Original Article

Journal of Intelligent Material Systems and Structures

Journal of Intelligent Material Systems and Structures
2016, Vol. 27(20) 2721–2743

DOI: 10.1177/1045389X16635845
jim.sagepub.com
SAGE

Modelling, characterisation and uncertainties of stabilised pseudoelastic shape memory alloy helical springs

Søren Enemark[1], Ilmar F Santos[1] and Marcelo A Savi[2]

Abstract

The thermo-mechanical behaviour of pseudoelastic shape memory alloy helical springs is of concern discussing stabilised and cyclic responses. Constitutive description of the shape memory alloy is based on the framework developed by Lagoudas and co-workers incorporating two modifications related to hardening and sub-loop functions designated by Bézier curves. The spring model takes into account both bending and torsion of the spring wire, thus representing geometrical non-linearities. Simplified models are explored showing that a single point in the wire cross section is enough to represent the global spring behaviour in spite of complex stress–strain distributions. The experiments are carried out considering different deflection amplitudes, frequencies and ambient temperatures, which influence the spring behaviour to different extents. The model is fitted against a calibration data set resulting in 1.3% residual standard deviation relative to the full range force. Compared to the validation data set, the errors are below 10% relative to the full range of the complex modulus. Uncertainty analysis of the model parameters using a Markov chain Monte Carlo technique shows low to high parameter correlation, and the relative uncertainties are less than $\pm 12\%$. Both the heat capacity and the convection coefficient are clearly identifiable from the performed experiments.

Keywords

shape memory alloys, helical springs, constitutive modelling, experiment, parameter estimation, parameter uncertainty, complex modulus

Introduction

A high level of attention has been given to shape memory alloys (SMAs) because of their extraordinary characteristics especially in terms of mechanical hysteresis, variable stiffness and strong thermo-mechanical coupling. Seismic applications are an example where Speicher et al. (2009) designed a device based on an SMA helical spring for retrofitting buildings. In the field of rotating machines, Enemark et al. (2015) showed from an experimental approach significant vibration suppression and adaptable critical speeds using SMA helical springs as foundation in a rotor-bearing system, and Ma et al. (2014) described the design, manufacturing and testing of a rotor support made from entangled SMA wire elements able to provide adaptable stiffness and damping characteristics.

Plastic strains accumulate during the initial loading cycles of SMAs. Therefore, a thermo-mechanical training process is required to get repeatable and stabilised behaviour. The transient training period comprises a number of loading cycles in the order of 100–400 depending on the material, its thermal treatment (shape setting) and the mechanical loading (Morin et al., 2011; Sakuma and Suzuki, 2007; Tobushi et al., 1992; Wolons et al., 1998). When the SMA element is stabilised, the dissipation capability as well as the average stiffness during a loading cycle depends on the loading amplitude, the ambient temperature and the loading rate. The martensitic phase transformation is non-diffusive, which means it is inherently rate independent.

[1]Department of Mechanical Engineering, Technical University of Denmark, Kongens Lyngby, Denmark
[2]Center for Nonlinear Mechanics, Department of Mechanical Engineering, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

Corresponding author:
Ilmar F Santos, Department of Mechanical Engineering, Technical University of Denmark, Nils Koppels Allé, Building 404, Room 005, DK-2800 Kongens Lyngby, Denmark.
Email: ifs@mek.dtu.dk

Nevertheless, the loading rate affects the response due to a balance between exo- and endothermic characteristics of the phase transformations (latent heat) and the heat transfer associated with the process. This aspect was first investigated by Shaw and Kyriakides (1995) on virgin pseudoelastic SMA wires in different media resulting in distinct convective conditions. Later, He and Sun (2010), Morin et al. (2011) and Malecot et al. (2006) investigated the temperature evolution during cycling loading. Also, Sameallah et al. (2015) investigated the evolution of the mechanical behaviour during cyclic loading of SMA wires, but did not measure the temperature. Pathak et al. (2010) established the convective heat transfer coefficient for SMA wires using different wire diameters, surrounding media and convective conditions (free/forced). Malecot et al. (2006) reported a 35°C increase in the average temperature during a cycle after approximately 80 cycles with a loading rate of 5 Hz. Therefore, the SMA temperature may increase significantly, and it may take many cycles before the behaviour stabilises.

For dynamic purposes, it is convenient to quantify the SMA behaviour in terms of the complex modulus, which comprises the storage modulus (a measure of average stiffness) and the loss modulus (energy dissipated during a loading cycle). The complex modulus approach has been adopted by many authors to quantify the cyclic SMA behaviour depending on loading amplitude, loading rate, ambient temperature and, in some cases, also pre-tension and convective conditions. This includes Piedboeuf et al. (1998), Wolons et al. (1998), Gandhi and Wolons (1999), Malecot et al. (2006) and He et al. (2010) on SMA wires; Holanda et al. (2014) on SMA helical springs and Ma et al. (2014) on their proposed SMA metal rubber bearing. Piedboeuf et al. (1998) and He et al. (2010) reported that the energy dissipation has an optimum in terms of loading frequency caused by a balance between the latent heat and the convective properties. Wolons et al. (1998) and Gandhi and Wolons (1999) mentioned that the energy dissipation (per unit volume) of SMAs is up to 20 times higher than that of elastomers, which are usually used for energy dissipation.

The helical spring is a promising SMA element because it enables more design possibilities than a wire or bar for example. The design criteria are usually related to stiffness, actuation force, length (deformation) and direction (tension/compression), which can be used to determine the spring dimensions, namely, wire diameter, spring diameter, number of coils and potentially also the initial pitch angle. There is a diverse range of possible applications of SMA helical springs. Liang and Rogers (1993) highlighted four different conceptual uses of them in vibration control related to the internal friction of martensite, the hysteresis of phase transformations, the stiffness difference of martensite and austenite and the principle of temperature actuation. Borges et al. (2013) used SMA springs together with a fuzzy controller to reduce vibrations of a rotor-bearing system when crossing a critical speed. Han et al. (2006) strengthened a slender structure towards buckling using an SMA spring. Thermal actuation of SMA springs has been used for a wireless earth-like micro robot (Kim et al., 2006) and for a propulsion system for a micro-robotic fish (Cho et al., 2008). De Sousa and De Marqui (2014) modelled and investigated numerically the aero-elastic behaviour of a typical aerofoil integrated with SMA springs. Aguiar et al. (2010) and An et al. (2012) provide good overviews over SMA helical spring applications.

Because the material is non-linear, it is not straightforward to model the behaviour of SMA helical springs. Geometrical non-linearities may also be introduced (Enemark et al., 2014; Savi et al., 2015) because SMAs can withstand large strains. This is very clear in some of the force–deflection tests performed by Sakuma and Suzuki (2007) and Savi et al. (2015), where the martensitic stiffness appears higher than the austenitic stiffness. Several authors proposed equivalent one-dimensional (1D) models (Aguiar et al., 2010; An et al., 2012; Enemark et al., 2014; Liang and Rogers, 1993), where An et al. (2012) also accounted some geometrical non-linearities. Generally, the 1D models are able to represent experimental behaviour well. Moreover, finite element models have been successfully implemented and the cross-sectional strain and stress distributions have been found very complex (Lagoudas et al., 2012; Mirzaeifar et al., 2011; Saleeb et al., 2013; Savi et al., 2015). Mirzaeifar et al. (2011) and Savi et al. (2015) were able to directly reproduce the experimental results and Saleeb et al. (2013) took into account the training process in their model. Mirzaeifar et al. (2011) also used two simpler models, and they concluded that these models perform acceptably in comparison with the experiments.

In the framework of reliable design of machine components and accurate predictions, uncertainties in SMA behaviour have to be quantified. However, to our knowledge, there are only very few works related to uncertainty quantification of SMAs: Oehler et al. (2012) determined the sensitivity of the optimal location of SMA flexures on variable geometry chevrons as a consequence of uncertain material and geometry parameters. Crews and Smith (2014) evaluated the uncertainties of estimated heat transfer parameters of an SMA bending actuator using the Bayesian framework (a Markov chain Monte Carlo (MCMC) method). The uncertainty analysis was carried out to establish a robust controller for the bending actuator (Crews et al., 2013). Oliveira et al. (2014) carried out an uncertainty analysis related to tensile tests of a pseudoelastic SMA wire to evaluate whether modelling predictions were within the uncertainty range of the experimental data. For similar experiments, Enemark and Santos (2015)

determined model parameters and their uncertainties using maximum likelihood estimation.

