

# Deviation Models Revisited

Peter Struss (struss@in.tum.de, struss@occm.de)

Computer Science Dept., Technische Universität München, Boltzmannstr. 3  
D-85748 Garching, Germany

and

OCC'M Software GmbH, Gleissentalstr. 22  
D-82041 Deisenhofen, Germany

## Abstract

Tasks like diagnosis, failure-modes-and-effects analysis (FMEA), and therapy proposal involve reasoning about variables and parameters deviating from some reference state. In model-based systems, one tries to capture this kind of inferences by models that describe how such deviations are emerging and propagated through a system. Several techniques and systems have been developed that address this issue, in particular in the area of qualitative modeling. However, to our knowledge, a rigorous mathematical foundation and a “recipe” for how to construct such compositional deviation models has not been presented in the literature, despite the widespread use of the idea and the techniques. In this paper, we briefly revisit the concepts and use of deviation models and related techniques. We reveal (sometimes implicit) presumptions and limitations and present a general mathematical formalization of deviation models. Based on this, aspects of constructing libraries of deviation models, their properties, and their application in consistency-based diagnosis and prediction-based FMEA in a component-oriented framework are analyzed.

## Introduction

Several tasks which are addressed by knowledge-based systems involve reasoning about variables and parameters deviating from some reference state. In diagnosis, one has to localize or identify the reasons for a device behavior that deviates from the normal, or intended, one. Failure-mode-and-effect analysis (FMEA) determines the behavior deviations caused by component faults. And therapy generation attempts to identify influences on a system that remove or reduce deviations from a healthy state. In a model-based approach, one tries to capture this kind of inferences by models that describe how deviations from some reference state or behavior are emerging and propagated through a system. Several techniques and systems have been developed that address this issue, in particular in the area of qualitative modeling.

In our previous work on the theory and applications of consistency-based diagnosis, we often employed so-called deviation models quite successfully. In this context, we face the requirement that the device or system model be **compositional**, i.e. can be generated by aggregating local models of the system constituents (e.g. components) that are stored in a library. However, to our knowledge, a rigorous mathematical foundation and a “recipe” for how to construct such compositional deviation

models has not been presented in the literature, despite the widespread use of the idea and the techniques.

After a brief look at related work and its underlying assumptions, we will propose a formalization of deviation models based on relational models in section 3 and 4 and then analyze properties of such models in a compositional, component-oriented modeling framework for diagnosis and fault analysis. Section 6 discusses some issues concerning the use of the models for predictive FMEA and consistency-based diagnosis

## Previous Work

Very early work on *Incremental Qualitative Analysis* (IQ analysis) [de Kleer 79] and *Differential Qualitative Analysis* (DQ), or comparative analysis [Weld 88] aimed at expressing how a disturbance or parameter shift affects the behavior of a single system. These techniques compare two models that are structurally identical. Obviously, this is different from our intention to consider deviations between any two behaviors, including behaviors that are represented by models of a different structure (e.g. of a normal and a broken device).

In our own previous work, we used “qualitative deviation models” in applications of consistency-based diagnosis. For instance, in [Struss-Sachenbacher-Dummert 97], we based fault localization in an anti-lock braking system (ABS) of a car on such a model, assuming that the observations are given as qualitative deviations from some unspecified nominal behavior expressing statements like “The wheel rotates faster than it should”.

The models we constructed expressed constraints on the deviations of system variables and parameters from the nominal behavior. For instance, the model of a valve is given by a constraint:

$$[\Delta q] = [A] * ([\Delta p_1] - [\Delta p_2]) + [\Delta A] * ([p_1] - [p_2]) - [\Delta A] * ([\Delta p_1] - [\Delta p_2])$$

on the signs of the deviations of pressure ( $[\Delta p_i]$ ), flow ( $[\Delta q]$ ), and area ( $[\Delta A]$ ), where  $[x]$  means  $\text{sign}(x)$ . This constraint allows, for instance, to infer that an increase in  $p_1$  ( $[\Delta p_1] = +$ ) will lead to an increase in the flow ( $[\Delta q] = +$ ), if  $p_2$  and the area remain unchanged ( $[\Delta p_2] = 0$ ,  $[\Delta A] = 0$ ) and the valve is not closed ( $[A] = +$ ). We also provided a definition and construction of such qualitative deviation models from equation models of components. For each system variable and parameter  $v_i$ , we define the deviation as the difference between the actual and a reference value:

$$\Delta v := v_{\text{act}} - v_{\text{ref}}$$

Then algebraic expressions in an equation can be transformed to deviation models according to rules such as

$$a + b = c \Rightarrow \Delta a + \Delta b = \Delta c$$

$$a - b = c \Rightarrow \Delta a - \Delta b = \Delta c$$

$$a * b = c \Rightarrow a_{act} * \Delta b + b_{act} * \Delta a - \Delta a * \Delta b = \Delta c$$

$$a / b = c \Rightarrow (b_{act} * \Delta a - a_{act} * \Delta b) / (b_{act} * (b_{act} * \Delta b)) = \Delta c$$

