

## **A SYSML-BASED SIMULATION MODEL AGGREGATION FRAMEWORK FOR SEEDLING PROPAGATION SYSTEM**

Chao Meng  
Sojung Kim  
Young-Jun Son

Chieri Kubota

Systems & Industrial Engineering  
The University of Arizona  
Tucson, AZ 85721, USA

School of Plant Sciences  
The University of Arizona  
Tucson, AZ 85721, USA

### **ABSTRACT**

This paper proposes a Systems Modeling Language (SysML)-based simulation model aggregation framework to develop aggregated simulation models with high accuracy. The framework consists of three major steps: 1) system conceptual modeling, 2) simulation modeling, and 3) additive regression model-based parameter estimation. SysML is first used to construct the system conceptual model for a generic seedling propagation system in terms of system structure and activities in a hierarchical manner (i.e. low, medium and high levels). Simulation models conforming to the conceptual model are then constructed in Arena. An additive regression model-based approach is proposed to estimate parameters for the aggregated simulation model. The proposed framework is demonstrated via one of the largest grafted seedling propagation systems in North America. The results reveal that 1) the proposed framework allows us to construct accurate but computationally affordable simulation models for seedling propagation system, and 2) model aggregation increases the randomness of simulation outputs.

### **1 INTRODUCTION**

Grafting is a horticultural technique whereby tissues from one plant are joined with another to obtain a combination of rootstock and scion genotypes that are more desirable than those contained in a single plant. The advantage of the vegetable grafting is to improve fruit yield and quality as well as to reduce the environmental impact through reducing chemical fumigants utilized in traditional cropping systems. According to Kubota et al. (2008), many seedling propagators in North America have strong interests in introducing grafting into their seedling propagation systems. For a seedling propagation system, uncertainties and complexity involved in high resource (e.g. machines and labors) management and in environmental factors (e.g. humidity and temperature) make the system performance difficult to predict. Discrete event simulation has become one of the most used analysis tools for large scale systems like seedling propagation system because it is one of the few tools that can take randomness into account, and it can address aggregate as well as very detailed models. A simulation model representing a seedling propagation system can be used to predict system performance, supporting various decisions in design or operation (e.g. production scheduling in a peak season) of the seedling propagation system. In simulation, fidelity refers to the faithfulness with which model behavior reflects modeled system behavior (Kim, McGinnis and Zhou 2012). It has been recognized that the level of detail of simulation has a significant impact on the fidelity (Persson 2002; Vasudevan and Devikar 2011; Venkateswaran and Son 2004). For example, Venkateswaran and Son (2004) demonstrated that the dynamics and inventory level of supply chain are highly dependent on the level of detail of simulation models even though the models simulate the same process. One principal challenge that we faced while developing a highly detailed simulation model (i.e. low level simulation model) for the seedling propagation system in this work, was the low execution speed due to the large number of entities in the model. Such high computational overhead issues

have also been recognized by Harmonosky and Robohn (1995). It is important to develop a simulation model with an appropriate accuracy without being overly detailed to avoid wasting computational resources (Huang et al. 2009).

One way to balance the execution speed and fidelity is to aggregate the model. There are two primary issues that need to be resolved for simulation model aggregation. One is to narrow down the set of decision variables. The other one is to reduce the total number events during simulation execution. For the first issue, large number of decision variables not only increases the computation time of simulation-based optimization, but also sometimes prevents users from focusing on more significant decision variables. Therefore, using sensitivity analysis techniques such as ANOVA to exclude insignificant ones from the set of decision variables and to aggregate several decision variables can help decision makers or optimization programs find the optimal or near optimal decisions more efficiently (Webster and Padgett 1984; Cassandras et al. 2000; Rodriguez et al. 2008). For the second issue, the number of simulation events should be reduced by conceptually aggregating multiple activities into one (Madan et al. 2005; Huang et al. 2009). This is often achieved by aggregating model building blocks into one or a few and objects into that of a larger size during the modeling stage. In this way, the simulation provides what would happen “on the average” instead of modeling all the detailed activities to look into what would happen “explicitly”. One important requirement for such aggregation is that the aggregated model must maintain the same behaviors of the low level model. Since the first issue has been resolved using sensitivity analysis approach in Webster and Padgett (1984), this paper mainly focuses on resolving the second issue.

Although simulation model aggregation has been addressed in the past, the literature lacks formal modeling approaches to simulation model aggregation. Without a formal system conceptual model, the practice of model aggregation is *ad hoc* (Madan et al. 2005). Currently, commonly used languages in systematic and formal information modeling for simulation include EXPRESS, IDEF, UML and SysML (Son et al. 2003; Wang et al. 2011; Meng et al. 2013; Cetinkaya et al. 2010). In addition to the properties shared with other formalisms, SysML enables modelers to efficiently capture functional and performance requirements via Requirement diagram and to precisely define performance and quantitative constraints via Parametric diagram. Besides, a common practice for estimating parameters for a higher level model (aggregated model) available in the literature is to mostly use the mean values of lower level model parameters. Since the inputs and outputs of a complex system tend to have highly nonlinear relationships, the high level model based on this approach is likely to result in inaccurate outputs, where significant statistical information is lost during the aggregation process (Cassandras et al. 2000). Multivariate nonlinear regression-based approach (e.g. additive regression model) is known to be capable of accurately identifying both linear and nonlinear representation of predictors and responses (Fox 2000). This category of statistical approach can be used to estimate the aggregated model parameters via conducting simulation experiments. Since only limited information about the relationships between predictors and responses is available before conducting the regression analysis, obtaining an accurate form of nonlinear regression model is computationally expensive in terms of the required sample size. Therefore, an additive regression model with nonlinear partial-regression function can be used as it is also able to accurately estimate the relationships between predictors and responses, but is less computationally demanding. In this paper, a SysML-based simulation aggregation framework is proposed for developing an aggregated simulation model for a seedling propagation system. First, a formal system conceptual model is constructed in SysML to define system structure and activity. The simulation models conforming to the system conceptual model are then constructed using Arena software. An additive regression model-based approach is proposed to estimate parameters for the higher level model. The framework is finally demonstrated via modeling one of the largest grafted seedling propagation systems in North America.

