On the Complexity of Parameter Calibration in Simulation Models

Marko Hofmann
Institute for Technology of Intelligent Systems (ITIS)
University of the Federal Armed Forces, Munich
85577 Neubiberg, Germany
marko.hofmann@unibw.de

Model calibration is the task of adjusting an already existing model to a reference system. In general, this is done by adjusting model parameters to a set of given samples from the reference system. Model calibration is often regarded to be necessary for complex simulation models in order to create a homomorphic ("structurally equivalent") abstraction of (a special aspect of) reality. This paper introduces a formal approach to model calibration. Within the frame of this formalism it is shown that the computational complexity of model calibration is NP-complete. The practical implications of these theoretic results are presumably of minor importance for most single models. However, for huge model federations the complexity of parameter calibration could draw a serious line with respect to the validation of the federation and its cost-benefit ratio.

Keywords: Validation methodology, military applications, computer science, complexity theory

1. Introduction: Model Calibration

Model calibration is the task of adjusting an already existing model to a reference system (or, if system data is not available, to a trusted reference model). This is usually done by adjusting the (internal) parameters of the model according to input-output sets of the system (or reference model).

The use of the term "calibration of simulation models" dates to the origins of computer simulation itself; see, for example, Barker and Watson [1]. The adjustments done by parameter calibration are necessary because the models are based on abstractions, idealization, and many disputable assumptions. Thus, in order to get trustworthy results from the model, input-output pairs of the model are fine-tuned to input-output samples of the reference system; see Figure 1.

The importance of model calibration for practical work is highlighted in many publications. Closely related to the practical considerations of this paper are Davis et al. [2], Davis and Hillestad [3], Davis [4], and Davis and Bigelow [5, 6]. General discussions of model calibration can be found, for example, in Ören [7], Wigan [8], and Goldberg and Kaczka [9]. The practical importance of calibration is controversial. Often, truly reliable data is not available and even the reference system is not clearly defined. In Davis et al. [2], the calibration of most military models is therefore described as something more humble, approximate, and selective than fine-tuned calibration. Critical remarks on calibration with respect to validity can be found in Hemez [10] and Bayarri et al. [11]. Despite these qualifications there are many simulation models (especially technical simulations) that can be and have to be calibrated in the classical sense of fine-tuning.

The remainder of this paper is organized as follows: section 2 presents a theoretical model of parameter calibration. Within this model it is shown that parameter calibration is NP-complete. Section 3 discusses the formal model, the assumptions, and the results of section 2. Section 4 considers the practical implications of the result, and section 5 concludes the paper by summing up its contributions and making some conclusions.

2. Theoretical Considerations

This section provides a formal approach to model calibration and further to the complexity of model calibration. The formalism is tailored to the special issue of calibration addressed here. It is therefore necessarily incomplete. In section 1 a transformation of the reasoning into the DEVS formalism (Discrete Event System Specification; see Zeigler [12, 13]) is outlined.

JDMS, Volume 2, Issue 4, October 2005 Pages 217–226

2.1 Basic Definitions and Assumptions

2.1.1 Models and Systems

DEFINITION 1. A *model, M,* is a triple $(X^M, \varphi^M(P^M), Y^M)$ with

X^M : Set of possible values for the model input variables $(x_1^M,...,x_m^M)$; $X^M = (X_1^M,...,X_m^M)$; X_i^M = Domain for each variable.

$\varphi^M(P^M)$: Model transformation function $\varphi^M : X^M \times P^M \longrightarrow Y^M$ with model parameters $(p_1^M,...,p_n^M) = (p_1,...,p_n)$ and their domain sets $(P_1^M,...,P_n^M) =: P^M$.

Y^M : Set of possible values for the model output variables $(y_1^M,...,y_k^M)$; with $(y_1^M,...,y_k^M) = \varphi^M(x_1^M,...,x_m^M, p_1^M,...,p_n^M)$ and $Y^M = (Y_1^M,...,Y_k^M)$.

According to this definition a model is simplified to its input, output, and the transformation between them. The transformation function depends on the model parameters.

DEFINITION 2. A *reference system, S,* is a triple (X^S, φ^S, X^S) with

X^S : Set of possible values for the system input variables $(x_1^S,...,x_q^S)$; $X^S = (X_1^S,...,X_q^S)$; X_i^S = Domain for each variable.

φ^S : System dynamics function $\varphi^S : X^S \longrightarrow Y^S$.

Y^S : Set of possible values for the system output variables $(y_1^S,...,y_r^S)$; with $(y_1^S,...,y_r^S) = \varphi^S(x_1^S,...,x_r^S)$ and $Y^S = (Y_1^S,...,Y_r^S)$.

According to this definition a system is simplified to its input, output, and the transformation between them. In contrast to the model transformation function φ^M the system transformation function φ^S is, in general, not known.

DEFINITION 3. A *model application,* $\tilde{M}$, is a triple $(\tilde{X}^M, \varphi^M(\tilde{P}^M), \tilde{Y}^M)$ with

$\tilde{X}^M$: Tuple of input *values* $(\tilde{x}_1^M,...,\tilde{x}_m^M)$ from $(X_1^M,...,X_m^M)$,

$\varphi^M(\tilde{P}^M)$: Concretization of $\varphi^M(P^M)$ with a tuple of *values* for the model parameters $(\tilde{p}_1^M,...,\tilde{p}_n^M)$ from $(p_1^M,...,p_n^M)$, and

$\tilde{Y}^M$: Tuple of output *values* $(\tilde{y}_1^M,...,\tilde{y}_k^M)$ from $(Y_1^M,...,Y_k^M)$.

In other words, the term *model application* denotes the assignments of concrete values to the model.

