Validation of System Dynamics Models – a Case Study

Justyna Lemke*, Małgorzata Łatuszyńska**

Abstract

The purpose of this article is the analysis of the system dynamics model validation illustrated by the example of a model of the manufacturing resource allocation. In the first part of the article the authors present an overview of the definitions of validation and verification that can be found in the reference literature. Also, they emphasize the role which validation and verification play in the modeling process. Furthermore, they discuss the techniques of system dynamics model validation with particular focus on tests of the model structure, behavior and policy implications. The second part of the article contains an example of the validation process of a system dynamics model simulating manufacturing resource allocation in an electronic company. The purpose of the model is to assess the long-term effect of assigning workers to individual tasks on such production process parameters as efficiency or effective work time. The authors focus their particular attention on that part of the model which deals with a storehouse, one of the company production units. They conduct tests of its structure and behavior. When validating the structure the authors make use of the information obtained in a series of interviews with the company staff. They also refer to the generally accepted knowledge found in the reference literature. The results generated by the model in the course of the behavior tests are compared with the real data. The authors evaluate both the logic of the system behavior and the level of accuracy of the output data in reference to the real system.

Keywords: *validation of simulation models, system dynamics*

Introduction

Nowadays simulation is commonly used in many areas of business management. It is applied to forecasting as well as to understanding mechanisms within companies. Simulation models can be particularly helpful for minimizing wrong decision-making. It must be noted, however, that the model that is to satisfy the user's requirements has to meet quality standards regarding both the software and the accuracy of its representation of reality. This is why among many stages of creating a simulation we can find its verification and validation (Maciąg, Pietroń & Kukla, 2013, p. 161). It is an

* Justyna Lemke, MA, Maritime University of Szczecin, j.lemke@am.szczecin.pl, Wały Chrobrego 1-2, 70-500 Szczecin.
** Małgorzata Łatuszyńska, Professor, University of Szczecin, mlat@wneiz.pl, Mickiewicza 64, 71-101 Szczecin.

essential, but controversial (Barlas, 1996, p. 183) and still unsolved (Martis, 2006, p. 39) aspect of modeling methodology. The quality of a study conducted by means of a simulation model largely depends on its validation.

The purpose of this article is the analysis of the system dynamics model validation illustrated by the example of a model of the manufacturing resource allocation.

Literature review

One of the difficulties hindering the process of verification and validation lies in the understanding of both terms. Some authors claim that there is no need to differentiate between the two – for instance Pidd sees verification and validation as synonymous (Pidd, 1998, p. 33). But, as shown in Table 1, most authors recognize verification as different from validation.

Table 1. Verification and validation by selected authors

Verification	*Validation*	*Source*
Testing if the symbolic (formal) model has been properly transformed into its operational form (e.g. a computer program).	Proving that in the experimental environment the accuracy of the operational model (usually a computer one) is satisfying and in keeping with its intended use.	(Maciąg, Pietroń & Kukla, 2013, p. 162)
Testing if the computer program of the computerized model and its implementations are correct.	Proving that within its application domain the computerized model has a satisfying level of accuracy which is in keeping with its intended use.	(Sargent, p. 166)
Testing a seemingly correct model by its authors in order to find and fix modeling errors.	An overview and assessment of the model operation performed by its authors and by experts in the field in order to find out if the model with a satisfying level of accuracy represents the real system.	(Carson, 2002, p. 52)
The process of ascertaining if the implementation of the model (the computer program) represents precisely the concept authors' description and specification.	The process of deciding on the assessment method as well as the very assessment of the level to which the model (its data) represents the real world from the perspective of its intended use.	(Davis, 1992, pp 4-6)

Verification is regarded as a necessary, yet insufficient stage of model assessment, while validation is considered – in its narrower sense – to be one of the assessment stages, or – in a broader approach – the very assessment itself (Balcerak, 2003, pp 27-28).

Despite the differences, the above quoted definitions have something in common. Verification is typically conducted by the author of a model and refers to a computer program. In other words, it is the process of checking if the program is free from formal, 'technical' faults, while validation is a more complex issue – it results in determination if and how well the model represents the reality.

