International Journal of Production Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713696255

Selection of modules for mass customisation

C. da Cunha [a]; B. Agard [b]; A. Kusiak [c]
[a] Laboratoire IRCCyN, École Centrale de Nantes, France [b] CIRRELT, Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, Montréal (Québec), Canada [c] Intelligent Systems Laboratory, Department of Mechanical and Industrial Engineering, The University of Iowa, Iowa, USA

First published on: 28 January 2009

To cite this Article da Cunha, C., Agard, B. and Kusiak, A.(2010) 'Selection of modules for mass customisation', International Journal of Production Research, 48: 5, 1439 — 1454, First published on: 28 January 2009 (iFirst)
To link to this Article: DOI: 10.1080/00207540802473989
URL: http://dx.doi.org/10.1080/00207540802473989

International Journal of Production Research
Vol. 48, No. 5, 1 March 2010, 1439–1454

RESEARCH ARTICLE

Selection of modules for mass customisation

C. da Cunha[a], B. Agard[b]* and A. Kusiak[c]

[a]Laboratoire IRCCyN, École Centrale de Nantes, Nantes Cedex 03, France; [b]CIRRELT, Département de Mathématiques et de Génie Industriel, École Polytechnique de Montréal, Montréal (Québec), Canada; [c]Intelligent Systems Laboratory, Department of Mechanical and Industrial Engineering, The University of Iowa, lowa, USA

(*Received 18 September 2007; final version received 31 August 2008*)

Satisfying customer expectations is of paramount importance in today's markets. A customer expects to receive a product that meets as best as possible his/her expectations. To satisfy diversified requirements companies may focus on product families. For a specific customer the company has to select within a product family what will make a valid product. Then, in a just-in-time environment, the suppliers may have to provide the exact subassemblies corresponding to each product in a predefined time. This integration takes place by designing modules for the supplier. This paper proposes to extract customers' behaviour patterns, in terms of the components required, using data-mining and entropy maximisation when selecting modules to be manufactured. Different methods for selection of modules are proposed. Computational tests are performed to evaluate performance of the selection methods with respect to the specified assembly time/resource level.

Keywords: modular design; mass customisation; assemble-to-order; data mining; customer demand; pattern extraction; entropy maximisation

1. Introduction

Transformation from the 'demand-driven economy' to the 'offer-driven economy' has increased corporate competitiveness. To support the 'offer-driven economy', companies tend to enlarge their product portfolios in pursuit of increased market shares. The challenge is then to determine a trade-off between the cost of offering a wide range of products and the gains that such portfolio provides.

Modularity is useful in realising the assemble-to-order (ATO) policy when dealing with a large diversity of products. Indeed, a relatively small stock of modules permits to offer a large portfolio of final (modular) products. The designed modules should enable the best response to the customer needs. The latter can be accomplished by integrating the information about anticipated customer requirements at the product design phase.

Customers' behaviour is not easy to elicit. Selection of a particular product by a customer is the result of a cognitive process including its objective and subjective evaluation. The information obtained from the sales history is useful in capturing

*Corresponding author. Email: bruno.agard@polymtl.ca

ISSN 0020–7543 print/ISSN 1366–588X online

DOI: 10.1080/00207540802473989
http://www.informaworld.com

this phenomenon. Data mining techniques extracting patterns from large data sets describe some of the purchase patterns and characterise the products sold.

This paper emphasises the role of knowledge extracted from data in the selection of modules in an ATO environment. The importance of information availability at an early design phase is addressed. Selection methods based on data mining are proposed in Section 3. Information obtained from data mining is used with different extends. The computational results are discussed in Section 4. The computational experiments focus on analysis of the performance of different methods when considering the time/resource needed to respond to a demand for the final product assembled from the modules in stock. One of the major research issues is determining a trade-off between the gain from the information integration and the cost of information retrieval.

2. Literature review and problem statement

Mass production is perceived as an enabler of quality, uniformity, and productivity. Implementation of these principles requires systematic engineering and planning of the production process with the aim to improve production. Product standardisation is a way to pursue the economy of scale principle, where the fixed costs are lower owing to a larger number of units produced.

In recent decades, industry has begun pursuing product diversification strategies rather than the 'single model for all customers' strategy. Shapiro (1979) stated: 'the customisation degree of a product line or of a service tends to become the most important production variable for industrials goods' producer'. Mass customisation has then become a new trend for companies to compete in the market (Pine 1993). It is necessary to offer a large product portfolio while maintaining production efficiency that is realised through product families. The product families may be developed around product platforms that take into account different indicators such as product and process commonality as well as component and product cost (Simpson *et al*. 2006).

Although a large diversity of products offered may attract customers, nevertheless relatively few product configurations may represent the majority of sales. For example, Kocher and Rolland (1995) reported that for two different car models (Peugeot 306 and 405) 60% and 90%, respectively, of the demand is met with only 20% of the product diversity. It is therefore necessary to define a product management policy for a large product portfolio and a relatively small number of different vehicles sold. In a managed product diversity environment, the cost of offering a large portfolio should not exceed the gain resulting from customer satisfaction. Evaluating the cost of diversity is not easy owing to multiple cost sources. Nevertheless, it is essential to find the level of diversity that optimises the inventory holding cost and customer satisfaction (Tarondeau 1998).