Furthermore, for any monotonically growing (section of a) function  $y = f(x)$ , we obtain  $[\Delta y] = [\Delta x]$  as an element of a qualitative deviation model. This way, the deviation model of the valve stated above has been obtained by a well-defined transformation from the respective equation describing the behavior of a valve in absolute terms terms (with a positive constant c)

$$q = \text{sign}(p_1 - p_2) * c * A * \sqrt{|p_1 - p_2|} .$$

It is quite convenient that such model fragments often state direct relationships among deviations, independently of actual and reference values. In other cases (such as multiplication and division) we need information about the actual values (but possibly only their qualitative abstraction for qualitative deviation models).

We used the same kind of models in order to capture different intuitions which correspond to a different choice of the reference. In [Malik-Struss 96], we chose the equilibrium of a system (a controlled electric motor) to define the reference. Furthermore, we exploited deviation models to express constraints on changes over time. This means, we can select a previous (or future) state of a system as the reference. Then the deviation model of the valve can be interpreted like “if the pressure  $p_1$  has increased while  $p_2$  and  $A$  are identical, then the flow must have increased, as well”. This corresponds to the perspective of IQ analysis at a global level. In fact, if a (numerical) model contains monotonic functions only, then the local results of IQ analysis carry over to the entire domain, and the resulting models are identical to deviation models.

In contrast to IQ and DQ analysis, deviation models reflect the idea to compare two arbitrary behaviors. However, we never provided a general formal definition of such models, and it turns out that the way we constructed them (as indicated above) was based on some assumptions that limit their applicability. A simple example can illustrate this: We claimed that, from a constraint

$$(1) a + b = c$$

over the absolute values, we would derive the deviation equation

$$(2) \Delta a + \Delta b = \Delta c .$$

But what is the basis for obtaining (2) from (1)? We find that

$$\begin{aligned} \Delta a + \Delta b &= (a_{act} - a_{ref}) + (b_{act} - b_{ref}) \\ &= (a_{act} + b_{act}) - (a_{ref} + b_{ref}) = c_{act} - c_{ref} = \Delta c \end{aligned}$$

exploits the assumption that both the actual and the reference values satisfy the same constraint (1). More generally and more formally: if  $R$  is a model relation over a vector of variables  $\underline{v}$ , then the respective deviation model is given by the relation

$$(3) \quad \Delta R := \{ \Delta \underline{v} \mid \Delta \underline{v} := \underline{v}_{act} - \underline{v}_{ref} \wedge \underline{v}_{act}, \underline{v}_{ref} \in R \}$$

However, it is “self-referential” in the sense that it captures deviations w.r.t. the **same model**, and, hence, definition (3) does not properly reflect our intention to describe behavior deviations in comparison to the **nominal model**.

A simple example illustrates the consequences: The OK model of a closed valve states a zero flow:

$$(4) \quad f = 0$$

which leads to the sign-based deviation model

$$(5) \quad [\Delta f] = 0$$

This expresses the fact that, for a properly working closed valve, any deviation in the pressure drop does not cause a deviation in the flow, it remains 0. However, (4) also models a clogged pipe. But, in this case, the sign-based deviation model should be

$$(6) \quad [\Delta f] = -[f_{ref}]$$

stating that the zero flow represents a deviation which is opposite to the flow under the reference conditions.

The example shows that

- the same equation may lead to different deviation models.
- it may be impossible to determine the deviation locally: (6) does not fix the deviation of the flow; rather, this depends on other components in the device.

Of course, having given definition (3), it is obvious how to correct the bug. We will do this in the next section and then discuss the consequences.

## Formalizing Deviation Models

In general, it is straightforward to provide a proper definition for deviation models. For this purpose, we consider behavior models to be represented by a relation over a vector  $\underline{v} = (v_1, \dots, v_n)$  of system variables with the domain

$$\text{Dom}(\underline{v}) = \text{Dom}(v_1) \times \dots \times \text{Dom}(v_n),$$

i.e.  $R \subset \text{Dom}(\underline{v})$ , which may be implemented as a set of constraints. At this stage, we do not restrict the domains. They may be real numbers, intervals, signs, etc. The only condition is that distances

$$d_i: \text{Dom}(v_i) \times \text{Dom}(v_i) \rightarrow \text{Dom}(\Delta v_i)$$

are defined, e.g. as subtraction “-” on the domains mentioned above, and their composition

$$d = (d_1, \dots, d_n): \text{Dom}(\underline{v}) \times \text{Dom}(\underline{v}) \rightarrow \text{Dom}(\Delta \underline{v}).$$

A deviation model has to characterize the possible distances of the tuples w.r.t. a reference model relation,  $R_{ref}$ .