In this regard, the original contribution of this article has several aspects: (a) constitutive modelling: modifications to the existing SMA constitutive model by Lagoudas et al. (2012) to better describe the hardening and sub-loop behaviour relevant to springs, (b) helical spring modelling: a two-dimensional (2D) SMA helical spring model is established taking into account geometrical non-linearities, (c) experimental verification: experimental tests are developed and compared with numerical simulations and (d) uncertainty analysis: the model uncertainty is evaluated to give robust estimates of model parameters and their uncertainties. The proposed SMA spring model is fitted to and validated against experimental cyclic tensile tests highlighting the stabilised behaviour of SMA helical springs. Deflection amplitude, frequency and ambient temperature are varied, while the pre-tension length is kept constant. The behaviour under investigation only contains partial phase transformations (sub-loops) and not complete phase transformations, which motivates the implementation of a sub-loop function to the existing model by Lagoudas et al. (2012). Simple 1D models are able to properly represent the behaviour of SMA helical springs as highlighted in the presented literature review. However, to our understanding, it requires that some parameters have to attain non-physical values to counteract model deficiencies (Enemark et al., 2014). Here, the presented spring model accounts for material and geometrical non-linearities using a rigorous approach and it still maintains simplicity to aid transparency and to maintain computational efficiency compared to finite element models. A study is carried out highlighting different modelling approaches and their usability covering 1D and 2D models and different levels of approximations of the complex stress distributions in the spring wire cross section. An MCMC method is used to quantify model parameter uncertainties after fitting the model to the experimental results.

Experimental framework

This section covers the method for SMA spring fabrication and the test-bench for performing tensile tests.

Spring dimensioning

A helical spring has basically three dimensioning parameters, namely, wire diameter, helix or coil diameter and number of windings. The parameters should be chosen based on the application, but since SMA springs are highly non-linear, this may not be a simple task. In this work, two simple equations are used to dimension the spring based on the requirements related to actuation length and stiffness magnitude. It is assumed that the spring is pseudoelastic, which means that the austenitic phase is stable at room temperature if the spring is not stressed. Follador et al. (2012) also describe dimensioning of SMA springs.

Assuming the spring pitch angle is small, the maximum shear strain of the wire cross section (i.e. at the circumference) $\gamma(c)$ is related to the spring deflection u by

$$\gamma(c) = \frac{c}{2\pi N r_0^2} u = \frac{1}{l} u \tag{1}$$

where c is the wire radius, r_0 is the helix radius and N is the number of windings. A schematic of the spring may be seen in Figure 1. For dimensioning, it is further assumed that the shear stress is given by $\tau = G_A \gamma$, where G_A is the austenitic shear modulus. This is only satisfied when stresses are small such that the material is elastic. The relation between spring force F and spring deflection u becomes

$$F = \frac{\pi c^3}{2 r_0} \tau(c) = \frac{G_A c^4}{4 N r_0^3} u = k_A u \tag{2}$$

Requirements on the length measure l and the stiffness measure k_A can be used to determine the geometric properties. The stiffness k_A can be considered an upper bound for the tangential stiffness of the spring, since the tangential stiffness during phase transformation is usually significantly lower than the austenitic, elastic stiffness.

Fabrication of SMA helical spring

The fabrication and the shape setting of SMA helical springs are described by several authors

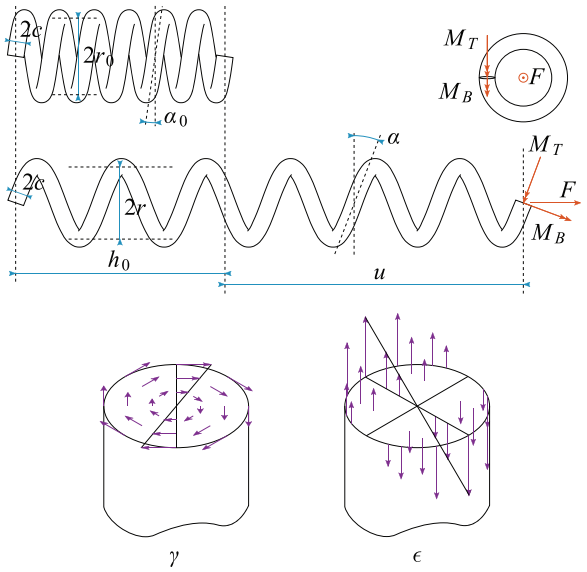

Figure 1. Schematic of spring, highlighting initial and deformed dimensions and the spring force F, based on the torsional moment M_T and the bending moment M_B. The torsion results in the shear strains γ in the wire cross section, and the bending results in the normal strains ε.

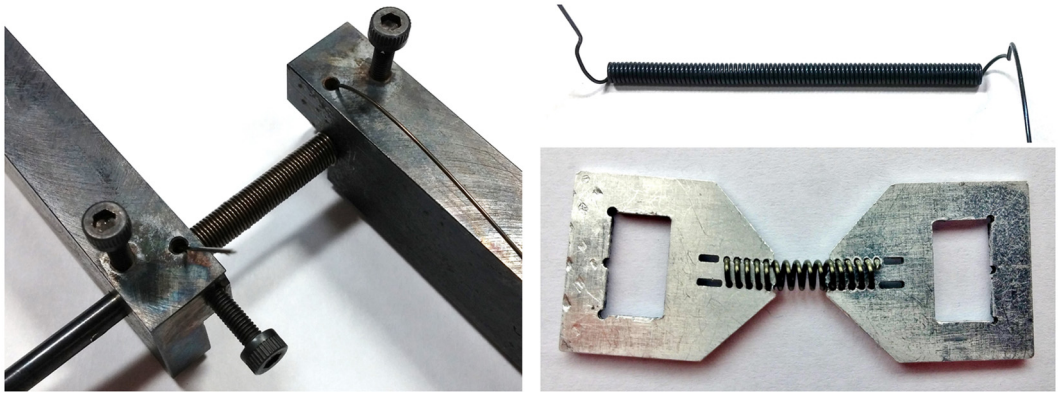

Figure 2. Spring fabrication: (a) SMA wire fixture used to keep the wire as a helical spring with inner diameter of 4 mm during shape setting; (b) SMA wire after second shape setting having an inner diameter of 2 mm (top) and SMA spring in its grips used when performing tensile tests after releasing it from the fixture (bottom).

(Follador et al., 2012; Liu et al., 2008; Sakuma and Suzuki, 2007; Savi et al., 2015; Tobushi et al., 1992). Elahinia et al. (2012) described shape setting in relation to medical applications. The production of the springs used in this work is greatly inspired by these articles.

The SMA wires have been annealed and straightened and have a light oxide finish according to the manufacturer. A steel fixture (Figure 2(a)) is used to wind up the SMA wire on a cylinder having a diameter of 4 mm. To set the shape, the wire and the fixture are heated in a pre-heated oven at 500°C for 18 min and quenched in water afterwards. The spring is then wound further onto a 2-mm cylinder and fixed, and the shape setting is repeated. The reason for using a two-step procedure is to avoid significant plastic deformation that would happen if the wire was wound directly onto the 2-mm cylinder (Elahinia et al., 2012). The finished spring may be seen in Figure 2(b), where it is also wound onto its two grips which are used for attaching the spring in the tensile test-bench.

Tensile test-bench

A test-bench (Figure 3) has been designed to perform tensile tests of the SMA springs. One end of the SMA spring is attached to a strain gauge–based force transducer (S2M/100N from HBM) to measure the spring force. The other end of the spring is attached to a crank mechanism. Both the pre-tension length (slider position) and deflection amplitude (crank radius) can be controlled. The angular position of the crank wheel is measured using an encoder (HEDS-9140#A00 and HEDS-5140#A13 from Avago Technologies), and this measurement is used to calculate the displacement of the piston arm holding the spring. A motor (3268G024BX4 from Faulhaber) and its controller use a feedback loop to control the speed of the crank wheel in the range from 0.01 to 20 Hz. The SMA spring is surrounded by a heat chamber to facilitate control of the ambient temperature. There is an inlet for a modified heat gun blowing approximately 6.5 L s^{-1} of heated air. The electric power to the gun heater is controlled, and together with a thermocouple (type T) and a feedback control loop, it is possible to maintain a constant ambient temperature within ±0.3°C. Since the force transducer is temperature sensitive, a heat wall and a long arm are used to hinder heat conduction to the sensor.

Constitutive SMA model

The thermo-mechanical behaviour of the SMA is described by the model presented by Lagoudas et al. (2012). In order to increase resemblance to the experiments, two related modifications are used, namely, a modification to the hardening function and the inclusion of sub-loop functions. Both modifications are described in the sections below.

