The Use Of Meta Models To Interpret The Results Of Simulation Studies

Vladimir Jerz, Marian Kralik
Faculty of Mechanical Engineering
Slovak University of Technology
Bratislava, Slovakia
vladimir.jerz@stuba.sk – marian.kralik@stuba.sk

Abstract—This article discusses some possibilities of correctly interpreting the results of simulation experiments, carried out using simulation systems. Metamodels can be used for the transparent expression of the influence of various process parameters on the resulting characteristics of these processes. These are models that can be created on the basis of the evaluation of sequences of experiments with simulation models of systems and processes.

Metamodel, simulation, Witness

I. INTRODUCTION
The method of systems simulation is now widely used for the analysis of the behavior of technical systems and processes. Given the difficulty of mastering work with top simulation systems and the complexity of correctly interpreting the results of simulation experiments, this issue is usually dealt with by professionals who specialize in the related optimization methods and the selected simulation software. Customers and those interested in the study tend to be mostly the owners and managers of companies, usually aiming to achieve unspecified improvements, possibly including saving workers, increasing production, reducing costs, changing the product range, saving production areas, reducing transit times, achieving lean manufacturing and so on. Proposals for improvements are based on detailed analyses of the processes, often using exact methods. The effects of the implementation of these measures can be checked using simulation procedures alone. The correct interpretation of the results of simulation experiments is a prerequisite for the selection of the recommendations that lead to the best solution. A clear, formal expression of the results of simulation studies is useful not only for experimenters themselves, but especially for customers who generally do not know the details used in the simulation models and experiments, but must be convinced of the appropriateness of the proposed measures and the anticipated results once implemented.

Metamodels can be used for the transparent expression of the influence of various process parameters on the resulting characteristics of these processes. These are models that can be created on the basis of the evaluation of sequences of experiments with simulation models of systems and processes. These can be formulated in the form of equations – mathematical dependencies – or can be provided in a clear graphical form. They can be regarded as descriptive predictive models, because they reflect certain behavioral characteristics of the modeled system in present state, but mostly in the state anticipated in the future.

II. SUMMARY OF THE SIMULATION MODEL AND SIMULATION OPTIMIZATION MODEL
For the purposes of the following considerations, the simulation model can be expressed as a function:

\[ y = f(p_1, p_2, \ldots, p_n) \quad (1) \]

or, alternatively,

\[ Y = f(p_1, p_2, \ldots, p_n). \]

where \( Y = (y_1, y_2, \ldots, y_m) \) \( (2) \)

or by the mapping

\[ m: P \rightarrow Y \quad (3) \]

The \( p_i \) variables will be designated as "input parameters" of the model, and parameters to be obtained at the output of the simulation experiment will be denoted as "monitored output variables" of model.

Function \( f \) cannot be formulated as a mathematical function. It is realized through simulation experiments using the simulation model. Values of the monitored output variables cannot be calculated, but they are obtained at the conclusion of the experiment using the simulation model. In certain cases they may be acquired as a result of a subsequent calculation by which they may enter the values...
of the input parameters of the model or miscellaneous constants, along with at least one output variable.

Input parameters often have the character of random variables. They are therefore input through theoretical or empirical probability distributions and their parameters.

In some cases, in terms of the appreciation of the results of the simulation experiment, not just the resulting value of the output variable (e.g. average queue length) is interesting, but also the dynamics of changes in the values of certain variables during a simulation run (e.g. time-dependent queue length). It is clear that the output variables also are of the nature of random variables, which must be taken into account in their interpretation [4].

III. METAMODELS

In the process of finding optimal solutions, or more often the best solutions in terms of the objectives pursued, usually some variants of the simulation model are compared in terms of the values of the observed output variable (or multiple output variables). If the individual variants differ from each other only by input parameter values, or if the differences in the structure can be expressed by values of the input parameters, the model can transform the input parameters to an output variable generally expressed by functional dependence (1) or (2).

As already mentioned, in the simulation models the value of an output variable is obtained by performing the simulation experiment, via statistical processing of the output values obtained from a sequence of experiments. Generally, it has the character of a random variable. The value of output variable cannot be calculated considering that in most cases the function cannot be formulated mathematically. It is clear, however, that the mathematical expression of that dependence could serve as an appropriate basis for predicting the behavior of the modeled object. This has led to efforts to formulate mathematical notations for dependencies expressed by the functions known as metamodels. The basis for metamodel creation is simulation study – a sequence of experiments using variants of the simulation model.

