natural gas pipeline. Natural gas is a cheaper source of energy than fuel oil. Hence, the production of HRSC is allocated to these two mills in order to reduce the overall cost

Table 13 The total cost, total emission, proportion of energy usage and material usage for the cost and emission optimisation models

Opt. model	Total cost (Millions THB)	Total emission (Millions kgCO$_2$)	Energy usage (%)			Material usage (%)		
			Melting process	Pre-heating process		Pig Iron	Scrap	Slab
			Electricity	Fuel Oil	Natural Gas			
Cost (A)	59,420.27	1,297.80	64	18	18	7	59	34
Emission (B)	68,975.51	677.18	32	0	68	1	9	90
(A)-(B)	9-,555.24	+620.63	+32	+18	-50	+6	+50	-54

of the supply chain. The cost incurred at P3, the third highest cost percentage, is spent on processing imported slabs to make HRSC. Since P3 can access marine transportation, which results in lower cost than trucks to the customers, it is highly utilised compared to P4 and P5. The results also show that P5 is preferred over P4 to process slabs, as indicated by the higher cost percentage. This is perhaps due to the proximity of customers, not already served by P1 to P3, being closer to P5 than P4.

Table 14 also shows that the production of HRSC in P1 and P2 is greatly reduced due to the import of slabs as already mentioned. Moreover, P3 imports most of the slabs and processes them to HRSC in this model because the model tries to utilise the cheaper marine transportation to the customers. P5 is also preferred over P4 in this model because close proximity between P5 to the customers leads to lower CO_2 emission.

Table 14 Percentage distribution of HRSC production

Model	Plant				
	P1	P2	P3	P4	P5
CO (%)	25	40	21	2	12
EO (%)	0	10	77	2	12

5.2 Sensitivity analysis and operational strategies

The prices of different energy sources and raw materials are key in the decisions to be made in the CO and EO models. Using the data from the past five years, the prices of the energy sources and raw materials change with ± 35 % range. Three energy sources, electricity, fuel oil and natural gas, as well as two raw materials, steel scrap and slabs, are tested for +35 % (High or H) and −35 % (Low or L) change in prices from the nominal values used in Section 5.1. The pig iron price is not examined here because its price often changes similarly with the price of the slabs.

A notation HHHHH represents the high level of all prices of electricity, fuel oil, natural gas, scrap, and slabs, respectively, whereas LLLLL denotes the low level in all five factors. Since there are five factors, each with two levels (high and low), there is a total of 32 scenarios to be tested. The results are summarised in Fig. 2. Table 15 highlights the total costs and total emission of the HHHHH, LLLLL, HHLLL, and LLLHL scenarios. The scenario of HHHHH yields the highest total cost as expected since prices of all the factors are at the high level (or most expensive). Likewise, the LLLLL scenario results in the lowest anticipated total costs. Fig. 2 also shows two interesting scenarios which are HHLLL and LLLHL. They are both trade-off points between the total costs and the total emission.
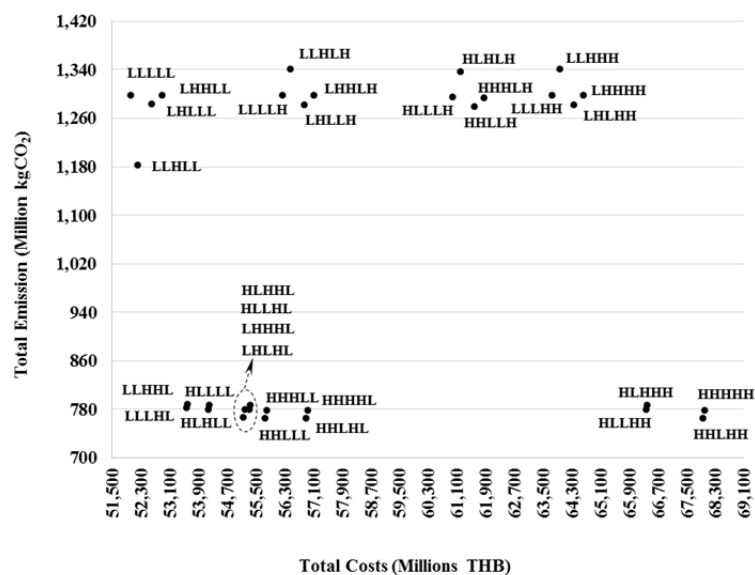

Fig. 2 Plot of total costs and total emission of 32 cases tested

Table 15 Total costs (in Millions THB) in the case of all high or low price levels

Case	Total Costs (Millions THB)	Total Emission (Millions kgCO$_2$)
HHHHH	68,017.83	778.51
LLLLL	52,015.28	1,297.80
LLLHL	53,562.74	782.42
HHLLL	55,759.24	765.00

In the LLLHL scenario in which all three energy sources are at their lowest price levels, the model may select any inexpensive source of energy to minimise the total costs. Comparing this scenario with the HHHHH situation, their total emission values are relatively similar but the total costs of the LLLHL scenario is less due to lower energy costs. As for the HHLLL scenario, both materials (scrap and slabs) are at their low price levels. Thus, the model tries to balance the usage of both materials to lower the total costs. Since only the natural gas factor is at the low level in the HHLLL scenario (the other two sources of energy are at the high levels), natural gas is preferred to keep the total costs low. However, the emission of the HHLLL scenario is lower than that of the LLLHL one.

We may conclude from these results and from Fig. 2 that the availability of inexpensive energy sources tends to lead to lower total costs. As for the total emission, it may not be totally conclusive. It should be noted, however, that the combination HL of the scrap and slabs, respectively, often leads to lower total emission.

In the HRSC supply chain, electricity is the primary source of energy in the production process. Increase in the electricity price will encourage imports of HRSC from overseas to avoid high production cost. Although not ideal from a metallurgical viewpoint fuel oil and natural gas could be used to melt scrap prior to casting into slabs. Increase in the prices of either fuel oil and natural gas may lead to reduction in scrap processing to keep the overall cost low. Increase in the domestic scrap price is likely to boost slab imports to be used as the raw materials for HRSC production. On the contrary, if the price of imported slabs increases, the steel mills may decide to increase utilisation of domestic scrap. Solutions that increase raw material imports are likely to decrease the total emissions in Thailand, but they may not yield the lowest total costs. To lower the total emission, the use of fuel oil should be avoided but the use of electricity and natural gas should be encouraged as the latter two are cleaner sources of energy.