Consequently, if a computer program is to be working properly and a model should represent the reality, a question must be asked when it should be tested in the course of modeling. The time to perform the tests is another question to be answered by a modeler. Various authors suggest, following Sargent's paradigm (Sargent, Verification and Validation of Simulation Models, 2010), that verification and validation should be run parallel during the modeling process and that validation should refer to data as well as to two models: the computer model and the conceptual one (see Figure 1).

The purpose of the conceptual model validation is to answer the question whether the model includes the appropriate number of details to meet the simulation objective, while the purpose of the data validation is to find out if the data used in the model is accurate enough. In case of a computer model, its validation, called the operational validation, answers the question whether the model adequately represents the real system. Robinson (Robinson, 1997, p. 54) recommends testing individual parts of the model (*White-box Validation*) and the model as a whole (*Black-box Validation*).

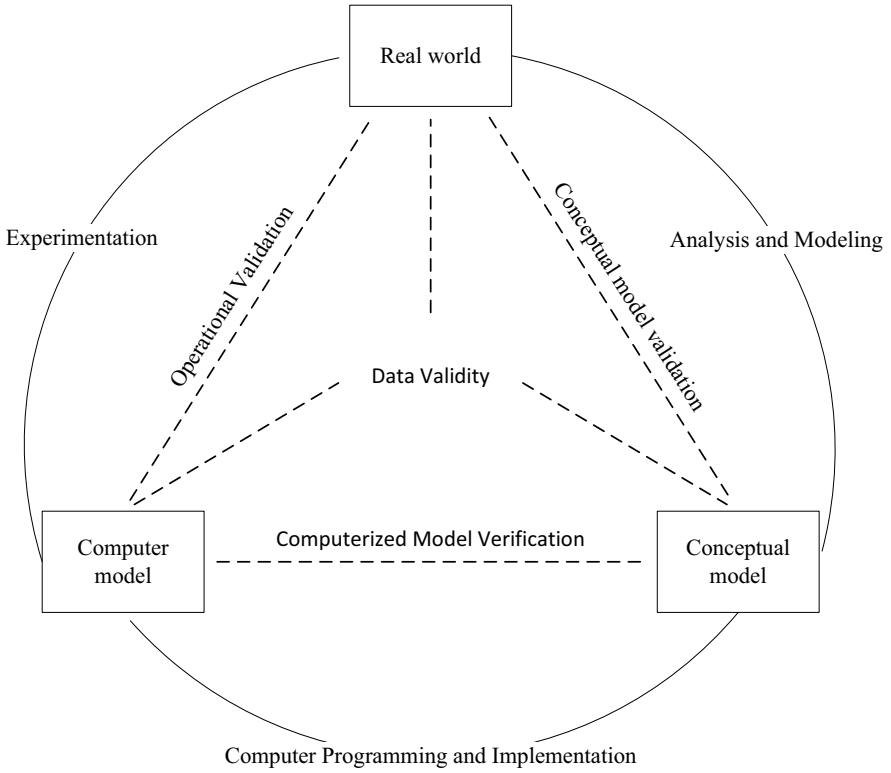

Figure 1. Verification and validation of a simulation model in the process of modeling

Source: own study on the basis of (Sargent, 2010, pp 169).

Another problem to be faced by those who are assessing the model is the reality representation. The question is what this means. According to Khazanch (Martis, 2006, p. 43), a conceptual model can be considered validated, hence adequately representing the real world, if it is:

- plausible,
- feasible,
- effective,
- pragmatic,
- empirical,
- predictive,
- inter-subjectively certifiable
- inter-methodologically certifiable.

It is essential to choose the validation methods that are adequate for the model purpose. Depending on the reference point for the data generated by

the system or on the modeling objective, various authors recommend testing the descriptive, predictive or structural validity (see: (Balcerak, 2003, p. 37); (Davis, 1992, pp 7-8)). It is worth noting that it is not necessary to test all the three conditions. Therefore, if, for instance, the purpose of the model is to find out why productivity in a company has been falling, there is no need to check if the data supplied by the model is going to correspond to the values that the company will generate in the future.