The stress–strain fields in the helical spring can be approximated to be 2D containing one normal and shear components. The strain vector is $\boldsymbol{\varepsilon} = \{\varepsilon, \gamma\}^{\mathrm{T}}$, where ε and γ are the normal strain and the engineering shear strain, respectively, and the stress vector is $\boldsymbol{\sigma} = \{\sigma, \tau\}^{\mathrm{T}}$, where σ and τ are the normal and shear stresses, respectively. The symbol T denotes the transpose. Vector and matrix notation is used in this article opposed to tensor notation. The governing equations of the model are

$$\boldsymbol{\varepsilon} = \mathbf{S}\boldsymbol{\sigma} + \boldsymbol{\varepsilon}_t \tag{3}$$

$$\dot{\boldsymbol{\varepsilon}}_t = \boldsymbol{\Lambda}\dot{\xi} \tag{4}$$

where parameters are defined as follows

$$\mathbf{S} = \mathbf{S}_A + (\mathbf{S}_M - \mathbf{S}_A)\xi = \mathbf{S}_A + \Delta\mathbf{S}\xi$$

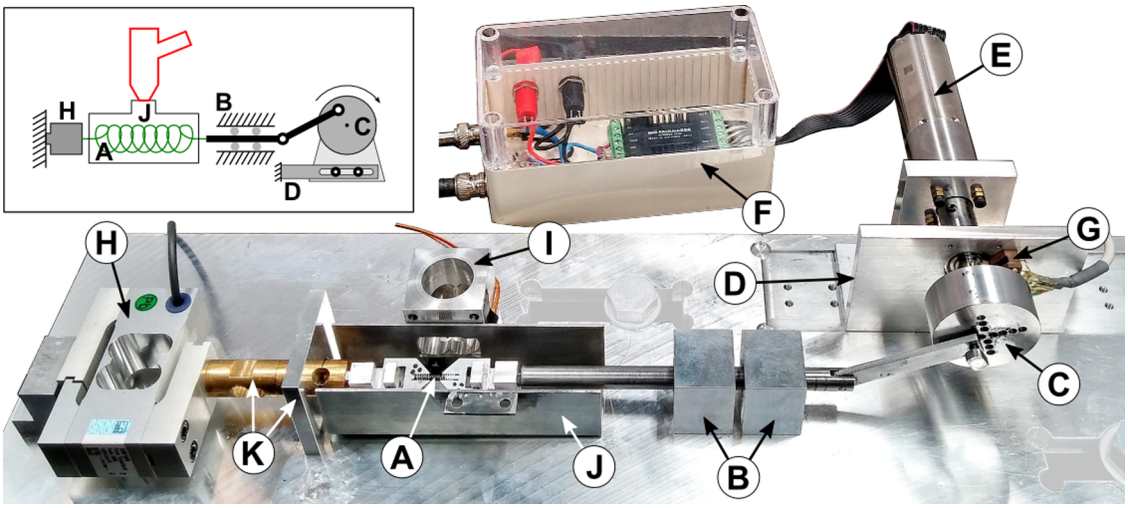

Figure 3. Tensile test-bench.
(A) SMA spring, (B) linear ball bearings holding the piston arm, (C) crank wheel controlling the deflection amplitude, (D) slider determining the pre-tension length, (E) brushless DC motor controlling speed, (F) motor controller, (G) encoder measuring crank wheel angle and speed, (H) force transducer measuring spring force, (I) heat gun inlet and thermocouple, (J) heat chamber (shown without its lid) and (K) heat wall and arm.

$$\mathbf{S}_i = \frac{1}{E_i}\begin{bmatrix} 1 & 0 \\ 0 & 2(1+\nu) \end{bmatrix} \quad (i = A, M)$$

$$\mathbf{\Lambda} = \begin{cases} \mathbf{\Lambda}_f = \frac{H}{\sqrt{\sigma^2 + 3\tau^2}} \begin{Bmatrix} \sigma \\ 3\tau \end{Bmatrix} & \dot{\xi} > 0 \\ \mathbf{\Lambda}_r = \frac{1}{\xi}\tilde{\boldsymbol{\varepsilon}}_t & \dot{\xi} < 0 \end{cases} \tag{5}$$

The linear thermal expansion is omitted because it is negligible compared to the phase transformation strain $\boldsymbol{\varepsilon}_t$. The compliance tensor (here matrix) is denoted $\mathbf{S}$ and it is a linear combination of the compliances of austenite and martensite, $\mathbf{S}_A$ and $\mathbf{S}_M$. The compliance matrices are composed by the elastic moduli E_i of martensite (subscript M) and austenite (subscript A) and by Poisson's ratio ν. The martensitic volume fraction is denoted as ξ, and it fulfils $0 \leq \xi \leq 1$. The phase transformation direction tensor (here vector) is $\mathbf{\Lambda}$. The forward phase transformation direction tensor, $\mathbf{\Lambda}_f$, depends on the maximum transformation strain H, which is considered constant in this investigation. The tilde ($\tilde{\boldsymbol{\varepsilon}}_t$, $\tilde{\xi}$) in the expression of the reverse phase transformation direction tensor, $\mathbf{\Lambda}_r$, indicates that the value should be taken at the latest point in time, where a change in phase transformation direction was encountered, that is, a turning point.

Two transformation functions, describing forward (Φ_f) and the reverse (Φ_r) transformations, are defined as follows

$$\Phi_f = (1-D)\mathbf{\Lambda}_f^{\mathrm{T}}\boldsymbol{\sigma} + \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{T}}\Delta\mathbf{S}\boldsymbol{\sigma} + \rho\Delta s_0 T - \rho\Delta u_0 - f_f\left(\zeta_f(\xi, \xi_0)\right) - Y \tag{6a}$$

$$\Phi_r = -(1+D)\mathbf{\Lambda}_r^{\mathrm{T}}\boldsymbol{\sigma} - \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{T}}\Delta\mathbf{S}\boldsymbol{\sigma} - \rho\Delta s_0 T + \rho\Delta u_0 + f_r(\zeta_r(\xi, \xi_0)) - Y \tag{6b}$$

It is assumed that the difference in specific heat capacity between martensite and austenite is negligible. This is a common engineering assumption (Lagoudas et al., 2012). The constants Y and D are related to the location of the transformation surfaces in the temperature–stress plane. The mass density is ρ, and Δs_0 and Δu_0 are the differences between austenite and martensite in the specific entropy and the specific internal energy at the reference state, respectively. The temperature is denoted T and the functions f_f and f_r are called hardening functions and together with the sub-loop functions ζ_f and ζ_r, they control the evolution of the martensitic volume fraction during transformation. These functions are modifications to the original model. The parameters in the preceding equations are related to common SMA parameters in the following way

$$\begin{aligned} \rho\Delta s_0 &= -\frac{2C_M C_A H}{C_M + C_A} \\ D &= \frac{C_M - C_A}{C_M + C_A} \\ a_1 &= -\rho\Delta s_0(M_s - M_f) \\ a_2 &= -\rho\Delta s_0(A_f - A_s) \\ a_3 &= -\frac{1}{20}a_1(3n_1^f - 3n_2^f + 5) + \frac{1}{20}a_2(3n_1^r - 3n_2^r + 5) \\ Y &= \frac{1}{2}\rho\Delta s_0(M_s - A_f) - a_3 \\ \rho\Delta u_0 &= \frac{1}{2}\rho\Delta s_0(M_s + A_f) \end{aligned} \tag{7}$$

Here, M_s and M_f are the start and finish temperatures of the martensitic (forward) transformation,

respectively; A_s and A_f are the start and finish temperatures of the austenitic (reverse) transformation, respectively; C_A and C_M are the slopes of the austenite and martensite transformation surfaces in the temperature–stress diagram at zero stress, respectively[1]; the parameters a_1, a_2, a_3, n_1^f, n_2^f, n_1^r and n_2^r are related to the hardening functions described in section 'Hardening function'. The expression for a_3 is specific for the Bézier hardening function (Enemark and Santos, 2015), but the remaining constants in equation (7) are identical to the expressions in the original model (Lagoudas et al., 2012).

At all times, the transformation functions should fulfil the two Kuhn–Tucker conditions

$$\dot{\xi} \geq 0, \quad \Phi_f \leq 0, \quad \Phi_f \dot{\xi} = 0 \tag{8a}$$

$$\dot{\xi} \leq 0, \quad \Phi_r \leq 0, \quad \Phi_r \dot{\xi} = 0 \tag{8b}$$

These conditions control whether or not phase transformations take place.

Energy equation

The SMA model can be coupled to the energy equation to explore the influence of latent heat due to phase transformations as described by Lagoudas et al. (2012). In this case, the material temperature T is not an independent variable, but it is determined via the energy equation. Disregarding the effect of linear thermal expansion and the differences in heat capacity between martensite and austenite, the reduced energy equation reads

$$\rho c_p \dot{T} + \chi \dot{\xi} = \rho \tilde{r} - \mathrm{div}(\mathbf{q})$$

where c_p is the specific heat capacity, $\tilde{r}$ is the rate of internal heat generation, $\mathbf{q}$ is the heat flux and the function χ is given by

$$\chi = -\boldsymbol{\Lambda}_d^{\mathrm{T}} \boldsymbol{\sigma} - \frac{1}{2} \boldsymbol{\sigma}^{\mathrm{T}} \Delta \mathbf{S} \boldsymbol{\sigma} + \rho \Delta u_0 + f_d(\zeta_d(\xi, \xi_0))$$

where subscript d indicates the forward (subscript f) or the reverse (subscript r) component depending on the sign of $\dot{\xi}$.