Hence, the operational strategies could be summarised as the followings.

- The combination of high price of domestic scrap and low price of domestic slabs encourages lower total emission of the HRSC supply chain.
- When the domestic electricity price is high, HRSC should be imported to avoid high production cost.
- When the prices of fuel oil and natural gas are high, scrap processing should be reduced.
- When the price of the domestic scrap increases, scrap imports could be increased to lower the total costs; and domestic CO_2 emission is consequently decreased.
- To lower the total emission, fuel oil should be avoided.

Moreover, in the steel supply chain, there could be a sudden increase of imported raw materials, especially slabs, from "price dumping", particularly when the oversea suppliers have excess stocks. This can weaken the competitiveness of local suppliers if the situation becomes prolonged. In this situation, the government may impose temporary measures to counterbalance the price of the imported materials. To encourage the use of cleaner sources of energy, the government may issue policies to subsidise certain sources of energy that has low CO_2 emission rate, or impose additional taxes to those energy sources with high rate of CO_2 emission.

6. Conclusion

This research develops a bi-objective model of the HRSC supply chain in Thailand. The objectives are to minimise the total costs (CO model) of the HRSC production as well as to minimise the total CO_2 emission (EO model) in order to reduce environmental impact. The HRSC supply

chain consists of four ports, five steel mills (P1 to P5), and 13 customers. There are three sources of energy, four types of products, and two modes of transportation, full-load and no-load trips, and two sources of materials. To demonstrate the use of the model, numerical data from the HRSC supply chain in Thailand are applied.

The results show that different objectives yield different production and energy usage. In the solution of the CO model, P2 and P1 should produce 40 % and 25 % of the total production, respectively, while electricity is used as the primary source of energy in the production of about 2/3 of all the energy required. But in the EO model, the total domestic production is reduced to cut down emissions. Among the other three steel mills, the majority or 77 % of the additional production is assigned to P3 which is located closest to the customers, resulting in reduction of the emissions caused by transportation. The steel mill P1, on the other hand, should limit its production since it is located furthest from the customers. This, however, may be difficult to implement in certain situations as there is a minimum production requirement to keep the mill open and operating efficiently.

Strategies for certain circumstances are suggested depending upon the objective of minimising the total costs or the total emissions or both. Moreover, during the initial period of price dumping, the government may temporarily impose import tax to safeguard competitiveness of the local manufacturers. Governmental policies and support to promote the use of cleaner energy sources are highly encouraged to stimulate sustainability of the industry.

The model presented in this research is a specific example of a more general three-echelon supply chain model, specifically suppliers (or ports in this case)-manufacturers-customers. Future research may consider inclusion of uncertainty, such as via demands and costs.

Acknowledgement

References

[1] Allwood, J.M., Cullen, J.M., Milford, R.J. (2010). Options for achieving a 50% cut in industrial carbon emissions by 2050, *Environmental Science & Technology*, Vol. 44, No. 6, 1888-1894, doi: 10.1021/es902909k.
[2] The Association of Southeast Asian Nations (2016). ASEAN Member States, from *http://asean.org/asean/asean-member-states*, accessed March 24, 2017.
[3] International Iron and Steel Institute (IISI) (1997). IISI life cycle inventory study for steel industry products, from *https://www.worldsteel.org/*, accessed March 25, 2017.
[4] Geilen, D., Moriguchi, Y. (2002). CO_2 in the iron and steel industry: An analysis of Japanese emission reduction potentials, *Energy Policy*, Vol. 30, No. 10, 849-863, doi: 10.1016/S0301-4215(01)00143-4.
[5] Ruth, M., Amato, A. (2002). Vintage structure dynamics and climate change policies: The case of US iron and steel, *Energy Policy*, Vol. 30, No. 7, 541-552, doi: 10.1016/S0301-4215(01)00129-X.
[6] Zhang, J., Wang, G. (2008). Energy saving technologies and productive efficiency in the Chinese iron and steel sector, *Energy*, Vol. 33, No. 4, 525-537, doi:10.1016/j.energy.2007.11.002.
[7] Sohaili, K. (2010). The impact of improvement in Iran iron and steel production technology on environment pollution, *Procedia Environmental Sciences*, Vol. 2, 262-269, doi: 10.1016/j.proenv.2010.10.032.
[8] Tongpool, R., Jirajariyavech, A., Yuvaniyama, C., Mungcharoen, T. (2010). Analysis of steel production in Thailand: Environmental impacts and solutions, *Energy*, Vol. 35, No. 10, 4192-4200, doi: 10.1016/j.energy.2010.07.003.
[9] Johansson, M.T., Söderström, M. (2011). Options for the Swedish steel industry – Energy efficiency measures and fuel conversion, *Energy*, Vol. 36, No. 1, 191-198, doi: 10.1016/j.energy.2010.10.053.
[10] Lee, Y.H., Kim, S.H. (2002). Production-distribution planning in supply chain considering capacity constraints, *Computers & Industrial Engineering*, Vol. 43, No. 1-2, 169-190, doi: 10.1016/S0360-8352(02)00063-3.
[11] Byrne, M.D., Bakir, M.A. (1999). Production planning using a hybrid simulation-analytical approach, *International Journal of Production Economics*, Vol. 59, No. 1-3, 305-311, doi: 10.1016/S0925-5273(98)00104-2.
[12] Zamarripa, M.A., Aguirre, A.M., Méndez, C.A., Espuña, A. (2012). Improving supply chain planning in a competitive environment, *Computers & Chemical Engineering*, Vol. 42, 178-188, doi: 10.1016/j.compchemeng.2012.03.009.
[13] Dehghanbaghi, N., Sajadieh, M.S. (2017). Joint optimization of production transportation and pricing policies of complementary products in a supply chain, *Computers & Industrial Engineering*, Vol. 107, 150-157, doi: 10.1016/j.cie.2017.03.016.

[14] Gholamian, M.R., Heydari, M. (2017). An inventory model with METRIC approach in location-routing-inventory problem, *Advances in Production Engineering & Management*, Vol. 12, No. 2, 115-126, doi: 10.14743/apem 2017. 2.244.

[15] Koç, U., Toptal, A., Sabuncuoglu, I. (2017). Coordination of inbound and outbound transportation schedules with the production schedule, *Computers & Industrial Engineering*, Vol. 103, 178-192, doi: 10.1016/j.cie.2016.11.020.