As the example at the end of section 2 shows that determination of a deviation may require an explicit specification of a reference value, we include the reference in one version of the definition.

**Definition 3.1 (Deviation Model)** Let  $R_m, R_{ref} \subset \text{Dom}(\underline{v})$  be two behavior models.

$$\Delta_c(R_m, R_{ref}) := \{ (d(\underline{v}_m, \underline{v}_{ref}), \underline{v}_{ref}) \mid \underline{v}_m \in R_m \wedge \underline{v}_{ref} \in R_{ref} \} \subset \text{Dom}(\Delta \underline{v}) \times \text{Dom}(\underline{v})$$

is called the **complete deviation model** of  $R_m$  w.r.t.  $R_{ref}$ .

$$\Delta_p(R_m, R_{ref}) := \{ d(\underline{v}_m, \underline{v}_{ref}) \mid \underline{v}_m \in R_m \wedge \underline{v}_{ref} \in R_{ref} \} = \prod_{\Delta \underline{v}} (\Delta_c(R_m, R_{ref})) \subset \text{Dom}(\Delta \underline{v}),$$

where  $\prod_{\Delta \underline{v}}$  denotes the projection to the distance vector, is called the **pure deviation model** of  $R_m$  w.r.t.  $R_{ref}$ .

A complete deviation model  $\Delta_c(R_m, R_{ref})$  is called **redundant** if

$$\Delta_c(R_m, R_{ref}) = \Delta_p(R_m, R_{ref}) \times \text{Dom}(\underline{v}).$$

The redundancy property expresses that the deviation constraints do not depend on the reference value; the pure

deviation model produces the same distance predictions as the complete model, as for the addition relation over the real numbers  $\mathfrak{R}$ :

$$R_m = R_{\text{ref}} = R_+ = \{(a, b, a + b) \mid a, b \in \mathfrak{R}\}$$

$$\Delta_c(R_+, R_+) = \{(\Delta a, \Delta b, \Delta a + \Delta b) \mid \Delta a, \Delta b \in \mathfrak{R}\} \times \mathfrak{R}$$

In other words, the deviations of  $a$  and  $b$  sum up to give the deviation of  $c$  regardless of the specific value of  $a$  and  $b$ .

Although Def. 3.1 is very general, it still contains an important presumption, namely that both relations are defined over the **same variables** (and domains). This is really a restriction, if you consider, for instance, process-oriented diagnosis and therapy proposal as described in [Struss et al. 03]. In that framework, the normal process of water treatment and the one resulting from some disturbance may lead to the different models that include different sets of processes, objects, and, hence, variables.

But even in a component-oriented model, the presumption of Def. 3.1 may be violated. The OK model of a pipe may refer to pressure and flow only, whereas its fault models may include parameters such as the size of a leakage, the resistance due to partial clogging, etc. In this case, we can project both relations to the set of variables they have in common and apply the analysis to the resulting relations. Therefore, in order keep different problem dimension separate, we maintain this presumption for the investigation in this paper. Also, we continue to talk about **model relations in general** to avoid mixing the general problem of defining, constructing, and exploiting deviation models with the problem of generating appropriate **qualitative** deviation models. Finally, we restrict the following investigations to component-oriented models for tasks like (consistency-based) diagnosis, diagnosability analysis, and FMEA.

### Deviation Models for Component-oriented Fault Analysis Deviation Models

More formally, we can characterize the interesting class of models in the following way. A model of the entire device is composed of its components' behavior models. Each model fragment  $R_{ij}$  describes the behavior of a component  $C_i$  under a mode

$$m_{ij} \in \text{modes}(C_i) = \{\text{OK}, F_{i1}, \dots, F_{ik}\}$$

which is the (unique) correct mode or a particular fault. For easier reading, we will denote the model relation for the OK mode by  $R_{i,\text{OK}}$ . Thus, for each mode assignment

$$\text{MA} := \{m_{ij} \mid m_{ij} \in \text{modes}(C_i)\}$$

which assigns a unique mode to each component, we obtain the model relation as the join of the respective component model relations

$$R_{\text{MA}} = \bigotimes_{m_{ij} \in \text{MA}} R_{ij}$$

The class of tasks we want to consider can simply be characterized by the choice of  $R_{\text{ref}}$  which is given by the behavior relation of the assignment of the mode OK to all components:

$$R_{\text{ref}} = R_{\text{OK}} := \bigotimes_i R_{i,\text{OK}}$$

This is different from other possible uses of component models. For instance, in model-based design, the reference is given by some behavior specification rather than the correct behavior of the components of the design at a certain stage. We just mention that even the "natural" choice of  $R_{\text{OK}}$  as the reference is based on some assumptions that restrict the scope of the respective problem solvers: it assumes that the intended behavior is achieved if all components work properly. This means w.r.t. the device that it is well-designed, and w.r.t. the task that no structural faults that introduce additional interactions (such as bridge faults) can be diagnosed explicitly.

