CART for supply chain simulation models reduction

Philippe Thomas, Marie-Christine Suhner, André Thomas

▶ **To cite this version:**

HAL Id: hal-01387317

https://hal.inria.fr/hal-01387317

Submitted on 25 Oct 2016

CART for Supply Chain Simulation Models Reduction:

Application to a Sawmill internal Supply Chain

Philippe THOMAS[1,2], Marie-Christine SUHNER[1,2], André THOMAS[1,2]

[1]Université de Lorraine, CRAN, UMR 7039, Campus Sciences, BP 70239, 54506 Vandœuvre-lès-Nancy cedex, France
[2]CNRS, CRAN, UMR7039, France
{philippe.thomas, marie-christine.suhner, andre.thomas}@univ-lorraine.fr

Code de champ modifié

Abstract. Evaluation of supply chain or workshop management is often based on simulation. This simulation task needs models which are difficult to design. The aim of this work is to reduce the complexity of simulation model design and to partially automate this task by combining discrete and continuous approaches in order to construct more efficient and reduced model. Model design focuses on bottlenecks with a discrete approach according to the theory of constraints. The remaining of the workshop is modeled in a less precise way by using continuous model in order to describe only how the bottlenecks are fed. This used continuous model is a regression tree algorithm. For validation, this approach is applied to the modeling of a sawmill workshop and the results are compared with results obtained previously by using a neural network model.

Keywords: Decision tree; CART; reduced model; simulation; supply chain

1 Introduction

Planning or scheduling scenario evaluation is an important issue for internal/external Supply Chains (SC) control. Simulation is useful to perform this evaluation and allows to highlight evolution of the resources states, work in process and queues. This information allows to perform a "Predictive scheduling" [10] which concerns MPS which is initially established with the Manufacturing Planning and Control System. This first approach must not be confused with "Reactive scheduling" which gives new MPS solution, established after significant events occur on the shop-floor. The real time systems performing production reporting lead to obtain information into the management system very quickly [6]. However, it is difficult to use this huge of information to take a decision [13, 14]. At this level of planning, load/capacity balancing is obtained via the "management of critical resource capacity" function or Rough-Cut Capacity Planning which essentially concerns bottlenecks [19]. Goldratt and Cox [4] propose to manage all the supply chain by bottlenecks control and call it Theory of Constraints (TOC). For this purpose, the use of dynamic discrete event simulation of material flow is helpful [16]. Simulation models of real industrial cases are very com-

plex and can lead to problems of scale [12] that is why it is interesting to use simplest (reduced/aggregated) models of simulation [2, 3, 20]. To design these reduced model, different approaches may be used as using continuous flow model to approximate discrete manufacturing environments [5, 15], or using metamodels (linear regression, splines, Kriging…) in order to perform simulation model [7]. Neural networks are also been used to perform this task [17, 18]. Another tool, regression tree, is able to find a suitable model useful for the model reduction. Classification and regression trees are machine learning methods able to fit a model to data. Models are obtained by partitioning data space and thanks to the association of a simple prediction model to each subspace [1]. The main goal of this paper is to evaluate the ability of regression trees to be used in order to reduce simulation model. In the second part, the proposed approach of reduction model and the regression trees are presented. The third part will be devoted to the presentation of an industrial application which is a sawmill flow shop case. In order to evaluate the pertinence of the proposition, the reduced simulation model of sawmill internal SC case is presented and, before to conclude, is compared to the same work performed by using neural network in the last part.

2 The reduced model

2.1 The algorithm

Model complexity is relative to the number of elements, connections and model calculations [21]. Different approaches have been proposed to reduce model. To see a brief overview, see [18]. The proposed reduction algorithm is a modification of those presented by Thomas and Thomas [17]. The proposed approach is based on the association of discrete event models and continuous models (regression tree) in order to design a simulation model. Our objective is to maximize the bottleneck utilization rate and, at the same time, simplify simulation model design for modelers. Its main steps are recalled and explained below:

1. Identify the structural bottleneck (work center (WC) which for several years has been mainly constrained in capacity).
2. Identify the conjunctural bottleneck for the bundle of Manufacturing Orders (MO) of the MPS under consideration.
3. Among the WC not listed in 1 and 2, identify the one (synchronisation WC) satisfying these two conditions:
   (a) present at least in one of the MO using a bottleneck,
   (b) widely used considering the whole MO.
4. If all MO have been considered go to 5 if not go to 3.
5. Use regression tree for modelling the intervals between WC that have been found during preceding steps (figure 1).