[16] Jung, K.H., Lee, S.K. (2006). New paradigm of steel mills in the supply chain of automotive sheets, *Supply Chain Management: An International Journal*, Vol. 11, No. 4, 328-336, doi: 10.1108/13598540610671770.

[17] Organisation Internationale des Constructeurs d'Automobiles (2016). Production statistics, from *http://www.oica.net/category/production-statistics*, accessed March 24, 2017.

[18] South East Asia Iron and Steel Institute (2013). Steel statistical yearbook 013, from *https://www.worldsteel.org/*, accessed March 24, 2017.

[19] Zadeh, A.S., Sahraeian, R., Homayouni, S.M. (2014). A dynamic multi-commodity inventory and facility location problem in steel supply chain network design, *The International Journal of Advanced Manufacturing Technology*, Vol. 70, No. 5-8, 1267-1282, doi: 10.1007/s00170-013-5358-2.

[20] The Office of Industrial Economics (2012). Annual report of Thailand economics situation, from *https://www.bot.or.th*, accessed January 30, 2017.

[21] Department of Alternative Energy Development and Efficiency (2011). Annual report of Thailand energy situation 2011, from *http://weben.dede.go.th/webmax/*, accessed January 30, 2017.

A study of the impact of ergonomically designed workplaces on employee productivity

Leber, M.[a,*], **Bastič, M.**[b], **Moody, L.**[c], **Schmidt Krajnc, M.**[d]

[a]University of Maribor, Faculty of Mechanical Engineering, Maribor, Slovenia
[b]University of Maribor, Faculty of Economics and Business, Maribor, Slovenia
[c]Coventry University, Faculty of Arts and Humanities, Coventry, United Kingdom
[d]University of Maribor, Faculty of Education, Maribor, Slovenia

A B S T R A C T

Ergonomics principles help designing the workplace in a way that makes work more efficient and safer. Employee satisfaction increasingly affects the productivity of a process, which also includes disabled people and represents an important source of human resources. In the framework of the EU-project ERGO WORK a survey-based research was conducted to measure the satisfaction of people with disabilities (PWD) in their workplace and asses how their satisfaction was perceived by employers in UK, Poland and Slovenia. Three hundred and three respondents were involved in the survey. Results show that PWD place a great emphasis on the satisfaction in the workplace. PWD in Slovenia are more satisfied than PWD in Poland, whereas the employers' perception of the satisfaction of PWD and other employees in Poland, Slovenia and UK does not vary. A general adaptation of the workplace significantly and positively influences the satisfaction of persons with disability and that the adaptation of the workplace to the needs of PWD is better if employers have access to knowledge, special equipment and financial resources.

1. Introduction

1.1 Theoretical frame

In a time of high unemployment, employers are generally less interested in fulfilling employees' individual needs. This trend is especially visible in poorer European countries for example in Poland where employees are often underpaid with poor job agreements [1]. In a situation where non-disabled employees are struggling, with many agreeing to work in very unfavourable conditions, people with disabilities are even more challenged to enter the labour market. In Poland, discrimination against persons with disabilities (12.2 % of the population) has been tackled in numerous ways, however whilst their role is seen to be increasing, PWD are still largely invisible in public [2]. Companies in Poland with at least 25 employees are legally obliged to employ a minimum of 6 % of employees who have a disability.

More than 15 % of people living in the European Union have a disability. In Slovenia, the percentage of PWD is between 12 % and 13 %, not significantly less than in the wider EU (according to estimates). In 2014, PWD represented approximately 4 % of the work force in Slovenia. The number of employed PWD is slowly but steadily increasing compared to the previous years in regular working environments [3]. Regular working environments include all employments in

the public and private sector in the mandatory quota system for employing PWD; this means that in relation to the total number of employees the employer has to employ a certain number of PWD as notes Statistical Office of the Republic of Slovenia 2014.

In the UK, almost one in five persons has a disability [3]. The employment rate of working disabled employees is 47.8 % compared with 75.9 % of non-disabled people, and disabled people are nearly four times as likely to be unemployed or involuntarily out of work as non-disabled people [4]. In the UK, the Equality Act 2010 covering disability (as well as age, sexual orientation, religious beliefs) legally protects people from discrimination in the workplace and in wider society. It requires equal treatment in access to employment; and employers and service providers are obliged to make reasonable adjustments to the workplace to overcome barriers experienced by disabled people. Reasonable adjustments should be made in a range of ways for example through workplace furniture, training, adjusted hours, changes to policies, and provision of assistive technology.

1.2 Literature overview

It is recognised that ergonomics in a business environment improve the operational performance (production and product efficiency or the quality of services), productivity and create wellbeing of employees in the workplace [3]. For improving business performance organisations have to create an environment that facilitates operational performance and addresses the employees' wellbeing (health, safety). Considering ergonomics principles in designing workplaces and the adaptation of tasks to the physical and mental abilities of workers contributes to the creation of such an environment which can contribute to a better climate that fosters success and wellbeing at the same time [5].

Ergonomics as a scientific discipline tries to increase effectiveness by designing workplaces or adapting work, and by eliminating processes without added value as well as threats which increase the risk of developing illnesses and injuries of employees [6]. The application of ergonomics principles in a business environment can ensure a direct benefit for employees and organisations by easing physical and psychological burdens, reducing the risk of developing occupational diseases and injuries, and increases work efficiency [7]. Based on an EU directive, institutions in individual member states demand a risk and threats assessment for individual jobs. Ergonomics is a recognised discipline used to assess whether work, equipment and environment match the required performance of persons involved [8].

When designing jobs and products the aggregated information on processes, tools, machines, subjects of work, tasks and operators must be taken into account, limitations, which are often conflicting, must be met and a design must be generated, which will be acceptable for all parties involved [9]. The successful optimization of products is only possible if ergonomic aspects are systematically embedded into the product life cycle and if an Ergonomic evaluation is carried out of prototype solution variants. The optimisation process brings the most appropriate foam thickness and its stress-strain relationship to produce low contact pressures while maintaining high level of stability of the product grasp [10].

In a two-year period, a group of researchers [5] has taken a look at larger production plant and examined to what extent ergonomics measures for increasing operational performance and wellbeing of employees influence the incidence of diseases, injuries and pain. Based on measurements and evaluations they found out that the number of injuries and diseases recorded was smaller if the organisation applied ergonomics principles for facilitating the employees' efficiency and wellbeing.