For traditional production strategies (make-to-order or make-to-stock), once the production has been launched (in response to a demand in the make-to-order approach or a stock replenishment in the make-to-stock strategy), all steps needed to produce a product are realised. An alternative to this production approach is the configure-to-order strategy, which has two forms: assemble-to-order (ATO) or pick-to-order (PTO). The ATO strategy is commonly used in automotive industry.

Standardisation can be an effective tool in product diversity management. Some products sold could include additional options not selected by a customer. The price of this 'product upgrade' is shared between the supplier and the producer. The question that arises is then, which standard product/component to produce? This problem has been handled by offering free upgrades (Briant and Naddef 2004). Besides, this sustainability of an 'upgrade' approach is questionable. The producers are reluctant to pay for what they perceive is non-adaptability of the suppliers. Therefore new design and management solutions enabling production of the right products at the right time are needed. The ATO policy with distributed production of modules and on-site final assembly is such a solution. This solution offers benefits stemming from low labour costs at locations where the initial production is performed and enables synchronisation with prime contractor's production lines. The main issue is the design and the stocking of modules used in the final assembly that guarantee an acceptable lead time (time between an order placement and the delivery date).

The concept of modularity has been used in many industries. Modular production is defined as the capability to design and produce components that can be combined in a meaningful way (Starr 1965). The modular design approach implies rethinking the design process in a company (Kusiak 1999). Huang and Kusiak (1998) developed a method to integrate functional constraints in the modular designs. A modular architecture was proposed to design product families. It aimed at optimising modules to match customers' requirements. Gonzalez-Zugasti *et al.* (1998, 2000) optimised a trade-off between the expected technical performance of products and the global cost of product families. Some researches considered the time evolution of customer needs (Zamirowski and Otto 1999; Dahmus *et al.* 2001; Agard and Kusiak 2004). The methods developed permit the design of a product family using a platform integrating the stable requirements and modules realising the evolving customer needs. Different types of modules have been used in industry, e.g., Vanilla boxes in the automotive industry that can only be assembled with basic components (Swaminathan and Tayur 1998) or Cadillac boxes (modules containing most of the features that can be disassembled or used for downward substitution) (Rao *et al.* 2004)

The information about product demand is necessary to optimally design a production system. Numerous studies have shown that major savings can be achieved from the use of pertinent information (Gavirneni *et al.* 1999; Lee and Whang 2000). When the number of references (i.e., different final products that are possible to provide) is much greater than the number of products sold, the product-mix information has to be carefully considered. Furthermore, the information about the consumer demand (ultimate consumers' behaviour) may not be available, non existent, or non reliable. This situation is more likely to occur when a manufacturer is not the final supplier. The model describing the consumer behaviour patterns is not only *stochastic* but also its parameters are *not known*. Information about the requirements of each customer is usually not existent or sparse and imprecise. Nevertheless, some information is available in the sales history data-bases that may provide a complete description of the products purchased by the customers, in terms of the elements (or features) included. The Enterprise Resource Planning (ERP) and supply chain management systems offer data describing the past and the future production environment. In a contractor-supplier context it is about dealing with the problem of the supplier that provides a set of integrated components that vary for each product sold by the contractor.

The decisions made at the design stage impact manufacturing, transport, and usage stages. The challenge is then to find a trade-off between the different life-cycle criteria. To make the right decision, different product designs should be evaluated. The major issue is in the analysis of the scattered data and extracting usable knowledge for the supplier.

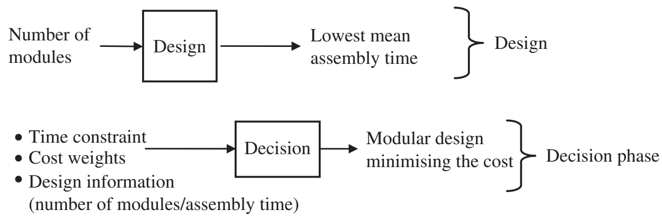

Figure 1. Decision process.