Model testers have at their disposal various techniques that help answer the above questions. Balci (Balci, 1986, p. 6) divided these techniques into two basic groups: the statistical and the subjective ones. The former include the variance analysis and the linear regression, while the latter comprise the comparison with other models, degeneration tests, event validation, historical data validation or the Schellenberg criteria. The lists and descriptions of these methods can be found in: (Martis, 2006; Balci, 1986; Jaszkiewicz, 1997, pp 193-199; Davis, 1992, pp 18-25). Further in the article its authors present those techniques that are recommended in the literature as useful for the validation of system dynamics models.

Research methods

As mentioned above, the choice of methods depends primarily on the purpose of the model. Therefore, when recommending tests for models built in the System Dynamics (SD) convention, we should first of all define their characteristics.

System Dynamics is a method of continuous simulation developed by J. W. Forrester and his associates at Massachusetts Institute of Technology during the late 1950s and early 1960s. Information about its assumptions and application can be found in numerous publications, such as (Campuzano and Mula, 2011, pp 37-48; Łatuszyńska, 2008, pp 32-77; Maciąg, Pietroń and Kukla, 2013, pp 182-212; Ranganath, 2008; Tarajkowski, 2008, pp 33-163). As far as the choice of validation methods is concerned the most important thing, apart from the continuity of the modeled systems, is that the modeler's principal aim is the examination of their dynamic properties. At the same time, it should be noted that the above mentioned dynamics results from the system structure and from periodic regulatory procedures (Maciąg, Pietroń i Kukla, 2013, p. 184).

In view of the above mentioned properties of the system dynamics models, Sterman (Sterman, 1984, p. 52; Sterman, 2000, pp 858-889) suggests validating a model in the context of its structure, behavior and the implications of the user's policy. Tests recommended for each of the groups and the related problems are presented in Table 2.

Table 2: Validation tests of system dynamics models

Test	*Problem*
Tests of Model Structure	
Structure Verification	Is the model structure consistent with the present state-of-the- art?
Parameter Verification	Are the parameters consistent with the present state-of-the-art?
Extreme Conditions	Does every equation make sense even if the inputs reach extreme values?
Boundary Structure Adequacy	Does the model contain the most important issues addressing a given problem?
Dimensional Consistency	Is every equation dimensionally consistent without the necessity to use parameters that are non-existent in the real world?
Test of Model Behavior	
Behaviour Reproduction	Does the model endogenously generate the symptoms of the problem, the behavior of modes, phases, frequencies and other characteristics of the real system behavior?
Behaviour Anomaly	Do the anomalies occur when the model assumptions have been removed?
Family Member	Does the model represent the behavior of various instances of the same class objects when their input parameters have been entered?
Surprise Behaviour	Is the model able to identify 'new' behavior that has not been known in the real system?
Extreme Policy	Does the model behave properly when extreme input values have been entered or when an extreme policy has been implemented?
Boundary Behaviour Adequacy	Is the model behavior sensitive to the addition or the change of the structure which represents reliable alternative theories?
Behaviour Sensitivity	Is the model sensitive to reliable changes of parameters?
Statisitical Character	Do the model outputs have the same statistical characteristics as the real system outputs?
Tests of Policy Implications	
System Improvement	Has the real system been improved as a result of the application of the simulation model?
Behaviour Prediction	Does the model describe correctly the results of the new policy?
Boundary Policy Adequacy	Are the policy recommendations sensitive to the addition or the change of structure which represents possible alternative theories?
Policy Sensitivity	Are the policy recommendations sensitive to reliable changes of parameters?

Source: own study on the basis of Sterman, Appropriate Summary Statistics for Evaluating the Historical Fir of System Dynamics Models, 1984, p. 52.

It should be underlined that the choice of a test does not determine the technique by means of which it must be performed. Plenty of suggestions can be found in the works by Barlas (Barlas, 1996) or Sterman (Sterman, 2000). The authors of this article present some of these techniques on the example of the manufacturing resource allocation.

