Automated selection of an accurate model of a visco-elastic material

A.C. Capelo, L. Ironi, S. Tentoni
Istituto di Analisi Numerica del C.N.R., Pavia, Italy
liliana@supers1.ian.pv.cnr.it

Abstract

A major problem in the development of a computational environment that can reason about physical systems is its ability to formulate a model. The work here described is part of a research effort aimed at developing a comprehensive environment that automates the formulation of the constitutive law of an actual visco-elastic material. In outline, we approached the problem in two main stages: at first, a library of models of ideal materials is generated, and then an accurate model of an actual material, which explains the observed response of the material to standard experiments, is selected. The model library includes both models of ideal materials and their qualitative response to standard experiments. The models are generated in two different formalisms (Rheological Formulae (RF) and Ordinary Differential Equations (ODE)), by following an enumerative procedure and an approach which is grounded both on a component-connection paradigm and on internal state variables. A class of candidate models, i.e. a class of ODEs, for the material is selected from the model space through the comparison of the observed behavior, qualitatively interpreted, with the qualitative behaviors generated directly from the rheological structures. Then, the most "accurate" model for the real material is chosen within the selected class so that both the goodness of experimental data fitting and the number of parameters in the model are "reasonable". This paper mainly concentrates on the methods and algorithms, both qualitative and quantitative, of model selection.

Keywords: Mathematical modeling, qualitative reasoning, rheology, visco-elastic materials.

Introduction

A crucial issue in automated reasoning about a physical system is the automated formulation of an appropriate model of its behavior. Recently, a great deal of work in the Qualitative Physics framework has been addressed to the automated model formulation problem, and a number of methods and implemented systems has been proposed (Addanki *et al.*, 1991; Bradley, 1994; Crawford *et al.*, 1992; Capelo *et al.*, 1993; Falkenhaier and Forbus, 1991; Ironi and Stefanelli, 1994; Iwasaki, 1992; Low and Iwasaki, 1992; Iwasaki and Levy, 1994; Nayak, 1994; Rickel and Porter, 1994; Weld, 1990; Weld, 1992). In these approaches, the model formulation problem involves the selection, within a predefined model library, of a model in accordance with a set of assumptions or a user's query about the system behavior. The model library can contain either complete models of the system under study, each of them characterized by different assumptions (Addanki *et al.*, 1991), or pieces of knowledge about the physical systems, called model fragments, which are suitably selected and composed to construct the scenario model (Falkenhaier and Forbus, 1991). Whereas in the former case the formulation of an accurate model can be simply viewed as a search process through a graph of models, where nodes represent models of the system and arcs represent the assumptions that distinguish two connected nodes, in the latter one two different issues must be considered: the constructed model has to be both adequate to describe the physical situation and as simplest as possible (Iwasaki and Levy, 1994; Rickel and Porter, 1994).

This paper describes our approach to the automated model formulation of an accurate quantitative model of the mechanical behavior of an actual visco-elastic material (Arridge, 1975) in accordance with the observed behavior of the material in response to standard experiments. The actual difficulties in building by hand models of materials motivated our work. Knowing the constitutive laws of materials, i.e. the relation between stress $s(t)$ and strain $e(t)$ and their time dependencies, allows us both to derive predictions of a material's behavior under the action of external forces and to associate the material with its mechanical properties. Moreover, the mechanical properties of a material may be correlated to some of its other properties (for example network structure, sensitivity to erosion, hy-

drophilia, capacity either to absorb or to release active ingredients, thyxotropy, and so on), whose knowledge is of fundamental importance to the assessment of the material. The measurement of these latter properties often presents more difficulties than the measurement of the mechanical ones. Hence, the rheological study of a material may often be both economic and important in the control of the industrial processing of products.

Although our approach has been tailored on a specific application domain, most ideas and techniques underlying it can be applied to other physical domains when the goal is to build an accurate quantitative model explaining a set of experimental data. The whole formulation process occurs in two main stages: at first, an exhaustive library of complete models of ideal visco-elastic materials, which differ from each other in structure, is automatically generated, and then an accurate model of an actual material is built in accordance with the observed response of the material to either creep or relaxation experiments.

As far as model library is concerned, our work is distinguished in that it *automatically* generates the ODE models of materials with complex rheological structures. Such structures are automatically enumerated by analogy with mechanical devices where components which reproduce the fundamental elastic and viscous responses are combined either in series or in parallel. The symbolic description of such structures is called Rheological Formula (RF). Then, by exploiting suitable connection rules and mathematical models of the basic components, which are expressed through internal state variables, the mathematical model of each RF is generated. Adequate filter procedures, based on the algebraic properties of the connection operators and on the mechanical equivalence which is captured by the ODEs, allow us to control the combinatorial explosion of the model generation process. In a previous paper (Capelo *et al.*, 1993) we gave a characterization of the generated ODE models and proved that four classes of ODEs represent the space of possible linear models of visco-elastic materials. Moreover, it can be proved by induction that the dimension of the model space is equal to $2n$, where n is the maximum number of components the RFs are made up. This result is important as understanding the space of possible models is an essential step in the construction of computational environments which aim at selecting the most appropriate model.

This paper focusses on the model selection problem and describes an approach which results to be a mixture of qualitative and quantitative techniques with both symbolic and numeric computations. More precisely, we exploit qualitative reasoning to select the class of ODEs which describe the qualitative behavior of the material, and then, within the selected class we identify the equation, namely the order of the ODE and the numeric values of its parameter, which refines the quantitative properties of the material. The selection of the plausible class of models occurs on the ground of the qualitative comparison of the simulated behaviors of the models in the library with the observations. To this end, algorithms for both qualitative simulation of the response of materials to creep and relaxation experiments, and qualitative interpretation of experimental data have been implemented.

The selection of the most accurate ODE model is easily performed when the expert knows the number of either retardation or relaxation times (Ferry, 1970; Whorlow, 1980) as the order of the ODE can be correlated with such a number. Therefore, in this case, the selection problem is restricted to a parameter identification problem. When this information is not available, the order of the equation is determined through a technique, borrowed from Statistics, which consists in minimizing a functional which expresses the balance between both the goodness of fitting, which increases when the number of parameters grows, and the significance of the numerical values of the parameters themselves, which diminishes when their number grows.

This paper is organized as follows: the next section deals with the approach used for selecting a plausible class of models, namely qualitative reasoning methods for simulation and data interpretation. Then, the algorithm for selecting the most accurate model together with the Akaike (Akaike, 1974) method, which has been adapted for solving our optimization problem, are presented.