Inclusive employment is one of the key priorities of the European Commission and its ten-year Europe 2020 strategy in the field of work and growth of the employed population. In order to avoid social exclusion of PWD it is important to adopt additional measures to meet their needs and at the same time consistently implement the principle of equal opportunities and the principle of non-discrimination due to disability [11].

Taking into consideration the fact that every one of us is different, the need for individual ergonomic support in the workplace becomes a necessity. Employees have special ergonomic

needs, whether it is due to disability, age or other special personal circumstances. The aging of the population and prolonging working lives contribute to an increasing number of employed persons with disability, whether it is a sight, hearing, movement or other type of impairment.

A holistic ergonomics access in companies supports human factors and the increasing diversity of employees; as such it contributes to the company's economic performance. It is recognised that ergonomics measures reduce absenteeism or absence from work (less work-related injuries and diseases) and raise the satisfaction and efficiency of employees. Taking this into consideration intensive campaigns oriented towards decision makers in organisations / companies must be promoted and supported [11].

Many factors influence the satisfaction of employees at work. They can be divided into organisational, group and personal factors, today we are considering further understanding of ergonomics in the workplace and the creation of workplaces for ergonomic principles [3]. Individuals will be satisfied if results are achieved at work that are in line with their personality, work values, needs and aspirations. Organisational factors such as content of work, distribution of working time, work conditions (adaptation of workplace, safety, and removal of disruptive factors), career prospects etc. influence the satisfaction as well. The following factors must not be overlooked: working atmosphere, relationships at work, treatment of people at work, which prevails in an undertaking [12]. On the path to achieving social inclusion in working environments it is important that people with disabilities are included in appropriate and meaningful work which leads to independence, creativity and enables an increasing self-respect and a positive self-image [13]. Removing and reducing obstacles at workplaces, understanding the needs of employed people with disabilities and creating opportunities for social participation and inclusion are important factors for the wellbeing and satisfaction of employed PWD [14].

2. Purpose of empirical research and hypotheses

The aim of this study was to research the satisfaction of PWD in the workplace and to find out how their satisfaction is perceived by employers in Great Britain, Poland and Slovenia. The following hypotheses were developed in order to achieve the aim of the research:

H1: PWDs' satisfaction varies according to the country of their workplace.
H2: The employers' perception on the satisfaction of PWD in the workplace is the same in all countries examined.
H3: The higher the perceived adaptation of the workplace for PWD or the better the knowledge of PWD on ways of adapting the workplace, the greater is the satisfaction of PWD in the workplace.
H4: The better the employers' knowledge of PWD-related legislation and their understanding of PWDs' needs are, and the better the employers' access to the resources is, the better the employers can adapt the workplace to PWDs' needs.

The study was limited to three European countries; and in the framework of the questionnaire to employers and employed PWD (the research did not consider higher education institutions). In the statistical processing of responses the study was limited to three indicators (adaptation of workplace to PWD, knowledge on legislation covering persons with disability and their needs, access to resources).

3. Research method

We conducted an online survey to obtain data from PWD and employers. We used Kruskal-Wallis H test to verify H1 and H2 and multiple regression analysis to test H3 and H4.

3.1 Sample

The study is an integral part of a wider research in the framework of the international ERGO WORK project (*www.ergo-work.eu*). Target population in ERGO WORK research was PWD and their employers in six countries: Belgium, UK, Italy, Poland, Slovenia, and Spain. Companies and

institutions have been invited to participate - we chose those that are employing more PWD. We tried to cover production companies from different branches and different sizes. The statements of employers on topics under research were obtained by self-administered questionnaire. We applied different means of communications with PWD regarding their level of disability. Some of them obtained self-administered questionnaire with additional oral interpretation of questions and extra time to complete it. We also used an interview based on the questions from the questionnaire to elicit the statements of PWD who were not able to fill in the questionnaire on their own. The questionnaires were distributed in printed version by post (approximately 250) and on line (approximately 270).

Of the 520 invited participants 480 questionnaires were successfully completed. However, we analysed and discussed only data obtained by respondents in three partner countries: UK, Poland and Slovenia. The response rates in other partner countries (Belgium, Italy and Spain) were too low. From all three countries (UK, Poland and Slovenia), we obtained 303 usable questionnaires.

A total of 303 individuals responded to this research. 29 % of respondents were PWD and 67 % were employees without disability (Table 1). The major part of respondents came from Slovenia (47.9 %), followed by respondents from Poland (28.4 %) and UK (23.8 %).

Table 1 Profiles of the samples

Country		Disability	Without disability	No information	Total
Poland	Count	33	53	0	86
	% of Total	10.9	17.5	0.0	28.4
Slovenia	Count	37	100	8	145
	% of Total	12.2	33.0	2.6	47.9
UK	Count	18	50	4	72
	% of Total	5,9	16.5	1.3	23.8
Total	Count	88	203	12	303
	% of Total	29.0	67.0	4.0	100

3.2 Data collection procedure and questionnaire

The questionnaire consisted of 70 questions (items) grouped into four sections. Considering the aims of this research, we selected 24 out of 70 items. The concept 'adaption of workplaces for PWD' was measured by 7 items, 'feelings on the workplace' by 5 items, 'familiarity with ergonomic design' by 4 items, and 8 items were selected to measure 'employers' knowledge of legislation covering PWD and their access to funding for adapting workplaces for PWD'. In addition, the respondents were asked to provide some demographic data. The original questionnaire was in English and it was translated to the language of each participating country. The online questionnaire was run using Bristol Online Surveys (BOS). This allowed easy distribution of the survey through a web link across Europe. Items from the questionnaire measuring specific concepts are presented in Table 2.

Table 2 Description of concepts and variables

Code	Description of concepts and variables
	Adaptation of workplace for persons with disability
Q14	How well does your workplace accommodate people with disabilities?
Q17	This workplace is well adapted for people with disabilities - generally.
Q18	This workplace is well adapted for people with visual impairments.
Q19	This workplace is well adapted for people with hearing impairments.
Q20	This workplace is well adapted for people with physical impairments.
Q21	This workplace is well adapted for people with mental health needs.
Q22	This workplace is well adapted for people who have intellectual disabilities.
	Feelings of persons with disability
Q26	People work well together in this workplace.