DEFINITION 4. A *reference system sample,* $\tilde{S}$, is a pair $(\tilde{X}^S, \tilde{Y}^S)$ with

$\tilde{X}^S$: Tuple of input *values* $(\tilde{x}_1^S,...,\tilde{x}_q^S)$ from $(X_1^S,...,X_q^S)$, and

$\tilde{Y}^M$: Tuple of output *values* $(\tilde{y}_1^S,...,\tilde{y}_r^S)$ from $(Y_1^S,...,Y_r^S)$, corresponding to $(\tilde{x}_1^S,...,\tilde{x}_q^S)$.

A reference system sample is the result of one concrete measurement in the reference system.

DEFINITION 5. A set of l reference system samples is defined as

$$\{\tilde{S}_1,...,\tilde{S}_l\} := \{(\tilde{X}_1^S, \tilde{Y}_1^S), (\tilde{X}_2^S, \tilde{Y}_2^S), ..., (\tilde{X}_l^S, \tilde{Y}_l^S)\}.$$

Assumption 1. It is assumed that the input X^S of the reference system can be transformed into the input X^M of the model and that the output of the model Y^M can be (re-)transformed into the output of the reference system Y^S by invertible functions $\psi : X^S \longrightarrow X^M$ and $\omega : Y^M \longrightarrow Y^S$ so that

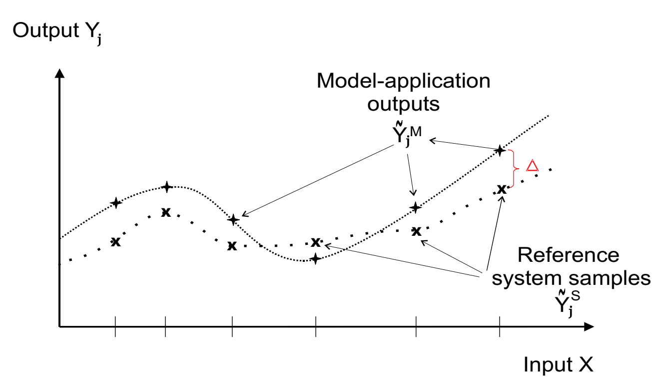

Figure 1. Model output and empirical data

$$\forall (x_1^S,\ldots,x_q^S)\in X^S, \exists (x_1^M,\ldots,x_m^M)\in X^M:$$
$$\psi(x_1^S,\ldots,x_q^S)=(x_1^M,\ldots,x_m^M)$$

and

$$\forall (\hat{y}_1^S,\ldots,\hat{y}_r^S)\in Y^S, \exists (y_1^M,\ldots,y_k^M)\in Y^M:$$
$$\omega(y_1^M,\ldots,y_k^M)=(\hat{y}_1^S,\ldots,\hat{y}_r^S).$$

Thus, every possible input of the system can be transformed into a unique model input as well as every model output (re-)transformed into a well-defined system output. Note that $\hat{y}_i^S$ denotes the result of the transformation of a model output into the domain of a system output, not the system results y_i^S itself:

$$\varphi^S(x_1^S,\ldots,x_q^S)=(y_1^S,\ldots,y_r^S),$$
$$\omega(\varphi^M_{p_1^M,\ldots,p_n^M}(\psi(x_1^S,\ldots,x_q^S)))=(\hat{y}_1^S,\ldots,\hat{y}_r^S).$$

Figure 2 sketches the formal model presented so far.

Assumption 2. The functions ψ and ω are regarded as fixed for every system-model connection.

They *could* be considered as variable, too. The interpretation of the model input and output would then be time and context dependent. This is sometimes relevant in practice. However, variable transformation functions ψ and ω would only increase the complexity of calibration.

2.1.2 Congruence

The purpose of parameter calibration is to assure that a model reproduces adequately observed behavior of a reference system. In other words, the models

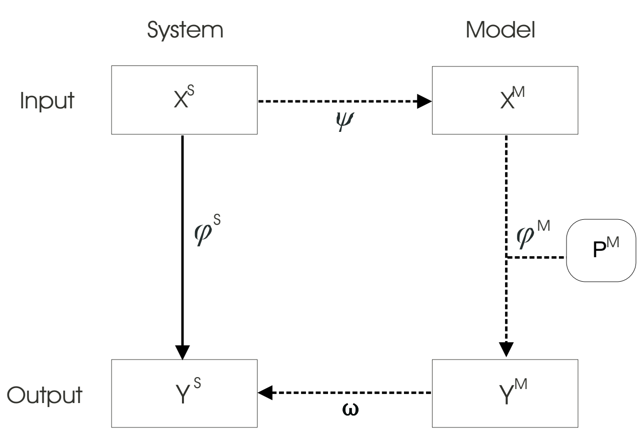

Figure 2. Value domains in modeling and transformations between them

should produce the same or at least similar outputs to a given input as the corresponding systems do. This congruence of model and system can be defined in two straightforward ways:

Definition 6. *Weak congruence* between a reference system, S, and a model, M, is defined as

$$\forall (x_1^S,\ldots,x_q^S)\in X^S \exists (p_1^M,\ldots p_n^M)\in P^M:$$
$$\begin{aligned}\varphi^S(x_1^S,\ldots,x_q^S)=(y_1^S,\ldots,y_r^S)&=\omega(\varphi^M(\psi(x_1^S,\ldots,x_q^S),p_1^M,\ldots,p_n^M))\\&=\omega(\varphi^M(x_1^M,\ldots,x_m^M,p_1^M,\ldots,p_n^M))\\&=\omega(y_1^M,\ldots,y_k^M)\\&=(\hat{y}_1^S,\ldots,\hat{y}_r^S).\end{aligned}$$

This formal definition denotes that for every possible system input ($\forall (x_1^S,\ldots,x_q^S)\in X^S$) there exists at least one parameter configuration of the model ($\exists (p_1^M,\ldots p_n^M)\in P^M$), so that the outcome of the model application equals the output of the system.