The remainder of the paper is structured as follows. In Section 2, the seedling propagation system is described. In Section 3, the system conceptual model is constructed in SysML. Then the additive regression model-based parameter estimation approach is proposed. The proposed framework is demonstrated in Section 4 for one of the largest seedling propagation system in North America. In addition, the simulation results obtained from the low level model and aggregated model are compared to the considered sys-

tem. In Section 5, conclusions and directions for future extensions to the proposed framework are discussed.

## 2 GRAFTED SEEDLING PROPAGATION SYSTEM

In a grafted seedling propagation system considered in this paper, grafted vegetable seedlings are produced using scion and rootstock seeds via a variety of stages including: 1) seeding, 2) germination, 3) pre-sorting grow on, 4) sorting, 5) pre-grafting grow on, 6) grafting, 7) healing, and 8) post-healing grow on. As depicted in Figure 1, both scion and rootstock seeds are first seeded into trays via a seeding machine in the seeding stage. The seeded trays (both scion and rootstock) are then moved into germination chambers in a selected germination area of the greenhouse. The germinated scions/rootstocks are moved into different compartments of the greenhouse and grow on until the cotyledon's full emergence. As soon as cotyledons emerge, the scions/rootstocks are transplanted into different trays via a sorting machine based on their sizes. The sorted scions/rootstocks are moved back to greenhouse for further grow-on before grafting. Once the size of scions/rootstocks is suitable for grafting, they are moved to grafting department where they are grafted by grafting workers or grafting robots. The grafted seedlings are then put into healing chambers for healing. Finally, the healed seedlings are moved back to greenhouse for further grow-on to reach the size that vegetable producers require.

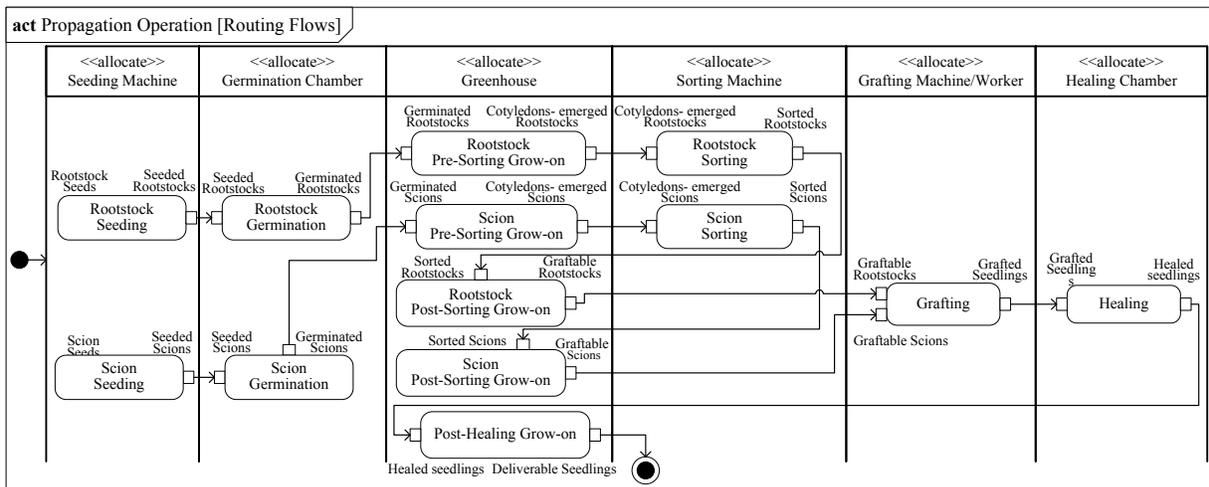


Figure 1: SysML activity diagram for material flow in seedling propagation system

## 3 SIMULATION MODEL AGGREGATION FRAMEWORK

### 3.1 Overview of Framework

The proposed simulation model aggregation framework involves three main steps (see Figure 2): 1) system conceptual modeling, 2) simulation modeling, and 3) parameter estimation. In the first step, a SysML-based system conceptual model is developed to describe system structures (e.g. departments, cells and machines/workers), activities (e.g. object loading/unloading, transportation and processing) and parameters (e.g. loading/unloading time and processing time) in a hierarchical manner. In the second step, different levels of simulation models are constructed conforming to the constructed conceptual model. The number of levels varies depending on the structure of the considered system. In this paper, three levels are used as the seedling propagation system can be conceptually described at each of the department, cell, and machine levels (high, medium and low levels). In the third step, different levels of simulation model parameters (e.g. mean value  $u$  and randomness  $w$ ) are used to formulate experiment settings, and then used as simulation inputs for experiments. The simulation inputs ( $u$ ,  $w$ ) and outputs ( $f(u, w)$ ) are

both used to construct regression models. Finally the regression models are applied to estimate the higher level model parameters ( $u'$ ,  $w'$ ). This procedure continues until the highest level model parameters are estimated.