**Table 2** Description of concepts and variables (continuation)

Q27	Management & staff work well together.
Q28	I feel included by my workmates and part of a team.
Q29	Most employees are happy here.
Q30	I am generally happy here.
	Knowledge of workplace design
Q31a	Familiarity with Ergonomic Design?
Q31b	Familiarity with Universal Design?
Q31c	Familiarity with Inclusive Design?
Q31d	Familiarity with Accessible Design?
	Knowledge of legislation covering persons with disability and access to resources needed for workplace adaptations
Q52	Do you know where to get advice from experts about adapting your workplace?
Q53	Do you know how to find specialised equipment?
Q54	Do you have access to funding for adapting your workplace?
Q55	How well do you understand the legislation applicable to employing people with disabilities?
Q56	How well do you understand the legislation applicable to adapting your workplace for PWD?
Q57	How far does your organisation consider the needs of the people who are going to do a job or work process, when you set up a new process or make changes?
Q61	When making changes to our workplace we pay attention to the needs of people with disabilities.
Q62	When changing work processes we pay attention to the needs of persons with disability.

3.3 Data processing procedures

Data analysis was carried out with statistical package IBM Statistical Package for Social Sciences Statistics (SPSS 21). We used Kruskal-Wallis H test to verify H1 and H2 and multiple regression analysis to test H3 and H4. Before we tested research hypotheses H3 and H4, exploratory factor analysis was applied to detect the presence of meaningful patterns among measured variables. The principal component analysis, as the extraction method, and varimax rotation procedure were applied. The factors' reliability was tested with Cronbach's alpha.

4. Results

To test Hypothesis H1 – PWDs' satisfaction varies according to the country of their workplace - the PWDs' attitudes about the item 'I am generally happy at this workplace' (Variable Q30) were taken into account. The attitudes were measured on a five point Likert scale, where 1 means 'Strongly agree' and 5 'Strongly disagree'. The respondents were offered the option 'I don't know enough about it to say', but it was not included in data analysis. The distributions of variable Q30 are presented in Fig. 1. The majority of PWD in Poland chose 2 and 3 (86.7 %), in Slovenia 1 and 2 (85.7 %), and 73.3 % of PWD in UK expressed their attitudes about this item with 2 and 3 on the five point scale.
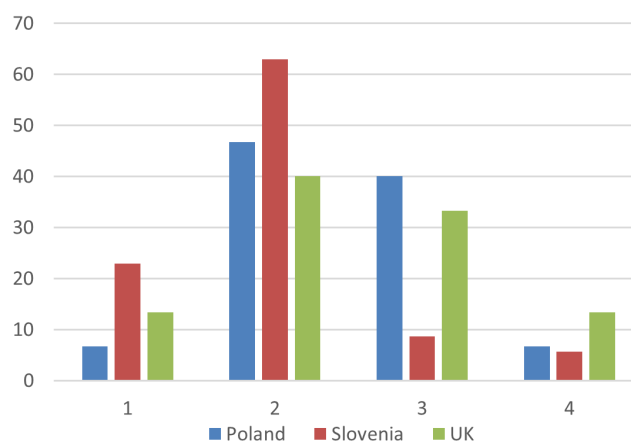
Fig. 1 Distributions of Q30 for PWD in three countries

A Kruskal-Wallis H test was conducted to determine if there were differences in the level of satisfaction (variable Q30) among three groups of PWD with their workplaces in different countries: Poland (n = 33), Slovenia (n = 37), and UK (n = 18). Distributions of Q30 scores were not similar for all groups, assessed by visual inspection of boxplot. Values are mean ranks unless otherwise stated. The mean ranks of Q30 scores were statistically significantly different between groups, $\chi 2(3)$ = 14.468, p = 0.002. Pairwise comparisons were performed using Dunn's procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed statistically significant differences in Q30 scores between Slovenia (32.50) and Poland (47.03) (p = 0.018), but not between any other group combination. The significant difference in PWDs' satisfaction confirms H1.

To verify hypothesis H2 – The employers' perception on the satisfaction of PWD in the workplace is the same in all countries examined – the employers' attitudes about the item 'Most employees are happy at this workplace' (Q29) were analysed. The employers expressed their attitudes on the five point Likert scale where 1 means 'Strongly agree' and 5 'Strongly disagree'. The respondents were offered the option 'I don't know enough about it to say', which was not included in data analysis. The distributions of Q29 for employers in three countries are shown in Fig. 2. The majority of employers expressed their attitudes about the PWDs' satisfaction with two on the five-point scale: in Poland (73.2 %), in Slovenia (59.3 %), and in UK (55.6 %).

The Kruskal-Wallis H test was conducted to determine if there were differences in the level of the employers' perception about the PWDs' satisfaction among three groups of employers who worked in three different countries: in Poland (n = 41), in Slovenia (n = 91), and in UK (n = 45). Distributions of Q29 scores were similar for all groups, assessed by visual inspection of boxplot. Medians of Q29 were not statistically significantly different between three countries $\chi 2(2)$ = 0.264, p = 0.876. All these results confirm hypothesis H2.

Before verifying hypothesis H3 – which assumed that PWDs' perception of workplace adaptation, and their knowledge on ways of how workplaces can be adapted affect their satisfaction in the workplace – a factor and reliability analyses were applied to test the dimensionality and reliability of the following concepts: adaptation of the workplace, knowledge on ways of adapting the workplace and the satisfaction of PWD in the workplace. When deciding on the number of factors for each concept the eigenvalue, the percentage of the total variance explained by the factors, and interpretability of factors were taken into account.