Regarding the entire device, we can easily define the appropriate deviation model

#### Definition 4.1 (Diagnostic Deviation Model)

A *complete diagnostic deviation model* of a mode assignment  $\text{MA}$  is defined as

$$\Delta_c(\text{MA}) := \Delta_c(R_{\text{MA}}, R_{\text{OK}})$$

and a *pure diagnostic deviation model* by

$$\Delta_p(\text{MA}) := \Delta_p(R_{\text{MA}}, R_{\text{OK}}).$$

While this definition is straightforward, it does not directly provide a satisfactory way to construct the desired deviation model. We can certainly (automatically) construct the behavior relations  $R_{\text{MA}}$  and  $R_{\text{OK}}$  and then compute the deviation model (automatically) according to definition 3.1. Even if this is feasible (considering the potentially large number of variables and tuples in the relation), it would not be convenient for another reason: the requirement of **compositional** modeling. What we would like to do is to compose the deviation model of the entire device from local component models the same way we compose the absolute model,  $R_{\text{MA}}$ . This is what we actually did in the applications referenced above. Formally, we constructed deviation models locally for each component  $C_i$  according to definitions 4.1 and 3.1

$$\Delta_c(m_{ij}) := \{(d(\underline{v}_i, \underline{v}_{i,\text{OK}}), \underline{v}_{i,\text{OK}}) \mid \underline{v}_i \in R_{ij} \wedge \underline{v}_{i,\text{OK}} \in R_{i,\text{OK}}\}$$

and then combined them to establish a device model for a mode assignment  $\text{MA}$

$$\Delta'_c(\text{MA}) := \bigotimes_{m_{ij} \in \text{MA}} \Delta_c(m_{ij})$$

Actually, we often constructed pure deviation models:

$$\Delta'_p(\text{MA}) := \bigotimes_{m_{ij} \in \text{MA}} \{d(\underline{v}_i, \underline{v}_{i,\text{OK}}) \mid \underline{v}_i \in R_{ij} \wedge \underline{v}_{i,\text{OK}} \in R_{i,\text{OK}}\}$$

The question to be answered is how this model relates to the diagnostic deviation model given by definitions 3.1/4.1. We have to compare the compositional deviation model

$$\Delta'_c(\text{MA}) := \bigotimes_{m_{ij} \in \text{MA}} \{(d(\underline{v}_i, \underline{v}_{i,\text{OK}}), \underline{v}_{i,\text{OK}}) \mid \underline{v}_i \in R_{ij} \wedge \underline{v}_{i,\text{OK}} \in R_{i,\text{OK}}\}$$

with the global one:

$$\Delta_c(\text{MA}) = \{(d(\underline{v}, \underline{v}_{\text{OK}}), \underline{v}_{\text{OK}}) \mid \underline{v} \in \bigotimes_{m_{ij} \in \text{MA}} R_{ij} \wedge \underline{v}_{\text{OK}} \in \bigotimes_i R_{i,\text{OK}}\}$$

Intuitively,  $\Delta'_c(MA)$  appears weaker, because it combines the OK relations in a join, whereas the local restrictions  $R_{ij}$  are not explicitly joined, but only via the distance  $d$ . More precisely,

$$(d(\underline{v}, \underline{v}_{OK}), \underline{v}_{OK}) \in \Delta_c(MA)$$

if and only if

$$\underline{v} \in \bigotimes_{m_{ij} \in MA} R_{ij} \wedge \underline{v}_{OK} \in \bigotimes_i R_{iOK}.$$

If we denote the restriction of  $\underline{v}$  and  $v_{OK}$  to the local variables of  $C_i$  by  $\underline{v}|_i$  and  $\underline{v}_{OK}|_i$ , respectively, then this is equivalent to

$$\forall_i (\underline{v}|_i \in R_{ij} \wedge \underline{v}_{OK}|_i \in R_{iOK}).$$

This implies

$$\forall_i (d(\underline{v}|_i, \underline{v}_{OK}|_i), \underline{v}_{OK}|_i) \in \Delta_c(m_{ij})$$

which means

$$(d(\underline{v}, \underline{v}_{OK}), \underline{v}_{OK}) \in \bigotimes_{m_{ij} \in MA} \Delta_c(m_{ij})$$