Work Centers (WC) remaining in the model are either conjunctural or structural bottlenecks or WC which are vital to the synchronization of the MO. Other WC are incorporated in "aggregated blocks" upstream or downstream of the bottlenecks [18].
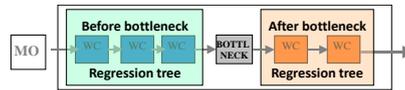
Fig. 11. Reduction model algorithm

2.2   Regression tree

Decision trees are sequential models, which logically combine a sequence of simple tests. A regression tree looks like a classification tree, except that the output variable is a continuous one and a regression model is fitted to each node to give the predicted value of the output [9]. The first regression tree (AID) has been proposed by Morgan and Sonquist [11]. The principle of this regression tree is to begin to the root node, and then, for each node, find the test on the input patterns which minimizes the sum of the node impurities in its two child nodes. Classification and Regression Tree (CART) algorithm works on the same principle [1]. The main difference between AID and CART relies with the pruning and estimation process. CART is based on the generalization of the binomial variance called Gini index. The growing procedure of the tree is recursive and the process is stopped when [8]:

— there is only one observation in each of the child nodes,
— all observations within each child node have the identical distribution of predictor variables, making splitting impossible,
— an external limit on the number of levels in the tree or on impurity decrease level threshold is reached.

The resulting tree presents generally overfiting problem and needs a pruning phase, based on the impurity term used during the growing phase (Gini) associated to a penalty term corresponding to the number of terminal nodes. The final tree selection is performed by using a cross-validation procedure [1].

3   Overview of the sawmill

The proposed approach is applied to build a simulation model of a sawmill workshop. In this actual case, managers need a tool to help them in their weekly decision-making Master Production schedule (MPS) process. They want:

— to evaluate the effectiveness of its MPS,
— to maximize its load rate, and so, its global productivity,
— to explain some unexplained congestion phenomena of the trimmer WC.

A first work [16] with a complete model has allowed to represent the congestion phenomena and to use this representation in order to improve the load rate. This model, has showed that a bottleneck load rate too high (higher than 60%) degrades productivity of bottleneck, and so, sawmill productivity. The difficulty is that bottleneck is the last sawmill WC but all influent factors on bottleneck productivity depend of the first

WC. However, this complete model is unusable on a real case for the MPS dynamic evaluation because of time needed to modify it.

Sawmill can be described from a process point of view. It can be represented by two linear parallel flows for main and secondary products. This fact, associated to variation of log dimensions lead process to be non-linear. Therefore, the physical industrial production system can be divided into three main parts. To explain the process, the trip of a log from its admission into the process to its exit in planks form will be described. The first process part is the Canter line presented figure 2. Dashed arrows indicate the products flow. The considered log is taken into the process by using conveyors RQM1, RQM2, and RQM3. According to its characteristics (determined by scanner MS), the log is driven to conveyors RQM4 or RQM5, which are used as input inventory for the Canter line. After that, logs go on the first canter's machine and later on the CSMK saw, which transforms logs into square-shaped parallelepipeds.
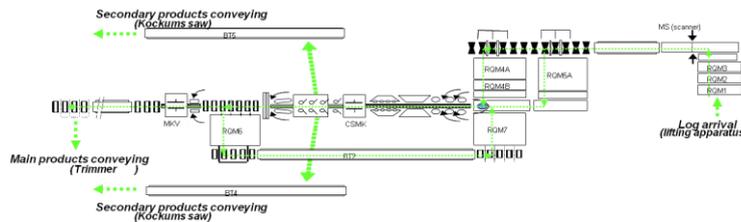

Fig. 2. The Canter line
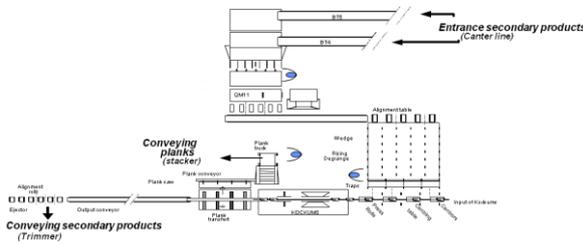
Fig. 3. The Kockums line

This step gives the two first sides of the parallelepipeds (the added value on the log) and produces two planks (secondary products), which are taken out of the Canter line by the BT4 and BT5 conveyors. The log is then driven on the RQM6 conveyors, rotated 90°, and stored in RQM7 awaiting its second passage on the CSMK saw. After the second passage, the squared is completed, and two other secondary products are taken out of the Canter line by BT4 and BT5 toward the second part of the process, the Kockums line. The squared log is cut on the MKV saw into three planks (called main products). These main products are driven to the third part of the process, the trimmer line.
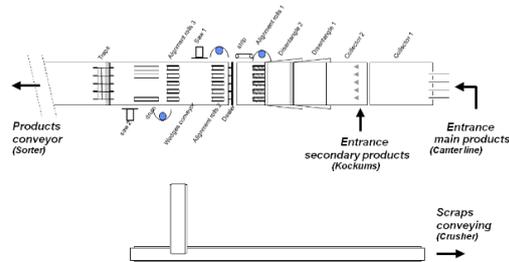
Fig. 4. The trimmer line

Figure 3 shows the second part of the process, where the main machine is the Kockums saw. Only secondary products are driven on this part. As previously said, the secondary products are taken in the line by the BT4 and BT5 conveyors. They are cut by the QM11 saw, after which they reach the Kockums saw, which optimizes the planks according to the products needed. The alignment table is used as the input inventory of the Kockums saw. The secondary products are finally sent to the third part of the process by the exit conveyor.