Analysis and study

The model, whose validation is to be found below, was built for an electronic company that deals mainly with manufacturing low-voltage condensers. The model purpose was to assess the long-term effect of the workforce allocation to individual production cells on production process parameters as the volume of the work-in-progress products and the productivity. On the basis of the data provided by the company the authors identified 19 production cells. Due to their sporadic workload some of the cells were not taken into account. The cells were grouped according to their tasks. As a result, a network of 9 overlapping cells was obtained (Table 3).

Table 3. A network of production cells in the company

To / From	G_BL	G_ IMP	G_ MON	G_PIER	G_WTOR	G_KJ	G_ MG	G_ NP	G_ZAL
G_BL	X	X	X	X		X		X	
G_IMP		X	X			X	X	X	
G_ MON	X	X	X	X	X	X	X	X	
G_PIER	X		X	X	X				
G_ WTOR	X		X	X	X				
G_KJ		X	X			X	X	X	X
G_MG	X		X	X	X		X	X	X
G_NP			X			X	X	X	X
G_ZAL			X			X	X	X	

Having surveyed the staff, the researchers realized that the company did not have a consistent methodology of task distribution. Each foreman was in charge of a group of workers and allocated tasks on their own authority. The absence of clearly defined mechanisms of controlling the transfer of semi-fabricated products from one work station to another turned out to be the key challenge in the process of the model creation. What is more, the company never registered the number of staff working in a cell. That information was eventually obtained through the analysis of operations

performed by individual workers and the allocation of the operations to individual production cells. This is why in the process of modeling the researchers sought the solutions that would represent as precisely as possible the number of workers employed in each cell as well as the number of semi-fabricated products leaving each cell.

Figure 2 shows a part of the model representing the operation of the MG cell (inventory). This part was validated, which conforms to the aforementioned methodology of validating individual parts of the model. Because their purpose was not to determine the number of workers assigned to a task, but to assess the applied policy, the researches set a one-month simulation time step. The analysis covered full 21 months. The number of workers was calculated by dividing the inventory working time by the working time of a single worker in a cell. In order to count the number of semi-fabricated and finished products leaving the inventory the researchers multiplied the number of batches in a given period of time by the average number of items in a batch in that period.

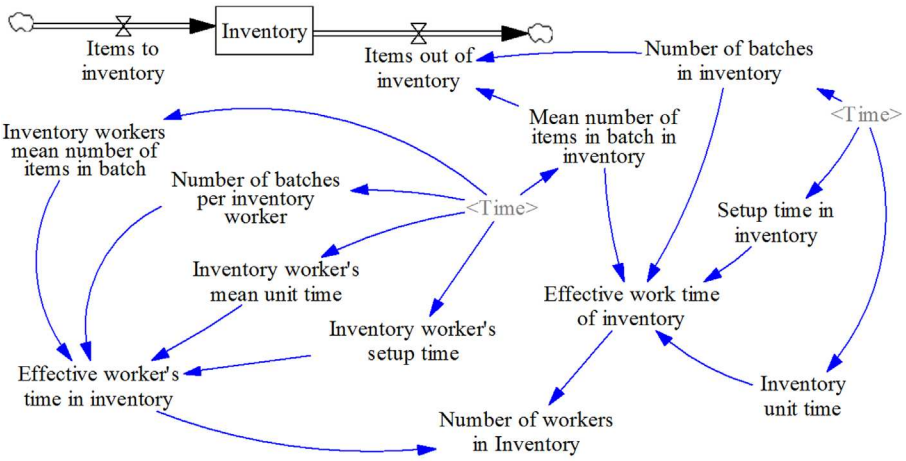

Figure 2. A part of the model – a production cell: inventory

The model was tested by means of the structure and behavior tests. Since the model has not been implemented yet, the test of policy implications cannot be performed.

When validating the structure the authors first of all used the data obtained in the course of survey among the company staff. They also based their validation on generally accepted knowledge from the reference literature.