Definition 7. *Strong congruence* between a reference system, S, and a model, M, is defined as

$$\exists (p_1^M,\ldots p_n^M)\in P^M \forall (x_1^S,\ldots,X_q^S)\in X^S:$$
$$\begin{aligned}\varphi^S(x_1^S,\ldots,x_q^S)=(y_1^S,\ldots,y_r^S)&=\omega(\varphi^M(\psi(x_1^S,\ldots,x_q^S),p_1^M,\ldots,p_n^M))\\&=\omega(\varphi^M(x_1^M,\ldots,x_m^M,p_1^M,\ldots,p_n^M))\\&=\omega(y_1^M,\ldots,y_k^M)\\&=(\hat{y}_1^S,\ldots,\hat{y}_r^S).\end{aligned}$$

This formal definition denotes that there exists at least one model parameter configuration ($\exists (p_1^M,\ldots p_n^M)\in P^M$), so that for all possible system inputs ($\forall (x_1^S,\ldots,x_q^S)\in X^S$) the outcome of the model application equals the output of the system.

In the first case (Definition 6), every model parameter is regarded as variable with respect to each special reference system sample i: (To distinguish between the values of different samples an additional sample index is used.)

$$(y_{1,i}^M,\ldots,y_{k,i}^M)=\varphi^M(\psi(x_{1,i}^S,\ldots,x_{q,i}^S),p_{1,i}^M,\ldots,p_{n,i}^M).$$

Weak congruence implies adjustment of the model to every new input, which is desirable only in special cases (like, for example, Bayes-learning). In all other cases a stronger definition is needed. Unfortunately, the demand of Definition 7 is extremely hard to prove and much too strong for almost every model of a complex real reference system. Thus, a more pragmatic approach is necessary. Instead of demanding perfect equality, deviations within a predefined range should be allowed between system *samples* and corresponding model outputs. Definition 8 introduces such a definition.

Definition 8. A model is *pragmatically congruent* if

$$| y^S_{j,i} - \hat{y}^S_{j,i} | \le \Delta_j \quad ; \quad \forall j \in \{1,...,k\}, \quad \forall i \in \{1,...,l\} .$$

Δ_j are the *problem-specific* "tolerated deviations" for the output variable j ; i is the index for the samples. Figure 2 and the congruence definitions may remind the reader of the concept of strong and weak consistency in multiresolution modeling (using models of different aggregation for similar purposes), as introduced in Davis and Bigelow [14]. Although the definition of congruence is independent from consistency, both share the same notion: the quantification of deviations between models and systems, respectively, between models of different resolution. The concept of pragmatic congruence is similar to what Davis and Bigelow have called strong consistency. The two approaches converge if the reference system (in congruence) is regarded as a special high-resolution model (in consistency) and if the tolerated deviations for each output variable (in congruence) are considered to be summarizable as one output variable (in consistency). Such a summarization of different output variables in one aggregated variable would, of course, reduce the overall complexity of calibration. Nevertheless, the problem would remain NP-complete: in the following reasoning one only has to renounce the double indexing both for samples *and* output variables.

However, there is one aspect in the work of Davis and Bigelow which is not addressed by congruence. The authors emphasize the importance of analytic outputs, which are projections (transformations) of direct outputs [14]. Such a projection might be an average over time, individual battles, etc. These outputs are often insensitive to details of the simulation. Such "purposeful aggregations" are neglected in this formalism, since they are out of the reach of a complete formalization.

Definition 9. To simplify the notation a selection function ω_j is introduced:

$$\omega_j : Y^M \longrightarrow Y^S_j \ ,$$

$$\omega_j(y^M_1, ..., y^M_k) := y^S_j \ .$$

Definition 10. Given a set of reference system samples $\{S_1, ..., S_l\}$, *model calibration* is the task of adjusting the values of the parameters, $(p^M_1, ..., p^M_n)$, so that

$$| y^S_{j,i} - \omega_j(\varphi^M(\psi(x^S_{1,i}, ..., x^S_{q,i}), p^M_1, ..., p^M_n)) | \le \Delta_j ;$$
$$\forall j \in \{1,...,k\}, \quad \forall i \in \{1,...,l\}$$

$$\Leftrightarrow$$

$$| y^S_{j,i} - \hat{y}^S_{j,i} | \le \Delta_j \quad ; \quad \forall j \in \{1,...,k\}, \quad \forall i \in \{1,...,l\}.$$

In short, calibration means adjustment of a model's parameters, so that all deviations between reference system sample variables and model output variables remain within predefined ranges of tolerance. With this definition it is now possible to formally address the problem of calibration complexity.

2.3 Complexity of Calibration

Assumption 3. All model parameters, p_i, are Boolean parameters: $p_i \in \{true, false\} \sim \{1,0\}$.

Definition 11 is only a reminder of some fundamental concepts and terms of Boolean logic.

Definition 11.

1) A logic variable p_i (a "Boolean parameter") or its negation $\overline{p}_i$ is called a literal.
2) A disjunction (OR-connection) of literals is a clause.
3) A conjunction (AND-connection) of clauses is a conjunctive form.
4) If every Boolean parameter occurs at least once in the conjunctive form it is called a conjunctive normal form (CNF)[1].
5) The problem of deciding whether a given Boolean formula (for example in CNF) is satisfiable (the whole term is true), is called the satisfiability problem (SAT).

Example 1.

$$(p_1 \vee p_2 \vee p_3) \wedge (\overline{p}_1 \vee \overline{p}_2 \vee \overline{p}_3)$$
$$\wedge (\overline{p}_1 \vee p_2 \vee p_3) \wedge (\overline{p}_1 \vee \overline{p}_2 \vee p_3)$$

is a Boolean formula in conjunctive normal form which is satisfiable, for example, with $(p_1, p_2, p_3) = (0,1,1)$ (whereas $(p_1, p_2, p_3) = (0,0,0)$ would fail to satisfy the first clause).