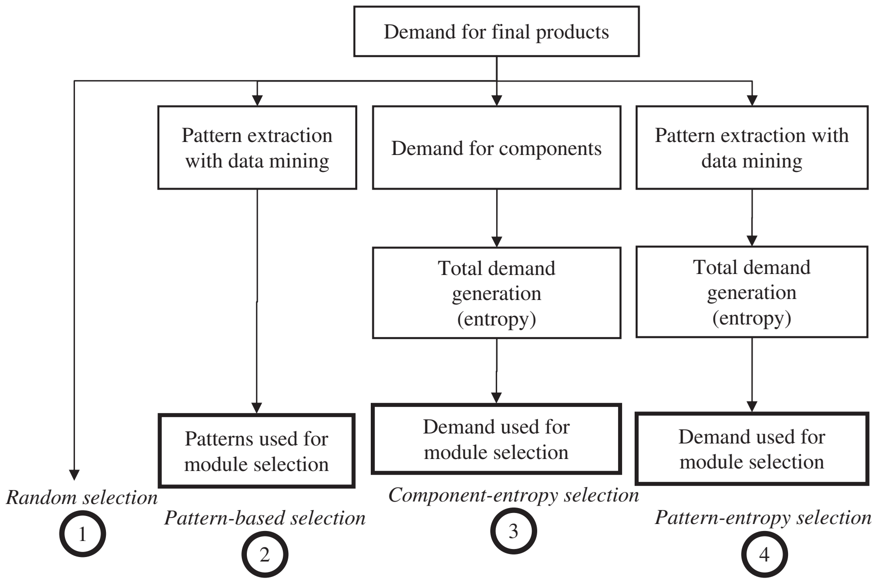

Figure 2. Experiment protocol for the different methods.

Data-mining offers a useful solution to this challenge. The selection of modules should optimise the trade-off between the time to assemble a product and the number of different stocked modules. Therefore, it is important to find a product design that results in the lowest assembly time for a given number of modules. The final decision can then be made, provided that the weights of the criteria are known (Figure 1).

3. Selection of modules

In this section, four different methods are proposed to select the best set of modules to manufacture for the supplier. The supplier receives from the contractor the set of components to provide for each individual product. Some of the selection methods proposed exploit more or less (or not at all) the patterns existing in the data set, describing the sets of components to be manufactured for each product.

The protocol for computational experiments with the four methods is shown in Figure 2. To be able to compare the results, identical demand (same sets of components to be delivered) was used in each experiment.

To test the applicability of the four methods, the demand for final products is generated using predefined rules (representing the hidden patterns in the data sets). These rules can be exclusive or inclusive relationships between components or functions, e.g., $A \Rightarrow \bar{B}$ and/or $A \Rightarrow C$ (see Section 4.1). The random selection used here as the base method for comparison is efficient in time and resources.

In the *random selection* method, the modules to be manufactured are randomly selected. The second method, *pattern-based selection*, involves extracting patterns from product demand data using data mining. The *component-entropy selection* is the third method tested. It uses the knowledge about components' demand to determine modules to stock. The fourth method: *pattern-entropy selection*, involves two stages. At the first stage, the patterns extracted from data are used to generate a coherent final demand. At the second stage, this generated demand is processed to define the modules. The last three methods are discussed in the subsequent sections of the paper.

3.1 *Pattern-based selection*

The *pattern-based selection* method uses rules extracted with data-mining. The existence of patterns (through inclusive or exclusive relations between options and alternatives) permits to restrict the set of potential modules. Actually, only the modules satisfying the rules are considered, then the search space is smaller than with the *random selection* method. The selection of modules to be manufactured is done randomly from this set. Note that, the extraction of rules can be time consuming, depending on the support and confidence required (see Section 4.2). Small values for support and confidence lead to a large number of rules with small probabilities and required more computing time to explore the search space. However, the determination of the modules satisfying these rules and the random picking within this set is not.

The following steps describe the *pattern-based selection*:

(1) Extract association rules describing customers' behaviour from the data set.
(2) The search space for module selection includes all individual component plus modules previously generated with association rules.
(3) Random module selection.

3.2 *Maximum entropy principle*

Da Cunha (2004) stressed the importance of information when selecting modules. The proposed approach (Da Cunha 2004) considered the generation of a demand for final products (from partial information using the maximum entropy principle). However, the information retrieval aspect was not considered.

Entropy was originally used in thermodynamics. It has been found useful as an indicator of chaos in data (Shannon 1948).

The entropy H of a random variable X taking values in $x_i \in I$ is defined as $H(X) = \Sigma_i - p_i \log p_i$, where $p_i = P(X = x_i)$.

Jaynes (1957a 1957b) stated that maximising entropy leads to the least biased probability distribution for given information. Using the principle of entropy maximisation enables generation of a minimally biased demand for the products using partial information.

Denoting p_i the probability that a given order is for product i, the problem of total information generation can be expressed as follows

$$\begin{gathered}\text{Max } H(D) = \sum_i -p_i \log p_i \\ \text{s.t. Available information}\end{gathered}$$

This model can be solved using an approximation method known as Uzawa's algorithm (Fortin and Glowinski 1983).

To illustrate the above model, consider a case where three different basic components a, b, c are used to assemble seven different products (A, B, ..., ABC). Assume that product A contains only component a, product AB contains components a and b, and so on. Only partial information is available, i.e., the probabilities that a demanded product contains a given component.

$$\text{Available information} \begin{cases} P(a) = 0.8 \\ P(b) = 0.8 \\ P(c) = 0.8 \end{cases}$$

To obtain the total information about the demand, the maximum entropy principle is used.