This proves

**Lemma 4.1** *The global diagnostic deviation model  $\Delta_c(MA)$  is stronger than the compositional diagnostic deviation model  $\Delta'_c(MA)$ :*

$$\Delta_c(MA) \subset \Delta'_c(MA).$$

Are they equal? In the above sketch of the proof, we have implications in both directions, except for the one that is highlighted. Can we reverse this inference, as well? Not in general: if, for  $\underline{d}_i \in \text{Dom}(\Delta \underline{v}_i)$ ,

$$\forall_i (\underline{d}_i, \underline{v}_{iOK}) \in \Delta_c(m_{ij})$$

then, **locally**, there exists a  $\underline{v}_i$  that is consistent with  $\underline{d}_i$  and  $\underline{v}_{iOK}$ :

$$\forall_i \exists \underline{v}_i \in R_{ij} \quad d_i = d(\underline{v}_i, \underline{v}_{iOK}).$$

However, it is not guaranteed that these local tuples  $\underline{v}_i$  can be combined to form a global one,  $\underline{v}$ , that is consistent with  $MA$ , i.e. it is not guaranteed that there is a  $\underline{v} \in R_{MA}$  such that for all  $i$

$$\underline{v}_i = \underline{v}|_i$$

holds. The simplest counterexample is the following: assume the normal behavior includes two components  $C_1$ ,  $C_2$  that both fix a variable  $x$  to be positive. Over the sign domain, this means

$$R_{1OK} = R_{2OK} = \{ (+) \} \in \text{Dom}(x) = \{ -, 0, + \}.$$

Furthermore, assume a fault of  $C_1$  makes  $x$  zero, while a fault in  $C_2$  would turn  $x$  negative:

$$R_{11} = \{ (0) \}, R_{21} = \{ (-) \}.$$

Then in both components, the deviation models determine the deviation of  $x$  to be negative:

$$\Delta_c(m_{k1}) = \{ (-, +) \}, k=1,2,$$

and, hence,

$$\Delta'_c(\{m_{11}, m_{21}\}) = \{ (-, +) \}.$$

However, since  $\{m_{11}, m_{21}\}$  is inconsistent, we have

$$\Delta_c(\{m_{11}, m_{21}\}) = \emptyset.$$

This shows

**Lemma 4.2** *In general, global and compositional diagnostic deviation models are not equivalent:*

$$\Delta_c(MA) \neq \Delta'_c(MA).$$

This means, in general, that we have to trade an important practical requirement, compositionality of the model, against another practically relevant feature, its completeness, i.e. its ability to detect all inconsistencies.

Fortunately, we can recover at least partially from this dilemma, if we impose a restriction on the distances  $d_i$ . The idea is that, if a given reference value  $\underline{v}_{ref}$  and distance tuple  $\underline{d}_0$  determine a **unique**  $\underline{v}_0$  such that  $d(\underline{v}_0, \underline{v}_{ref}) = \underline{d}_0$ , then each globally consistent tuple  $(\underline{d}_0, \underline{v}_{ref})$  corresponds to a unique vector  $\underline{v}_0$  which must also be globally consistent. This property is satisfied for real numbers, if we define  $d(x, y) := x - y$ , but not for the sign domain or, more generally, for interval domains. So we can state

**Lemma 4.3** *If for all  $i$  and for all  $y_i \in \text{Dom}(v_i)$  the distance  $d(x, y_i)$  determines  $x$  uniquely, i.e. the function*

$$d y_i(x) := d(x, y_i)$$

*is injective, then the compositional diagnostic deviation model is equivalent to the global one:*

$$\Delta'_c(MA) = \Delta_c(MA).$$

*In particular, this holds for  $\text{Dom}(\underline{v}) = \mathcal{R}^n$  and*

$$d(x, y) := x - y.$$

Before we discuss the practical consequences of this result, we complete the analysis by investigating the strength of pure deviation models, i.e. models that comprise only constraints on deviations, ignoring the reference values. The question is whether complete deviation models can be redundant in the sense of Def. 3.1.