Selection of the appropriate class of models

Qualitative methods both for simulation and interpretation of observations play a key role in the first stage of the model formulation problem. The qualitative interpretation of observations gathered from ad hoc designed experiments allows us to highlight some properties of the studied material, and consequently to make a guess for a set of its plausible models. The selection of the plausible ODEs occurs on the ground of the qualitative comparison of the simulated behaviors with the observations (Figure 1).

For the sake of clarity and completeness, we briefly recall the basic assumptions, definitions and methods underlying our previous work.

Suitable modeling assumptions are made about materials, acting forces and processes. A material is assumed to be a continuous, homogeneous and isotropic medium, and processes take place in isothermal condition in order to decouple the thermodynamics aspects from the mechanical ones. Only stable materials are considered and therefore the deformation of a body solely occurs when mechanical energy has been provided. As many rheological experiments are carried out in one dimension, that is mechanical energy is supplied through longitudinal traction or compression forces, we only consider one-dimensional deformation processes.

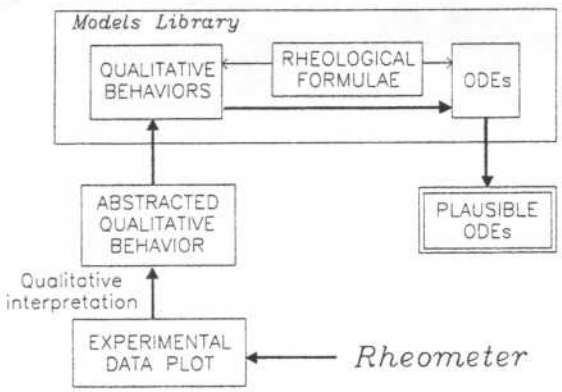


Figure 1: First stage in model selection: a class of models is selected from the model library in agreement with the qualitative interpretation of data. Thick line arrows track data flow, while thin ones denote interactions within library items

The method adopted for building the model of a material is based on a component-connection paradigm and on internal state variables. Each fundamental response, in our case elasticity and viscosity, corresponds to an ideal material which can be represented by a mechanical analogous device. More precisely, the purely elastic response is associated with a material H analogically represented by a spring whose response is described by the Hooke's law of linear elasticity $s = Ee$, where E is a constant which depends on the material. Similarly, the purely viscous response corresponds to a material N analogically represented by a dashpot, and is described by the Newton's law of linear viscosity $s = \eta\dot{e}$, where η, the viscosity coefficient of the material, is a positive constant and the dot denotes the time derivative. Although the constitutive law of a material may be non-linear and may contain non-constant coefficients, we consider linear visco-elastic models, i.e. models whose visco-elastic behavior is described by an ODE with constant coefficients. Nevertheless, most materials show a linear time dependent behavior in the limit of infinitesimal deformation and even in finite deformation as long as the strain remains below a certain limit, which varies from material to material.

Models of complex materials are built by analogy with mechanical devices, which are obtained by suitably assembling, either in parallel or in series, components which represent the fundamental mechanical properties so that the whole device behaves analogously to an actual material. The symbolic description of an analogical structure, RF, represents a model of its corresponding material at the lowest level of description.

The model library is automatically generated and includes (1) RFs of all non-equivalent structures made up of n components, (2) the corresponding ODE models, and (3) the simulated qualitative responses of the generated models to standard experiments. At first, the rheological formulae are recursively built and grouped with respect to equivalence relations: only one representative for each class is kept in the library. Then, the set of formulae is mapped to its corresponding set of mathematical models by exploiting the ODE formalism, the basic component models and suitable connection rules. More precisely, when components C_1 and C_2 (not necessarily basic components) are connected in parallel $(C_1|C_2)$, they undergo the same elongation while the total stress gets distributed among the components, that is: $s = s_1 + s_2\ ,\quad e = e_1 = e_2$, where e_i, s_i $(i = 1, 2)$ are the *internal variables*, whose time evolution is expressed either directly by the basic models or by differential equations obtained by the recursive application of the connection rules starting from the basic models. If the components are connected in series $(C_1 - C_2)$, each component takes the same load and the total elongation is the sum of the elongation of each component: $s = s_1 = s_2\ ,\quad e = e_1 + e_2$. By exploiting the constitutive equations of the basic components and the connection laws, the mathematical model of an arbitrarily complex RF can be recursively derived, as described in (Capelo *et al.*, 1993). ODE models can be further grouped with respect to their mechanical behavior by considering the structure of their constitutive equations.

Let us call *Formal Equation (FE)* the symbolic ODE obtained by giving unitary value to all non-zero coefficients of the constitutive equation. In (Capelo *et al.*, 1993) we proved (**Theorem 1**) that the following four classes of equations gather all of the admissible FEs :

$$(FE_1, m) \qquad \sum_{i=0}^{m} D^i s = \sum_{i=0}^{m} D^i e$$

$$(FE_2, m) \qquad \sum_{i=0}^{m} D^i s = \sum_{i=1}^{m+1} D^i e$$

$$(FE_3, m) \qquad \sum_{i=0}^{m} D^i s = \sum_{i=0}^{m+1} D^i e$$

$$(FE_4, m) \qquad \sum_{i=0}^{m+1} D^i s = \sum_{i=1}^{m+1} D^i e$$

where $m \geq 0$, and D^i denotes the $i-$th time derivative operator. If n is the maximum number of basic components the RFs are made up, it can be proved by induction that the ODE models space is equal to $2n$. The number of equations in each class (FE_i, m) is between $n/2$ and $n/2 + 1$.

The first stage in the model selection process consists in taking the class of (FE_i, m) which exhibits the same qualitative behavior of the real material out of the set $\{(FE_i, m), i = 1, 2, 3, 4\}$.

The qualitative behavior of the material in response to either a creep or relaxation experiment is characterized by the presence or absence of either strain or stress

properties, respectively. Therefore, the simulation algorithm, as well as the observations interpretation algorithm, generates qualitative profiles which highlight such physical properties.

Qualitative simulation

The simulation algorithm (Algorithm B), which is an extension and a generalization of the algorithm (Algorithm A) previously defined only for creep experiments (Capelo *et al.*, 1993), generates the responses of the ideal materials in the library to creep and relaxation experiments. It operates at the lowest level of description, i.e. on the rheological formula, as symbolic integration procedures could be heavy or unfeasible to be applied, and QSIM-like algorithms (Kuipers, 1986) would fail because of both the need to introduce a large number of auxiliary variables in writing the qualitative version of the ODEs and the consequent difficulty in controlling the proliferation of the predicted behaviors.