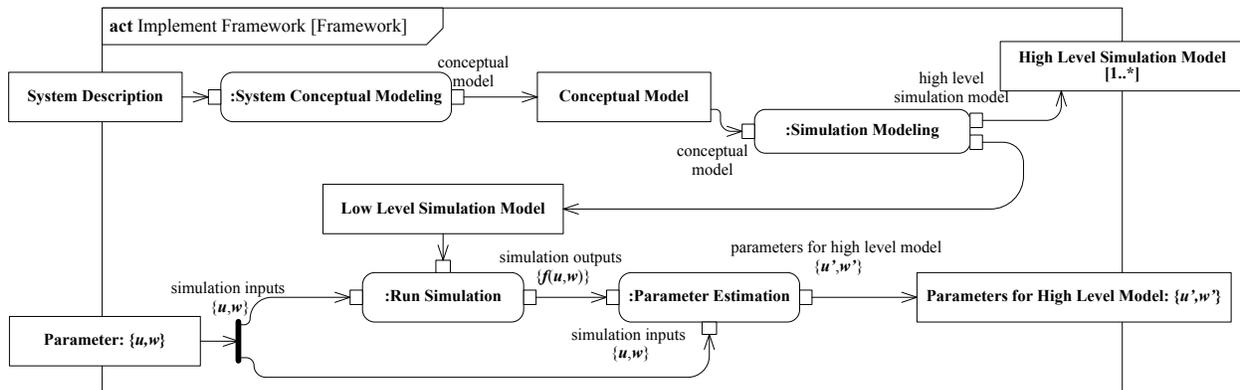


Figure 2: SysML activity diagram for the proposed aggregation framework

### 3.2 SysML-based System Conceptual Model

As mentioned in Friedenthal et al. (2012), accurate presentation of system requirements and behaviors is essential in a robust system description. For simulation model aggregation, it is even more important to model the system structure and activity in a hierarchical way such that different levels of simulation models can be constructed and parameters of higher level models can be estimated via lower level ones. Therefore, the SysML-based system conceptual model in this work consists of (1) a structure model and (2) an activity model. The structure model is used to define the resources (e.g. grafting worker) and entities (e.g. trays), and the activity model is used to define the system activities for constructing the simulation model logic. Both models are constructed via the SysML block definition diagram. In SysML, block is the modular unit of structure that is used to define a system component, an item that flows through the system, activities, external entities or other logical abstractions. The block definition diagram is used to define blocks and the relationships between them (Friedenthal et al. 2012). Therefore, block definition diagram are used in our work to define the hierarchy of the system of our interest.

For illustration purposes, the system conceptual models are explained in a top-down manner as follows. Figure 3 shows the structure model that has a hierarchy of three levels (i.e. high, medium and low levels) of system components. In the high level, the seedling propagation system consists of six production components including greenhouse, seeding department, germination department, sorting department, grafting department, healing department, and miscellaneous worker and shelf. Shelf is the logistics equipment that miscellaneous workers use to transport trays of seeds/seedlings between departments. Each production component performs a distinct production function, and miscellaneous workers are responsible for connecting those production components together. For example, greenhouse is the component in which seedlings grow on (see Figure 1) to certain sizes under required temperature, humidity and light intensity. In this level, the unit of material flow is order. From the high level to the medium level, the departments are further decomposed into a variety of cells that have the same production function as the department but realized in a smaller scale. Instead of order, the unit of material flow in the medium level model is tray as the trays are distributed into different cells for processing. The number of cells each department possesses varies significantly. For example, the seeding department may only have two cells since the cycle time is relatively short (e.g. 20 seconds/tray). While germination department may have more cells since the germination time involves days. Therefore more cells are needed to handle the trays from the upstream departments. In the low level, the production cells are further decomposed into atomic components such as machines, chambers and workers. The seeding cell, sorting cell and grafting cell

consists of corresponding machine/robot and workers respectively, and the basic material flow is seedling. The germination cell and healing cell consists of several chambers, and the basic material flow is tray.

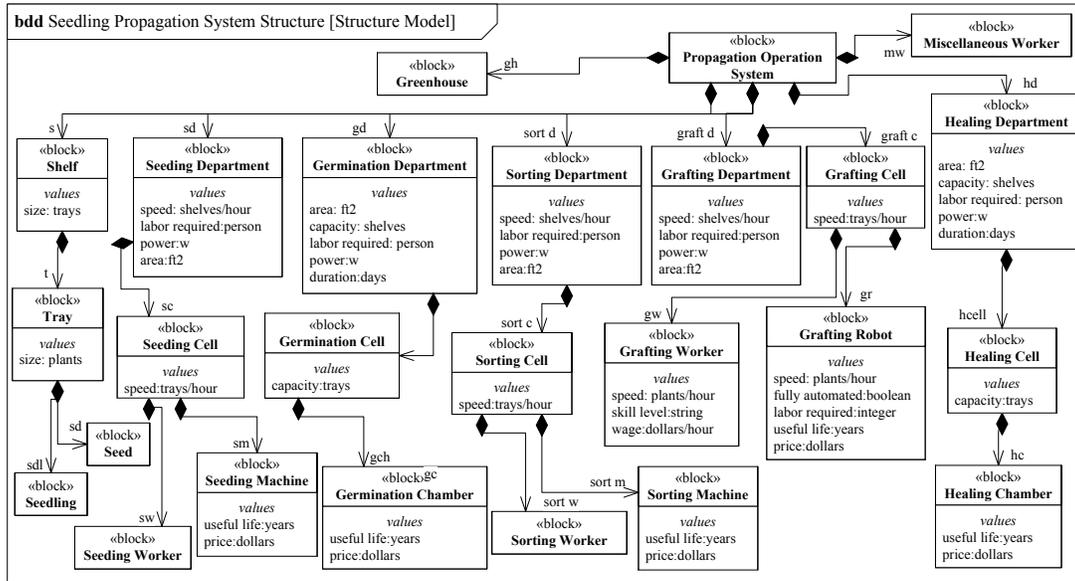


Figure 3: SysML block definition diagram for a structure model of seedling propagation system

While the structure model depicts the system hierarchy and the relationships between the system components, the activities of the system also have to be captured in a corresponding hierarchical manner. As depicted in Figure 4, the activities of each department are also modeled in three levels (i.e. low, medium and high levels) conforming to the structure model.