The results of the factor analysis for seven variables measuring the adaptation of the workplace to PWD showed two factors with eigenvalue higher than one (Table 3). The first dimension (factor F1), measured by variables Q18, Q19, Q21 and Q22, was named as 'adaptation of the workplace to specific types of disabilities' and the second (factor F2), measured by variables Q14, Q17 and Q20, as 'general adaptation of the workplace'. The majority of the factor loadings were higher than 0.7, which indicates a very well defined structure [15]. Both factors together explained approximately 66 % of the total variance and Cronbach's alphas were higher than 0.7. All this confirms their reliability.
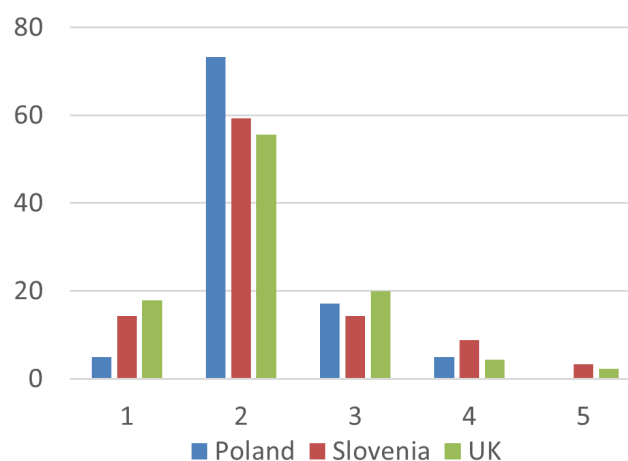

Fig. 2 Distributions of Q29 for employers in three countries

Table 3 Factor loadings, communalities, explained variance and Cronbach's alpha for the adaptation of the workplace

Variable	F1	F2	Communalities
Q18	0.644	0.327	0.522
Q19	0.700	0.279	0.568
Q21	0.803	0.131	0.622
Q22	0.831	0.114	0.704
Q14	0.065	0.834	0.699
Q17	0.273	0.837	0.774
Q20	0.370	0.758	0.711
Eigenvalue	3.497	1.14331	
% of explained variance	49.963	16.333	
Cronbach's alpha	0.782	0.810	

Remark: The meaning of variables Q14 to Q20 is explained in Table 2

Table 4 Factor loadings, communalities, explained variance and Cronbach's alpha for the factor 'People with disabilities' feelings in the workplace'

Variable	F1	Communalities
Q26	0.825	0.681
Q27	0.740	0.547
Q28	0.853	0.727
Q29	0.848	0.718
Q30	0.857	0.735
Eigenvalue	3.408	
% of explained variance	68.16	
Cronbach's alpha	0.854	

Remark: The meaning of variables Q26 to 30 is explained in Table 2

The results of the factor analysis (Table 4) for variables measuring the PWDs' feelings in the workplace (variables Q26 to Q30) have shown that one factor had eigenvalue higher than one and it explained 68.16 % of the total variability. All factor loadings were higher than 0.7 and the reliability test with Cronbach's alpha confirms the reliability of this factor.

As indicated in Table 5, the results of the factor analysis have shown that it makes sense to merge all variables measuring the various possibilities for adapting a workplace (variables Q31a to Q31d) into one factor, which explains the 72.26 % of the total variance. All factor loadings as well as Cronbach's alpha were high, thus indicating the reliability of this factor.

Hypothesis H3 – which assumed that the PWD' feelings in workplace depends on the workplace adaptation for PWD or their knowledge on ways of adapting the workplace – was verified through a multiple regression analysis in which the dependent variable was the factor 'PWD' feelings in the workplace' and factors 'adaptation of the workplace to individual types of disabilities', 'general adaptation of the workplace' and 'knowledge on ways of adapting the workplace' were independent variables. The multiple regression analysis was carried out on data provided by PWD ($n = 76$).

The regression coefficient of the factor 'general adaptation of the workplace' was 0.34 and it was significantly different from zero at a significance level of 0.05. By improving the general adaptation of the workplace the PWD' feelings in the workplace increases too. For the other two factors, the regression coefficients were not significant. The general adaptation of the workplace explained 6 % ($R^2 = 0.06$) of the variance in the PWD' feelings in the workplace. As the factor 'general adaptation of the workplace' takes a significant and positive impact on the PWD' feelings, hypothesis H3 is confirmed.

Table 5 Factor loadings, communalities, explained variance and Cronbach's alpha for the factor 'Knowledge on ways of adapting the workplace'

Variable	F1	Communalities
Q31a	0.765	0.585
Q31b	0.871	0.759
Q31c	0.897	0.804
Q31d	0.862	0.743
Eigenvalue	2.8904	
% of explained variance	72.26	
Cronbach's alpha	0.871	

Remark: The meaning of the variables Q31a to Q31d is explained in Table 2

Table 6 Factor loadings, communalities, explained variance and Cronbach's alpha for the variables, which measure the level of knowledge on the legislation covering persons with disability and their needs as well as the employers' access to resources needed for workplace adaptations

Variable	Factor loadings		Communalities
	F1	F2	
Q52	0.185	0.818	0.703
Q53	0.227	0.835	0.748
Q54	0.141	0.732	0.556
Q55	0.754	0.372	0.706
Q56	0.736	0.350	0.665
Q57	0.730	0.230	0.586
Q61	0.820	0.184	0.706
Q62	0.813	0.021	0.661
Eigenvalue	4.015	1.316	
% of explained variance	50.19	16.4	
Cronbach's alpha	0.852	0.762	

Remark: The meaning of variables Q52 to Q62 is explained in Table 2

Before verifying the hypothesis H4 factor analysis was conducted to test the dimensionality of concepts 'the employers' knowledge on the legislation and needs of PWD' and 'employers' access to knowledge and resources needed for workplace adaptations'.

The results of the factor analysis (Table 6) showed that the variables Q52 to Q57 and Q61 and Q62 measured two dimensions, which were represented by two factors with eigenvalue higher than one. The first factor (F1), which was consisted of variables Q55 to Q57 and Q61 and Q62, was named 'employer's knowledge on the legislation and PWD's needs'. It explained 50.19 % of the total variance. The second factor (F2), which composed of variables Q52 to Q54 explained 16.4 % of variance was named 'employers' access to resources needed for workplace adaptations'. By resources, we mean knowledge, specialised equipment and financial resources.

Hypothesis H4 – which assumed that the adaptation of the workplace for PWD depends on the employers' accessibility to resources needed for workplace adaptations as well as their knowledge of PWD-related legislation and understanding of PWDs' needs – was tested with two regression models because two dimensions were detected for the concept adaptation of the work place to PWD. The dependent variable in the first regression model was the factor 'general adaptation of the workplace for PWD' and the factor 'adaptation of the workplace to specific types of disabilities' was dependent variable in the second model. In both regression models, the factors 'knowledge on the legislation and PWDs' needs' and 'access to resources needed for workplace adaptations for PWD' were independent variables. Both regression analyses were conducted on data obtained by employers (n = 177).