In order to define both the qualitative creep (QB_c) and relaxation (QB_r) response let us remind that a *creep* experiment consists of applying an external force on the material and observing the caused deformation, whereas a *relaxation* test consists of imposing a deformation and measuring the corresponding produced stress. According to the input signal shape, we further distinguish in static and dynamic experiments. Static tests involve the instantaneous imposition of a constant stress (or strain) and the observation of the subsequent evolution over time of the strain (or stress). Dynamic tests involve the application of an oscillatory input signal. Static experiments highlight the qualitative viscoelastic properties we are interested in better than dynamic tests, as in the former case the response of the material can be analyzed even in the very initial phase of the experiment, whereas in the latter one it provides useful information only after a transitory phase. For this reason we only consider static tests. The form of the applied excitation is suggested by criteria of theoretical and experimental simplicity. Standard *static* excitations are mathematically modelled by step functions of the type:

$$c_0[H(t-t_0) - H(t-t_1)]$$

where t_0 and t_1 (loading and unloading instants) are the significant time-points, $H(t)$ is the Heaviside function, and c_0 is constant (Figure 2).

Creep and relaxation are dual aspects of the same phenomenon: the molecular rearrangements occurring inside a material subjected to external forces depend upon time. When the stress is regarded as the cause, the molecular rearrangements appear at a macroscopic level as a retardation of strain: the time required can be very short if the tested material is elastic, very long if it is viscous, or finite – of the order of the scale of the experiment – in the intermediate cases. Similarly, when the stress is regarded as the effect, a stress relaxation is produced macroscopically over a more or less

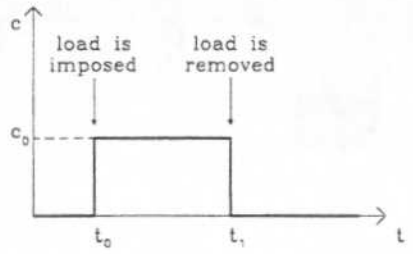


Figure 2: Standard static test: a stress or a strain is suddenly imposed and then held constant for a time $\Delta t = t_1 - t_0$

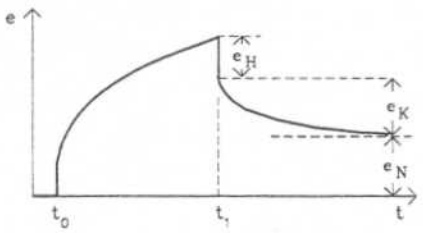


Figure 3: A Typical strain response to a stress step excitation (creep test)

long time, depending on the mechanical properties of the material.

The strain response to a step stress excitation results from the superposition of three basic components, namely an elastic instantaneous e_H deformation, a delayed (still elastic) e_K one, and a viscous irrecoverable e_N deformation:

$$e = e_H + e_K + e_N \ .$$

Therefore, for example, a purely viscous material, which dissipates all the deformation energy as heat through viscous forces, undergoes an irrecoverable deformation and is characterized by $e = e_N \neq 0$, $e_H = e_K = 0$. More generally, a visco-elastic material might store part of the deformation energy elastically as potential energy and dissipate the remaining one viscously as heat, featuring a strain response like the one shown in Figure 3, which is characterized by $e_H,\ e_K,\ e_N \neq 0$. Such a material would recover a part of the deformation instantaneously, a part more slowly, but would also undergo a permanent deformation to a certain extent.

The stress response to a step strain excitation results from the superposition of three stress components:

$$s = s_H + s_M + s_N \ ,$$

each one associated with a mechanical property of the tested material. More precisely: $s_H \neq 0$ means that part of the undergone stress does not relax at all in the time interval (t_0, t_1) and denotes the ability of the material to store potential energy; $s_M \neq 0$ means

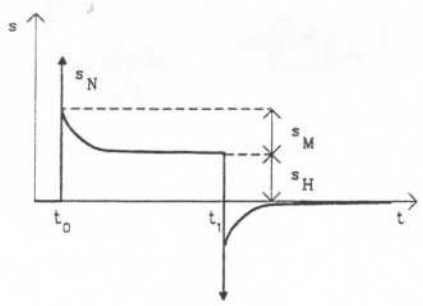


Figure 4: A typical stress response to a strain step excitation (relaxation test)

that part of the undergone stress relaxes slowly during (t_0, t_1); $s_N \neq 0$ means that part of the undergone stress relaxes instantaneously at $t = t_0$ and the deformation energy is partially dissipated through viscous forces.

As an example, Figure 4 illustrates a stress response characterized by s_H, s_M, $s_N \neq 0$. Let us remark that $s_N \neq 0$ theoretically corresponds to a Dirac delta function and is graphically represented by a vertical arrow pointing either upward or downward according to the stress impulse sign.

Therefore, both QB_c and QB_r can be described by just three logical parameters (α, β, γ) which are associated with either the strain or stress properties, respectively, and may take on either the value TRUE (T) or FALSE (F) in accordance with the presence or absence of the corresponding property in the material. For example $QB_c = (T, T, T)$ means that the material exhibits instantaneous elasticity $(e_H \neq 0)$, delayed elasticity $(e_K \neq 0)$ and viscosity $(e_N \neq 0)$. In order to define QB_c and QB_r for any complex formula, let us remind that, according to the connection rules, strains are added in series and stresses are added in parallel. Therefore:

if $QB_c[C_1] = (\alpha_1, \beta_1, \gamma_1)$, $QB_c[C_2] = (\alpha_2, \beta_2, \gamma_2)$, then

$$QB_c[C_1 - C_2] = QB_c[C_1] \vee QB_c[C_2] =$$
$$= (\alpha_1 \vee \alpha_2,\ \beta_1 \vee \beta_2,\ \gamma_1 \vee \gamma_2);$$

if $QB_r[C_1] = (\alpha_1, \beta_1, \gamma_1)$, $QB_r[C_2] = (\alpha_2, \beta_2, \gamma_2)$, then

$$QB_r[C_1|C_2] = QB_r[C_1] \vee QB_r[C_2] =$$
$$= (\alpha_1 \vee \alpha_2,\ \beta_1 \vee \beta_2,\ \gamma_1 \vee \gamma_2),$$

where $\vee$ denotes the logical OR operator.