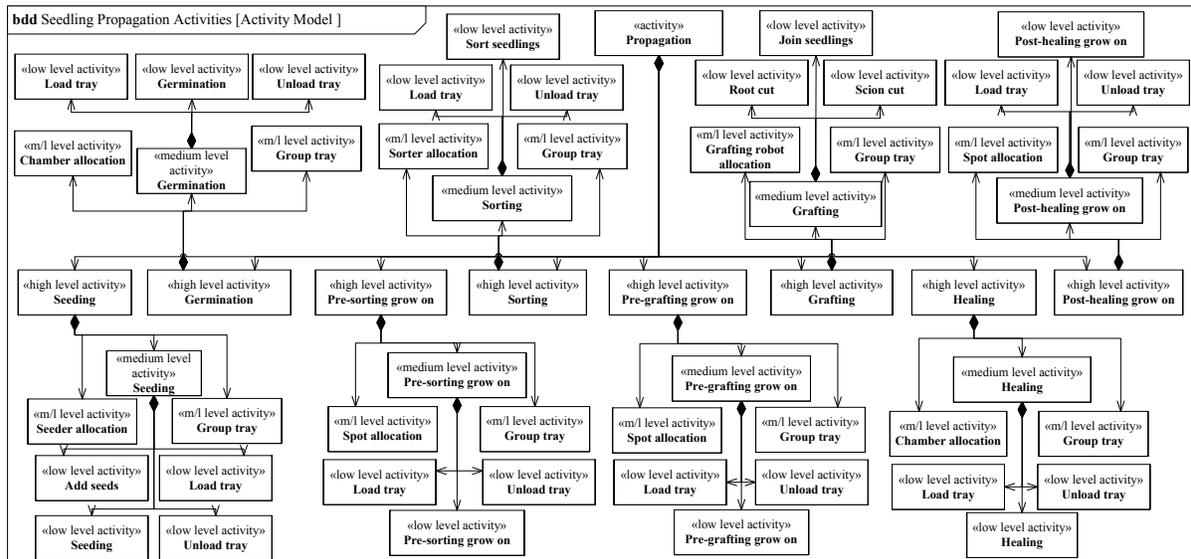


Figure 4: SysML block definition diagram for activity model of seedling propagation system

The high level activities are stereotyped by "high level activity", and include seeding, germination, pre-sorting grow on, sorting, pre-grafting grow on, grafting, healing and post-healing grow on. The activities in this level correspond to the department level of the structure model. In the medium level, the high level activities are further decomposed into more detailed activities that are stereotyped by "medium level

activity" or "m/l level activity". Here "m/l level activity" indicates that a particular activity exists in both medium and low level. In the medium level, since trays are distributed to various production cells (e.g. trays of seeds can be placed in different germination cells) and batched in shelves for moving to the next department, resource allocation-related activities (e.g. *seeder allocation*, *chamber allocation*, *sorter allocation*, *spot allocation* and *grafting robot allocation*) and *group tray* are considered. The activities in this level correspond to the production cell level of the structure model. Finally in the low level, more detailed activities are modeled to conform to the functionalities of machines/workers within a specific production cell. For example, the *seeding* activity in the medium level is further decomposed into *add load tray*, *add seeds*, *seeding* and *unload tray*.

### 3.3 Additive Regression Model-based Parameter Estimation

Regression modeling is one of the most widely used statistical techniques since it identifies the relationships between independent variables (also called predictors) and dependent variables (also called responses) (Montgomery et al. 2012). In this paper, an additive regression model is used to estimate simulation model parameters. As depicted in Figure 5, the regression model formulation procedure can be described as follows. The predictors  $x_i = \{x_{ij}, j=1, \dots, n\}$  are specified as simulation inputs to obtain corresponding target response  $t_i$  (simulation outputs) via running a low level simulation model. Here the target response is defined as the higher level model parameter that we want to estimate. The predictors  $x_i$  and target response  $t_i$  are then used to formulate the regression model  $f(\cdot)$ . Finally, the response  $y_i$  obtained from the formulated regression model  $f(\cdot)$  is used as the corresponding higher level model parameter.

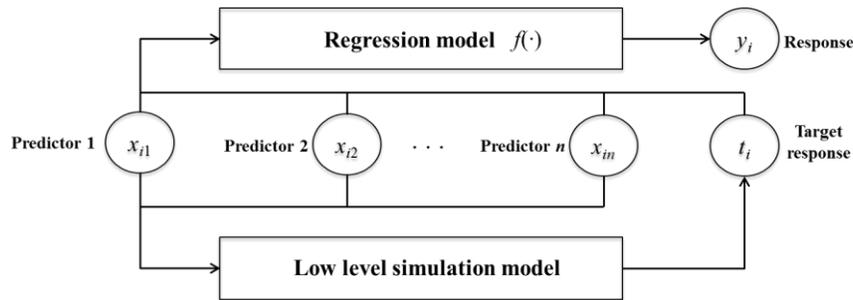


Figure 5: Regression model formulation procedure

The additive regression model with nonlinear partial-regression function  $f_j(\cdot)$  is formulated as follows:

$$y_i = \alpha_i + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_{J_i}(x_{iJ_i}) + \varepsilon_i, \varepsilon_i \sim N\left(0, \sum_{j=1}^{J_i} \sigma_{ij}^2\right) \quad (1)$$

$$f(x_{ij}) = \beta_{ij1}(x_{ij}) + \beta_{ij2}(x_{ij})^2 + \dots + \beta_{ijp_j}(x_{ij})^{p_j} \quad (2)$$

where,  $y_i$  is the response,  $x_{ik}$  is the  $k$ th predictor for the  $i$ th of  $n$  observations and  $\varepsilon_i$  is an error of the model.  $\beta_{ijp_j}$  is a weight for the  $j$ th predictor with  $p_j$ th order and  $\alpha_i = \sum_{j=1}^{J_i} \beta_{ij0}$  is a constant value. To obtain the coefficients of the model, we need to find an appropriate order  $p_j$  for each partial-regression function  $f_j(\cdot)$  with weight vector  $\beta_{ij} = [\beta_{ij0}, \beta_{ij1}, \dots, \beta_{ijp_j}]^T$  under the additive regression model. For this, we further define the augmented data matrix and response as follows:

$$X_{ij} = \begin{bmatrix} x_{ij1}^0 & x_{ij1}^1 & x_{ij1}^2 & \cdots & x_{ij1}^{p_j} \\ x_{ij2}^0 & x_{ij2}^1 & x_{ij2}^2 & \cdots & x_{ij2}^{p_j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ijN}^0 & x_{ijN}^1 & x_{ijN}^2 & \cdots & x_{ijN}^{p_j} \end{bmatrix}, y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iN} \end{bmatrix} \quad (3)$$

where  $N$  is the number of simulation output samples. Thus, the response  $y_i$  is subject to normal distribution as follows:

$$y_i \sim N\left(\sum_{j=1}^{J_i} X_{ij}\beta_{ij}, \sum_{j=1}^{J_i} \sigma_{ij}^2\right) \quad (4)$$

Since we use additive regression model, the weight  $\beta_{ij}$  can be represented as:

$$\hat{\beta}_{ij} = (X_{ij}^T X_{ij})^{-1} X_{ij}^T y_i \quad (5)$$

and its maximum likelihood can be expressed as:

$$\log L_{ij} = -\frac{N}{2} \log(2\pi) - \log(\sigma_{ij}) - \frac{1}{2\sigma_{ij}^2} (y_i - X_{ij}\beta_{ij})^T (y_i - X_{ij}\beta_{ij}) \quad (6)$$

$$\frac{\partial \log L_{ij}}{\partial \beta_{ij}} = \frac{1}{\sigma_{ij}^2} (X_{ij}^T y_i - X_{ij}^T X_{ij} \beta_{ij}) = 0 \quad (7)$$

Similarly,  $\sigma_{ij}^2$  can be represented as:

$$\hat{\sigma}_{ij}^2 = \frac{1}{N} (y_i - X_{ij}\hat{\beta}_{ij})^T (y_i - X_{ij}\hat{\beta}_{ij}) \quad (8)$$

and the maximum likelihood can be expressed as:

$$\frac{\partial \log L_{ij}}{\partial \sigma_{ij}} = -\frac{N}{\sigma_{ij}} + \frac{1}{\sigma_{ij}^3} (y_i - X_{ij}\hat{\beta}_{ij})^T (y_i - X_{ij}\hat{\beta}_{ij}) = 0 \quad (9)$$

$$\hat{\sigma}_{ij}^2 = \frac{1}{N} (y_i^T y_i - y_i^T X_{ij}\hat{\beta}_{ij}) \quad (10)$$

Thus, the estimated response  $\hat{y}_i$  is subject to normal distribution as follows:

$$\hat{y}_i \sim N\left(\sum_{j=1}^{J_i} X_{ij}\hat{\beta}_{ij}, \sum_{j=1}^{J_i} \hat{\sigma}_{ij}^2\right) \quad (11)$$

In addition, we adopt  $K$ -fold cross validation (CV) to find the most promising order  $p_j$  of each partial-regression function  $f_j(\cdot)$  with estimated response  $\hat{y}_i$ . The  $K$ -fold CV splits the data into  $K$  equally sized blocks, and each block takes its turn as a validation set for a training set comprised of the other  $K-1$  blocks (Rogers and Girolami 2011). Although the  $K$ -fold CV is computation demanding, it prevents the model over-fitting in terms of the order  $p_j$  and also gives better accuracy than a bootstrap method. Figure 6 shows the overall process of the proposed parameter estimation approach. Moreover, in this research, the Leave-One-Out Cross Validation (LOOCV) is used to find the most probable which has the lowest mean LOOCV error. The following equation shows the average squared validation loss for LOOCV:

$$L_i^{CV} = \frac{1}{N} (y_i - \hat{y}_i)^T (y_i - \hat{y}_i) \quad (12)$$

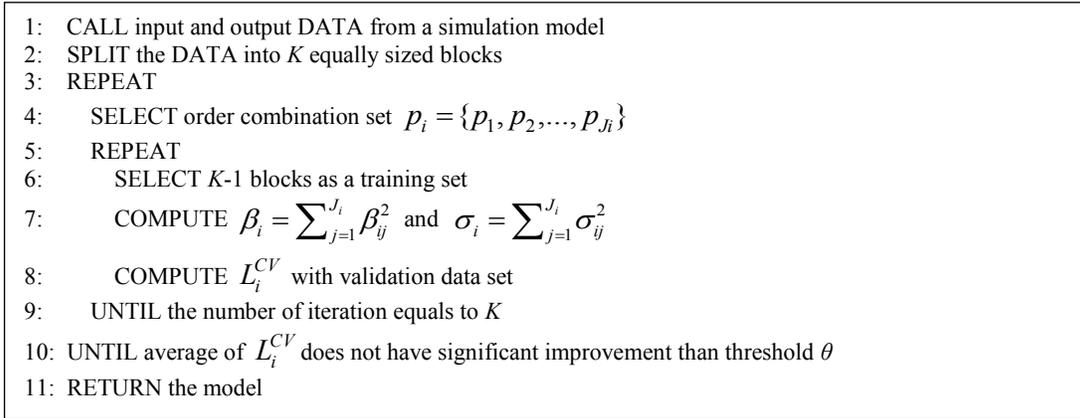
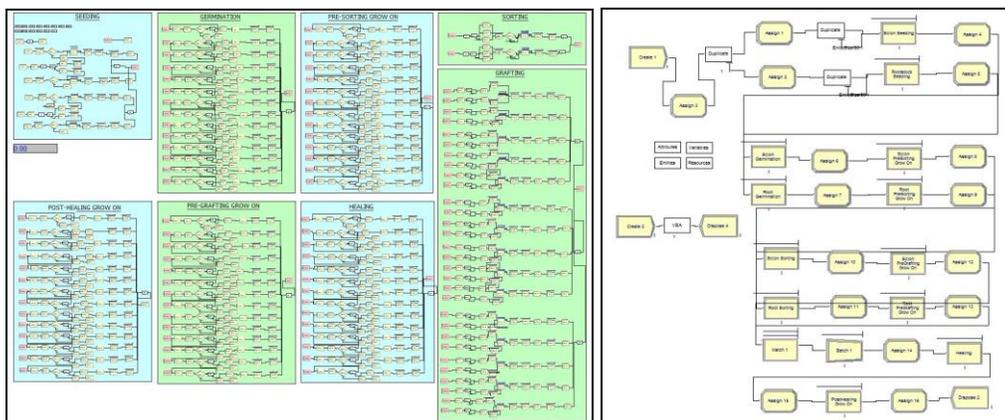