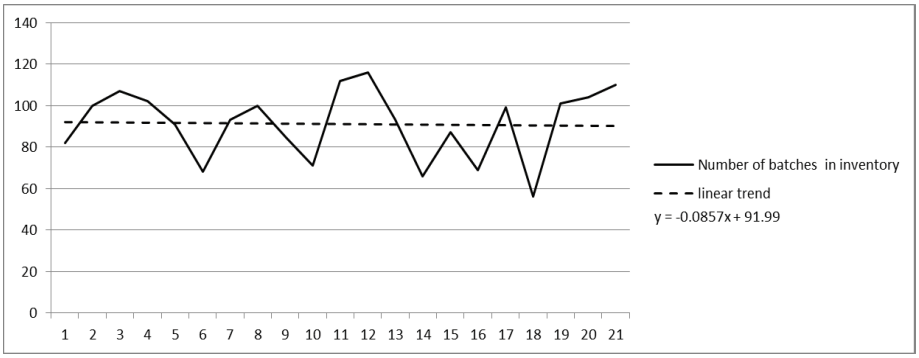

Figure 3. Characteristics of the flow of batches through the inventory (the variable: number of batches)

It should be noted that the structure validation tests are some of the most difficult ones to formalize and perform (Barlas, 1996, p. 190). The information which is indispensable at this stage of validation cannot be presented just as a set of figures. The tools used by the authors of this paper to evaluate the structure accuracy fit the requirements of Barlas' standards.

The following input parameters of the system were tested: the number of batches in the inventory (*the number of batches in inventory*) and the unit time of assignments performed in the inventory (*Unit time inventory*). Following the preliminary assessment of the charts (see examples in Figures 3 and 4) and the calculations run for different types of trends, the researchers, basing on linear trends, decided to determine the values necessary to set the effective working time for each production cell and the working time of one production cell worker. The trend equations are shown in Table 4.

Table 4. Values of the trends and the mean squared error for the parameters of the MG production cell

Variable	*Trend*
Single worker's working time parameters	
Inventory worker's mean number of items in batch	Y=-0.4092t+49.699
Inventory worker's mean unit time	Y=1e-05t+0.0015
Number of batches per inventory worker	Y=-0.0493t+16.142
Inventory worker's setup time	Y=-0.0062t+0.3119
Parameters of production cell effective working time	
Mean number of items in batch in inventory	Y=-2.4413t+103.76

Variable	*Trend*
Inventory unit time	Y=1e-05t+0.0078
Number of batches in inventory	Y=-0.0857t+91.99
Setup time in inventory	Y=-0.002t+0.0909

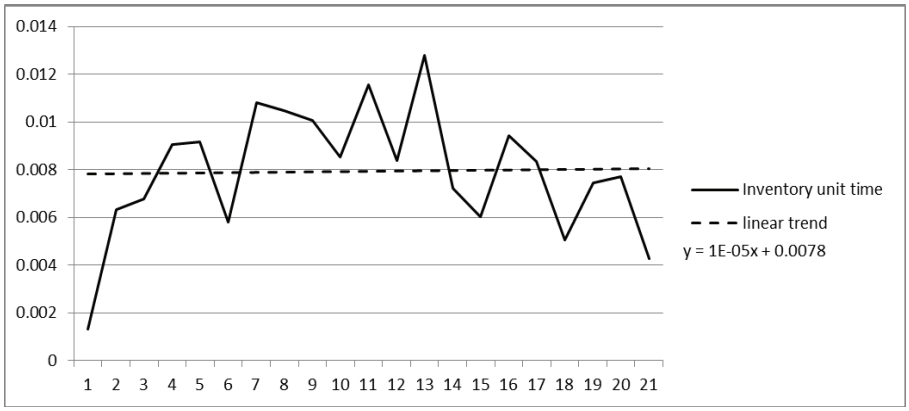

Figure 4. Characteristics of the unit time of assignments in the inventory (the variable: Inventory unit time)

Although some fluctuations do appear in the real system, they are not of seasonal nature. Additionally, the model is supposed to clarify the relations between the number of workers in individual production cell and the parameters of the production process rather that to forecast the value of individual parameters. Therefore, in a view to the lengthy period covered by the analysis, the above fluctuations can be considered hardly significant.