It is obvious that QB_c and QB_r would be straightforward defined for any given complex formula C if it could be possible to use one of its equivalent formulae which is characterized either only by the series operator or only by the parallel one. To this end, in the following we give theorems which state a one-to-one correspondence between the classes of ODEs and suitable classes of rheological formulae (reference classes), and the mechanical equivalence of any complex formula with a formula in one of the reference classes. More precisely, let us denote by Ω the mapping of a RF to its respective FE, by K and M the models defined by the formulae $H|N$ and $H - N$, respectively known in the literature as Kelvin and Maxwell models, and by K_m (M_m respectively) the m-th order generalized Kelvin (Maxwell respectively) models:

$$K_m = \underbrace{K - K - \ldots - K}_{m} \qquad (M_m = \underbrace{M \mid M \mid \ldots \mid M}_{m}),$$

and assume, conventionally, that $K_0 = M_0 = \emptyset$. It can be easily proved by induction on the connection laws that for any integer $m \geq 0$:

Theorem 2. The formulae $K_m - H$, $K_m - N$, K_{m+1}, $K_m - M$ map to (FE_1, m), (FE_2, m), (FE_3, m), (FE_4, m), respectively :

$$\Omega(K_m - H) = (FE_1, m)$$
$$\Omega(K_m - N) = (FE_2, m)$$
$$\Omega(K_{m+1}) = (FE_3, m)$$
$$\Omega(K_m - M) = (FE_4, m)$$

Theorem 3. The formulae $M_m|H$, $M_m \mid N$, $M_m \mid K$, M_{m+1} map to (FE_1, m), (FE_2, m), (FE_3, m), (FE_4, m), respectively :

$$\Omega(M_m \mid H) = (FE_1, m)$$
$$\Omega(M_m \mid N) = (FE_2, m)$$
$$\Omega(M_m \mid K) = (FE_3, m)$$
$$\Omega(M_{m+1}) = (FE_4, m)$$

The formulae mentioned in Theorems 2, 3 are the *reference formulae*.

An important consequence of Theorems 1, 2, 3 is the following *duality property*:

Theorem 4. Given any arbitrarily complex rheological formula C, it is possible to find a mechanically equivalent formula C_- obtained by combining in series a suitable number of components H, N, K, as well as it is possible to find an equivalent formula $C_|$ obtained by combining in parallel a suitable number of components H, N, M.

[*Proof.*
If $\Omega(C) = (FE_1, m)$ then set $C_- = K_m - H$, $C_| = M_m|H$.
If $\Omega(C) = (FE_2, m)$ then set $C_- = K_m - N$, $C_| = M_m|N$.
If $\Omega(C) = (FE_3, m)$ then set $C_- = K_{m+1}$, $C_| = M_m|(H|N)$.
If $\Omega(C) = (FE_4, m)$ then set $C_- = K_m - (H - N)$, $C_| = M_{m+1}$.]

Let us call C_- and $C_|$, chosen as in the above theorem proof, equivalent *dual* (respectively *series* and *parallel*) *representations* of C. By denoting with $\mathcal{M}$ the set of all and none but the mechanically distinct RFs,

by Theorem 4 we can state that $\mathcal{M}$ is isomorphic to the dual sets:

$$\mathcal{M}_- = \{K_m - H,\ K_m - N,\ K_{m+1},\ K_m - M\}$$

$$\mathcal{M}_| = \{M_m | H,\ M_m | N,\ M_m | K,\ M_{m+1}\}$$

Therefore, the set of models $\{H,\ N,\ K\}$ plays the role of a "basis" of components for $\mathcal{M}_-$ with respect to the series operator, as well as the set $\{H,\ N,\ M\}$ respectively plays a similar role for $\mathcal{M}_|$ with respect to the parallel operator.

Theorem 4 suggests a natural way of defining both QB_c and QB_r associated with any given formula C by considering its equivalent representation either C_- in $\mathcal{M}_-$ or $C_|$ in $\mathcal{M}_|$, respectively.

The algorithm for determining both QB_c and QB_r associated with any given formula C is defined as follows:

$$QB_c[C] = \begin{cases} \text{If } C \in \{H, N, K\} \text{ then :} \\ \quad QB_c[H] = (T, F, F) \\ \quad QB_c[N] = (F, F, T) \\ \quad QB_c[K] = (F, T, F) \\ \text{else, let } C \text{ be replaced by its series equivalent representation } C_- : \\ \quad QB_c[C_-] = \begin{cases} QB_c[K_m - H] \\ QB_c[K_m - N] \\ QB_c[K_{m+1}] \\ QB_c[K_m - (H - N)] \end{cases} \end{cases}$$

$$QB_r[C] = \begin{cases} \text{If } C \in \{H, N, M\} \text{ then :} \\ \quad QB_r[H] = (T, F, F) \\ \quad QB_r[N] = (F, F, T) \\ \quad QB_r[M] = (F, T, F) \\ \text{else, let } C \text{ be replaced by its parallel equivalent representation } C_| : \\ \quad QB_r[C_|] = \begin{cases} QB_r[M_m | H] \\ QB_r[M_m | N] \\ QB_r[M_m | (H \mid N)] \\ QB_r[M_{m+1}] \end{cases} \end{cases}$$

The algorithm strictly depends on the application domain. However, its lack of generality is well compensated by its completeness and soundness. As a matter of fact, the given simulation algorithm is the qualitative transcription of the connection rules which mathematically express, through the internal variables, the links between either strain or stress and their own respective components. With reference to the RFs, the strain components are added in presence of a series operator, whereas the stress ones in presence of a parallel operator. Therefore, the proof of the soundness of the algorithm derives straightforward from the connection rules for the series and parallel operators. The proof of its completeness is given by Theorems 1-4. More precisely, Theorem 1 defines the elements (FE) of the set of the admissible ODE models; Theorems 2 and 3 respectively state a bijective correspondence between FEs and RFs (elements of both sets $\mathcal{M}_-$ and $\mathcal{M}_|$), which are built by connecting either in series the elements of the set $\{H,\ N,\ K\}$ or in parallel the elements of the set $\{H,\ N,\ M\}$. Finally, Theorem 4 associates any complex given formula, by exploiting its FE, with its equivalent representation both in $\mathcal{M}_-$ and in $\mathcal{M}_|$.

Although in terms of computational efficiency Algorithms A and B are comparable, version B marks a significant improvement. In fact, Algorithm B is fully justified from the formal point of view as it is based on sound arguments such as the connection rules and FEs, whereas Algorithm A, which recursively builds the response of the material starting from the qualitative behavior of the elements H and N, is partly suggested by the connection rules and partly by intuitive physical arguments. Moreover, Algorithm B works for both creep and relaxation tests.

Finally, let us observe that, although Algorithm B has been given, for the sake of simplicity, for input signals represented by step functions, it also holds for input signals represented by a summation of step functions. This is ensured by the linearity of the ODEs or, equivalently, by the Boltzmann principle of superposition.

Qualitative interpretation of experimental data

Qualitative interpretation of experimental data, i.e. their characterization in terms of relevant qualitative physical features, is an important step in view of model selection (Forbus, 1987; DeCoste, 1991; McIlraith, 1989).