Obviously, if all local deviation models are redundant then so are the compositional deviation models. This yields

**Lemma 4.4** *A compositional complete diagnostic deviation model is redundant if all local deviation models are redundant.*

However, if any local deviation model is not redundant, we must expect that dropping the reference values from the model makes the overall model strictly weaker. For instance, a multiplicative constraint  $a * b = c$  yields a complete deviation model specified by  $\Delta a * \Delta b + b_{ref} * \Delta a + a_{ref} * \Delta b = \Delta c$ , which does not allow to drop the reference values. The reference values have to be determined for this local model which means the complete deviations models also have to be used for the other local models even though they might be redundant. To give again the simplest example, assume there is a component enforcing an equality constraint  $a = b$  connected to the multiplication component. Then its deviation model is redundant, and the qualitative pure deviation model states  $[\Delta a] = [\Delta b]$ . However, the multiplicative deviation model cannot restrict the sign of  $[\Delta c]$  from  $[\Delta a] = [\Delta b] = +$  alone, but it could, for instance,

infer  $[\Delta c] = +$ , if  $[a] = [b] = +$  is known from the complete deviation model of the equality constraint.

### Practical Consequences of the Analysis

The theoretical analysis above may appear fairly abstract. However, the results have a tremendous impact on the use of deviation models in practice which we discuss in this section in a fundamental way, before we turn to specific issues that are related to the exploitation of deviation models in consistency based diagnosis.

The step **from self-referential to diagnostic** deviation models **sacrifices** a lot of the simplicity of the models. As an abstract example, consider  $b = k \cdot a$  with  $0 < k < 1$  constant as the fault model that is compared to the OK model  $b = a$ .

Note that the resulting deviation model  $\Delta b = \Delta a + (k-1) \cdot a_{ok}$  does not only contain a reference to  $a_{ok}$ . It also implies that determining the sign of  $\Delta b$  requires determining whether  $\Delta a > (k-1) \cdot a_{ok}$  which can only be done at the numerical level. Because the result depends on  $a_{ok}$ , there exists no finite set of landmarks for  $\Delta a$  that would allow determining the sign of  $\Delta b$ . A sign-based deviation model would be totally ambiguous in this case as indicated by Table 1.

The example at the end of section 4 shows that we **usually have to use complete deviation models**, i.e. include and determine reference values in the deviation model. To clearly emphasize the practical impact: when we work with a compositional model, even if only a single local model  $\Delta_c(m_{ij})$ , is not redundant, **all** local models must be used in their complete form, even if they are redundant. This is because they may be required to restrict  $v_{i,ok}$  which in turn is needed to derive a conclusion from  $\Delta_c(m_{ij})$ . In other words,  $R_{ok}$  has to be computed globally **in addition to**, and for enabling the computation of deviations.

**Table 1.** (Part of) the deviation model for the abstract example

$\Delta a$	$\Delta b$	$a_{ok}$	$b_{ok}$
-	-	+	+
0	+	+	+
+	+	+	+
+	0	+	+
+	-	+	+
+	+	0	0
0	0	0	0
...	...	...	...

Let us discuss this from another perspective. In the above definitions and analysis, we included the reference,  $R_{ok}$ , in the deviation model. Of course, due to the anti-symmetry of the distances  $d_i$ ,

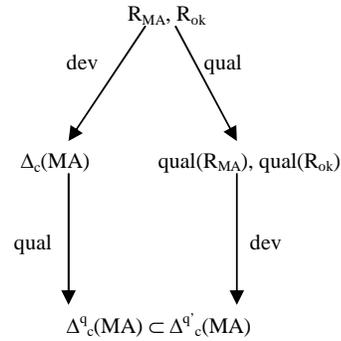
$$d_i(v_i, v_{i,ok}) = -d_i(v_{i,ok}, v_i)$$

we obtain the respective definitions and results for deviation models that include the mode assignment model,  $R_{MA}$ , instead of  $R_{OK}$ , in the deviation model. Then the above results imply that one has to apply the absolute behavior model of the mode assignment and additionally compute the

deviations. But this means: reasoning about deviations does **not substitute** the use of the absolute model as we had hoped, but is simply additional effort to derive a convenient and intuitive description, e.g. for interpreting FMEA results.

Furthermore, compositional qualitative deviation models usually generate weaker results than global ones. This means that besides the usual incompleteness of the qualitative (interval-based) calculus, there is an **additional** loss of completeness in the step of **constructing** the qualitative deviation model.

Finally, because the difference of qualitative values introduces ambiguity, the qualitative deviation models should be constructed by abstraction from numerical deviation models (left-hand path in Fig. 1) rather than as deviation models of the qualitative abstractions of the numerical models (right-hand path).



**Figure 1.** Deviation models of qualitative models are weaker than the qualitative versions of numerical deviation models.

So far, we analyzed properties of the relations representing deviation models, especially under the aspect of compositionality. In the next section, we discuss their use in (consistency-based) problem solvers for fault analysis.

### Using Deviation Models in Model-based Systems

Sometimes, the entire model relation is used to compute a solution (e.g. the diagnosability analysis, see [Dressler-Struss 03]). In most systems, the model is applied to derive a more complete description of a situation or scenario based on a partial description and/or check the consistency practical description with the model.