Assumption 4. For every output variable of every reference system sample there exists a parameter combination, *expressible as a formula in propositional logic*, so that

$$| y^S_{j,i} - \omega_j(\varphi^M(\psi(x^S_{1,i}, ..., x^S_{q,i}), p^M_1, ..., p^M_n)) | \le \Delta_j ;$$
$$\forall j \in \{1,...,k\}, \quad \forall i \in \{1,...,l\}$$

1 Product of sums (PoS)

is fulfilled.

Example 2. Assume that

$$
\begin{aligned}
&| y_{1,1}^M - y_{1,1}^S | \leq \Delta_1 \text{ is true for } p_1 \wedge \overline{p}_4 \vee p_{n-1}, \\
&| y_{2,1}^M - y_{2,1}^S | \leq \Delta_2 \text{ is true for } p_5 \vee \overline{p}_8 \vee p_n, \\
&\vdots \\
&| y_{k,1}^M - y_{k,1}^S | \leq \Delta_k \text{ is true for } \overline{p}_2 \vee \overline{p}_3 \wedge p_9 \vee \overline{p}_n, \\
&| y_{1,2}^M - y_{1,2}^S | \leq \Delta_1 \text{ is true for } p_1, \\
&| y_{2,2}^M - y_{2,2}^S | \leq \Delta_2 \text{ is true for } p_3 \vee \overline{p}_4 \otimes p_{n-1} (\otimes := XOR), \\
&\vdots \\
&| y_{k,2}^M - y_{k,2}^S | \leq \Delta_k \text{ is true for } p_7 \vee \overline{p}_{n-1} \vee p_n, \\
&\vdots \\
&| y_{k,l}^M - y_{k,l}^S | \leq \Delta_k \text{ is true for } p_4 \otimes \overline{p}_{n-2} \quad .
\end{aligned}
$$

That would lead to

$$| y_{j,i}^S - \hat{y}_{j,i}^S | \leq \Delta_j \quad ; \quad \forall j \in \{1,...,k\}, \quad \forall i \in \{1,...,l\}$$

is true if

$$(p_1 \wedge p_4 \vee p_{n-1}) \wedge (p_5 \vee \overline{p}_8 \vee p_n) \wedge \cdots \wedge (p_4 \otimes \overline{p}_{n-2})$$

is satisfiable.

Proposition 1. Every formula in propositional logic is truth-equivalent to a unique CNF formula.

The proof can be found in any book about propositional logic: for example, in Kleine Büning et al. [15].

Conclusion 1: Any successful model calibration can be formulated as a conjunctive normal form (CNF) of literals of the model parameters. In other words, the model calibration is successful if and only if the corresponding CNF is satisfied.

Proposition 2. The satisfiability problem (SAT) of Boolean terms in CNF is NP-complete.

This is S. A. Cook's famous theorem, proven in [16] and elaborated upon by Karp [17]. An introduction to the satisfiability problem can also be found in Kleine Büning et al. [15].

NP stands for *nondeterministic polynomial* and denotes the class of *nondeterministic polynomial time solvable problems*, where nondeterministic is just a fancy way of talking about guessing a solution. A problem is in NP if one can quickly (in polynomial time[2]) test whether a solution is correct (without worrying about how hard it might be to find the solution). The theory of NP-completeness is a solution to the practical problem of applying complexity theory to individual problems. NP-complete problems are defined in a precise sense as the hardest problems in NP. Even though we don't know whether there is any problem in NP that is not in P (polynomial time solvable problems), we can point to an NP-complete problem and say that if there are any hard problems in NP, that problem is one of the hard ones. At present, all known algorithms for NP-complete problems require time that is exponential in the problem size. It is unknown whether there are any faster algorithms; see Hopcroft and Ullman [18] for an introduction into complexity theory.

Conclusion 2: With the assumption made, the problem of model calibration is NP-complete.

To summarize this section: at present, the complexity of the best algorithm for finding a solution in the formalized model calibration problem requires time that is exponential in the problem size (number of output variables and number of reference system samples). Hence, there is definitely a problem size that is out of reach of computability.

3. Discussion of Formal Model, Assumptions, and Results

3.1 Formal Model

The formal model is a simplification because it ignores initial and internal states as well as time-flow mechanisms. However, the formalism is not intended to be a new abstract model description technique. Its sole function is to define, as simply as possible but as complexly as necessary, a framework for the definition of model calibration. The model is also based on the assumption that only φ^S is unknown, not the significant input parameters for a certain phenomena. Referencing the nomenclature of system specification hierarchy used in Zeigler et al. [13] (which is based on Klir [19]), the observation frame (Level 0) of the system is taken as given. Levels 3 and 4, the state transition and the internal coupling, are ignored. Thus, the formal model is reduced to represent only I/O behavior (Level 1) and part of the I/O function (Level 2) of a system (or model). The representation of Level 2 is incomplete because initial states are left out. "Congruence" only requires that model and reference system are in line at Level 1. In the nomenclature of Zeigler et al. [13] that is called "replicative validity." This term has been avoided on purpose in this paper, since validity should

2 The computational time required to solve a problem of size n is limited by a polynomial function $o(n) = a_0 + a_1 x^1 + a_2 x^2 + ... + a_n x^n$.

only be assessed in context with the objectives of the modeling (within an "experimental frame" [13]).

3.1.1 Reference to the DEVS Formalism

The DEVS (Discrete Event System Specification) formalism was introduced by B. Zeigler in a seminal publication in 1976 [12]. Since then his formalism has been extended in various directions [13] and is now considered to be the standard framework for theoretical work about discrete modeling and simulation. For that reason, an informal transformation of the reasoning in the formalism used in section 2 and the DEVS formalism is outlined.