The problem at hand is then

$$\text{Max } H(D) = \sum_i -p_i \log p_i$$

s.t.

$$\begin{gathered} P(A) + P(AB) + P(AC) + P(ABC) = 0.8 \\ P(B) + P(AB) + P(BC) + P(ABC) = 0.8 \\ P(C) + P(AC) + P(BC) + P(ABC) = 0.8 \end{gathered}$$

Uzawa's algorithm determines consistent probability of the demand for the final product.

$$\text{Total information} \begin{cases} P(A) = 0.0343 \\ P(B) = 0.0343 \\ P(C) = 0.0343 \\ P(AB) = 0.1314 \\ P(AC) = 0.1314 \\ P(BC) = 0.1314 \\ P(ABC) = 0.5029 \end{cases}$$

When the demand for final products is available (regardless of whether extracted or generated), the method based on usage frequency performs well with a cost-minimisation criterion.

The two methods presented next, the *component-entropy selection* and the *pattern-entropy selection* use the entropy principle.

3.3 *Component-entropy selection*

The demand information for components can be obtained by monitoring their inventory level. This information at disposal can be used to generate a demand for final products (see the example presented in Section 3.2). The modules are selected using decreasing usage frequency, i.e., the first module selected is the most demanded one, the second module is the second most demanded, and so on.

The following steps describe the *component-entropy selection*:

(1) Generate a demand for final products with maximum entropy principle.
(2) Select modules using decreasing usage frequency.

3.4 *Pattern-entropy selection*

The *pattern-entropy selection* integrates the result provided by data-mining in the deterministic selection of modules and the maximum entropy principle.

This integration is accomplished by the following procedure.

(1) Extract association rules describing customers' behaviour from the data set.
(2) Use of the maximum entropy principle to generate a final demand consistent with the obtained rules.
(3) Select modules using decreasing usage frequency.

To illustrate this procedure, consider the previous three component case (Section 3.2). If Step 1 leads to the following rule: components a and b are exclusive (they are never simultaneously present in a customer demand), Step 2 will allow to find a coherent final demand; a constraint of the optimisation problem of the PME will be $P(AB)+P(ABC)=0$. Then Step 3 will result in the selection of modules with the highest frequency (as the frequency of AB and ABC is 0, these modules will not be chosen).

3.5 *Objective and performance indicator*

To compare the quality of modules selected by the different methods, a common objective is proposed. When the assemble-to-order strategy is implemented in synchronous manufacturing, the delay between an order receipt and the delivery of the requested product is a key point. Indeed, this delay has to verify the terms of the customer-supplier contracts (Calvi *et al*. 2000, Cachon 2003).

The evaluation of different sets of modules is therefore based on the mean final assembly time (T), i.e., the mean time required to assemble a product from the selected set of modules. Selected modules are manufactured and stocked waiting a demand to arrive, and therefore the time to manufacture the modules is not considered. The indicator is defined as follows

$$T=\frac{1}{n}\times\sum_{i=1}^{n}t_i\times p_i$$

In the above expression:

- n is the number of different final products to manufacture,
- t_i is the time of final assembly of product i from the modules in stock,
- p_i is the demand probability for product i.

The worst case scenario (the longest assembly time) could also be considered, besides in the industrial case study considered, a small quantity of modules is manufactured in advance, so the longest assembly time is balanced with smaller ones. Also costs could be considered, the same approach applies but the objective function needs to be changed.

To evaluate the performance of the different methods, an indicator is needed. This indicator should be synthetic (i.e., should represent the performance of a given method for various number of modules) and it should enable a direct gauging of the performance.

The relative performance $Perf(\alpha, i)$ of a method α for a stock size i is computed as follows

$$Perf(\alpha, i) = \frac{t_i^+ - t_i^\alpha}{t_i^+ - t_i^-}$$

where:

- t_i^+ is the maximum time of final assembly for all methods, for a stock size i
- t_i^- is the minimum time of final assembly for all methods, for a stock size i
- t_i^a is the time of final assembly for the method α, for a stock size i

$Perf(\alpha, i) = 1$ if α is the best for a stock size i and 0 if it is the worst method.

Then, the performance of the method α for all stock sizes i from p to q is defined as

$$Perf(\alpha) = \frac{1}{q - p + 1} \sum_{i=p}^{q} Perf(\alpha, i)$$

Based on this indicator, the better performing method has the performance of 1 and the worst the performance of 0.

4. Computational results

Computational tests were performed to evaluate performance of the selection methods. The tests are based on a wire harness design introduced in Agard and Tollenaere (2002). Wire harness is an important assembly in a vehicle. It is a set of wires and connectors transmitting power and information (see Figure 3). The functions (airbag release, power windows, headlights control, etc.) are performed over different wires and connectors.

For each car to be manufactured, the contractor sends the set of relevant elements required to each supplier. Then each supplier ships the set of components needed by each car.