In the simplest case, the dependency of single mentioned output variable \( y \) consists of only one relevant input parameter ( \( n = 1 \) ). Graphically, such a dependence can be expressed by a curve. If it is obvious that it will be a linear function, it is sufficient to perform simulation experiments with two values of the parameter. This gives 2 points, which can define a line – in other words, they are sufficient to calculate the coefficients \( a \) and \( b \) in the equation of straight line: \( y = a + bp \).

Another option is to create such a model based on a sequence of experiments with multiple input parameter values – for example, by the method of least squares. In real terms, however, it is questionable to what extent the assumption of linearity is justified. If we use the second of these methods and the help of the graphical representation of the sequence of experiments, we can assess whether there might be a linear relationship. Whether such an approximation is sufficient depends on the intended use of the generated metamodel. In any case, if the model is used to predict the behavior of real objects, it is preferable to express it in a form \( y = a + bp + \varepsilon \) where \( \varepsilon \) is used to express the general stochastic effects in creating a metamodel. It is a classical linear regression model. In addition to the classical linear regression model, in the literature other models are analyzed which can be considered linear [1, 3] and non-linear.

The veracity and credibility of the metamodel can be increased if in its creation the stochastic character all its parameters is taken into account.

IV. DESCRIPTION OF THE SIMULATION MODEL CREATED FOR THE PURPOSE OF THIS PRESENTATION

The simulation system Witness was used for the creation of the simulation model for the purpose of this presentation (Fig. 1). The purpose of these experiments was to find a suitable number of workers and pallets on circular conveyor. The monitored output variable was the productivity of the system (number of products produced in a given time), and in a modified version of the experiment it was profit (revenues minus costs).

Fig. 1. Layout of the created simulation model

Using the function \( f \), the dependence of the observed output variable can be written as follows:

\[ y = f(p_1, p_2) \]

For obtaining the data for the formulation of the aforementioned dependency a sequence of experiments with different suitable values of the input parameters needs to be done. Experiments can be performed sequentially - each tone following the assignment of specific values of the input parameters. However, the simulator Witness provides a Scenario Manager module to simplify experimentation and an Optimizer module that is designed to directly search for optimal combinations of the parameters, but in cases with a smaller number of discrete values of the considered input parameters the "All Combinations" method can be used to significantly accelerate the obtaining of the necessary results. The
Optimizer module was used so that each combination of the values of the input parameters was simulated several times in order to eliminate the impact of fluctuations in the values of random variables in the model on the credibility of the results.

The first sequence of experiments was performed in order to obtain the dependence of productivity ($y$) on the number of pallets on the conveyor ($p_2$) given a constant number of workers ($p_1 = 5$). It is clear that this is not a linear dependence. It is simple to create a metamodel in graphical form (Fig. 2). Under these conditions, it is clear that increasing the number of pallets beyond 8 does not lead to an increase in productivity. Conversely, a significant increase in the number of pallets would cause conveyor deadlock (in cases of significantly greater numbers of pallets, as was shown in the experiment).

Using Excel, it is relatively easy to formulate a mathematical metamodel with sufficient accuracy. Across the points of dependence listed above one of the curves provided by Excel (exponential, logarithmic, power law or polynomial curve up to 6th degree) can be drawn. In this case, a 4th degree polynomial seems to be appropriate. Excel offer this directly in both graphical and mathematical form (Fig. 3).

If the intention is to create a metamodel which can inform us about the values several output parameters depending on some input parameter, the graph can contain more curves which can be differentiated by line type or color. In view of the different units in which the values of different output parameters are given, it is possible to use a graph with several different scales on the vertical axis.

The target of second sequence of experiments was to find the dependence of productivity on the two aforementioned input parameters simultaneously. The graphic representations of these dependencies which are used in the case of continuous input parameters are shown in the literature and they are based on visual display by means of contours or on a three-dimensional view (Fig. 4).