Quite generally, physical features can be captured through the identification of characteristic shapes in the experimental data plot. Therefore, in order to reason qualitatively about the observed response, graphical data are first abstracted to a qualitative representation: a qualitative curve description is automatically provided in terms of regions which are homogeneous with respect to such graphical features as steepness, convexity and linearity. This process of data segmentation takes its basic ideas from pattern recognition and qualitative physics. We would like to remark that, though here applied to a specific domain, the adopted approach is quite general.

In the following, for the sake of simplicity and to better focus on methodologies, we limit our attention to creep experimental data. As a matter of fact, analysis of relaxation data can be carried out in quite a similar way by suitably extending the numeric descriptors and thresholds spaces as well as the vocabulary of qualitative curve attributes, in order to account for relaxation-specific graphical features.

Suitable numeric curve descriptors, namely the strain,

Table 1: Quantity spaces of the numeric curve descriptors.

Numeric descriptor d	Quantity space $\mathcal{Q}(d)$
e	$\{0, +s, +l, +\infty\}$
e'	$\{-\infty, -l, -s, 0, +s, +l, +\infty\}$
e''	$\{-, 0, +\}$
r^2	$\{0, 1^-, 1\}$

its first and second order time derivatives (e', e'') and the linear correlation coefficient (r), are chosen and a set of thresholds, whose values are obviously domain-dependent, is suitably defined so as to provide, through a mapping to qualitative values, a quantization of the continuous world.
After the definition of quantity spaces for the numeric descriptors, segmentation of the experimental time range into significant time points and time intervals is performed.

More precisely, let $\mathcal{D} = \{e, e', e'', r^2\}$ denote the set of quantitative descriptors; for any $d \in \mathcal{D}$ let $\mathcal{Q}(d)$ be its appropriate *quantity space* (see Table 1), i.e. the set of all the qualitative values of d necessary to reason qualitatively about the graphical features, and $\mathcal{S}(d)$ the set of the intervals determined by suitable numeric thresholds which allow to map the numeric values of d to qualitative values in $\mathcal{Q}(d)$ according to the following diagram:

$$\begin{array}{ccc} & d & \\ \swarrow & & \searrow \\ \mathcal{S}(d) \ni [s_1(d,l), s_2(d,l)] & \longleftrightarrow & l \in \mathcal{Q}(d) \end{array}$$

Moreover, let *time point* denote the time neighborhood of a measurement point, and *time interval* an interval whose width is not negligible with respect to the experimental range. For example, if t_0 is the first creep instant then $\tau_0 = [t_0, t_0 + \varepsilon[$, with $0 < \varepsilon << 1$, is the time point associated with it. The distinction is made necessary in order to reason about instantaneous properties in presence of measurement errors.
Among time intervals, in order to cope with asymptotic reasoning, we will also enclose intervals extending beyond the experimental range up to $+\infty$; such intervals will be denoted by an ∞-subscript.

Then, given $d \in \mathcal{D}$ and a time point or interval τ, the following mapping is defined

$$q \ : \ d, \tau \ \rightarrow \ qd(\tau) \in \mathcal{Q}(d)$$

which associates d with its qualitative value in τ, by a suitable numeric approximation of d in τ and mapping it onto $\mathcal{S}(d)$:

$$qd(\tau) = l \ \Leftrightarrow \ d(\tau) \in [s_1(d,l), s_2(d,l)] \ .$$

For example, if τ_0 is a time point we say that e' takes on the qualitative value 0 in τ_0 when the slope of the least squares line through the experimental points in τ_0 is $\in [s_1(e', 0), s_2(e', 0)]$, where $s_1(e', 0), s_2(e', 0) \in \mathbb{R}$ are the threshold values corresponding to the qualitative 0.
Quite similarly we say that $qr^2(\tau) = 1$ if r^2 , computed on $\{t_i\}_{i=1,k} \in \tau$, takes on a numeric value very close to 1, while $qr^2(\tau) = 1^-$ if the same quantity has a value close to 1, but less tightly:

$$qr^2(\tau) = 1 \ \Leftrightarrow \ s_1(r^2, 1) < r^2 < s_2(r^2, 1) = 1$$

$$qr^2(\tau) = 1^- \ \Leftrightarrow \ s_1(r^2, 1^-) < r^2 < s_2(r^2, 1^-)$$

where $s_1(r^2, 1^-) \leq s_1(r^2, 1)$ (e.g. $s_1(r^2, 1^-) = .995$, $s_2(r^2, 1^-) = s_1(r^2, 1) = .998$).

Table 2: Qualitative characterization of graphical properties featured in a time point or interval τ by means of numeric descriptors.

Qualitative curve attribute	Characterization
vertical	$qe'(\tau) = \pm\infty$
steep	$qe'(\tau) \leq -l \vee qe'(\tau) \geq +l$
linear	$qr^2(\tau) = 1$
linear & growing	$qr^2(\tau) = 1 \wedge qe'(\tau) > 0$
weakly linear	$qr^2(\tau) = 1^-$
weakly linear & growing	$qr^2(\tau) = 1^- \wedge qe'(\tau) > 0$
concave	$qe''(\tau) < 0$
loosely concave	$qe''(\tau) \leq 0$
convex	$qe''(\tau) > 0$
loosely convex	$qe''(\tau) \geq 0$
asymptotically positive horizontal	$qe(\tau_\infty) > 0$
asymptotically largely positive horizontal	$qe(\tau_\infty) > +s$

A general approach to the characterization of graphical features by means of the numeric descriptors is shown in Table 2. The experimental time range is partitioned into time points and time intervals (Figure 5) where symbolic properties, characterizing specific graphical features, are satisfied (right column in Table

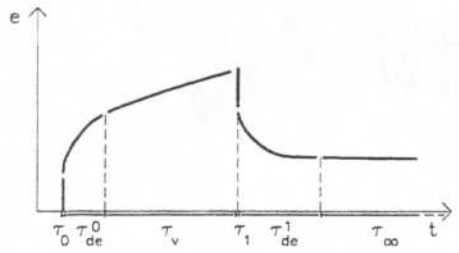


Figure 5: Data segmentation in view of qualitative curve description

2); the experimental data plot is then described by the sequence of the qualitative curve attributes (from the vocabulary in the left column) which are appropriate for each time segment.

Assessment of the observed qualitative behavior

In the case of creep data, two experimental stages can be distinguished: creep, which is related to loading imposition and holding time $[t_0, t_1[$, and recovery, related to loading removal $[t_1, +\infty[$. As a matter of fact, either stage can provide enough information to assess instantaneous elasticity, delayed elasticity and viscosity. However, since data are always affected by measurement errors, reasoning about data can take advantage of both creep and recovery curve analysis. For the same reason, in the definition of clauses we require that each property is weakly satisfied in both creep and recovery and strongly in at least one of them.
In the light of the above considerations, the key clauses for the assessment of each mechanical feature are now described.