The results of the multiple regression analysis for the dependent variable 'general adaptation of the workplace for PWD' showed that both independent variables together explained 46 % of variance of variable 'general adaptation of the workplace to the needs of PWD' (R^2 = 0.46). The regression coefficients of variables 'employers' knowledge on the legislation and PWDs' needs' and employers' access to resources needed for workplace adaptations' were 0.209 and 0.461, respectively. Both coefficients were significantly different from zero at $p < 0.05$ or $p < 0.001$, respectively. The higher regression coefficient for the factor 'employer's access to resources

needed for workplace adaptations' showed that this factor has a greater impact on improving the general adaptation of the workplace to the needs of PWD than the factor 'employer's knowledge on the legislation and PWDs' needs'.

A multiple regression was run to predict impact of 'knowledge on the legislation and PWD's needs' and 'access to resources needed for workplace adaptations for PWD' on 'adaptation of the workplace to specific types of disabilities' (ADAPT_SPEC). The multiple regression model statistically significantly predicted ADAPT_SPEC, adj. R^2 = 0.088. Only the regression coefficient of the factor 'employer's access to resources needed for workplace adaptations' has a positive and significant value (b = 0.296, p = 0.023), which demonstrates its impact on adapting the workplace to specific types of disability. The results of both regression analyses confirmed hypothesis H4.

5. Discussion

The research was carried out to measure the satisfaction rate of PWD and employers in relation to the workplace and to analyse whether differences exist in the three studied countries. An additional purpose of the research was to examine the factors (general and specific adaptation of the workplace, knowledge on universal, inclusive and ergonomic workplace) which affect the satisfaction level in the workplace from the viewpoint of PWD and employers.

The results of the research support hypothesis H1, which assumed that PWD's satisfaction varies according to the country of their workplace. In addition, we found out that PWD in Slovenia are more satisfied than PWD in Poland.

Based on the current situation in Slovenia which shows that difficulties in employing PWD still exist, we could conclude that this group of PWD was satisfied to even have a job given the current situation. Likewise, it is possible that the difference is on the one hand conditioned by personal characteristics, namely that these were friendly, reliable persons with a positive self-image, and on the other hand by perceiving social inclusion and the sense of belonging to the company, as evidenced by several international studies [13, 14, 16]. In general, the satisfaction among employees is also affected by minor ergonomic improvements in the workplace which, are mostly initiated by employees (including PWD) and can be carried out at low cost and are based on good local practices [17].

Hypothesis H2 assumed that in Poland, in the UK and in Slovenia the employers' perception regarding PWDs' satisfaction in the workplace does not differ. The hypothesis was confirmed.

Our assumption that employers apply similar criteria for job satisfaction of their employees and that their evaluations are usually between 2 and 3 on five-point scale where 1 means very satisfied or between 3 and 4 when 5 means very satisfied was confirmed. It is not expected that they will evaluate the job satisfaction of their employees very bad (bad evaluation of employers' work) or extremely good (lack of self-criticism – there are always possibilities for improvements). Employers assessed the satisfaction of PWDs on the basis of regular annual interviews with them. Detection of satisfaction of PWDs is similar in all three countries. This is probably the consequence of the implementation of the Strategy Europa 2020, carried out in these countries. Employers consider social responsibility and the ergonomics principles to improve the process to social inclusion of PWD in the workplace.

Hypothesis H3 was also verified when we saw that the PWD's satisfaction in the workplace depends on the general adaptation of the workplace to the person with disability.
Employees are often focused on the protection of safety and health at work, including (but not restricted to) reducing risks of injuries and diseases and promoting a work-life balance – factors improving the satisfaction in the workplace and the quality of life [5].

We are of the opinion that working environments from which people with disabilities from our research come from gradually respect and meet needs, capabilities, potentials and advantages of PWD which is one of the goals of the Europe 2020 strategy. The success of future human-systems integration efforts requires the fusion of paradigms, knowledge, design principles, work places, tools and devices and methodologies of human factors and ergonomics with those of the science of complex adaptive systems as well as modern systems integration [22]. To perform a task not only the tools and/or adaptations are needed but also human support and

supportive services in order to overcome barriers faced by PWD in working environments. Appropriate adaptations and a concept of universal design must be implemented in the labour market to its fullest extent so as to give the PWD concrete options for finding and keeping a job. The success of future human-systems integration efforts requires the fusion of paradigms, knowledge, design principles, work places, tools and devices and methodologies of human factors and ergonomics with those of the science of complex adaptive systems as well as modern systems integration [18].

In examining the H4 hypothesis, it was concluded that the better the employers' access to knowledge on adapting the workplace for PWD, to special equipment for adapting the workplace for PWD and to financial resources, the better is the adaptation of the workplace to the needs of PWD.

It is clear that certain conditions must be met to provide an appropriate ergonomic-friendly environment in an organisation containing in particular knowledge on ergonomics, access to ergonomic solutions when designing equipment and last but not least the provision of appropriate financial resources. This finding is also similar to the results of a study conducted in a larger Swiss company which carried out an analysis of ergonomic work conditions in the workplace. The conditions significantly improved after a training programme was carried out, and at the same time the individual knowledge level and interest as well as a sense of self-responsibility in the field of ergonomics increased [19]. This includes that employees and employers must obtain appropriate knowledge and information on principles of ergonomic workplace design adapted to PWD. We believe, that a programme of education and training was prepared in the framework of the "*ERGO WORK – Joining academia and business for new opportunities in creating ergonomic workplaces*" project and it would be reasonable to use it in practice and in educational institutions across the EU in future. The aim of the project was to improve the ergonomic design of workplaces for PWD and at the same time promoting social inclusion through the improvement and adaptation of workplaces so as to accommodate the various needs of PWD. Macro ergonomics offers a perspective as well as methods and tools for more successful human factors and ergonomic design, development, intervention, and implementation. Thus the present human factors can be used engineers, psychologists or ergonomist to achieve better results or can expand their cooperation [20].

6. Conclusion

The research which was conducted in the framework of the ERGO WORK project as a survey among decision-makers and PWD provides guidelines for an ergonomic workplace design strategy and strengthens the importance of PWDs' satisfaction in the workplace. Because the integration of PWD in the workplace is an EU priority the need for better cooperation among EU member states, and knowledge and practices transfer in this field is starting to develop. Therefore, it will become ever more important (as in the ERGO WORK project) to lay the focus on understanding and eliminating the obstacles in integration and in resolving these questions through training and cooperation between the academia and industry.