To illustrate the diversity of assemblies, consider a standard wire harness in a middle range car. This wire harness connects 15 different elements. Depending on the silhouette and the engine type, different sets of these elements are used (up to 9). Potential diversity is about 7 million of different wire harnesses for a unique car model (Agard and Tollenaere 2002).

By considering inclusive or exclusive relations between elements, e.g., the passenger air-bag may require the driver air-bag, the potential diversity can be reduced.

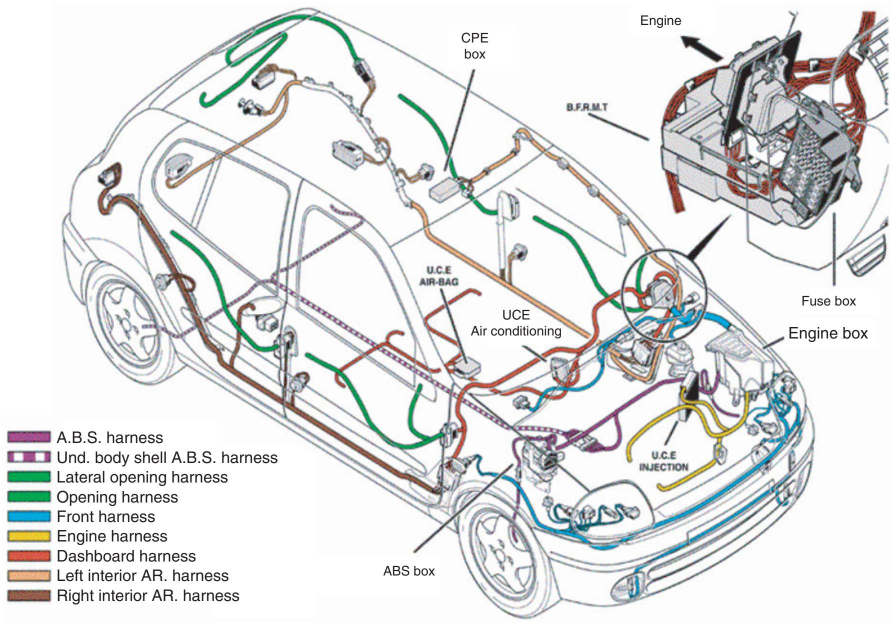

Figure 3. Wire harness in a car (Source: Valeo Connective Systems).

A data set representing a demand was generated (cf. Section 4.1) (i.e., set of components required for each car). Note that this unique demand example illustrates the basic concept, and the results presented here are representative of numerous tests that have been conducted.

A data mining algorithm was used to extract association rules from this data set (cf. Section 4.2). Finally, the four different selection methods were tested (Section 4.3).

4.1 *Input data*

Consider 9 components ($n=9$), i.e., 511 potential final products ($\#pdt=511$). Three rules are given that represent some mechanical or marketing constraints.

- Constraint 1: Components 1 and 2 are exclusive
- Constraint 2: Components 3 and 4 are exclusive
- Constraint 3: Component 6 requires component 8

These relations can be forwarded to the designers and integrated as constraints in the selection of modules. Table 1 summarises the input data used by the different design methods.

Constrains appear in the dataset in the way that: for constraint 1 when 'Comp $1=1$' then 'Comp $2=0$' and when 'Comp $2=1$' then 'Comp $1=0$'. The same applies for constrain 2. To respect constrain 3 then 'Comp $6=1$' then 'Comp $8=1$' (here it is not symmetric).

In the evaluation, the input data are the methods to be evaluated and the demand (the same demands were used by the four methods).

Table 1. Input data for the different design methods.

Random selection	Pattern-based selection	Component-entropy selection	Pattern-entropy selection
The selection process does not consider information about customers' behaviour	Extracted rules from customers' data and the three constraints	Demand generated from the components' demand data and the three constraints using the PME	Demand generated with the extracted rules and the three constraints using the PME

4.2 *Data mining implementation*

Anand and Büchner (1998) referred to data mining as the discovery of non-trivial, implicit, previously unknown, and potentially useful and understandable patterns from large data sets. The patterns extracted from the historical data can assist in designing product families (Agard and Kusiak 2004).

Often many rules are extracted from a database, however, only some of them may be of interest. To identify interesting rules several metrics have been developed. The most common are support, lift, and confidence. For example, for the association rule ($L \Rightarrow R$), the three metrics are defined as follows:

$$\text{Support} = P(L \wedge R)$$
$$\text{Lift} = \frac{P(L \wedge R)}{P(L)P(R)}$$
$$\text{Confidence} = \frac{P(L \wedge R)}{P(L)}$$

Consider 100 sale records with 50 of them having the characteristic L, 20 with the characteristic R, 10 records having both. Then the rule ($L \Rightarrow R$) has support $= 10/100 = 0.1$, lift of $(10/100)/(50/100 \times 20/100) = 1$, and the confidence of $(10/100)/50/100 = 0.2$.