A classic DEVS (an "atomic DEVS model") is a structure

$$M = \langle X, S, Y, \delta_{int}, \delta_{ext}, \lambda, ta \rangle$$

where

X	is the set of input values;
S	is a set of states;
Y	is the set of output values;
$\delta_{int} : S \rightarrow S$	is the internal transition function;
$\delta_{ext} : Q \times X \rightarrow S$	is the internal transition function, where $Q = \{(s,e) \mid s \in S, 0 \leq e \leq ta(s)\}$ is the total state set and e is the time elapsed since transition;
$\lambda: \ S \rightarrow Y$	is the output function; and
$ta : S \rightarrow R^{+}_{0,\infty}$	is the time advance function.

DEVS models can be coupled to each other, and they are closed under coupling. (A coupled model is equivalent to an atomic one.) Complex DEVS models are based on the recursive coupling of simpler DEVS models. The composition of DEVS models is made much easier with the DEVS-with-ports extension, where $X := \{(p,v) \mid p \in InPorts, v \in X_p\}$ is the set of input ports and values and $Y := \{(p,v) \mid p \in OutPorts, v \in Y_p\}$ is the set of output ports and values. This extension enables the modeler to use input and output vectors. Detailed explanations of the DEVS formalism and many examples of its capacity can be found in Zeigler et al. [13].

In contrast to the formalism introduced in section 2 (in the following denoted as "calibration formalism"), the DEVS formalism explicitly describes internal states (S) and time advance (ta, respectively, Q). Instead of one transition function between input and output ($\varphi : X \rightarrow Y$ in the calibration formalism) DEVS uses three different functions, for internal and external transitions and the output generation. In exchange, the calibration formalism uses a set of parameters (P^M) to describe the variability of the input-output transformation, at the same time neglecting all internal states.

A transformation of the reasoning in section 2 to the DEVS formalism can be outlined as follows: the binary parameters P^M of the calibration formalism are regarded as indices numbering different possible transition and output functions $\delta_{int}, \delta_{ext}, \lambda$ of atomic DEVS-with-ports models. For the functions actually used to adjust the model to an existing input-output pair of a reference system sample the parameter is set to 1; all others are set to 0. In this setting the complexity of the calibration task increases with the number of the ports and the number of atomic DEVS-with-ports used to construct the whole model. A representation of internal system states and time advance could be easily added to the calibration formalism; however, that would simply not affect the calibration problem as stated in Definition 10, in which internal states and time advance are irrelevant. To a certain extent the models described by the calibration formalism can be seen as black box snapshots from DEVS models.

For a complete integration of the reasoning in this paper into the DEVS formalism it would be necessary to discuss the *mathematical dynamic system theory* (MDST) in which DEVS specifies a subclass of systems. In this theory, levels of structure and behavior are well-defined, as are homomorphisms, including approximate ones that relate to the concepts of weak, strong, and pragmatic congruences presented here. An in-depth explanation of morphisms, approximate morphisms, and levels of structure and behavior within the MDST can be found in chapters 1, 2 (levels of structure and behavior), 12, 13, and 14 (morphisms) of Zeigler et al. [13]. Further general introductions into different facets of the MDST include Klir [19, 23], Luenberger [20], Padulo and Arbib [21], and Beltrami [22].

3.2 Assumptions

Assumption 1 is more critical than it may seem. The measured input of complex system samples is not always directly applicable for models. Often, the data has to be aggregated and completed. In fact, to find an appropriate input and output transformation between system and model can be very challenging. As an example, take attrition processes in aggregated combat simulation models like Lanchester-based models. In general, there is no straightforward way to get the Lanchester coefficients directly from the measurements, and the model output of Lanchester models often needs human interpretation. However, all these adaptations have to be based on reproducible transfer rules. In addition, it is not imperative to have invertible functions, but it facilitates model analysis a lot.

Assumption 2 simplifies the problem. To do without would only increase complexity.

To treat *all* parameters as Boolean (Assumption 3) may seem a rather strong simplification. However, in order to check whether a parameter positively affects the model calibration, modelers often simply switch it on and off. Hence, treating all parameters as Boolean can be regarded as a first approximative approach, similar to what is done in practice. A formal justification of the assumption can be made as follows: the range of non-Boolean parameters is segmented into as many intervals as needed for the calibration. The intervals are taken as new (modified) parameters. If the value of the original parameter falls into a certain interval, the corresponding modified parameter is set to be 1; all others remain 0. These refinements can be as fine-granular as necessary. Hence, less restrictive value domains would only increase the complexity.

Assumption 4 is in fact very weak. Every Boolean term is expressible in propositional logic. Thus, the assumption only demands that a successful parameter combination for the adjustment exists. If not, the model simply would be too limited for the phenomena.

3.3 Results

Note that the conclusions are not only based on the four assumptions but also on what has been defined as model calibration in Definition 10 and on the formal model itself. Subsequently, the results are only as useful as the formal approach matches reality. There is no doubt that practical model calibration comprises more than the task defined in Definition 10. However, the fundamental problem seems to be unaffected by those additional tasks or by special methods and techniques of calibration.

Much more fundamental critiques of the results can be raised in an epistemological context. First, most models of complex systems are only rough approximations of their reference system. They are not intended to be homomorphic to something in reality. Often such models are only thought triggers or means of communication for the modelers. Second, one seldom has all the data needed to do fine-tuned system-model calibration, even if all the data could be retrieved, in principle. Third, some of the most rewarding models deal with massive system inherent uncertainty. Precise knowledge of reference inputs and outputs is lacking in such cases. In order to deal with these problems a completely different approach to model validation has to be used. These approaches are based on conceptual criteria that are hard to formalize; for example, usefulness for exploration and insight, comprehensibility, explainability, and capacity to deal with uncertainty; see Davis and Bigelow [5]. The results of this paper can be interpreted as another argument to intensify research in that direction: even if fine-tuned system-model calibration is possible, it is limited in problem size.