In the case of discrete values of the input parameters, Excel spreadsheets can be used as visual displays, which in later versions of Excel provide the possibility of conditionally formatting cells using color scales (Fig. 5). For this obvious figure, it is clear that the highest productivity is obtained with a minimum of five workers.
and nine pallets. Increasing the number of either doesn’t produce a better result.

<table>
<thead>
<tr>
<th>number of workers</th>
<th>number of pallets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

Fig. 5. Dependence of productivity on the number of workers and pallets

The modified metamodel is based on the same simulation model, but the output value is profit. Profit is calculated as the difference between the revenues of manufactured products and the cost of the given numbers of workers and pallets within a specified time period (Fig. 6). It is obvious that in terms of this criterion a combination of 3 workers and 6 pallets is optimal.

<table>
<thead>
<tr>
<th>number of workers</th>
<th>number of pallets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Fig. 6. Dependence of profit on the number of workers and pallets

The purpose of these metamodels is not only to find the optimal combination of input parameter values in terms of the selected criteria (an output of “Optimizer” module would suffice in that case) but also to provide an estimate of the output values for any combination of values of the input parameters that could be transiently used for various reasons (dropout of workers, maintenance of technical equipment, etc.).

In all these examples, the displayed value of each output variable was obtained by averaging the outputs of a sequence of several experiments, differing only by using different streams of random numbers in the model. This means that increasing the number of experiments only produces a more accurate average value, i.e. the narrowing of the confidence interval around each outcome. If the purpose of the metamodel is to predict results which the simulated system could achieve, it seems to be more appropriate to express each datum in the chart or designed table as the boundary of prediction interval.

In fig. 7 is an example of a metamodel that provides information on the time required to service the given number of demands in a queuing system (it is a different model than was used in the previous discussion). From the nature of the model it follows that the dependence is linear, so a straight line was drawn across the points obtained from a sequence of simulation experiments by the method of least squares. For outermost values of the input parameter (number of demands), 10 runs were performed with different streams of random numbers and the boundaries of the prediction intervals were calculated with a probability of 95%. The shaded area shows how much time will be necessary for the processing of the given number of demands with a probability of 95%.

Fig. 7. Time required to service the given number of demands with 95% probability [2]

The question arises as to whether it would be possible to use the so-called error bars that are available in Excel for graphical representations for these and similar purposes. There is an option to enter an absolute deviation or percent deviation from a certain value. It is clear that the use of error bars is only possible if over the whole considered range of the independent variable the width of the prediction interval has about the same percentage or absolute deviation from the mean of the observed output variable. In the example above, however, the absolute deviation of the outermost points is in the interval between 20.8 and 30, and in percentage terms, between 13% and 6%. It is clear that the use of error bars in this case it would be inappropriate.

The question arises as to how, in the case of nonlinearity, the estimated values of the output variable with a given probability could be expressed mathematically and also displayed in a graph.

If for these reasons it isn’t possible to use the error bars to generate a graphical representation, it would possible to depict this area using 3 curves. The first would be formed by fitting a curve across the points representing the mean of the output variable. Curves bounding that area from above and below could be created in a similar way – by fitting curves across the points representing the upper and lower bounds of the prediction intervals.

V. CONCLUSION AND FUTURE SCOPE

There are more possibilities where the metamodels in presentation of results of simulation studies would be used. Of course the choice of specific tool depends on the problem to be solved. Metamodels provide major opportunities for presenting the results of simulation experiments and therefore should be used to a greater extent, especially when communicating with professionals who aren’t experts in simulation. Therefore it is necessary to continue to raise awareness of the possibilities of metamodels, but at the same time promote opportunities for
expansion of simulation software by the tools to facilitate
the use of metamodels for outputs from simulation studies.

This work was supported by scientific grant agency of
the Ministry of Education, Science, Research and Sport of
the Slovak Republic under the contract No. VEGA
1/1056/12 Research of progressive methods and devices in
automation of production.

REFERENCES


[2] Jerz, V.: Simulation and optimization models – the
basic tool for processes improvement. In.: Posterus –
internet journal (www.posterus.sk), 3/11 - 2010, ISSN
1338-0087

0-07-298843-7, ISBN-10 0-07-298843-6

Strengths and Weaknesses of Simulation
Optimization. World Academy of Science,
Engineering and Technology 65/2012, pp. 892-896,
www.waset.org/proceedings.php, eISSN 2010-3778