The support reflects the frequency of (L and R) to be observed in the dataset. Confidence reflects the ratio of (L and R) when L is present. High support and confidence give strong rules; this means that there is a strong relation between L and R. A lift smaller than 1 implies that $L \Rightarrow \text{not}(R)$. For the same support, a rule with a higher confidence proves a stronger relation.

TANAGRA software (Rakotomalala 2005) was selected as the rule extraction tool. The association rule algorithm used in the paper is presented in Agrawal and Srikant (1994).

A data set with 250 000 demand configurations was generated (TANAGRA's limit on the data set size for extraction of association rules). A configuration was represented as a binary-vector, representing the presence (1) or absence (0) of a component (see Table 2). A generated demand vector represents an incoming order, and therefore it satisfies the constraints described in Section 4.1. A vector was generated as follows: each dimension (from 1 to 9) was the result of a random selection (a number from the interval[0;1]). If the variable was greater than 0.5, then the dimension was assigned value 1, if not 0. Non-feasible vectors were discarded.

Table 2. Examples of demand configurations.

Comp 1	Comp 2	Comp 3	Comp 4	Comp 5	Comp 6	Comp 7	Comp 8	Comp 9
0	1	0	1	0	0	1	0	**0**
0	0	1	0	1	0	1	1	**0**
1	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**1**

Table 3. Examples of extracted rules.

No	Antecedent	Consequent	Lift	Support	Confidence
1	'comp 2 = 1'	'comp 8 = 1'	1	0.574	**0.967**
2	'comp 3 = 0'	'comp 8 = 1'	1	0.58	**0.968**
3	'comp 1 = 0'	'comp 8 = 1'	1	0.613	**0.969**
4	'comp 3 = 0'	'comp 1 = 0'	2	0.586	**0.978**
5	'comp 3 = 0'	'comp 9 = 0'	1	0.590	**0.984**
6	'comp 4 = 1'	'comp 1 = 0'	2	0.567	**0.988**
7	**'comp 4 = 1'**	**'comp 3 = 0'**	**2**	**0.571**	1

Some demand configurations appeared more than once to reflect identical customer behaviour. For the data set, a 250,000 × 9 matrix, the product-mix was established and the weight (or probability of demand) of each product (being the corresponding number of orders divided by 250,000) was obtained.

4.3 *Results*

4.3.1 *Extracted rules*

The data described in Table 2 was mined to extracted rules. Support of 0.5 and confidence of 0.95 were chosen after different experiments. The extracted rules were post-processed for redundancy. Some of the generated association rules were trivial, even in the presence of high support and confidence. Domain expertise should be an integral part of the data mining process to eliminate trivial rules and determine potential associations that should be tested. Table 3 shows examples of the extracted rules.

For example, line number 1 on Table 3 means that if *component* 2 is present then *component* 8 is also present, with a Lift = 1, Support = 0.574 and Confidence = 0.967.

After the domain expert analysis only 13 rules remained. It considers only positive associations ('if comp $x = 1$ then comp $y = 1$') with the largest support.

Those 13 rules are used in two methods: the *pattern-based selection* and the *pattern-entropy selection*.

4.3.2 *The method of interest*

If the supplier requires a specific maximum mean final assembly time, for synchronous delivery for example, then it is possible to compute the minimal set of modules with the different methods.

Table 4. The number of modules guaranteeing final assembly in 0.5 time units.

Random selection	Pattern-based Selection	Component-entropy selection	Pattern-entropy selection
288	150	35	24

For example, for the expected mean assembly time of 0.5 time units, all four methods can be used to find a set of modules that will meet this constraint. Table 4 summarises the number of modules generated by the four methods when the maximal mean final assembly is 0.5 time units.

The *random selection* method determines a stock of 288 modules so that the final assembly time would not exceed the limit of 0.5 t.u.

Data mining extract rules describing the customers' behaviour. This information can be used to improve the selection of the modules. With the *pattern-based selection* the random selection of the modules is then limited to the ones that respect the rules. For example, rule 1 (cf. Table 3) reduces the potential stocked modules to the ones that satisfy the implication (component 2 $\Rightarrow$ component 8), i.e., modules that include component 8 but not component 2 will not be stocked. Such rules restrict the selection of modules respecting the characteristics of the final products sold; therefore, the final assembly time constraint is met while stocking fewer modules. That method proposes 150 modules for a final assembly time of 0.5 t.u.

To eliminate the random characteristic within the selection of the modules, a choice based on the demand for final products was selected. The maximum entropy enables to generate final demand configurations that are consistent with the partial information at disposal. If the only information available is the demand for components (obtained from the record of components' stock depletion), then the *component-entropy selection* method produced 35 modules. Using both, the information about the demand for the different components and the rules extracted, the demand generated is closer to the actual one. The stock determined with the *pattern-entropy selection* method is 24 modules.

The results produced by the four methods are of importance. More modules lead to additional costs. The relationship between the assembly time and the supporting number of modules is a powerful tool for decision makers. It can be used to negotiate lateness fees. Finding a balance between the lateness cost and savings owing to reduced stock is an interesting problem.