As mentioned in the Article 27 of the UN Convention on the Rights of Persons with Disabilities and the EU's Strategy for Europe 2020, employment and employment opportunities are key priorities of all EU governments. The inclusion of PWD in the open labour market is high on the EU's priority list. Addressing the special needs of PWD through tailor-made ergonomic solutions and the adaptation of workplaces is one of the possible solutions to achieve this target.

The purpose of the research was to provide further support in developing and innovating the employment sector in all partner countries of the ERGO WORK project and in EU's industry in general. Regardless of the field, in discussions on disability the quality of life must play a central role when decisions are being made. It is of utmost importance to recognise that in the employment sector each and every person has different needs in terms of support and different life goals. Therefore, it is necessary to support and respect the individual decisions as far as possible.

Acknowledgement

References

[1] Kiełbasiński, A. (2014). Dlaczego umowy śmieciowe są złe. Pięć grzechów głównych według "Wyborczej", from *http://wyborcza.pl/1,75478,15262523,Dlaczego_umowy_smieciowe_sa_zle__Piec_grzechow_glownych.html*, a*ccessed January 7, 2016.*

[2] Bojarski, L. (2013). Poland – Country report non-discrimination 2013, European network of legal experts in gender and non-discrimination, from *https://www.equalitylaw.eu/country/poland*, accessed January 7, 2016.

[3] Moody, L., Saunders, J., Leber, M., Wójcik-Augustyniak, M., Szajczyk, M., Rebernik, N. (2017). An exploratory study of barriers to inclusion in the European workplace, *Disability and Rehabilitation*, Vol. 39, No. 20, 2047-2054, doi: 10.1080/09638288.2016.1217072.

[4] Papworth Trust. Disability in the United Kingdom 2014, Facts and Figures, from *http://www.papworthtrust.org.uk*, accessed January 7, 2015.

[5] Hoffmeister, K., Gibbons, A., Schwatka, N., Rosecrance, J. (2015). Ergonomics climate assessment: A measure of operational performance and employee well-being, *Applied Ergonomics*, Vol. 50, 160-169, doi: 10.1016/j.apergo.2015.03.011.

[6] Wickens, C.D., Lee, J.D., Liu, Y., Gordon-Becker, S.E. (2004). *An Introduction to Human Factors Engineering*, 2nd edition, Pearson Prentice Hall, Upper Saddle River, New Jersey, USA.

[7] Sanders, M.S., McCormick, E.J. (1993). *Human Factors in Engineering and Design,* 7th edition, McGraw-Hill Education, New York, USA.

[8] McAtamney, L., Corlett, E.N. (1992). Ergonomic workplace assessment in a health care context, *Ergonomics*, Vol. 35, No. 9, 965-978, doi: 10.1080/00140139208967376.

[9] Feyen, R., Liu, Y., Chaffin, D., Jimmerson, G., Joseph, B. (2000). Computer-aided ergonomics: A case study of incorporating ergonomics analyses into workplace design, *Applied Ergonomics*, Vol. 31, No. 3, 291-300, doi: 10.1016/S0003-6870(99)00053-8.

[10] Harih, G., Borovinsek, M., Ren, Z., Dolsak, B. (2015). Optimal products' hand-handle interface parameter identification, *International Journal of Simulation Modelling*, Vol. 14, No. 3, 404-415, doi: 10.2507/IJSIMM14(3)3.285.

[11] Ergo Work. Ergo work recommendations for the system and policy makers, from *http://www.ergo-work.eu/site/english_site/link_en.html#den*, accessed January 7, 2016.

[12] Kopelman, R.E., Brief, A.P., Guzzo, R.A. (1990). The role of climate and culture in productivity, In: Schneider, B. (Ed.), *Organizational climate and culture,* Jossey-Bass, San Francisco, USA, 282-318.

[13] Mihanović, V. (2011). Disability in the context of a social model, *Hrvatska revija za rehabilitacijska istraživanja*, Vol. 47, No. 1, 72-86.

[14] Alborno, N., Gaad, E. (2012). Employment of young adults with disabilities in Dubai – A case study, *Journal of Policy and Practice in Intellectual Disabilities,* Vol. 9, No. 2, 103-111, doi: 10.1111/j.1741-1130.2012.00341.x.

[15] Hair Jr., J.F., Black, W.C., Babin, B.J., Anderson, R.E. (2010). *Multivariate data analysis: A global perspec*tive, 7th edition*, Pearson Education, Upper Saddle River, New Jersey, USA.

[16] Andrews, A., Rose, J.L. (2010). A preliminary investigation of factors affecting employment motivation in people with intellectual disabilities, *Journal of Policy and Practice in Intellectual Disabilities,* Vol. 7, No. 4, 239-244, doi: 10.1111/j.1741-1130.2010.00272.x.

[17] Kogi, K. (2012). Practical ways to facilitate ergonomics improvements in occupational health practice, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 6, 890-900, doi: 10.1177/0018720812456204.

[18] Karwowski, W. (2012). A review of human factors challenges of complex adaptive systems: Discovering and understanding chaos in human performance, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 6, 983-995, doi: 10.1177/0018720812467459.

[19] Menozzi, M., Von Buol, A., Waldmann, H., Kundig, S., Krueger, H., Spieler, W. (1999). Training in ergonomics at VDU workplaces, *Ergonomics,* Vol. 42, No. 6, 835-845, doi: 10.1080/001401399185324.

[20] Kleiner, B.M. (2008). Macroergonomics: Work system analysis and design, *Human Factors*: *The Journal of the Human Factors and Ergonomics Society*, Vol. 50, No. 3, 461-467, doi: 10.1518/001872008X288501.

Calendar of events

- 7th International Conference on Industrial Technology and Management, Oxford, UK, March 7-9, 2018.

- The 8th International Conference on Key Engineering Materials, Osaka, Japan, March 16-18, 2018.

- 2nd International Conference on 3D Printing Technology and Innovations, London, UK, March 19-20, 2018.

- The 2018 International Conference on Robotics and Intelligent Control, Nagoya, Japan, April 27-30, 2018.

- 2018 The 5th International Conference on Manufacturing and Industrial Technologies, Hefei, China, May 18-21, 2018.

- The 3rd International Conference on Advanced Functional Materials, San Francisco, USA, August 3-5, 2018.

- 3rd International Conference on Material Engineering and Smart Materials, Okinawa, Japan, August 11-13, 2018.

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 *Production Engineering Institute*, *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
Production Engineering Institute
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. 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.

Contents