4. Practical Considerations

4.1 Models in General

The crucial question after almost every theoretic treatment of complexity is: Is, and how is, practical work affected? The traveling salesman problem (TSP), for example, is also NP-complete. Nevertheless, there are numerous heuristics (for example the Lin-Kernighan algorithm [24]) that find optima or almost optimal solutions for huge (with thousands of cities) TSP within a few seconds. But the TSP has one great practical advantage over the SAT problem: *almost* optimal solutions are, in general, sufficient. For the SAT problem approximations are useless, since they do not satisfy the given CNF. Fortunately, in general, model calibration is not expected to be perfect in practice. If a model output variable of minor importance differs from the system sample output more than the demanded Δ, that might be tolerable. Indeed, in some cases it might even be tolerable that only the most important output parameter is congruent to the system sample. However, such a model can hardly be called "congruent" to the real system. The model's application is "safe" only in the hands of users who conceive the model's limits. That *should* include the model developers, but comprehensibility is a very subjective matter. However, with sufficient knowledge of the initial problem, the reference system, and the internal model it is often possible to get valuable results even from a "bad" ("bad" being defined as non-congruent) model.

This is especially true for explorative analysis, training simulation, and preparative experiments. (For further utility of "bad" models see Hodges [25], an emphatically recommended paper.) In the case of decision support tools (DSTs) the estimation has to be made with more prudence. Although the famous quote from Richard Hamming [26], "The purpose of computing is insight, not numbers"—see also Vazsonyi [27], Huntington et al. [28], and Samet [29]—is most likely appropriate for most decision support models, too, trust should be established on the foundation of measurable success whenever possible. (There are some exceptions in explorative analysis.) Hence, congruence, which can be interpreted as affirmed prediction, is going to be decisive in future DSTs. The situation is comparable to weather forecast models. However, for single models used for the special purpose for which they have been designed, calibration seems to be manageable, even if used for decision support. The situation is different for huge model federations.

4.2 Simulation Model Federations

The first trigger for this paper has been a study on the "Standardization of Command and Control Modules for the German Army" [30] which we performed in 2003 at our Institute. The purpose of the study was to check whether the command and control modules (used in semi-automated forces, for example) developed for three German high-resolution combat simulations systems (HORUS, SIRA, and COSIMAC [31–34]) could be (re)used in each other or whether they could be taken as templates for further standardization. During the preparatory work of this study we experimented with the three combat simulation systems and tried to fine–tune them to each other. The goal of this tuning was to show, in principle, consistency (in the sense of Davis and Bigelow [14]) among them. Although we concentrated on only a few model outputs (global and local force ratios and maneuverability) and limited the manipulation to basic model parameters (hit and hit/kill probabilities, reconnaissance, and vehicle movement parameters), it became clear that the interdependencies among these parameters were too strong to allow quick calibration (in this case between models and not between a model and its system; for an introduction into "mutual calibration" of models see Hofmann [34] and the closely related work in Zeigler [36]). It was relatively easy to adjust the parameters so that one output value became consistent, but almost impossible to do it in reasonable time for sets of output values. Facing this (for the study, secondary) problem, we renounced calibration completely and concentrated on the proper standardization problem: Is it possible to standardize command and control modules for different combat simulation systems? Some of the results of the study [30] have been published in English, too [37].

In a federation of models with totally independent model parameters, model calibration could be handled very simply: as a sequence of independent calibrations. The partial solutions lead directly to the global solution. A federation of models with totally consistent parameters could be treated as a monolithic model. Unfortunately, in general, the model parameters of different models are neither totally independent nor totally consistent. As an example, take the following:

Example 3. A combat simulation federation comprises the combat models M^1 and M^2. In the aggregated model, M^1, reconnaissance is based upon a general range-dependent detection probability; M^2 is a sector scan model, which uses a line-of-sight algorithm to determine "exactly" if a detection occurs. The models simulate friendly forces, cooperating via shared perceived situations. The task is to adjust the reconnaissance performance of this federation to realistic samples. The parameter of interest in model M^1 is simply the global detection probability p_D^{M1}. In M^2 there are several parameters that have to be considered. In general, they are hidden in the concrete realization of the line-of-sight algorithm. To simplify matters, assume that the detection calculation in M^2 is completely based upon a visibility parameter p_V^{M2}. The parameters in the two models are obviously different, but both affect the same output (reconnaissance performance). There is only a very small range in which independent calibration of the parameters p_R^{M1} and p_R^{M2} would be useful. Independent adjustments to the realistic samples would fail, since the perceived situations are communicated between the two models. Any change in the reconnaissance performance of the federation could only stem from the communication delay between the models. Hence, the reconnaissance performance of the federation can be reduced or improved significantly only if both parameters are changed in line.

As a further simplification consider that the different dependent parameters (p_D^{M1} and p_V^{M2} in the example) are Boolean and that they are perfectly related ($p_D^{M1} = 1 \Leftrightarrow p_V^{M2} = 1$). With these assumptions the task of federation calibration fits the formal model from section 2, since the parameters are interchangeable. The crucial difference between a single model and a model federation with respect to model calibration is that the initial problem, the original reference system, and the internal model details are not completely known to the federation users. Furthermore, federations are seldom used for a single purpose. It is therefore much harder to use "bad" federations than "bad" single models. Without a high amount of congruence to reality a federation is much more risky than a single model (assuming that the single model is used by experts!). Thus, a kind of congruence like the one formalized in Definition 8 would be an ideal, albeit sometimes impractical, demand for model federations.

Much more about the challenges of model composition in the domain of military simulations can be found in Hofmann [37, 39–41] and Davis and Anderson [38].