It appears that the more knowledge is exploited from a data set, the smaller the number of modules is.

4.3.3 *Performance evaluation*

For the *pattern-based selection* method, the modules to manufacture were randomly selected within the set of modules that satisfy the data mined rules (cf. Section 4.3.1). The assembly time generated by this method cannot be guaranteed as the selection of modules to manufacture is random.

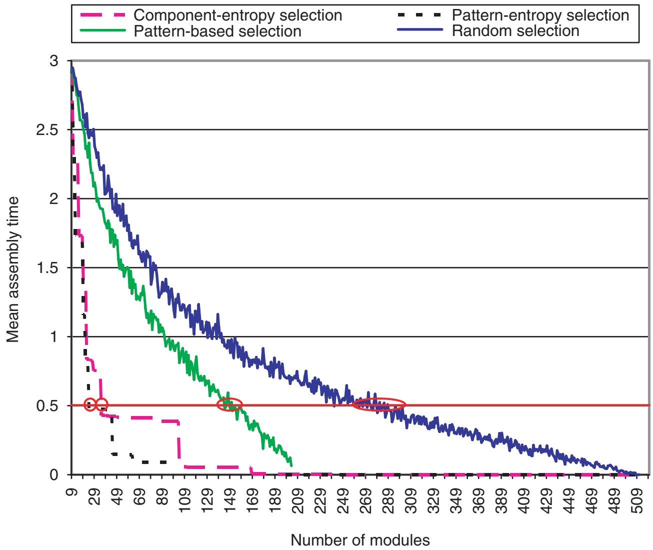

Figure 4. Results generated by the four methods.

To compare the *random selection* and *pattern-based selection* methods with the other two determinist methods the average assembly time $\overline{T_i}$ for a stock size equal to i was considered (100 tests have been done for each stock size).

$$\overline{T_i} = \frac{1}{100}\sum_{j=1}^{100} T_{i,j}$$

The four different methods are compared in Figure 4: *random selection*, *pattern-based selection*, *component-entropy selection*, and *pattern-entropy selection*. For every stock size from 9 to 511 (that represent the number of modules that are manufactured), a selection of modules is realised that respects the final assembly time.

Whatever the method is, larger number of modules lead to smaller times of final assembly. It is easy to see that if the time available for the final assembly is 0.5, we get back the results described in Table 4, 24 modules for the *pattern-entropy*, 35 modules for the *component-entropy*, around 150 modules for the *pattern-based* and around 280 for the *random selection*. Unlike the *pattern-based selection* method, the *component-entropy* and *pattern-entropy* do not incorporate randomness, i.e., when used with the same data, they always provide the same results.

The curves clearly convey the difference between the systematic and random methods. Indeed, the irregularity observed for the curves of *pattern-based selection* and *random-selection* are induced by the use of a sample of 100 random iterations.

The values of the synthetic performance indicator (Section 3.5) for the different methods are given in Table 5.

The *pattern-based* method takes advantage of information provided by data mining. Modules represent patterns in customers' requirements; the stronger the rule is the more products could be satisfied, then the mean time for final assembly will decrease

Table 5. Performance of the methods subject to the demand type.

Random selection	Pattern-based selection	Component-entropy selection	Pattern-entropy selection
0	0.75	0.95	0.99

more rapidly. If small values are selected for support and confidence when extracting rule, more rules will lead to more modules to explore. An extreme will be to consider support and confidence equal zero, the search space will then consider all possible modules (without any patterns), which leads to the *random selection* method, and less performance.

The *component-entropy* selection takes advantage of the maximum entropy principle; this proves to be very efficient in such problems, components and module with high probabilities (in the generated demand) are favourites. The *pattern-entropy* selection exploits both patterns extraction with data mining and maximum entropy principle; this last methods leads to fewer modules for the same level of performance.

The last two methods prove to perform the best. The more information about the demand is exploited, the more efficient is the selection of modules.

5. Conclusion

In this paper, four different methods were developed to select the best set of modules that permit to decrease the mean final assembly time for a product family in an assembly-to-order context.

As some information about required component is available, data-mining was applied to improve the selection of modules. Using association rules to extract patterns from the demand in the selection of modules allowed an important decrease in the number of modules to manufacture (and still insure the same performance in the final assembly process). The results presented in the paper demonstrated that the integration of the association rules may generate significant savings. Depending on the parameters selected for association rule extraction (support and confidence), some rules may be omitted as well as non real rules may appear. This will influence the number of proposed modules, further study should consider the influence of those parameters when extracting rule, on the pattern-based methods.

Entropy maximisation was proposed to enrich the selection of the modules. This approach gave excellent results. A combination of association rules and entropy maximisation could even improve previous results.

The information used by data mining is usually created by other applications. Such information is routinely collected in many industries, e.g., the information about each part of an aircraft, from design to production, is stored as long as the plane operates. The investment needed to perform data-mining is therefore relatively small in comparison to the potential benefits.