Figure 6: Pseudo code for simulation parameter estimation procedure

#### 4 EXPERIMENT

The proposed framework is demonstrated in this section via one of the largest grafted seedling propagation system in North America. As discussed in Section 3.2, three levels of simulation models were constructed in Arena conforming to the structure and activity hierarchies described in Figures 3 and 4, and Figures 7(a)-(b) compare the complexity of the low level model (left figure) with the high level model (right figure). Then two steps of parameter estimation experiments were conducted in this section. The first step was to estimate the medium level model parameters using the low level models. Similarly, the second step was to estimate the high level model parameters using the medium level model whose parameters were estimated in the first step. In the experiments, our primary focus was on time-related parameters since the outputs (e.g. throughput and cost) of our interests mainly depend on time. The low level model parameters that were used to estimate the medium level model parameters are specified in Table 1. In the table, stochasticity refers to the ratio of the standard deviation (std.) to the mean. For example, the ratio of std. and mean of tray preparation time in seeding department was set to 0.1, 0.3 and 0.5 in the experiment. For each simulation parameter (or predictor in regression model), three levels were selected based on real data, and  $3^3=27$  experiment settings were formulated for each department.



(a) Low level model

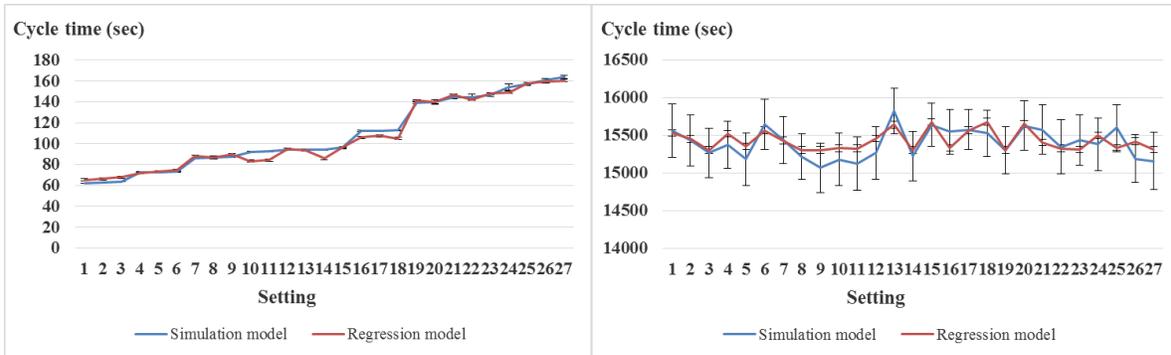
(b) High level model

Figures 7: Model complexity comparison between low level and high level models

Table 1: Parameter specification for low level model

Seeding dept.	Tray preparation time (sec/tray)	Stochasticity of tray preparation time	Seeding speed (sec/tray)
Parameter values	(5, 7, 10)	(0.1, 0.3, 0.5)	(15, 20, 25)
Germination dept.	Tray transportation time (sec/tray)	Stochasticity of tray transportation time	Germination duration (day)
Parameter values	(5, 7, 10)	(0.1, 0.3, 0.5)	(2.5, 3, 3.6)
Greenhouse dept.	Tray transportation time (sec/tray)	Stochasticity of tray transportation time	Grow on duration (day)
Parameter values	(5, 7, 10)	(0.1, 0.3, 0.5)	(6, 7, 8)
Sorting dept.	Tray loading/unloading time (sec/tray)	Stochasticity of tray loading/unloading time	Sorting speed (sec/plant)
Parameter values	(5, 7, 10)	(0.1, 0.3, 0.5)	(0.7, 1, 1.5)
Grafting dept.	Scion/rootstock cut speed (sec/plant)	Plant joining speed (sec/plant)	Stochasticity of scion/rootstock cut speed
Parameter values	(3, 4, 5)	(9, 12, 15)	(0.1, 0.3, 0.5)
Healing dept.	Tray transportation time (sec/tray)	Stochasticity of tray transportation time	Healing duration (day)
Parameter values	(5, 7, 10)	(0.1, 0.3, 0.5)	(2.5, 3, 3.6)

Based on the simulation outputs obtained from the low level and medium level simulation models, six additive regression models (that correspond to six types of cells/departments) were formulated in each level using the procedures described in Figure 6. The outputs from regression models were validated against corresponding simulation results. Figure 8(a) depicts the sorting cell cycle time for one tray, and Figure 8(b) depicts the sorting department cycle time for one order (i.e. 180 trays). The paired *t*-test results (with the significance level of 0.05) showed that there was no significance difference in the results from the regression model and the simulation model (*p*-value=0.1120 for sorting cell cycle time; and *p*-value=0.2173 for the sorting department cycle time). Thus, the formulated additive regression models for the considered case were accurate enough to estimate higher level model parameters.