In the second group of tests proposed by Barlas, i.e. the valuation of the model behavior, the authors examined, among others, the system behavior logic.

In the above discussed example they checked if the single cell worker's working time did not exceed the value of the effective working time of the production cell itself. The chart (Figure 5), generated by the VENSIM program, shows that in this context the model can be considered reliable.

At the same time the parameter values generated by the model were compared with those generated by the real system. In order to perform the valuation the authors decided to use the Theil statistics proposed in the reference literature (Kasperska, 2005, p.137).

Table 5 shows the results of the Theil statistics for the following variables: *number of items in inventory* and *number of workers in inventory*.

Table 5. *The Theil statistics for the number of items leaving the inventory and the number of inventory workers*

Parameter	*Items out of inventory*	*No of workers in inventory*
UM	0.0027	0.0955
US	0.4447	0.8128
UC	0.5526	0.0917
U	1	1

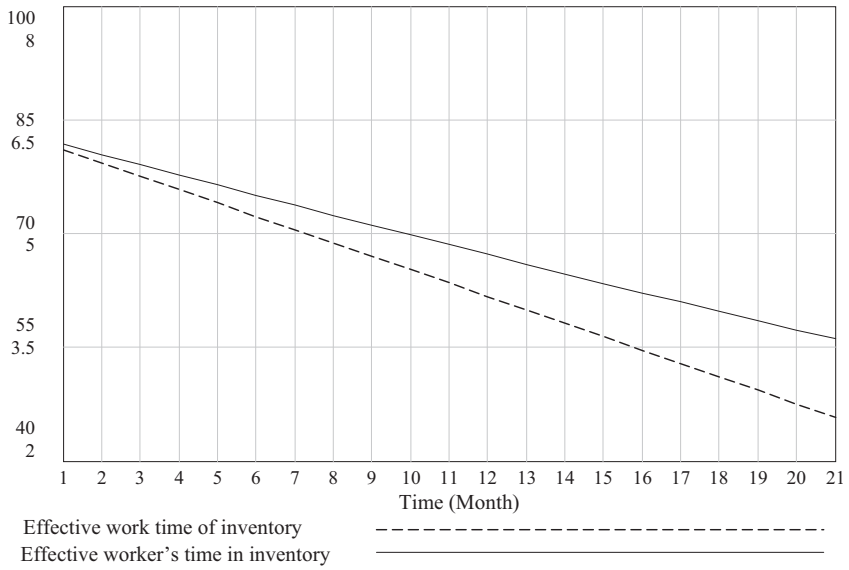

Figure 5. Comparison of the effective working time of a production cell with the effective working time of a single worker

The Theil statistics parameters were supplemented with the analysis of the charts where actual data was compared with the simulated ones (Figures 6 and 7).

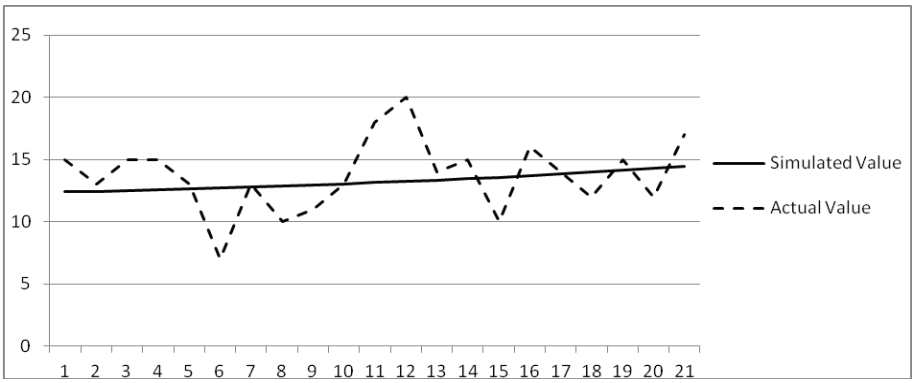

Figure 6. Comparison of the charts representing the actual and simulated data about the number of workers in a production cell (the variable: number of workers in inventory)

The characteristics of the set of the Theil statistics parameters for the variable *Number of workers in inventory* can be regarded as belonging to the (0, 1, 0) pattern. The disturbances in the real system, as indicated in the chart, are not taken into consideration in the model. Since, as it has been mentioned before, the purpose of the model is not to examine the periodic fluctuations, the error can be seen as non-systematic.