It is assumed that the SMA spring temperature is uniform, and that the heat conduction through the spring fixtures is negligible and much smaller than the heat convection. This assumption is reasonable because the interfaces between the spring and the spring fixtures are small, and because the spring fixtures are completely enclosed in the heat chambers and therefore have the same temperature as the surrounding air. The collected energy transfer, that is, $\rho\tilde{r} - div(\mathbf{q})$, is therefore assumed to have the form $-\hat{h}(A/V)(T - T_\infty)$, where $\hat{h}$ is the specific heat transfer coefficient, A/V is the ratio between surface area and volume of the spring wire and T_∞ is the ambient temperature. In the case of a helical spring, the surface-area-to-volume ratio is $A/V = ((2\pi cL)/(\pi c^2 L)) = 2/c$, where L is the spring wire length and c is the wire radius. The energy equation is rearranged

$$E = \rho c_p \dot{T} + \chi \dot{\xi} + \hat{h} \frac{2}{c} (T - T_\infty) = 0 \tag{9}$$

Hardening function

In this work, the cubic Bézier hardening function is used, which was first presented by Enemark and Santos (2015). The function is inspired by the smooth hardening function used in the original model (Lagoudas et al., 2012). Enemark and Santos (2015) showed that the cubic Bézier hardening resulted in better resemblance to the experiments of tensile tests of SMA wires.

The hardening function is designated by the cubic Bézier curve

$$\begin{Bmatrix} B_1(s) \\ B_2(s) \end{Bmatrix} = \begin{Bmatrix} -2(1-3\delta)s^3 + 3(1-3\delta)s^2 + 3\delta s \\ (3n_a + 3n_b - 2)s^3 + 3(1 - 2n_a - n_b)s^2 + 3n_a s \end{Bmatrix}$$

for $s \in [0, 1]$, and where $n_a, n_b \in (0, 1)$ are the curvature controlling parameters and $\delta \ll 1$ is an auxiliary variable used for numerical implementation. In this work, $\delta = 10^{-4}$ is used. The forward and reverse hardening functions are given by

$$f_f(\zeta_f) = a_1 f(\zeta_f) + a_3 \quad f_r(\zeta_r) = a_2 f(\zeta_r) - a_3 \tag{10}$$

where a_1, a_2 and a_3 are the constants, cf. equation (7), and

$$f(\zeta) = \{B_2(s) | B_1(s) = \zeta \wedge s \in [0, 1]\}; \quad \zeta \in [0, 1]$$

and ζ depends on the volume fraction of martensite ξ, and it is calculated using the sub-loop functions described in the next section. The equation $B_1(s) = \zeta$ is a third-order polynomial in s and the unique solution in the definition interval is

$$s = \frac{1}{2} - \sqrt{\frac{1-\delta}{1-3\delta}} \cos\left(\frac{1}{3} \arccos\left[\frac{1-2\zeta}{1-\delta} \sqrt{\frac{1-3\delta}{1-\delta}} \right] + \frac{\pi}{3} \right)$$

Since the forward and reverse transformations may evolve differently, the curvature controlling parameter set (n_a, n_b) is substituted by (n_1^f, n_2^f) in forward transformation and by (n_1^r, n_2^r) in reverse transformation. The inverse hardening function is

$$\zeta = f^{-1}(\tilde{f}) = \begin{cases} 0 & \text{for} \quad \tilde{f} < 0 \\ \{B_1(s) | B_2(s) = \tilde{y} \wedge s \in [0, 1]\} & \text{for} \quad 0 \leq \tilde{f} \leq 1 \\ 1 & \text{for} \quad 1 < \tilde{f} \end{cases} \tag{11}$$

where

$$\tilde{f} = \begin{cases} \frac{1}{a_1}\left[(1-D)\boldsymbol{\Lambda}_f^{\mathrm{T}}\boldsymbol{\sigma} + \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{T}}\Delta\mathbf{S}\boldsymbol{\sigma} + \rho\Delta s_0 T - \rho\Delta u_0 - Y - a_3\right] & \text{for } \dot{\xi}>0 \\ \frac{1}{a_2}\left[(1+D)\boldsymbol{\Lambda}_r^{\mathrm{T}}\boldsymbol{\sigma} + \frac{1}{2}\boldsymbol{\sigma}^{\mathrm{T}}\Delta\mathbf{S}\boldsymbol{\sigma} + \rho\Delta s_0 T - \rho\Delta u_0 + Y + a_3\right] & \text{for } \dot{\xi}<0 \end{cases} \tag{12}$$

The expression of $\tilde{f}$ is determined by equating the transformation function to 0 in equations (6) and then isolating the hardening function value. The inverse hardening function is used to determine the variable ζ if $\dot{\xi}$ changes sign, which is necessary in the numerical implementation. For further information about the cubic Bézier hardening and its inverse function f^{-1}, the reader is referred to Enemark and Santos (2015).

Sub-loop function

In order to better capture the behaviour of sub-loops or minor loops, it is necessary to introduce sub-loop functions to the constitutive model. The purpose of these functions is to alter the position of the transformation surfaces depending on where in the stress–temperature plane $\dot{\xi}$ changes sign or becomes 0 at the latest point in time (a turning point). This way the initialisation of phase transformation becomes smooth in the case of incomplete transformations (as long as the hardening function also has smooth transitions), which is not the case if sub-loop functions are not used. The Brinson model (Brinson, 1993) has sub-loop capability implemented by default, but this is not the case for the model by Lagoudas et al. (2012). However, the sub-loop behaviour was investigated and implemented in various ways by Bo and Lagoudas (1999) in an earlier version of the same model. This work uses sub-loop functions similar to the ones proposed by Enemark et al. (2014), which consist of polynomials. Enemark et al. modified the sub-loop function in the Brinson model to better capture the experimental behaviour of an SMA spring. The reader is referred to Enemark et al. (2014) and Enemark and Santos (2015) for additional information on this family of sub-loop functions.

The sub-loop function determines the variable ζ according to the direction of transformation depending on the martensitic volume fraction ξ and an auxiliary variable ξ_0, which is related to ξ at the latest turning point. In forward transformation, ζ increases monotonically in the interval $[0,1]$ as a function of $\xi \in [\xi_0, 1]$. A simple sub-loop function uses a linear relationship, but in order to control the sub-loop behaviour, more advanced functions have to be used. Here, a quadratic Bézier curve is proposed, and it is defined by the three consecutive points

$$Q_1 = \begin{Bmatrix} \xi_0 \\ 0 \end{Bmatrix}$$

$$Q_2 = \begin{Bmatrix} 1 \\ 1-\xi_0 \end{Bmatrix} + \frac{1}{2}(p+1)\left(\begin{Bmatrix} \xi_0 \\ \xi_0 \end{Bmatrix} - \begin{Bmatrix} 1 \\ 1-\xi_0 \end{Bmatrix}\right)$$

$$Q_3 = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix}$$

where $p \in (-1, 1)$ is a sub-loop controlling parameter. The end points are trivial, but Q_2 is constructed in such a way that the width of the sub-loop increases with increasing p. For $p = 0$, the curve becomes a straight line. The resulting Bézier curve is

$$\begin{Bmatrix} C_{f1}(s) \\ C_{f2}(s) \end{Bmatrix} = (1-s)^2 Q_1 + 2(1-s)sQ_2 + s^2 Q_3 = \begin{Bmatrix} (1-\xi_0)ps^2 + (1-\xi_0)(1-p)s + \xi_0 \\ (1-2\xi_0)ps^2 + (1+2\xi_0 p - p)s \end{Bmatrix}$$

for $s \in [0,1]$. The actual sub-loop function for forward transformation is

$$\zeta_f(\xi, \xi_0) = \left\{C_{f2}(s) \middle| C_{f1}(s) = \xi \wedge s \in [0,1]\right\}, \quad \xi \in [\xi_0, 1], \xi_0 \in [0,1) \tag{13}$$

It is necessary to solve a second-order polynomial to determine s and evaluate ζ_f. The solution is unique in the definition interval $s \in [0,1]$.