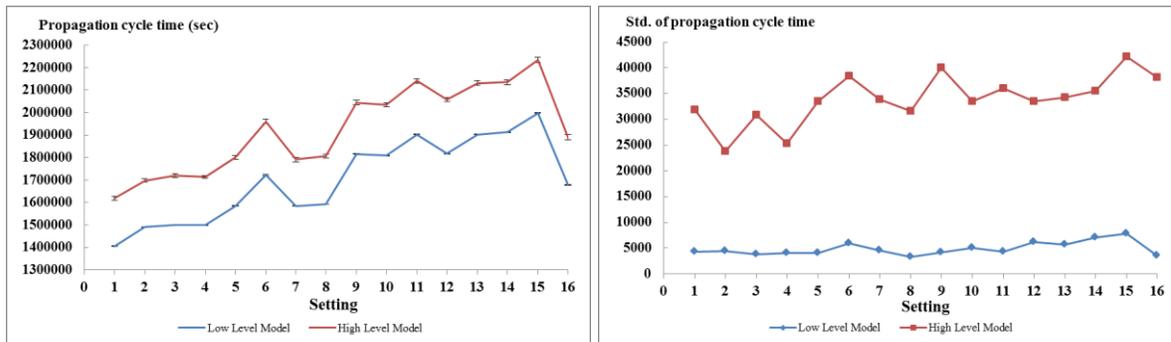
(a) Sorting cell cycle time (b) Sorting department cycle time  
 Figures 8: Output comparison between regression and simulation models

To validate the high level model and the estimated parameters under different experimental settings, the simulation results from the low level model were used as references. Table 2 shows the parameter values used for formulating the 16 experiment settings. Figures 9(a)-(b) show the means and standard deviation of propagation cycle time (i.e. time to complete the propagation operation) for 900 trays, respectively. As shown in Figure 9(a), two models produced the same trend of propagation cycle time over 16 experiment settings, and the results from high level model had higher mean (7.2% on average). The execution speed of the high level model was observed to be 10 times faster than that of the low level model on average. This observation demonstrates that although minor information loss (e.g. smaller size entity and more detailed parameters) during aggregation is not inevitable, the high level model maintains the same essential behavior as the low level model but achieves significant execution speed improvement. Besides, as shown in Figure 9(b), the high level model had higher randomness in terms of output standard deviation than that of the low level model. The reason is that as the entities were aggregated from small size (e.g.

one plant/tray in the low level model) to large size (e.g. 180 trays in the high level model), randomness within each department of the simulation model was accordingly increased (see Figures 9(a)-(b)). It implies that the high level model has higher uncertainty of prediction results than that of the low level model even though the both models can generate similar trend of results.

Table 2: Experiment setting parameters for high level model validation

Seeding dept.	Tray preparation time (sec/tray)	Stochasticity of tray preparation time	Seeding speed (sec/tray)
Parameter values	7	0.1	20
Germination dept.	Tray transportation time (sec/tray)	Stochasticity of tray transportation time	Germination duration (day)
Parameter values	7	0.1	(2.5, 3.6)
Greenhouse dept.	Pre-sorting grow on duration (days)	Pre-grafting grow on duration (days)	Post-healing grow on duration (days)
Parameter values	(2, 3)	(10, 14)	1
Sorting dept.	Tray loading/unloading time (sec/tray)	Stochasticity of tray loading/unloading time	Sorting speed (sec/plant)
Parameter values	7	0.1	1
Grafting dept.	Scion/rootstock cut speed (sec/plant)	Plant joining speed (sec/plant)	Stochasticity of scion/rootstock cut speed
Parameter values	3	12	0.1
Healing dept.	Tray transportation time (sec/tray)	Stochasticity of tray transportation time	Healing duration (day)
Parameter values	7	0.1	(1.5, 2.6)



(a) Mean of propagation cycle time (b) Standard deviation of propagation cycle time  
 Figure 9: Propagation cycle time comparison between low and high level simulation models

## 5 CONCLUSION

In this paper, a SysML-based simulation model aggregation framework is first proposed for vegetable seedling propagation system. In the proposed framework, SysML is used to build a system conceptual model in terms of system structures and activities in a hierarchical manner. Additive regression model-based approach is also proposed to estimate parameters for a higher level model. The proposed framework is then demonstrated via one of the largest vegetable seedling propagation system in North America. In the conducted experiments, the simulation model aggregated based on the proposed framework has produced output with same accuracy as that of a detailed simulation model, achieving significant execution speed improvement. Our future research works are to study the impact of parameter ranges on the accuracy of the estimation, and to develop algorithms to determine an appropriate aggregative level considering the required accuracy and computational requirements.

## ACKNOWLEDGMENTS

This work has been supported by U. S. Department of Agriculture (USDA) Specialty Crop Research Initiative (SCRI) under USDA SCRI Project No. 2011-51181-30963.