For instance, in FMEA, a scenario is described in terms of external conditions and a particular state of a device. Then, for each mode assignment corresponding to a (single) component fault, the respective model can be exploited to restrict the value of other variables with a focus on effects that represent a deviation from the function the device is supposed to perform. It is often quite natural to express both the fault and the relevant effects as deviations from the nominal state (“pressure lower than nominal”, “extension of the landing gear too slow”), which makes deviation models attractive for model-based FMEA generation.

In our relational formalism, a scenario is also expressed as a relation on the variables and parameters that define it and usually contains a single tuple. It specifies a context in which the resulting device behaviors in the nominal mode and under a (single) component fault have to be compared. This implies the assumption that the scenario relation is consistent with both behaviors. Otherwise, there would be a modelling bug w.r.t to the scenario or the behaviors. This is plausible because the scenario is supposed to describe **exogenous variables** and a **given state** of the system, and the behavior models are only correct if they cover the device response to any physically possible situation (We mention this because we face a different situation in diagnosis, as discussed below).

We may ask the question which kind of deviation model is more appropriate for the task, the one containing the reference model or the one containing the model of the mode assignment under consideration. If we use numerical models, it does not matter, because either one can be deduced from the other one using the deviations. However, if we apply qualitative deviation models, there might be reasons to use the one that includes  $R_{MA}$ . This is because it might allow stronger conclusions. For instance, a valve being stuck closed would be described by

$$\begin{array}{lll} \text{State} = \text{OPEN} & A = 0 & \Delta A = - \\ \text{State} = \text{CLOSED} & A = 0 & \Delta A = 0 \end{array}$$

and, thus, distinguishable from a fault mode where the valve is open but not enough:

$$\begin{array}{lll} \text{State} = \text{OPEN} & A = + & \Delta A = - \\ \text{State} = \text{CLOSED} & A = 0 & \Delta A = 0 \end{array}$$

In reference to the Ok mode, they would both map to

$$\begin{array}{lll} \text{State} = \text{OPEN} & A_{\text{ref}} = + & \Delta A = - \\ \text{State} = \text{CLOSED} & A_{\text{ref}} = 0 & \Delta A = 0 \end{array}$$

A consideration of computational cost may lead to a different choice, because in this case, the model computes the absolute values for each fault plus the deviations. The alternative is to compute the absolute values for the OK mode only and deviations for each fault.

The deviation model is then used in the following way:

- The **scenario** is described as a relation on absolute values of **exogenous and state variables** and **deviation 0** for all of them. This reflects the fact that the scenario applies to both the OK and the faulty mode.
- The model of the faulty device,  $R_{MA}$ , specifies **parameters** and their **deviations** which are **non-zero** only for (some) parameters in the model(s) of the faulty component(s) and 0 for all others.

The model (if strong enough) will compute the absolute behavior under this scenario and the deviations w.r.t. to the nominal mode. This illustrates again what was discussed in the previous section, namely that computing the deviations is only an additional effort that serves convenience. In other systems, this is done outside the model by comparing the results of two behavior predictions.

We now turn to the exploitation of deviation models in consistency-based diagnosis. Its basic task is to check whether a model of some behavior mode is consistent with a set of given observations in order to determine those modes that might describe the observed situation. In what way can a deviation model be exploited in this framework?

To answer this question, we first look at the construction of such models which corresponds to first joining three relations,  $R_{OK}$ ,  $R_{MA}$ , and  $R_{dev}$  which contains the definitional relations for all distances  $d_i$ , and then projecting the result to  $(\Delta \underline{y}, \underline{y})$ :

$$\Delta_c(MA) = \Pi_{\Delta \underline{y}, \underline{y}} (R_{dev} \bowtie R_{MA} \bowtie R_{OK})$$

The available observations relate to the current situation, and the goal is to check whether MA can consistently be assumed to be present in this situation. It is **not** the purpose to check whether or not the nominal mode is consistent. This has been done before with the result that it is inconsistent with the given observations (because otherwise there would be no reason to investigate the fault corresponding to MA). Hence,  $R_{OK}$  relates to a different situation and therefore has to be considered consistent with the observations. Furthermore,  $R_{dev}$  is just a collection of definitions which are not related to any situation and, hence, consistent with the observations. Therefore, if  $\Delta_c(MA)$  is inconsistent with observations, the origin must lie in  $R_{MA}$ . If we use the form of deviation model that includes  $R_{MA}$ , then we notice again that the deviation part of the model is just a convenient supplement and does not contribute to the core part of the problem solving.

However, it can be an important one, if the available observations are stated in terms of deviations, as is illustrated by the example of diagnosis of the ABS system based on qualitative driver observations.