Further research should consider the integration of the extracted knowledge at the early stages of modular design to provide the best product customisation. The selection of the appropriate module stock is not the only issue that has to be considered when implementing the assembly-to-order policy. Indeed, this change in production policy has to be aligned with other organisational measures. The characterisation of customers' behaviour proved to be of major interest for solving technical problems. For example, the consideration of dates and times of sales would permit for precise description of customers' behaviour. This information could be useful in designing promotions and work schedules.

References

Agard, B. and Kusiak, A., 2004. Data-mining based methodology for the design of product families. *International Journal of Production Research*, 42 (15), 2955–2969.

Agard, B., and Tollenaere, M., 2002. Design of wire harnesses for mass customization, *4th International Conference on Integrated Design and Manufacturing in Mechanical Engineering – IDMME 2002*, 14–16 May, Clermont-Ferrand, France.

Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules in large databases. *Proceedings of 20th International Conference on Very Large Data Bases*, 12-15 September, Santiago de Chile, Chile. Morgan Kaufman. pp. 478–499.

Anand, S. and Büchner, A., 1998. *Decision Support Using Data Mining*. London: Financial Times Pitman Publishers.

Briant, O. and Naddef, D., 2004. The optimal diversity management problem. *Operations Research*, 52 (4), 515–526.

Cachon, G., 2003. Supply chain coordination with contracts. *In*: S. Graves and A. de Kok, eds. *Supply chain management: design, coordination and operation*. Boston, MA: Elsevier.

Calvi, R., le Dain, M.A., and Harbi, S. 2000. Designing a performance measurement system to manage the customer-supplier partnership: a structured approach, *2nd International Conference on Performance Measurement*, Cranfield School of Management, Cambridge, pp. 19–21.

da Cunha, C., 2004. *Définition et gestion de produits semi-finis en vue d'une production de type assemblage à la commande*. Thesis (PhD). Institut National Polytechnique de Grenoble, Grenoble, France.

Dahmus, J., Gonzalez-Zugasti, J., and Otto, K., 2001. Modular product architecture. *Design Studies*, 22 (5), 409–424.

Fortin, M. and Glowinski, R., 1983. *Augmented Lagrangian Methods*. Amsterdam: North-Holland.

Gavirneni, S., Kapuscinski, R., and Tayur, S., 1999. Value of information in capacited supply chain. *Management Science*, 45 (1), 16–24.

Gonzalez-Zugasti, J., Otto, K., and Baker, J., 1998. A method for architecting product platforms with an application to interplanetary mission design. *Proceedings of 1998 DETC ASME Design Automation conference*, Atlanta, GA.

Gonzalez-Zugasti, J., Otto, K., and Baker, J., 2000. A method for architecting product platforms. *Research in Engineering Design*, 12 (2), 61–72.

Huang, CC. and Kusiak, A., 1998. Modularity in design of products and systems. *IEEE Transactions on Systems, Man and Cybernetics*, 28 (1), 66–77.

Jaynes, E., 1957a. Information theory and statistical mechanics. *The Physical Review*, 106 (4), 620–630.

Jaynes, E., 1957b. Information theory and statistical mechanics ii. *The Physical Review*, 108 (2), 171–190.

Kocher, I. and Rolland, P., 1995. Quand les structures dessinent les produits. *Gérer et Comprendre*, 39 (June), 38–48.

Kusiak, A., 1999. *Engineering design: products, processes, and systems*. San Diego, CA: Academic Press.

Lee, H. and Whang, S., 2000. Information sharing in a supply chain. *International Journal of Technology Management*, 20 (3/4), 373–387.

Pine, B., 1993. *Mass Customization: The new frontier in business competition*. Boston, MA: Harvard Business School Press.

Rakotomalala, R., 2005. TANAGRA: un logiciel gratuit pour l'enseignement et la recherché. In *5èmes Journées d'Extraction et Gestion des Connaissances, EGC-2005*. France: Paris, pp. 697–702.

Rao, U., Swaminathan, J., and Zhang, J., 2004. Multi-product inventory planning with downward substitution, stochastic demand and setup costs. *IIE Transactions*, 36 (1), 59–71.

Shannon, C., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27 (October), 379–423, 623–656.

Shapiro, B., 1979. Making money through marketing. *Harvard Business Review*, 57 (4), 135–142.

Simpson, T.W., Siddique, Z., and Jiao, R.J., eds, (2006). Product platform and product family design: Methods and applications. New York: Springer.

Starr, M., 1965. Modular production: a new concept. *Harvard Business Review*, Nov.–Dec., 131–142.

Swaminathan, J. and Tayur, S., 1998. Managing broader product lines through delayed differentiation using vanilla boxes. *Management Science*, 44 (12), 161–172.

Tarondeau, J.C., 1998. *Stratégie industrielle*. 2nd ed. France: Vuibert.

Zamirowski, E. and Otto, K., 1999. Product portfolio architecture definition and selection. *International Conference on Engineering Design,* ICED-99. Munich, Germany, 1515–1518.