The third part of the process is the trimmer line, which is presented in figure 4. This line performs the final operation of cross cutting. This operation consists in cutting up products to length. The inputs of the line are from collectors 1 and 2, which collect the secondary and main products from Kockums and Canter lines respectively. Saw 1 is used to perform default bleeding and Saw 2 cuts up products to length. A previous work [16] has shown that this last machine, the trimmer saw, is the bottleneck of the entire process, and, as said previously, the productivity of the trimmer depends to the decisions taken on the canter work center. So, the impact of bad decisions are seen too much late to be corrected. So in order to evaluate decisions, managers need a simulation tool.

4   The simulation models

As said previously, the bottleneck of this line is the trimmer. Consequently, modeling the function of the inventories RQM4, RQM5, and RQM7, and of the Canter and Kockums lines is unnecessary and gave no direct and useful information for the evaluation of the MPS. In fact, only the arrival times of the products in the trimmer queue are useful for simulating the load of this bottleneck. So, in the complete model presented figure 5, all the parts surrounded with dashed line is unnecessary and may be replaced by a continuous model, here a regression tree.

According to the modeling process recalled in part 2, the specific sawmill model could be designed. To build this model, we need to identify the input variables. Thomas *et al*. [18] collected the available input data which can be classified into three categories: Data related to the products (here the logs), data related to the process and data related to the bill of material or routing (here the cutting plan). The data related

to the products are mainly dimensional ones as length (lg) and three values for timber diameters (diaPB, diaGB, and diaMOY). The data related to the process are the process variables collected at the time of log arrival. In particular, we require the input stock and the utilization rate of the bottleneck, here the trimmer (Q_trim, and U_trim, respectively). The number of logs present in the process between the inputs and the exit of the Canter line (Q_RQM) and in the input inventories (Q_RQM4, Q_RQM5 and Q_RQM7) are needed.
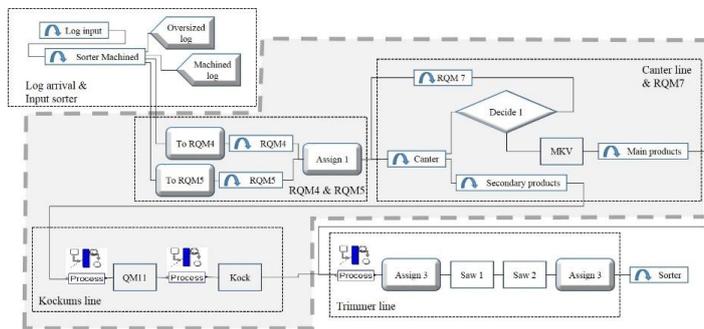

Fig. 55. Complete model

The data related to the routing correspond here to the information related to the cutting plan of the logs which must be cut into main and secondary products. This information is given by the type of part (T_piece).

Consequently, the model input variables are: Lg, diaGB, diaMOY, diaPB, T_piece, Q_trim, U_trim, Q_RQM, Q_RQM4, Q_RQM5 and Q_RQM7. In our application, 12775 products are simulated with the complete model. These data are used to fit the behavior of the reduced model to the complete one which serve as reference model.

The next step is to identify the output variable. Our objective is to estimate the delay (ΔT) corresponding to the duration of the throughput time for the 12775 products. ΔT is measured between the process input time and the trimmer queue input time. For regression problems, the goal is to reduce the distance between the model and the data. Hence, the RMSE criterion is suitable, and is used in this work:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \hat{y}_n)^2} \tag{1}$$

where N denotes the number of data, y_n is the nth actual data point, and $\hat{y}_n$ is the predicted value.

Figure 6 presents the evolution of the RMSE in function of the number of nodes in the tree. This figure shows that classically, even if RMSE always decreases on the learning data set, it is not the same on the validation data set. It begins by decreasing before to increase. This fact illustrates the overfitting problem and the selection of the

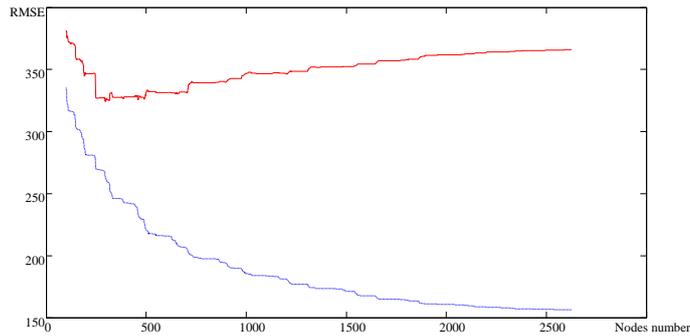best model must be performed by using a cross validation strategy. In our case, the selected model comprises 296 nodes.