Instantaneous elasticity:

This property consists in the material ability at yielding a prompt deformation to loading and unloading (see Figure 3). Strain jumps at t_0 and t_1 graphically characterize instantaneous elasticity. Due to the instrument limitations at coping with instantaneous events, namely the instrumental inertia combined with the measurement errors, this property is captured by reasoning about the curve steepness on time-points τ_0, τ_1:

```
if curve is
  vertical at τ0 AND steep at τ1
  OR steep at τ0 AND vertical at τ1
then
  property-1 is instantaneous-elasticity
else
  property-1 is NOT-instantaneous-elasticity
endif
```

(Of course, in accordance with Table 2, `vertical at` τ_0 `AND vertical at` τ_1 implies `property-1 is instantaneous-elasticity`.)

Viscosity:

The presence of a constant positive deformation rate in the late creep stage is due to energy dissipation. Correspondingly on recovery the material is not able to get back to the original length. Therefore, an eventually linear growth at late loading or, equivalently, a finite positive residual as t goes to infinity characterize this property.
As regards the creep stage, the significant time interval can be identified as follows:

$$\tau_v := \left[\min_t\{t \mid qr^2(t, t_1) = 1\}\ ,\ t_1\right],$$

where $\min_t\{t \mid qr^2(t, t_1) = 1$ denotes the first instant of the linear region. Similarly a time interval τ_v^- is defined to characterize a weakly linear region. On recovery, in order to reason about asymptotic attributes, a time interval $\tau_\infty := [T, +\infty[\ ,\ T >> t_1$ is implicitly introduced. The qualitative value of the asymptotic strain $qe(\tau_\infty)$ is derived by an approximation of $\lim_{t\to\infty} e(t)$ obtained by extrapolating the late recovery points under a simple exponential decay model.

By combining creep and recovery reasoning, viscosity is finally assessed as follows:

```
if curve is
  linear-and-growing at τv
  AND
  asymptotically-positive-horizontal
  OR weakly-linear-and-growing at τv⁻
  AND
  asymptotically-largely-positive-horizontal
then
  property-3 is viscosity
else
  property-3 is NOT-viscosity
endif
```

Delayed elasticity:

In order to reason about delayed elasticity the curve concavity must be ascertained, even when data are affected by measurement errors. The expert would perform a subjective visual smoothing which can be emulated in a number of ways. A simple and quite natural approach consists in reasoning by intervals rather than point-wise; first the time interval to analyze τ_{de} (τ_{de}^0 for creep and τ_{de}^1 for recovery, respectively) is divided into subintervals $\tau_{de} = \cup_h \tau_h$, and then for each subinterval $qe''(\tau_h)$ is taken as the most frequent of the pointwise computed signs of e''. Therefore:

$\tau_{de}^0 := [t_0, t_1[\ \setminus(\tau_0 \cup \tau_v^-)$;

```
if τde⁰ = ∅
then
  property-2 is NOT-delayed-elasticity
else
  perform interval partitioning τde⁰ = ∪h τh
  if curve is
    loosely-concave at any τh
```

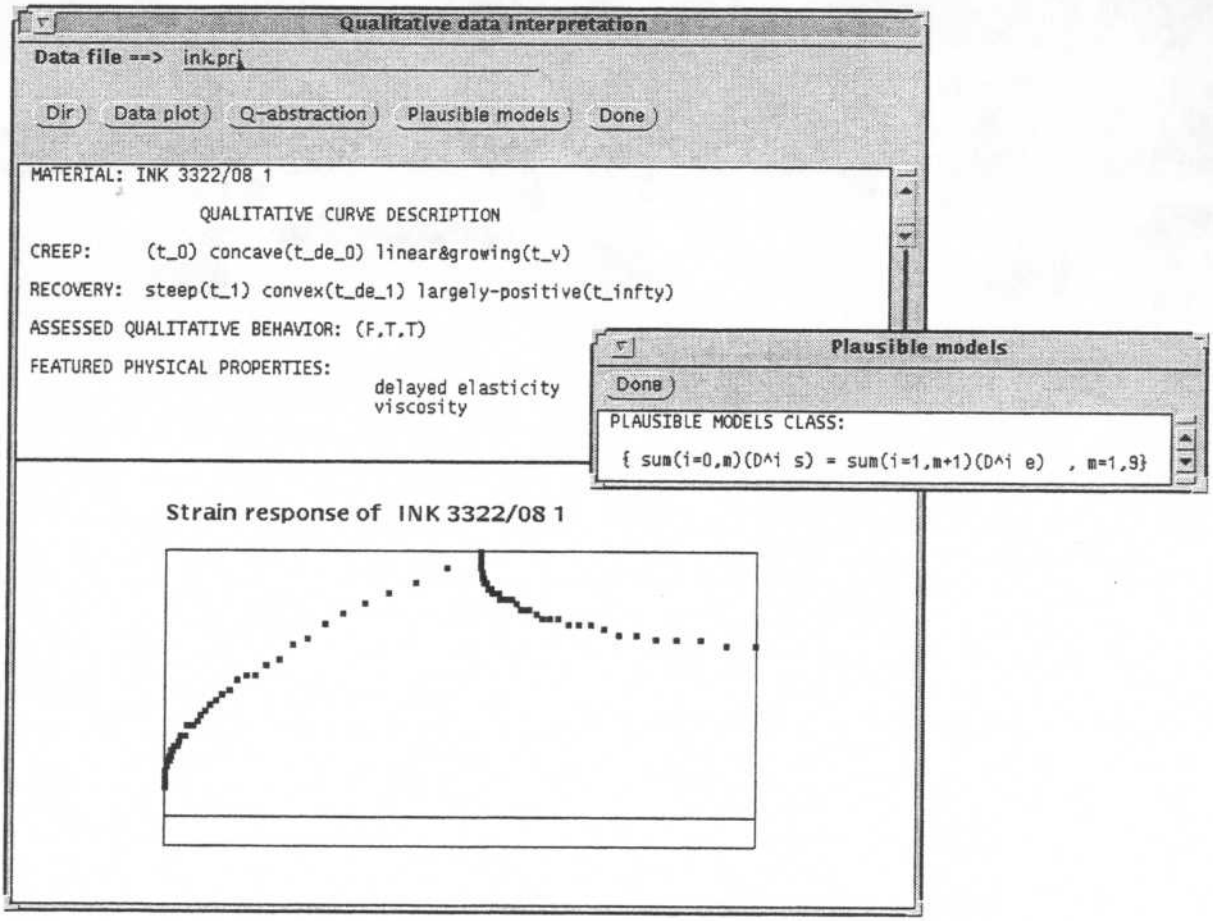


Figure 6: Selection of the plausible models class for an ink-like material

```
    AND
    concave for at least one τ_h
  then
    property-2 is delayed-elasticity
  else
    property-2 is NOT-delayed-elasticity
  endif
endif
```

The above clause is actually made more robust by including reasoning about the convexity of the recovery curve.