If the deviation model including the OK model is used, we face a problem: as already stated above, we must assume that the given set of observations cannot be related to this model, and the problem arises how to enter the observations in order to trigger predictions. Either observations contain absolute values and deviations; then the reference values could be computed (or, at least restricted). But in this case, we would probably use the model including  $R_{MA}$ . Otherwise, we have to express properly the idea that the deviations should express a comparison between different device responses to the same situation.

In order to allow the detection of inconsistencies, the observations cannot be restricted to exogenous variables, but have to include at least some dependent variables. The respective observations relate to the actual behavior and must be considered as potentially deviating from the values expected under nominal behavior. They will only be useful if there is some agent able to provide a deviation value for them which allows computing a reference value which is then added to the observations. Similar to the FMEA case, the exogenous (independent) variables can be considered as values for the OK model with their deviations being 0.

## Discussion

We proposed a rigorous way of formalizing the definition and use of models that capture how faults create a deviation of a behavior from a nominal behavior and analyzed properties of these models (particularly in a compositional modeling framework) and the benefits they might promise. The result of this analysis is somewhat conflicting with expectations or claims we had, the major problems being that

- the models expressing the deviations between two different model behaviors are more complicated than the “self-referential” ones, and
- pure deviation models (which do not include the prediction of either the absolute OK or fault mode behavior) will often be too weak,
- which means that computation at the deviation level is just additional effort, though one that may be convenient or even necessary.

When we consider process-oriented diagnosis (as in [Heller-Struss 02]), the problems become even harder. In this case, there is no notion of a fault model. We just have two different models, say of the actual and the normal situation. So far, the process models we use include a  $\Delta$ -version for each positive or negative influence. This is correct, but not enough. If a process is active in, say, the actual model which is not active in the normal model, then this should also imply a  $\Delta$ -influence for each of the newly introduced influences. In the same way, de-activated processes should result in a  $\Delta$ -influence of opposite sign for each eliminated influence. Including this in the theory and system is not straightforward, because it involves reasoning about the changed activity of processes. We could obtain a solution by applying Definition 3.1 to relations that include discrete variables like variables that represent the activity of processes and by introducing an appropriate definition of deviations for such variables which is a useful extension in general. But this involves an explicit comparison of the process structures, and we are far from having worked this out in detail.

Although this may be considered a fairly negative result, we reckon that such models will remain useful for many applications and expect that the foundation given here allows determining the preconditions for such applications and the expected gain.

## Acknowledgments

This work was supported in part by the CEC under the grant GRD1-2001-40133 (AUTAS). I would like to thank the MQM group at the Technische Universität München and the partners in the AUTAS project for discussions and valuable hints. Also, I highly appreciated the useful hints from several reviewers, although I am afraid I could respond to all their critical comments.

## References

- [de Kleer 79] de Kleer, J.: Causal and Teleological Reasoning in Circuit Recognition. TR-529, AI Lab, MIT, Cambridge Mass., 1979
- [Dressler-Struss 03] Dressler, O., Struss, P.: A Toolbox Integrating Model-based Diagnosability Analysis and Automated Generation of Diagnostics. In: Proceedings of the 14th International Workshop on Principles of Diagnosis (DX03), pages 99-104, Washington, D.C., USA, June 2003.
- [Heller/Struss 02] U.Heller, P.Struss: Consistency-Based Problem Solving for Environmental Decision Support. In: Computer-Aided Civil and Infrastructure Engineering 17 (2002) 79-92.
- [Struss et al. 03] P. Struss, M. Bendati, E. Lersch, W. Roque, P. Salles: Design of a Model-based Decision Support System for Water Treatment. In: Working Papers of the 17th International Workshop on Qualitative Reasoning (QR-03), pages 215-220, Brasilia, Brazil, 2003.
- [Malik-Struss 96] Malik, A., Struss, P.: Diagnosis of Dynamic Systems Does Not Necessarily Require Simulation. In: Workshop Notes of the 10th International Workshop on Qualitative Reasoning QR-96, AAAI Press, pp. 127-136, 1996. Also in: 7th International Workshop on Principles of Diagnosis (DX-96), Montreal, Canada, pp. 147-156, 1996.
- [Struss-Sachenbacher-Dummert 97] Struss, P., Sachenbacher, M. Dummert, F.: Diagnosing a Dynamic System with (almost) no Observations. Workshop Notes of the 11th International Workshop on Qualitative Reasoning, (QR-97) Cortona, Italy, June 3-6, pp. 193-201, 1997.
- [Weld 88] Weld, D.: Comparative Analysis. Artificial Intelligence, 36 (1988), pp. 333-374. 1988