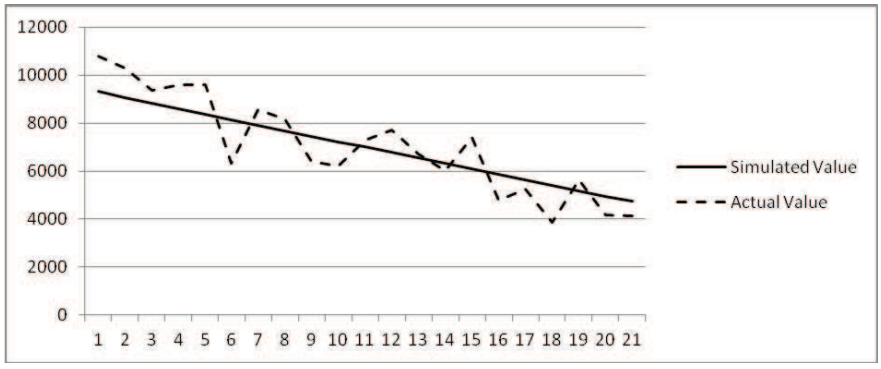

Figure 7: Comparison of the charts representing the actual and simulated data about the number of items leaving the inventory (the variable: Items out of inventory)

In case of the variable *Items out of inventory* the Theil statistics system (1, α, 1-α) as well as the chart analysis allow us to make an assumption that the real and simulated data have the same mean and trends, but they differ point by point and the deviations from the real data are the effect of a non-systematic error.

Therefore the analysis of the data and the comparison of the charts presenting the actual and simulated data lead to the conclusion that in this particular aspect the model is burdened with a non-systematic error, thus it can be regarded as reliable.

Discussion

The above example proves that the afore discussed part of the model can be regarded as workable as far as its representation of the real system is concerned. Admittedly, we can assume that if the model has turned out reliable in case of one production cell, the same computation scheme will be applicable to the others. But if we want to prove beyond any doubt that the whole model is reliable, we should expand the validation over all the remaining cells. Moreover, the elements that are linking individual cells should be tested as well. First of all, however, we should find out if the total number of workers in all the production cells does not exceed the number of workers in the company. In addition, the work-in-progress in any of the cells cannot reach negative values.

We should also make certain that the values adopted as data are obtained from the IT system currently operating in the company. In order to improve the reliability of computations it is worth considering the introduction of additional parameters - such as the time a single worker spends in a production cell - to the company records.

At the same time it seems worthwhile to prolong the period of time covered by the analysis – it could result in better leveling off the fluctuations occurring in the real system.

Conclusion

Summarizing, it is worth noting that the definitions of validation include such expressions as ‘a satisfying level of accuracy’ or ‘in keeping with the model’s intended use’, which are subjective phrases. It means that whether a model is considered reliable or not will largely depend on the judge’s impression (Sargent, 1998). What is more, neither verification nor validation is absolute, so we are not able to decide if the model has been verified or validated in 100%. Therefore, we cannot acknowledge that it is reliable unless we have run as many tests as necessary (Carson, 2002, p. 52). Although the model

can be tested at different stages of its creation by various people (such as the modeler or an independent expert), the truth is that it is its user who will eventually decide how effectively the model helps them in their decision-making.

References

Ranganath, B. (2008). *System dynamics: Theory and case sudies.* New Dehli: I. K. International Publishing House.

Balcerak, A. (2003). Walidacja modeli symulacyjnych – źródła postaw badawczych. *Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej, 74*(15), 27-44.