Fig. 6. Evolution of RMSE in function of nodes number – identification data set in blue dashed – validation data set in red

This best model must be compared with those obtained in preceding work which use neural network model [18]. These results are presented table 1. They show that regression tree outperforms neural model and allows to improve results of 20%.

RMSE	Identification	Validation
Neural model	408.45	413.93
Tree model	263.47	324.10

Table 1. Comparison of RMSE for Neural network and regression tree models

5   Conclusion

The use of regression tree in order to build a reduced simulation model is investigated here. The regression tree model is used to model the functioning of a part of the process that is not constrained in capacity. This approach has been applied to the modeling of a sawmill workshop. The results show that the complete and reduced models gave similar results even if the log arrival rule is changed. Moreover, the results are compared with results obtained with a neural network model and the comparison shows that regression tree outperforms neural network on this application. This means that it seems efficient to use a regression tree to model a part of a process instead of constructing the complete model. Assuming that the construction of a regression tree is a quasi-automated task, in which the modeler only collects and selects the input data set. It is faster and easier to design this kind of reduced model. This approach allows the modeler to focus on the management of bottlenecks. Our intentions for future works are the validation of this approach on different applications, particularly

on several external supply chains, such that at least one particular enterprise belongs to different supply chains.

6 References

1. Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984. *Classification and regression trees*, Chapmann & Hall, Boca Raton, USA.
2. Brooks, R.J., Tobias, A.M., 2000. Simplification in the simulation of manufacturing systems. *International Journal of Production Research*, 38(5): 1009-1027.
3. Chwif, L., Paul, R.J., Pereira Barretto, M.R., 2006. Discret event simulation model reduction: A causal approach. *Simulation Modelling Practice and Theory*, 14: 930-944.
4. Goldratt., E., Cox, J., 1992. *The Goal: A process of ongoing improvement*, North River Press; 2nd Revised edition, Great Barrington, USA.
5. Ho, Y.C., 1987. Performance evaluation and perturbation analysis of discrete event dynamics systems. *IEEE Transaction on Automatic Control*, 32(7), 563–572.
6. Khouja, M., 1998. An aggregate production planning framework for the evaluation of volume flexibility. *Production Planning and Control*, 9(2), 127–137.
7. Kleijnen, J.P.C., Sargent, R.G., 2000. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120, 14–29.
8. Lewis R.J., 2000. An introduction to classification and regression tree (CART) analysis. *Annual Meeting of the Society for Academic Emergency Medicine*, San Francisco, California, May 22–25
9. Loh, W.Y., 2011. Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 14–23.
10. Lopez, P., and Roubellat, F. 2001. *Ordonnancement de la production*, Hermès, Paris.
11. Morgan, J.N., Sonquist, J.A., 1963. Problems in the analysis of survey data, and a proposal. *J. Am. Stat. Assoc.*, 58, 415-434.
12. Page, E.H., Nicol, D.M., Balci, O., Fujimoto, R.M., Fishwick, P.A., L'Ecuyer, P., and Smith, R., 1999. An aggregate production planning framework for the evaluation of volume flexibility. *Winter Simulation Conference*, 1509-1520.
13. Pritsker, A., and Snyder, K., 1994. Simulation for planning and scheduling, *APICS*, August.
14. Roder, P., 1994. Visibility is the key to scheduling success, *APICS Planning and Scheduling*, August
15. Suri, R., Fu, B.R., 1994. On using continuous flow lines to model discrete production lines. *Discrete Event Dynamic Systems*, 4, 129-169.
16. Thomas, A., Charpentier, P., 2005. Reducing simulation models for scheduling manufacturing facilities. *European Journal of Operational Research*, 161(1): 111-125.
17. Thomas, P., Thomas, A., 2011. Multilayer Perceptron for Simulation Models Reduction: Application to a Sawmill Workshop. *Engineering Applications of Artificial Intelligence*, 24, 646-657.
18. Thomas, P., Thomas, A., Suhner, M.C., 2011. A neural network for the reduction of a Product Driven System emulation model. *Production Planning and Control*, 22, 767-781.
19. Vollmann, T.E., Berry, W.L., Whybark D.C., 1992. *Manufacturing, Planning and Systems Control*. The Business One Irwin.
20. Ward, S.C., 1989. Argument for constructively simple models. *Journal of the Operational Research Society*, 40(2): 141-153.
21. Zeigler, B.P., 1976. *Theory of modelling and simulation*. Wiley, New York.