Similarly, the sub-loop function during reverse transformation is a monotonically increasing function in the interval $[0,1]$ for $\xi \in [0, \xi_0]$. The proposed quadratic Bézier function is defined by the successive points

$$Q_4 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

$$Q_5 = \begin{Bmatrix} 1 \\ 1-\xi_0 \end{Bmatrix} + \frac{1}{2}(p+1)\left(\begin{Bmatrix} \xi_0 \\ \xi_0 \end{Bmatrix} - \begin{Bmatrix} 0 \\ 1-\xi_0 \end{Bmatrix}\right)$$

$$Q_6 = \begin{Bmatrix} \xi_0 \\ 1 \end{Bmatrix}$$

which result in the curve

$$\begin{Bmatrix} C_{r1}(s) \\ C_{r2}(s) \end{Bmatrix} = (1-s)^2 Q_4 + 2(1-s)sQ_5 + s^2 Q_6 = \begin{Bmatrix} -\xi_0 ps^2 + \xi_0(1+p)s \\ (1-2\xi_0)ps^2 + (1+2\xi_0 p - p)s \end{Bmatrix}$$

for $s \in [0,1]$. The sub-loop function for reverse transformation becomes

$$\zeta_r(\xi, \xi_0) = \{C_{r2}(s) | C_{r1}(s) = \xi \wedge s \in [0,1]\}, \quad \xi \in [0, \xi_0], \xi_0 \in (0,1] \tag{14}$$

Evaluation of ζ_r also requires solution of a second-order polynomial, and the solution is unique in the definition interval.

Algorithm 1. Algorithm for the inverse function is given in equation (15a).

if $\zeta_f = 0$ **then**
  $\xi_0 \leftarrow \xi$;
else if $\zeta_f \geq \xi$ **then**
  $\xi_0 \leftarrow 0$;
else if $p \neq 0$ **then**
  $p_1 \leftarrow p(1 + \zeta_f - 2\xi) - 1$;
  $p_2 \leftarrow \frac{1}{3} - \frac{p_1}{p^2}$;
  $p_3 \leftarrow \frac{p_1 + 3\zeta_f}{3p^2} - \frac{2}{27}$;
  $s \leftarrow \frac{1}{3} + 2\sqrt{\frac{p_2}{3}}\cos\left[\frac{1}{3}\arccos\left(\frac{3p_3}{2p_2}\sqrt{\frac{3}{p_2}}\right) + \frac{\pi}{3}\right]$;
  $\xi_0 \leftarrow \frac{\xi - \zeta_f}{(1 - ps)(1 - s)}$;
else
  $\xi_0 \leftarrow \frac{\xi - \zeta_f}{1 - \zeta_f}$;
end if

Algorithm 2. Algorithm for the inverse function is given in equation (15b).

if $\zeta_r = 1$ **then**
  $\xi_0 \leftarrow \xi$;
else if $\zeta_r \leqslant \xi$ **then**
  $\xi_0 \leftarrow 1$;
else if $p \neq 0$ **then**
  $p_1 \leftarrow p(2\xi - \zeta_r) - 1$;
  $p_2 \leftarrow \frac{1}{3} - \frac{p_1}{p^2}$;
  $p_3 \leftarrow \frac{p_1 - 3\zeta_r + 3}{3p^2} - \frac{2}{27}$;
  $s \leftarrow \frac{2}{3} - 2\sqrt{\frac{p_2}{3}}\cos\left[\frac{1}{3}\arccos\left(\frac{3p_3}{2p_2}\sqrt{\frac{3}{p_2}}\right) + \frac{\pi}{3}\right]$;
  $\xi_0 \leftarrow 1 + \frac{\xi - \zeta_r}{1 - (1 + ps)(1 - s)}$;
else
  $\xi_0 \leftarrow \frac{\xi}{\zeta_r}$;
end if

The inverse with respect to ξ_0 of the sub-loop functions has to be used in turning points. For forward and reverse transformations, they are

$$\xi_0^f(\zeta_f, \xi) = \begin{cases} 0 & \text{for} \quad \zeta_f \in [\xi, 1], \xi \in [0, 1) \\ \{\xi_0 | C_{f1}(s) = \xi \wedge C_{f2}(s) = \zeta_f \wedge s \in [0, 1]\} & \\ & \text{for} \quad \zeta_f \in [0, \xi), \xi \in [0, 1) \end{cases} \tag{15a}$$

$$\xi_0^r(\zeta_r, \xi) = \begin{cases} 1 & \text{for} \quad \zeta_r \in [0, \xi], \xi \in (0, 1] \\ \{\xi_0 | C_{r1}(s) = \xi \wedge C_{r2}(s) = \zeta_r \wedge s \in [0, 1]\} & \\ & \text{for} \quad \zeta_r \in (\xi, 1], \xi \in (0, 1] \end{cases} \tag{15b}$$

Both inverse functions have unique solutions, but solution of third-order polynomials in s is required; see Algorithms 1 and 2.

Numerical implementation

The numerical solution closely follows the one provided by Lagoudas et al. (2012). There are two additions, which are described in the following sections. The implementation is independent of the number of spatial dimensions, that is, 1D, 2D or three-dimensional (3D).

Coupled constitutive and energy equations. In every time step n, the strain $\boldsymbol{\varepsilon}_n$ is considered known, and the constitutive equation (3) and the energy equation (9) are solved simultaneously in two steps. Also, the variable ξ_0 is known and constant as long as $\dot{\xi}$ does not change sign, and this issue is treated in the next section. The energy equation (9) is written in a backward Euler sense, that is

$$E = \rho c_p \frac{T_n - T_{n-1}}{\Delta t} + \chi_n \frac{\xi_n - \xi_{n-1}}{\Delta t} + \hat{h}\frac{2}{c}(T_n - T_\infty) = 0 \tag{16}$$

where $\Delta t = t_n - t_{n-1}$ is the time difference between the time steps.

The first step is a thermo-elastic prediction, where it is assumed that $\dot{\xi} = 0$ resulting in $\xi_n = \xi_{n-1}$ and $\boldsymbol{\varepsilon}_{t_n} = \boldsymbol{\varepsilon}_{t_{n-1}}$. This means that it is straightforward to calculate $\boldsymbol{\sigma}_n$ and T_n from equations (3) and (19).

Subsequently, it is tested, whether the Kuhn–Tucker conditions (8) are satisfied. If they are not, a phase transformation is taking place and a transformation correction step has to be made to $\boldsymbol{\sigma}$, T and ξ. This is done iteratively using a return mapping algorithm, which basically is consecutive Newton step, because E has to be 0 at all times and Φ has to be 0 during transformation

$$\Phi + \partial_{\boldsymbol{\sigma}}\Phi\Delta\boldsymbol{\sigma} + \partial_\xi\Phi\Delta\xi + \partial_T\Phi\Delta T = 0 \tag{17a}$$

$$E + \partial_{\boldsymbol{\sigma}}E\Delta\boldsymbol{\sigma} + \partial_\xi E\Delta\xi + \partial_T E\Delta T = 0 \tag{17b}$$

$$\mathbf{S}\Delta\boldsymbol{\sigma} + (\Delta\mathbf{S}\boldsymbol{\sigma} + \boldsymbol{\Lambda})\Delta\xi = 0 \tag{17c}$$

where $\partial_y x$ denotes the partial derivative of x with respect to y. Equation (17c) comes from equation (3), where $\Delta\boldsymbol{\varepsilon} = \mathbf{0}$ in the correction step. The iterative step in ξ is found by solving the system of equations

$$\Delta\xi = \frac{(E\partial_T\Phi - \Phi\partial_T E)}{[\partial_T E\partial_\xi\Phi - \partial_\xi E\partial_T\Phi + (\partial_{\boldsymbol{\sigma}}E\partial_T\Phi - \partial_T E\partial_{\boldsymbol{\sigma}}\Phi)\mathbf{S}^{-1}(\Delta\mathbf{S}\boldsymbol{\sigma} + \boldsymbol{\Lambda})]} \tag{18}$$

This means that ξ is updated by $\xi_n^{(i+1)} \leftarrow \xi_n^{(i)} + \Delta\xi$, where i denotes the iteration. The boundaries $0 \leq \xi \leq 1$ have to be enforced when updating ξ. Next, $\boldsymbol{\sigma}$, T, Φ and E are updated, and iterations continue until convergence. It should be noted that the Kuhn–Tucker conditions specify that the transformation functions always should be below or equal to 0 only in the case that transformation in the appertaining direction is admissible and the relevant transformation function is defined. For example, this means that the computed value of Φ_f will exceed 0 when ξ becomes 1. Therefore, iterations during the forward phase transformation correction should be stopped if ξ persistently encounters its upper boundary, even though $|\Phi_f|$ does not meet its tolerances. This is similar in reverse transformation for $\xi = 0$.

Turning point. The aspect of turning points, that is, when $\dot{\xi}$ changes sign, is two-parted: detecting a possible turning point and updating and calculation of ξ_0 in the case of a turning point. The value of ξ_0 is constant (and known) as long as the sign of $\dot{\xi}$ is constant.