An example

As an example, let us consider the observed data from a creep experiment performed on an ink-like material. Figure 6 is a screen dump which displays a plot of the observations, and the featured physical properties attributed to the shape of the plot (delayed elasticity and viscosity). As a result of the match of the identified strain characteristics (e_H =F, e_K =T, e_N =T) against the qualitative simulated behaviors of ideal materials, a class of plausible candidate mathematical models of the material is automatically provided. For the material in the example, 9 candidate models, whose equations are of the form $\sum_{i=0}^{m} D^i s = \sum_{i=1}^{m+1} D^i e$ (where $m = 1, ..., 9$) are generated.

Selection of the accurate model

Given the class (FE_i, m) of ODE models which exhibit the same qualitative behavior as the actual material, the selection problem consists in identifying an equation E_i which refines the quantitative properties of the material. E_i is the equation of order k whose parametric coefficients have been assigned numeric values so that it better fits the experimental data. Therefore, in order to solve the problem we have to determine both the order k of the equation and the numeric values of its parametric coefficients, whose number n obviously depends on k ($n(k)$). It is clear that if the order of the equation is increased, and consequently the number of the parameters is increased, the goodness of fitting improves, but the significance of the numerical values of the parameters themselves diminishes. On the other hand, also the possibility to give a physical interpretation of these numerical values can fail, and even more important, the information about the number of retardation times (Ferry, 1970; Whorlow, 1980), which is a feature of the material and is strictly related to the order of the model equation, can be lost if we restrict the problem to the goodness of fitting.

Let us remind that the retardation times are parameters associated with the material state changes which subsequently occur. For example, in polymeric

materials they can be associated with the break of either the hydrogen or Van der Waals bounds which does not occur at simultaneous times. From the modeling point of view, the retardation times are specified in the arguments of exponential functions whose sum defines the solution of the ODE model. In some cases, the expert can estimate the number of state changes the material will undergo during the rheological tests, and then the number of retardation times. When the user has such a knowledge, the selection problem is restricted to a parameter identification one (Figure 7 (a)). By exploiting arguments based on the Laplace transform, it can be proved that there is a correlation between the number of retardation times and the order of the equation which describes the behavior of the material. Therefore, the order k, and consequently the number of parameters is straightforward fixed. More precisely, if k is the number of retardation times, the order of the most plausible ODE model is equal to k if the selected class is either FE_1 or FE_3, and equal to $k+1$ in the other two cases. Then, the most accurate model of the material can be determined by computing, through ad hoc implemented techniques of fitting of experimental data, the values of the parametric coefficients which appear in (E_i, k).

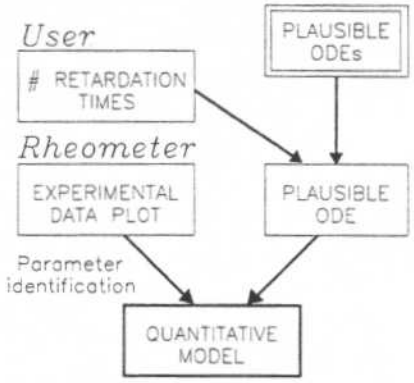
The problem is more complex if no information is provided about the ODE order. In such a case, the quantitative model is obtained through an optimization technique whose goal is to determine the optimal order k of the equation E_i so that both the goodness of fitting and the significance of the identified numerical values of parameters are guaranteed. The selected class of plausible ODEs is made up of equations E_i which differ in the order but not in the structure, i.e. $(E_i, k+1)$ differs from (E_i, k) only for the presence of terms which express the $(k+1)$-derivative of either stress or strain. An analogous situation can be found in the theory of time series where the problem to determine the order of the model (e.g. AutoRegressive Moving Average (ARMA) model) is quite important (Choi, 1992). In this context several methods have been proposed, and among those we consider the AIC (Akaike Information Criterion) method (Akaike, 1974). In our case, such a criterion can be formulated as follows:

(**Akaike Criterion**) The equation of order k which best fits the experimental data corresponds to the equation whose number of parameters $n(k)$ is the minimum of the function:

$$(1)\quad AIC(n(k)) = 2n(k) - 2\log[\text{maximized likelihood}].$$

It can be proved that the order k estimated in such a way is never lower than the actual order of the E_i which describes the behavior of the material. It can be straightforward proved that if, as it is plausible in our case, the experimental errors are independent and normally distributed, the maximum likelihood method

(a)



(b)

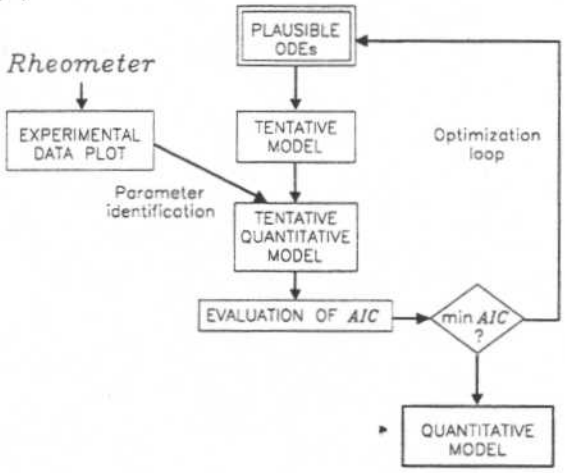


Figure 7: Second stage in model selection: (a) if the number of retardation times is supplied by the User, a single ODE is selected and the quantitative model is obtained by parameter identification; (b) if no information is provided about the ODE order, the quantitative model is obtained through an optimization technique

corresponds to the least squares method. Therefore, apart from constant additive terms, the function (1) becomes:

$$(2) \quad AIC(n(k)) = 2n(k) + m \log S^2(n(k)),$$

where m is the number of experimental data and $S^2(n(k))$ the sum of the squares of residuals when the fitting is performed with a model with $n(k)$ parameters.

From the algorithmic point of view, the value of $n(k)$ which minimizes the function (2) is calculated through a loop (Figure 7 (b)) whose main steps are:

- selection of an equation (E_i, k),
- parameter identification of the $n(k)$ parameters in (E_i, k),
- evaluation of $AIC(n(k))$,
- comparison of $AIC(n(k-1))$ with $AIC(n(k))$ and loop termination if $AIC(n(k-1)) < AIC(n(k))$, since in general the function (2) is convex.