4.3 Short Remark on "Tendentious Calibration"

The term "calibration" has a negative connotation for some people. They see it as a kind of "tendentious arbitrariness": the models can be tuned to produce whatever result you want. This suspicion cannot be ignored completely. (The German army has recently initiated a study in that direction. Serious doubts had been uttered about the objective (not tendentious) use of an acquisition model based on utility theory, despite the "calibration" of that model with two high-resolution combat simulation systems.) However, it is satisfying that within the formal framework presented above tendentious calibration would be easily detectable. The possible manipulations are restricted to the choice of the reference system samples and the tolerated deviations, Δ_i. Both of them should always be available.

5. Summary and Conclusion

5.1 Summary

The paper has introduced a formal model of what is generally known as rigorous model calibration. Especially in the domain of technical simulations, calibration is often taken for granted as inevitable and feasible. However, within the formal model presented it has been shown that with increasing model size (number of parameters) calibration becomes computationally challenging (NP-complete) even if excellent data is available. This is especially true when discussing federates of models, in which case the computational requirements for calibration easily exceed the limits of feasibility.

5.2 Conclusion

The computational complexity of many practical problems is difficult to assess. In order to get hard scientific results some simplifying or exaggerating assumptions are often necessary. Thus, the practical relevance of complexity considerations is controversial. However, the widespread influence of a paper from E. H. Page and J. M. Opper [42] about the complexity of composable simulations has clearly shown the benefits of computational complexity considerations for practical military modeling and simulation, too.

Although the formal approach to model calibration presented in this paper might be arguable, and although the NP-completeness of single model calibration might be of minor practical importance, model federations, especially huge federations, seem to be affected. Such federations have scores of internal parameters, which are connected by sometimes unknown dependency functions. No single person has an overview of the whole simulation, and the interpretation of unexpected results is extremely difficult. Hence, the validity of such federations heavily depends on comparisons from simulation runs with real-world samples. In addition, huge federations are seldom single purpose models (due to the efforts needed to create them). To summarize, it is necessary to compare a lot of output parameters from the model with a lot of system samples in order to get trustworthy results. Unfortunately, this is exactly the setting of the formal model presented, for which no polynomial solution algorithm exists. Thus, it is possible that there is a limit in practice for the size of model federations above which it is too costly to properly calibrate and further on validate in the traditional rigorous sense.

In order to overcome this problem the notion of "rigorous" calibration has to be attenuated. First, validation, in the sense of congruence, might be of negligible practical importance in many applications of federations, especially for training and experimentation. Validity, for these applications, should be defined in pragmatic terms of usefulness and appropriateness; see Hofmann [43] for an introduction of pragmatics into validation. Such modest goals of validation are in line with the fundamental possibilities of validation in most complex military models; see Hofmann and Pötzsch [44], Davis [45], and Davis and Bigelow [5]. Second, even for decision support models major deviations of the model output relative to the reference system may be tolerable, because insight is often much more important than sheer numbers. This is especially true when dealing with explorative analysis, for which the whole subject of "calibration" and "validation" should be rethought. (See Davis and Bigelow [5], where an attempt is made to broaden the concept of model validation itself.) However, theoretical results on computational complexity cannot simply be neglected. They draw a serious line for practicability, in this case for practical federation validation, at least if based on system comparisons.

6. References

[1] Barker, P. E., and H. K. Watson. "Calibrating the Simulation Model of the IBM System/360 Time Sharing System." In *Proceedings of the Third Conference on Applications of Simulation*, 130–141. 1969.

[2] Davis, P. K., J. H. Bigelow, and J. McEver. "Model Abstraction Techniques and Applications: Informing and Calibrating a Multiresolution Exploratory Analysis Model with High Resolution Simulation: The Interdiction Problem as a Case History." In *Proceedings of the 32nd Conference on Winter Simulation*, 316–325. 2000.

[3] Davis, P. K., and R. Hillestad. "Families of Models that Cross Levels of Resolution: Issues for Design, Calibration and Management." In *Proceedings of the 25th Conference on Winter Simulation*, 1003–1012. 1993.

[4] Davis, P. K. "Modeling Methodology: Dealing with Complexity: Exploratory Analysis Enabled by Multiresolultion, Multiperspective Modeling." In *Proceedings of the 32nd Conference on Winter Simulation*, 293–302. 2000.

[5] Davis, P. K., and J. H. Bigelow. "*Implications for Model Validation of Multiresolution, Multiperspective Modeling (MRMPM) and Exploratory Analysis.*" RAND, 2003.

[6] Davis, P. K., and J. H. Bigelow. "*Motivated Metamodels.*" RAND, 2003.

[7] Ören, T. I. "Concepts and Criteria to Assess Acceptability of Simulation Studies: A Frame of Reference. " *CACM* 24, no. 4 (1981):180–189.

[8] Wigan, M. R. "The Fitting, Calibration, and Validation of Simulation Models." *Simulation* 18, no. 5 (1972):188–192.

[9] Goldberg, A., and E. E. Kaczka. "Comparison of Models of Individual Choice in a Complex Setting." In *Proceedings of the 11th Conference on Winter Simulation, Vol. 1,* 181–185. 1979.

[10] Hemez, F. M. "The Myth of Science-Based Predictive Modeling." In *Foundations 2004 Workshop for Verification, Validation, and Accreditation in the 21st Century*, Arizona State University, Tempe, AZ. 2004.

[11] Bayarri, M. J., J. O. Berger, D. Higdon, M. C. Kennedy, A. Kottas, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C. H. Lin, and J. Tu. "A Framework for Validation of Computer Models." In *Foundations 2002 Workshop for Verification, Validation, and Accreditation in the 21st Century*, Johns Hopkins University Applied Physics Laboratory, Laurel, MD. 2002.

[12] Zeigler, B. P. *Theory of Modeling and Simulation*. New York: John Wiley and Sons, 1976.

[13] Zeigler, B. P., T. Kim, and H. Praehofer. *Theory of Modeling and Simulation. 2d ed.* New York: Academic Press, 2000.