The parameter ξ_0 enters into the expression of Φ. The value of Φ determines whether or not a transformation is taking place and it therefore determines the sign of $\dot{\xi}$. This means that it is somewhat a circular problem to determine, whether there is a turning point. However, this is solved by setting up some rules:

1. The transformation indicator d_n at time step n is defined by

$$d_n \equiv \begin{cases} \text{sign}(\xi_n - \xi_{n-1}) & \text{if} \quad \xi_n \neq \xi_{n-1} \\ 1 & \text{else if} \quad \xi_n = 0 \\ -1 & \text{else if} \quad \xi_n = 1 \\ -d_{n-1} & \text{otherwise} \end{cases}$$

   Forward transformation corresponds to $d_n = 1$, and reverse to $d_n = -1$.
2. Initially, the preferred direction of transformation $\tilde{d}_n$ between time step $n-1$ and n is given by

$$\tilde{d}_n \equiv \begin{cases} 1 & \text{if} \quad \xi_{n-1} = 0 \\ -1 & \text{else if} \quad \xi_{n-1} = 1 \\ d_{n-1} & \text{otherwise} \end{cases}$$

3. If $\tilde{d}_n \neq d_{n-1}$, then there is a turning point in time step $n-1$ and ξ_0 has to be calculated at time step n.
4. If the Kuhn–Tucker condition corresponding to the preferred direction of transformation $\tilde{d}_n$ at time step n is satisfied after the thermo-elastic prediction, a turning point takes place at time step $n-1$. In this case, $\tilde{d}_n \leftarrow -d_{n-1}$ and ξ_0 have to be calculated at time step n.

For example, if a forward transformation takes place between time step $n-2$ and $n-1$, Φ_f is calculated at first in time step n. Note that Φ_r is not defined before ξ_0 is calculated, and therefore, Φ_r is not calculated. Then, it should be checked if the Kuhn–Tucker condition related to forward transformation, equation (8a), is violated in time step n. If so, it is not a turning point, the transformation continues in the same direction and ξ_0 is unchanged. If on the other hand, the Kuhn–Tucker condition is satisfied, there is a turning point at time step $n-1$, ξ_0 has to be calculated, and now reverse transformation is the preferred direction of transformation. If there is no phase transformation taking place, d_n changes sign and ξ_0 is calculated in each time step. When $\xi_{n-1} = 0$ or $\xi_{n-1} = 1$, the situation is trivial, because then only one transformation direction is admissible and furthermore $\xi_0 = \xi_{n-1}$.

If time step $n-1$ is a turning point (which is detected in time step n), the transformation surface of the new preferred direction is moved to the temperature–stress state at time step n if possible. This means that $\Phi_d = 0$ has to be solved for ξ_0 for the values of $\boldsymbol{\sigma}$, T and ξ at time step $n-1$ using the new preferred direction. If there is no solution, then ξ_0 becomes ξ_{n-1}. This special case is included by the expressions of the inverse functions, equations (11), (12) and (15). Note that the procedure of detecting a turning point and resetting ξ_0 is not a correction to the former time step. It is a way to ensure continuity at time step n. By moving the transformation surface, smooth transitions are ensured, because the hardening function is smooth at the boundaries, $\zeta = 0$ and $\zeta = 1$. From time step $n-1$, all state variables are known, and therefore, it is possible to determine $\tilde{f} = f_d(\zeta_d)$ using the values of $\boldsymbol{\sigma}_{n-1}$ and T_{n-1} using equation (12). Then the inverse hardening function, equation (11), is used for determining ζ_d. Finally, ξ_0 is found using the inverse sub-loop function in equation (15) by inserting ζ_d and ξ_{n-1}. The overall numerical implementation is summed up in Algorithm 3.

Helical spring model

As mentioned in the introduction, there exist a variety of model approaches for SMA helical springs in the literature. Equivalent 1D models generally provide acceptable performance (Aguiar et al., 2010; An et al., 2012; Enemark et al., 2014), and they are advantageous because of their simplicity. The idea behind the 1D models is to choose a point in the spring wire cross section to govern the SMA behaviour and then extrapolate to obtain the global spring behaviour. Therefore, also formulas relating the spring deformation to the representative SMA strain and the spring force to the representative SMA stress are needed. The disadvantage of most presented 1D models is that they do not take into account large global deformations (large pitch angles) resulting in geometrical stiffening. If the pitch

Algorithm 3. Algorithm for using the constitutive model by Lagoudas et al. (2012) coupled to the energy equation and with the Bézier hardening function and sub-loop function.

1. Load ξ_{n-1}, $\xi_{0_{n-1}}$, ε_{n-1}, $\varepsilon_{t_{n-1}}$, T_{n-1} and d_{n-1} from the former time step and ε_n from the current time step.
2. Make a thermo-elastic prediction:
   (a) $\xi_n \leftarrow \xi_{n-1}$, $\xi_{0_n} \leftarrow \xi_{0_{n-1}}$, $\varepsilon_{t_n} \leftarrow \varepsilon_{t_{n-1}}$.
   (b) Calculate $\boldsymbol{\sigma}_n$ (3) and T_n (16).
3. If $\xi_n = 0$ then $\tilde{d}_n \leftarrow 1$, else if $\xi_n = 1$ then $\tilde{d}_n \leftarrow -1$, otherwise $\tilde{d}_n \leftarrow d_{n-1}$.
4. Calculate ζ_d (13, 14), f_d (10), $\boldsymbol{\Lambda}_d$ (5) and Φ_d (6) using $d = \tilde{d}_n$.
5. If both $\Phi_d < 0$ and $-\tilde{d}_n$ is admissible† do (**turning point**):
   (a) $\tilde{d}_n \leftarrow -d_{n-1}$ and calculate $\boldsymbol{\sigma}_{n-1}$ (3).
   (b) For time step $n-1$, calculate $\boldsymbol{\Lambda}_d$ (5), $\tilde{f}$ (12), $\zeta_d = \zeta$ (11) and ξ_0 (15) using $d = \tilde{d}_n$, $\boldsymbol{\sigma}_{n-1}$ and T_{n-1}.
   (c) For time step n, recalculate ζ_d (13, 14), f_d (10), $\boldsymbol{\Lambda}_d$ (5) and Φ_d (6) using $d = \tilde{d}_n$ and the new value of ξ_0.
6. $d_n \leftarrow \tilde{d}_n$.
7. If $\Phi_d > 0$ do (**transformation correction**):
   (a) Calculate the derivatives of Φ_d and E_d with respect to ξ_n, T_n and $\boldsymbol{\sigma}_n$.
   (b) Calculate $\Delta\xi$ (18) and enforce its boundaries. Update $\xi_n^{(i)} \leftarrow \Delta\xi + \xi_n^{(i-1)}$ and $\boldsymbol{\varepsilon}_{t_n}^{(i)} \leftarrow \boldsymbol{\Lambda}_d^{(i-1)}\Delta\xi + \boldsymbol{\varepsilon}_{t_n}^{(i-1)}$.
   (c) Calculate T_n (16), and $\boldsymbol{\sigma}_n$ (3).
   (d) Calculate ζ_d (13, 14), f_d (10), $\boldsymbol{\Lambda}_d$ (5), Φ_d (6) and E_d (16).
   (e) If both $|\Phi_d| < tol_1$ and $|E_d| < tol_2$ the solution has converged. Otherwise go to item 7a.

† $\tilde{d}_n = 1$ is admissible only if $\xi_n < 1$, and $\tilde{d}_n = -1$ is admissible only if $\xi_n > 0$.

angle is large, bending of the spring wire becomes significant (Wahl, 1944), and therefore, the stress and strain states become 2D. An et al. (2012) also considered large pitch angles of SMA springs but made an equivalent 1D model based on several simplifying assumptions.

A 2D model is presented here, in which a single point in the cross section is used to govern the SMA behaviour, as described above. If the spring index (ratio between coil and wire radii) is small, there is another phenomenon taking place, that is, distortion of the otherwise symmetric strain distribution in the wire cross section, because the spring wire is equivalent to a curved bar. This effect was taken into account for an SMA spring by Mirzaeifar et al. (2011), but it was concluded that it did not have a significant impact on the force–deflection relationship. For linear elastic springs, Wahl (1944) showed that the error of the force–deflection relationship is less than 3% when not correcting for the distortion for spring indexes as low as 2.7. Also, end effects are not considered in this model, and in practice they have been avoided (to a high extent) by the type of spring grip.