## REFERENCES

- Cassandras, C.G., C.G. Panayiotou, G. Diehl, W. Gong, Z. Liu, and C. Zou. 2000. "Clustering methods for multi-resolution simulation modeling." In *Proceedings of the conference on enabling technology for simulation science*:37-48.
- Celik, N., E. Mazhari, O. Kazemi, J. Candy, P. Sarfare, M. Alotaibi, and Y. J. Son. 2010. "Automatic partitioning of large scale simulation in grid computing for run time reduction." *International Journal of Operations Research and Information Systems*, 1(2), 64-90.
- Cetinkaya, D., A. Verbraeck, and M.D. Seck. 2010. "Applying a model driven approach to component based modeling and simulation." In *Proceedings of the 2010 winter simulation conference*, Edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, 546-553. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Fan, Y., and G. M. James. 2013. Functional additive regression. Accessed March 22. <http://www-bcf.usc.edu/~gareth/research/FAR.pdf>.
- Fox, J. 2000. *Nonparametric simple regression*. Thousand Oaks CA: Sage Publications.
- Friedenthal, S., A. Moore, and R. Steiner. 2012. *A Practical Guide to SysML: the Systems Modeling Language*. 2nd ed. Waltham, MA: Morgan Kaufmann.
- Harmonosky, C.M., and S.F. Robohn. 1995. "Investigating the application potential of simulation for real time control decisions." *International Journal of Computer Integrated Manufacturing*, 8, 126-132.
- Huang, Y., F. Ponci, and F. Dougal. 2009. "Towards automatic level selection for design-aimed simulation." In *Proceedings of the 2009 grand challenges in modeling & simulation conference*: 270-277.
- Kim, H., L.F. McGinnis, and C. Zhou. 2012. "On fidelity and model selection for discrete event simulation." *Simulation: Transactions of the Society for Modeling and Simulation International*, 88(1), 97-109.
- Kubota, C., M. A. McClure, N. Kokalis-Burelle, M. G. Bausher, and E. N. Roskopf. 2008. "Vegetable grafting: history, use and current technology status in North America." *HortScience*, 43(6), 1664-1669.
- Madan, M., Y. J. Son, H. Cho, and B. Kulvatunyou. 2005. "Determination of efficient simulation model fidelity for flexible manufacturing systems." *International Journal of Computer Integrated Manufacturing*, 18, 236-250.
- Meng, C., S. Nageshwaraniyer, A. Maghsoudi, Y. J. Son, and S. Dessureault. 2013. "Data-driven modeling and simulation framework for material handling systems in coal mines." *Computer and Industrial Engineering*, 64, 766-779.
- Montgomery, D. C., E. A. Peck, and G. G. Vining. 2012. *Introduction to linear regression analysis*, 5<sup>th</sup> ed. Wiley.
- Persson, J. F. 2002. "The impact of different levels of detail in manufacturing systems simulation models." *Robotics and Computer Integrated Manufacturing*, 18, 319-325.
- Plant propagation. [http://en.wikipedia.org/wiki/Plant\\_propagation](http://en.wikipedia.org/wiki/Plant_propagation), accessed in March 20, 2013.
- Rodriguez, J. D., J. O. Miller, K. W. Bauer Jr., and R. E. Neher Jr. 2008. "Methodologies for aggregating large hierarchical simulation models." *Proceedings of SPIE*, DOI: 10.1117/12.776702
- Rogers, S., and M. Girolami. 2011. *A first course in machine learning*. Boca Raton: CRC Press.
- Son, Y. J., A.T. Jones, and R.A. Wysk. 2003. "Component based simulation modeling from neutral component libraries." *Computers and Industrial Engineering*, 45, 141-165.
- Vasudevan, K., and A. Devikar. 2011. "Selecting simulation abstraction levels in simulation models of complex manufacturing systems." In *Proceedings of the 2011 winter simulation conference*, Edited by S. Jain, R.R. Creasey, J. Himmelsbach, K.P. White, and M. Fu, 2268-2277. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Venkateswaran, J., and Y. J. Son. 2004. "Impact of modeling approximations in supply chain analysis-an experimental study." *International journal of Production Research*, 42(15), 2971-2992.
- Wang, J., Q. Chang, G. Xiao, N. Wang, and S. Li. 2011. "Data driven production modeling and simulation of complex automobile general assembly plant." *Computers in Industry*, 62, 765-775.

Webster, D. B., and M. L. Padgett. 1984. "Determining the level of detail in a simulation model-a case study." *Computer and Industrial Engineering*, 8(3-4), 215-225.

#### **AUTHOR BIOGRAPHIES**

**CHAO MENG** is a Ph.D. student in the Department of Systems and Industrial Engineering at The University of Arizona. He received his Master degree in Logistics Engineering from Southwest Jiaotong University, China in 2011, and Master degree in Industrial Engineering from The University of Arizona in 2013. His research interests are in Data-driven simulation modeling, Model-driven architecture and Meta-heuristics. He is a member of IIE and INFORMS. He can be reached at [meng@email.arizona.edu](mailto:meng@email.arizona.edu).

**SOJUNG KIM** is a Research Assistant in the Department of Systems and Industrial Engineering at the University of Arizona. He received his Master in Industrial and Systems Engineering from Dongguk University, Republic of KOREA in 2010. His research interests include Agent based modeling & simulation, Human decision making and Behavior modeling. He is a member of INFORMS, and IIE. His email address is [sojungkim@email.arizona.edu](mailto:sojungkim@email.arizona.edu) and his web page is <https://sites.google.com/site/sojungk8/>.

**YOUNG-JUN SON** is a Professor of Systems and Industrial Engineering at the University of Arizona and Director of Advanced Integration of Manufacturing Systems and Technologies Center. His research focuses on the coordination of a multi-scale, networked-federated simulation and decision model needed for design and control in manufacturing enterprise, renewable energy network, homeland security, agricultural supply network, and social network. He has received several research awards such as the SME 2004 Outstanding Young Manufacturing Engineer Award, the IIE 2005 Outstanding Young Industrial Engineer Award, the IERC Conference Best Paper Awards (2005, 2008, 2009), and Best Paper of the Year Award in 2007 from IJIE. He can be reached at [<son@sie.arizona.edu>](mailto:son@sie.arizona.edu).

**CHIERI KUBOTA** is a professor in the School of Plant Sciences at the University of Arizona. Dr. Kubota received a Ph.D. in Horticulture Engineering from Chiba University, Japan in 1994. Dr. Kubota's research contributes to the development of science and technology in the area of controlled environment agriculture (CEA). Her research platform includes all CEA systems including greenhouses, growth chambers, postharvest storage and tissue culture. Her research interests include: 1) Plant responses in greenhouse controlled environment, 2) Controlled environment for enhancing greenhouse produce quality (overall and nutritional quality), 3) Environmental control in plant production under artificial lighting conditions, and 4) Modeling and simulation of plant growth and development. Her email address is [ckubota@ag.arizona.edu](mailto:ckubota@ag.arizona.edu).