Balci, O. (1986). Creditibility assessment of simulation results: The state of the art. *Proceedings of the Conference on Simulation methodology and Validation.* San Diego: The Society for Computer Simulation.

Barlas, Y. (1996). Formal aspects of model validity and validation in system dynamics. *System Dynamics Review, 12*(3), 183-210.

Campuzano, F. i Mula, J. (2011). *Supply chain simulation: a system dynamics approach for improving performance.* New York: Springer.

Carson, J. S. (2002). Model verification and validation. *Proceedings of the 2002 Winter Simulation Conference*, (strony 52-58). San Diego.

Davis, P. (1992). *Generalizing concepts and methods of verification, validation and accreditation (VV&A) for military simulations.* Santa Monica: RAND.

Jaszkiewicz, A. (1997). *Inżynieria oprogramowania.* Gliwice: Helion.

Kasperska, E. (2005). *Dynamika systemowa - symulacja i optymalizacja.* Gliwice: Wydawnictwo Politechniki Śląskiej.

Łatuszyńska, M. (2008). *Symulacja komputerowa dynamiki systemów.* Gorzów Wielkopolski: Wydawnictwo Państwowej Wyższej Szkoły Zawodowej.

Maciąg, A., Pietroń, R. i Kukla, S. (2013). *Prognozowanie i symulacja w przedsiębiorstwie.* Warszawa: PWE.

Martis, M. (2006). Validation of simulation based models: A theoretical outlook. *The electronic Journal of Business Research Methods, 4*(1), 39-46.

Pidd, M. (1998). *Computer simulation in management science.* Chichester: John Willey & Sons.

Robinson, S. (1997). Simulation model verification and validation: Increasing the users confidence. *Proceedings of the 1997 Winter Simulation Conference*, (strony 53-59). Atlanta.

Sargent, R. (1998). Verification and validation of simulations models. *Proceedings of the 1998 Winter Simulation Conference*, (strony 121-130). Washington.

Sargent, R. (2010). Verification and validation of simulation models. *Proceedings of the 2010 Winter Simulation Conference*, (strony 166-183). Baltimore.

Sterman, J. (1984). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamica, 10*(II), 51-65.

Sterman, J. (2000). *Business Dynamics System Thinking and Modeling for a Complex World.* New York: McGraw-Hill.

Tarajkowski, J. (2008). *Elementy dynamiki systemów.* Poznań: Wydawnictwo Akademii Ekonomicznej w Poznaniu.

Abstrakt (in Polish)

Celem niniejszego artykułu jest analiza procesu walidacji modeli zbudowanych w konwencji metody dynamiki systemów na przykładzie modelu alokacji zasobów produkcyjnych. W pierwszej części artykułu autorzy przedstawiają przegląd definicji walidacji i weryfikacji, które można znaleźć w literaturze przedmiotu. Ponadto podkreślają rolę, jaką odgrywa weryfikacja i walidacja w procesie modelowania. Omawiają także techniki walidacji modelu systemu, ze szczególnym uwzględnieniem testów struktury modelu, jego zachowania i implikacji decyzyjnych. Druga część artykułu zawiera przykład procesu walidacji modelu systemu alokacji zasobów produkcyjnych w firmie elektronicznej. Celem tego modelu jest ocena długoterminowego wpływu przypisania pracowników do poszczególnych zadań na takie parametry produkcyjne jak wydajność oraz efektywny czas pracy. Autorzy koncentrują swoją szczególną uwagę na części modelu dotyczącej magazynu, przeprowadzając badania jej struktury i zachowania. Podczas sprawdzania poprawności struktury autorzy korzystają z informacji uzyskanych w serii wywiadów z pracownikami firmy. Odnoszą się także do ogólnie przyjętej wiedzy, jaką można znaleźć w literaturze przedmiotu. Uzyskane wyniki badań są porównane z danymi rzeczywistymi. Autorzy oceniają zarówno logikę zachowania systemu jak i poziom dokładności danych wyjściowych w odniesieniu do systemu rzeczywistego.

Słowa kluczowe: *walidacja modeli symulacyjnych, dynamika systemów.*