The SMA spring is a helix with initial coil radius r_0, assumed constant wire radius c, initial height h_0, N the number of windings and assumed constant length $L = \sqrt{h_0^2 + (2\pi N r_0)^2}$, see Figure 1. The initial pitch angle is then $\alpha_0 = \arctan(h_0/(2\pi N r_0))$. It is assumed that the spring is restricted from rotating, which results in the condition

$$\frac{r}{r_0} = \frac{\cos\alpha}{\cos\alpha_0}$$

where r and α denote the coil radius and pitch angle of the deformed spring, respectively. This relation is used to simplify the expressions below. The deformed pitch angle is given by

$$\alpha = \arcsin\left(\frac{u}{L} + \sin\alpha_0\right)$$

where u is the spring elongation. The shear and normal strain distributions in the wire cross section are (Wahl, 1944)

$$\begin{aligned}\gamma(a) &= a\left(\frac{\sin\alpha\cos\alpha}{r} - \frac{\sin\alpha_0\cos\alpha_0}{r_0}\right)\\ &= \frac{a}{r_0}\cos\alpha_0(\sin\alpha - \sin\alpha_0)\end{aligned} \tag{19}$$

$$\begin{aligned}\varepsilon(y) &= y\left(\frac{\cos^2\alpha_0}{r_0} - \frac{\cos^2\alpha}{r}\right)\\ &= \frac{y}{r_0}\cos\alpha_0(\cos\alpha_0 - \cos\alpha)\end{aligned} \tag{20}$$

where $a \in [0, c]$ is a radial coordinate, and $y \in [-c, c]$ is a Cartesian coordinate perpendicular to the spring centre axis. These strain expressions are symmetric, and the distortion due to the fact that the spring wire is a curved bar is not considered as mentioned above. The strain distributions are illustrated in Figure 1.

The tensile spring force F comes from the torsional moment M_T and the bending moment M_B in the spring, cf. Figure 1

$$\begin{aligned}F &= M_T\frac{\cos\alpha}{r} + M_B\frac{\sin\alpha}{r}\\ &= \frac{\cos\alpha_0}{r_0}(M_T + M_B\tan\alpha)\end{aligned} \tag{21}$$

$$M_T = \int_0^c\int_{-\pi}^{\pi}\tau(\varepsilon(\theta, a))a^2\,\mathrm{d}\theta\,\mathrm{d}a \tag{22}$$

$$M_B = \int_0^c \int_{-\pi}^{\pi} \sigma(\varepsilon(\theta, a))a^2 \sin\theta \mathrm{d}\theta \mathrm{d}a \qquad (23)$$

Here, θ is the angular coordinate corresponding to the radial coordinate a. Note the geometrical stiffening effect on F due to the tan α factor.

It is not straightforward to evaluate the integrals in the expressions of the moments above, because the martensitic volume fraction varies through the wire cross sections and it couples the shear and normal stresses. Nevertheless, a constant martensitic volume fraction throughout the cross section is approximated, which is also the basis of the equivalent 1D models found in the literature, including Liang and Rogers (1993), Aguiar et al. (2010), An et al. (2012) and Enemark et al. (2014). This approximation is justified in the next section. Under shear behaviour, the constitutive equation reads $\tau = G(\gamma - \gamma_t)$ (equation (3)). From equation (19), we have that $\gamma(a) = (a/c)\gamma(c)$. The shear modulus G and the transformation shear strain γ_t are assumed constant in the cross section. By evaluating the integral in equation (22), the torsional moment becomes

$$M_T = \frac{2}{3}\pi c^3 G\left(\frac{3}{4}\gamma(c) - \gamma_t\right) = \frac{2}{3}\pi c^3 G(\gamma^* - \gamma_t) = \frac{2}{3}\pi c^3 \tau^*$$

Here, $\gamma^* \equiv (3/4)\gamma(c) = \gamma((3/4)c)$ is defined in order to use the governing equation and further define $\tau^* \equiv G(\gamma^* - \gamma_t)$. This means that the location $a^* = (3/4)c$ is representative for determining an equivalent constant martensitic volume fraction. To our understanding, the maximum shear strain and stress, corresponding to a point at the circumference ($a^* = c$), are used to govern the SMA behaviour in the most equivalent 1D models found in the literature. The difference is significant as will be shown in the following section.

A similar approach is made for the normal strain and stresses, where $\varepsilon(y) = (y/c)\varepsilon(c)$ and $\sigma = E(\varepsilon - \mathrm{sign}(\varepsilon)\varepsilon_t)$ (equation (3)).[2] The elastic modulus E and the transformation normal strain ε_t are assumed constant in the cross section. The bending moment becomes

$$M_B = \frac{4}{3}c^3 E\left(\frac{3\pi}{16}\varepsilon(c) - \varepsilon_t\right) = \frac{4}{3}c^3 E(\varepsilon^* - \varepsilon_t) = \frac{4}{3}c^3\sigma^*$$

This results in $\varepsilon^* \equiv (3\pi/16)\varepsilon(c) = \varepsilon((3\pi/16)c)$, $\sigma^* \equiv E(\varepsilon^* - \varepsilon_t)$ and $y^* \equiv (3\pi/16)c$.

The resulting force, equation (21), becomes

$$F = \frac{2}{3}\frac{c^3}{r_0}\cos\alpha_0(\pi\tau^* + 2\sigma^* \tan\alpha) \qquad (24)$$

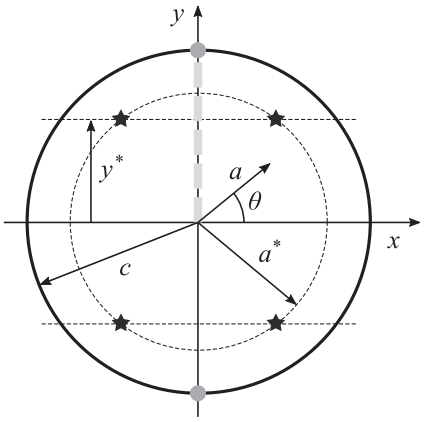

Figure 4. Illustration of different approaches to integrate the SMA behaviour in the wire cross section. The 2S approach uses the four star points to represent the SMA behaviour in the cross section, the similar 2C approach instead uses the two circumference points (solid circles) and the 2L approach integrates numerically the behaviour along the thick dashed line and extrapolates to the whole cross section.

To determine the spring force at a given deflection u, first the representative strain tensor $\boldsymbol{\varepsilon}^* = \{\varepsilon^*, \gamma^*\}^{\mathrm{T}} = \{\varepsilon(3\pi/16)c, \gamma((3/4)c)\}^{\mathrm{T}}$ is determined using equations (19) and (20). Then the representative stress tensor $\boldsymbol{\sigma}^* = \{\sigma^*, \tau^*\}^{\mathrm{T}}$ is determined using the SMA model and $\boldsymbol{\varepsilon}^*$. Finally, the force is calculated using equation (24).

Justification of modelling approach

In order to justify the approach for modelling the SMA helical spring, some comparisons to other approaches are provided. The different approaches are illustrated in Figure 4 and they are as follows:

- 2S The 2D model using the *star point*: both shear and normal strains are taken into account, and the four points represented by $a^* = (3/4)c$ and $y^* = (3\pi/16)c$ are used to govern the SMA behaviour.
- 2C The 2D model using the point of maximum strain (the *circumference point*): both shear and normal strains are taken into account, and the two points represented by $a^* = c$ and $y^* = c$ are used to govern the SMA behaviour.
- 2L The 2D model using *line* integration: shear and normal stresses are evaluated at integration points along a line from the wire cross section centre to the circumference at maximum normal strain as shown in Figure 4. The stresses are then extrapolated to the whole cross section and integrated numerically to evaluate the moments

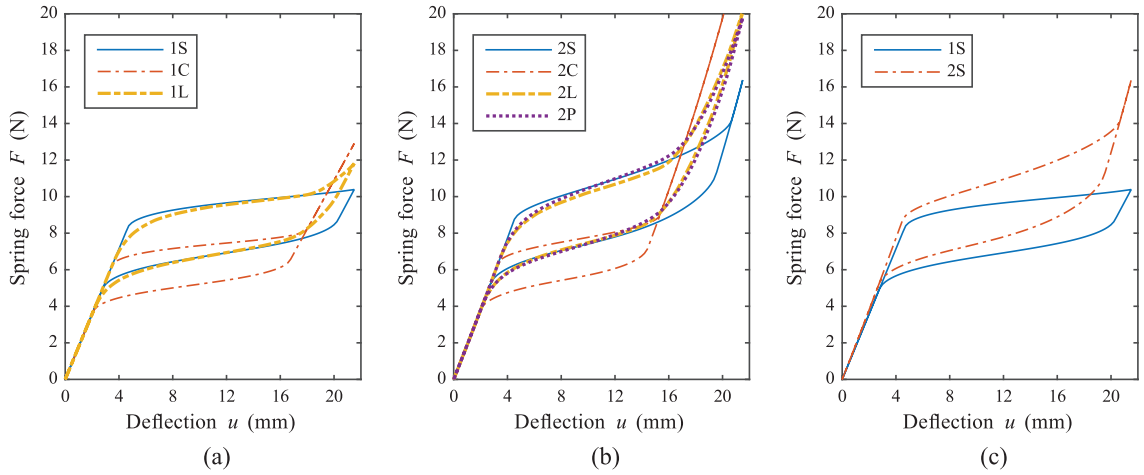

Figure 5. Force–displacement relationship for different model approaches when using the same spring and material properties. (a) One and (b) two dimensional model approaches and (c) comparison between the two.

$$M_{T,\text{line}} = 2\pi \int_0^c \tau(\varepsilon(y)) y^2 \mathrm{d}y$$

$$M_{B,\text{line}} = 4 \int_0^c \sigma(\varepsilon(y)) y \sqrt{c^2 - y^2} \mathrm{d}y$$

In this case, the shear stress is simplified to be independent of the angular position θ like the shear strain, and the normal stress is simplified to be independent of the horizontal position x like the normal strain. The stresses σ and τ are coupled due to the constitutive model and the 2D strain field.