[14] Davis, P. K., and J. H. Bigelow. *Experiments in Multiresolution Modeling(MRM).* RAND, 1998.

[15] Kleine Büning, H., T. Lettmann, and C. J. van Rijsbergen. *Propositional Logic: Deduction and Algorithms*. Cambridge, U.K.: Cambridge University Press, 1999.

[16] Cook, S. "The Complexity of Theorem Proving Procedures." In *Proceedings of the 3rd ACM Symposium on the Theory of Computing*, 151–158, 1971.

[17] Karp, R. "Reducibility Among Combinatorial Problems." In *Complexity of Computer Computations*, 85–103. New York: Plenum Press, 1972.

[18] Hopcroft, J. E., and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison Wesley, 1979.

[19] Klir. G. J. *Architecture of Systems Complexity*. New York: Sauders, 1985.

[20] Luenberger, D. G. *Introduction to Dynamic Systems*. New York: Wiley, 1979.

[21] Padulo, L., and M. A. Arbib. *System Theory*. Philadelphia: Saunders, 1974.

[22] Beltrami, E. *Mathematics for Dynamic Modeling*. New York: Academic Press, 1987.

[23] Klir, G. J. *Facets of System Science*. New York: Plenum, 1991.

[24] Lin, S., and B. W. Kernighan. "An Effective Heuristic for the Traveling Salesman Problem." *Operations Research* 21, no. 2 (1973): 498–516.

[25] Hodges, J. S. "Six (or so) Things You Can Do with a Bad Model." *Operations Research* 39, no. 3 (1991):355–365.

[26] Hamming, R. W. *Numerical Methods for Scientists and Engineers*. 2d ed. Dover Publications, 1987. (1st ed. 1962)

[27] Vazsonyi, A. "The Purpose of Mathematical Models is Insight, Not Numbers." *Decision Line* (December/January 1998).

[28] Huntington, G., J. Weyant, and J. Sweeney. "Modeling for Insights, Not Numbers: The Experiene of the Energy Modeling Forum." *Omega* 10 (1982): 449–462.

[29] Samet, P. A. "Insight, Not Numbers." *The Computer Journal* 15, no. 1 (1972): 88–92.

[30] Hofmann, H. W., M. Hofmann, and V. Poetzsch. *Standardisierung der Automatischen Führungssimulation Heer*. ITIS/ University of the Federal Armed Forces of Germany, Neubiberg, 2003.

[31] Information about HORUS, PAPST and KORA only available with permission: http://www.iabg.de/.

[32] Information about SIRA only available with permission: http://www.cae.com/ or http://offizierschule.de/hptzh/sira/.

[33] Hofmann, H. W., and M. Hofmann. "On the Development of Command & Control Modules for Combat Simulation Models on Batallion down to Single Item Level. " In *New Information Processing Techniques for Military Systems, RTO Meeting Proceeding MP-049*, Neuilly-sur-Seine, Cedex, France. 2001.

[34] Hofmann, M. *Zur Abbildung von Führungsprozessen in hochauflösenden Gefechtssimulationssystemen*. NG Dissertationsverlag, 2000.

[35] National Research Council. *Modeling and Simulation. Vol. 9 of Technology for the United States Navy and Marine Corps: 2000–2035*. National Academic Press, 1997.

[36] Zeigler, B. P. *Multifacetted Modelling and Discrete Event Simulation*. Ontario, Canada: Academic Press, 1984.

[37] Hofmann, M. "Essential Preconditions for Coupling Model-Based Information Systems." In *NATO Modeling and Simulation Group (NMSG) Conference on C3I and Modeling and Simulation Interoperability, RTO-MP-123*, Antalya, Turkey. 2003.

[38] Davis, P. K., and R. H. Anderson. *Improving the Composability of Department of Defense Models and Simulations*. 2003.

[39] Hofmann, M. "Component Based Military Simulation: Lessons Learned with Ground Combat Simulation Systems." In *Proceedings of the 15th European Simulation Symposium*, Delft, NL. 2003.

[40] Hofmann, M. "Challenges of Model Interoperability in Military Simulations." *Simulation* 80, no. 12 (December 2004): 659–667.

[41] Hofmann, M. "Model Composability: Experiences with Military Simulation Systems." In *Proceedings of the 1st Mediterranean Multiconference on Modeling and Simulation, Genoa, Conference on Conceptual Modeling and Simulation*. 2004.

[42] Page, E. H., and J. M. Opper. "Observations on the Complexity of Composable Simulation." In *Proceedings of the 1999 Winter Simulation Conference*, Phoenix, AZ, 553-560. 1999.

[43] Hofmann, M. "Introducing Pragmatics into Verification, Validation and Accreditation." In *Proceedings of the 2nd European Simulation Interoperability Workshop (EURO-SIW)*, SISO, Harrows, GB. 2002.

[44] Hofmann, M., and V. Pötzsch. „Validation: Possibilities and Limits in High Resolution Combat Simulation Models." In *Proceedings of the 1st European Simulation Interoperability Workshop (EURO-SIW)*, SISO, Harrows, GB. 2001.

[45] Davis, P. K. "Generalizing Concepts and Methods of Verification, Validation, and Accreditation for Military Simulations." Technical Report R-4249-ACQ, RAND, 1992.

Author Biography

Marko Hofmann *is Project Manager at the Institute for Technology of Intelligent Systems (ITIS) at the University of the Federal Armed Forces in Neubiberg, Germany. After his studies of computer science at the University of the Federal Armed Forces he served two years in an army battalion staff. From 1995 to 2000 he was research assistant at the Institute for Applied System Analysis and Operations Research (IASFOR), where he received his Ph.D. in computer science. Since April 2000 he has been responsible for basic research in applied computer science at the Institute for Technology of Intelligent Systems. He lectures at the University of the Federal Armed Forces in Munich and at the University of Applied Science in Kufstein.*