The output of the loop, i.e. the quantitative ODE model of order $k-1$ with $n(k-1)$ coefficients is an accurate model of the behavior of the material as it guarantees both the goodness of data fitting and the significance of the numerical values of its coefficients.

Discussion

We presented an approach to the automated formulation of an accurate quantitative model of the behavior of a visco-elastic material. Such a model can be exploited by the expert for a precise analysis of the material during its assessment phase. More precisely, it can be numerically simulated to predict the behavior of the material under any complex load, and the parameters in the model could be interpreted as a measure of some properties of the material itself. As far as the latter use of the model is concerned, there is a need for a rationale which allows us to correlate the parameters in the model and physical properties of materials. This issue could be a new challenge in the study of materials, which has been performed at a pure experimental level so far.

The model selection process occurs in two main stages: the first stage is performed at a pure qualitatively level and produces the class of models which qualitatively describe the material behavior, the second one, which occurs at a pure quantitative level by exploiting both statistical and numerical methods, generates the quantitative model of the material. The model is selected within a library of models which are automatically generated by connecting variously either in series or in parallel 20 basic components. The restriction of the number of components to 20 is reasonable as structures made up of 20 elements are associated with rather complex materials. As the model space dimension is equal to 40, it could seem reasonable to select directly the accurate quantitative model by extending the optimization loop in Figure 7 (b) to the whole set of models in the library instead of selecting first a class of plausible models. There are at least two reasons why both two steps are necessary. The most important reason is that there is no guarantee that the "best" model obtained through the data fitting by means of all the models results to be compatible, in qualitative terms, with the observations, or, in other words, that it does capture some important features of the material (e.g. instantaneous elasticity). Moreover, the computational cost of parameters identification grows significantly with the number of models. Such a number is actually not too large as we only consider linear models of visco-elastic materials but it could be quite high as soon as the library is extended both with non-linear models, which will be built starting from non-linear laws, e.g. $s(t) = k_1(\exp(k_2 e(t)) - 1)$, and $s(t) = k_1 \log(k_2 e'(t))$, for the elastic and viscous basic components respectively, and with models which take also plasticity into account. Therefore, the qualitative stage seems to be essential for both an efficient and physically correct approach to model selection.

The paper gives some contribution to qualitative reasoning methods: both qualitative simulation and data interpretation methods ad hoc implemented for our specific goal have been presented. The simulation algorithm is strictly domain-dependent but generates all and none but the actual physical behaviors. The data interpretation algorithm provides reasoning techniques to emulate the expertlike visual interpretation of experimental data. It has been tested on experimental data related to different polymeric materials, such as inks, rubbers, and drugs (sodium carboximethylcellulose) and its performance has been evaluated in accordance with the interpretation provided by the experts who supplied the data. In spite of its simplicity, the algorithm provides also useful information about the adequacy of the models in the library to describe the behavior of the material under study: a possible convex shape of the data during the creep phase denotes that the linear theory of elasticity and viscosity we adopted to build the model library is inadequate to study such a material. Although it has been designed for interpreting data in a specific domain, it results to be a domain-independent technique as it is capable to trasform a stream of observed data into a qualitative description that characterizes its shape, that is it highlights the qualitative properties of the numerical data such as monotonicity, convexity and linearity regions.

Acknowledgement

We would especially like to thank C. Caramella of the Department of Pharmaceutical Chemistry of the University of Pavia and B. Pirotti of the PABISCH S.p.A - Scientific Division who provided us with the experimental data.

References

Addanki, S.R.; Cremonini, R.; and Penberthy, J.S. 1991. Graphs of models. *Artificial Intelligence* 51:145–177.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* 19:716–723.

Arridge, R.G.C. 1975. *Mechanics of Polymers.* Clarendon Press, Oxford.

Bradley, E. 1994. Automatic construction of accurate models of physical systems. In *Proc. 8th International Workshop on Qualitative Reasoning*, Nara. 13–23.

Capelo, A.C.; Ironi, L.; and Tentoni, S. 1993. A model-based system for the classification and analysis of materials. *Intelligent Systems Engineering* 2(3):145–158.

Choi, B. 1992. *ARMA Model Identification.* Springer, New York.

Crawford, J.; Farquhar, A.; and Kuipers, B.J. 1992. QPC: A compiler from physical models into qualitative differential equations. In B. Faltings, P. Struss, editor 1992, *Recent Advances in Qualitative Physics.* MIT Press. 17–32.

DeCoste, D. 1991. Dynamic across-time measurement interpretation. *Artificial Intelligence* 51:273–342.

Falkenhaier, B. and Forbus, K.D. 1991. Compositional modeling: finding the right model for the job. *Artificial Intelligence* 51:95–143.

Ferry, J.D. 1970. *Viscoelastic Properties of Polymers.* John Wiley, New York.

Forbus, K.D. 1987. Interpreting observations of physical systems. *IEEE Transactions on Systems, Man, and Cybernetics* 17:350–359.

Ironi, L. and Stefanelli, M. 1994. A framework for building and simulating qualitative models of compartmental systems. *Computational Methods and Programs in Biomedicine* 42:233–254.

Iwasaki, Y. and Levy, A.Y. 1994. Automated model selection for simulation. In *Proc. AAAI-94*, Los Altos. Morgan Kaufmann.

Iwasaki, Y. 1992. Reasoning with multiple abstraction models. In Faltings, B. and Struss, P., editors 1992, *Recent Advances in Qualitative Physics.* MIT Press. 67–82.

Kuipers, B.J. 1986. Qualitative simulation. *Artificial Intelligence* 29:280–338.

Low, C.M. and Iwasaki, Y. 1992. Device modelling environment: an interactive environment for modelling device behaviour. *Intelligent Systems Engineering* 1(2):115–145.

McIlraith, S.A. 1989. Qualitative data modeling: application of a mechanism for interpreting graphical data. *Computational Intelligence* 5:111–120.

Nayak, P.P. 1994. Causal approximations. *Artificial Intelligence* 70:277–334.

Rickel, J. and Porter, B. 1994. Automated modeling for answering prediction questions: Selecting the time scale and system boundary. In *Proc. AAAI-94*, Los Altos. Morgan Kaufmann.

Weld, D.S. 1990. Approximation reformulations. In *Proc. AAAI-90*, Los Altos. Morgan Kaufmann. 407–412.

Weld, D.S. 1992. Reasoning about model accuracy. *Artificial Intelligence* 56:255–300.

Whorlow, R. 1980. *Rheological Techniques.* Ellis Horwood, Chichester.