2P The 2D model using *plane* integration: shear and normal stresses are evaluated at integration points spread out over the entire wire cross section, and the moments in equations (22) and (23) are evaluated numerically. This is the most accurate model approach, and also the most computationally demanding.

Using the same notation, the approaches 1S, 1C and 1L are also included, where normal strains (and therefore also normal stresses) are neglected. The 1L and 1P approaches give identical results, and 1P is therefore omitted. Simulated force–displacement tests are shown in Figure 5 for isothermal conditions. Generic SMA material properties are used, and the geometrical properties of the spring are given in Table 1. From Figure 5, it is evident that 1S and 2S perform very acceptably if deflections are of a moderate level (here less than 16 mm) when comparing to the precise 1L and 2P models. At large deformations, the errors become pronounced. On the other hand, the transformation surfaces are misplaced in the 1C and 2C approaches resulting in significant errors. There are also clear discrepancies between the 1S and 2S approaches as shown in Figure 5(c) even for relatively small deflections indicating that bending of the spring is an important aspect. The results for the 2L and 2P approaches are very similar for all levels of deflection. The chosen values of the elastic moduli for martensite and austenite are 30 and 40 GPa, respectively. Even so, the 2D models predict that the spring stiffness is highest when the SMA is in the martensitic phase. This is caused by geometric non-linearities, which have also been observed in the experiments (Enemark et al., 2014; Sakuma and Suzuki, 2007; Savi et al., 2015).

Table 1. Geometrical properties of SMA spring.

c (mm)	r_0 (mm)	h_0 (mm)	N (–)
0.28	1.28	7.0	6

From this example, it is clear that it is possible to significantly increase the predictability using the star points instead of the circumference points and the 2D model instead of the 1D model. The 2L model and especially the 2P model are considerably heavier computationally than the 2S model and the difference is small if the spring is subjected to moderate deflections. Also, if the SMA model is coupled to the energy equation, the 2L and 2P approaches become significantly more complicated, because heat conduction terms have to be added, which connect the integration points in the cross section.

It is possible to almost identically replicate the 2P model results using even the 1C model, if the model parameter values are changed. In this regard, it should be emphasised that many model parameters would obtain unrealistic values, and it is therefore not possible to give reliable predictions, if the unrealistic values are

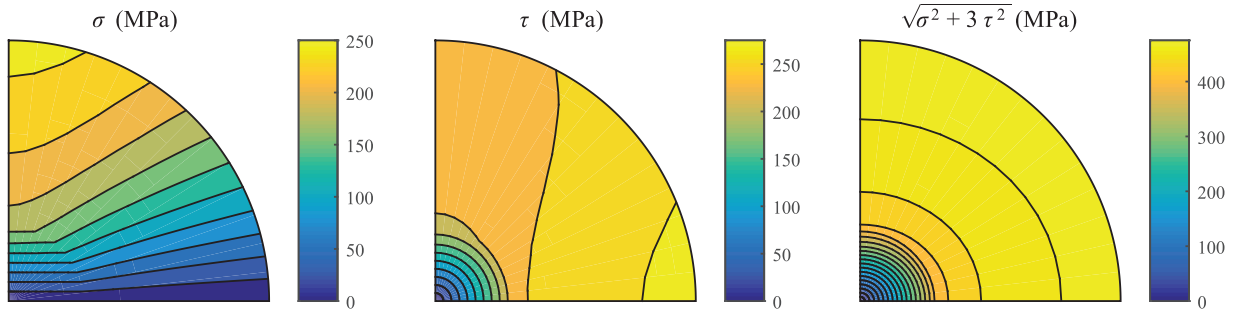

Figure 6. Normal, shear and von Mises stresses in a quarter of the cross section for a deflection of 12 mm. The intervals between the contour lines are 25 MPa.

not known a priori. Using the 2S model, only H and E_M have to be changed to increase predictability at large deformations, that is, up to a 40% decrease from their initial values in this case. But as long as deformations are moderate, parameter tuning of this level should not be necessary.

The importance of including the normal stress and the coupling between the two directions is also illustrated in Figure 6. Here, the normal, shear and von Mises stresses are shown in a quarter of the cross section for a deflection of 12 mm. The results are based on the 2P model. The magnitude of the normal stresses is comparable to the shear stresses. It is also clear that the normal stress varies with the horizontal x coordinate, even though the normal strain is independent of x. Similarly, the shear stress is not symmetric around the cross section centre even though the shear strain is. This is because of the coupling of the stresses, σ and τ, in the transformation functions, Φ_f and Φ_r. Nevertheless, the 2L model results are almost identical to the 2P results in Figure 5, so the approximation that the stress distributions have same symmetric properties as the strains can indeed be used without significant loss of accuracy. The tension–torsion coupling of SMAs is more thoroughly investigated by Lagoudas et al. (2012) and Mehrabi et al. (2014) for example.

Complex modulus representation

In order to use SMAs in dynamic systems, it is convenient to explore the dynamic properties in terms of the complex modulus. There exist several ways to quantify the complex modulus from either experiments or simulations (Gandhi and Wolons, 1999; Holanda et al., 2014; Malecot et al., 2006; Piedboeuf et al., 1998), and they result in slightly different values. The method used here is very similar to the one presented by Gandhi and Wolons (1999), but here the complex modulus is evaluated in terms of forces instead of stresses. This means the storage modulus, K, is a measure of average stiffness (measured in N m^{-1}) during a loading cycle. The loss factor is called η, and it is equivalent to linear viscous damping b by $\eta = (b\omega/K)$, where ω is the frequency of oscillation. The loss modulus equals $K\eta$, and the actual complex modulus is $K(1 + \mathrm{j}\eta)$, where j is the imaginary unit. The storage modulus and the loss factor are determined by assuming that the response of the dynamic system, in which the SMA spring is the active element, is harmonic: $x(t) = x_0 + A\sin(\omega t + \phi)$, where x_0 denotes the pre-tension length, A is the amplitude and ϕ is the phase. A response $\hat{x}(t)$ is defined by being equal to $x(t)$, but delayed by a phase angle of 90°, that is, $\hat{x}(t) = x_0 + A\sin(\omega t - (\pi/2) + \phi) = x_0 - A\cos(\omega t + \phi)$. The size of the equivalent force from the SMA element has the form $\tilde{F} = F_0 + K(x - x_0) + b\dot{x} = F_0 + K(x - x_0) - K\eta(\hat{x} - x_0)$, similar to a combined linear spring and a viscous or hysteretic damper plus a pre-tension force F_0. It is possible to extract the complex modulus by the two following integrals

$$I_1 = \oint F\mathrm{d}x = \int_{-\pi/\omega}^{\pi/\omega} F\dot{x}\mathrm{d}t \approx \int_{-\pi/\omega}^{\pi/\omega} \tilde{F}\dot{x}\mathrm{d}t = \pi A^2 K\eta$$

$$I_2 = \oint F\mathrm{d}\hat{x} = \int_{-\pi/\omega}^{\pi/\omega} F\dot{\hat{x}}\mathrm{d}t \approx \int_{-\pi/\omega}^{\pi/\omega} \tilde{F}\dot{\hat{x}}\mathrm{d}t = \pi A^2 K$$

The integral $\oint F\mathrm{d}x$ is the amount of energy dissipated during a cycle. Subsequently, the storage modulus and the loss factor are

$$K = \frac{I_2}{\pi A^2}, \quad \eta = \frac{I_1}{I_2}$$

The pre-tension force F_0 is determined by

$$\frac{\omega}{2\pi}\int_{-\pi/\omega}^{\pi/\omega} F\mathrm{d}t \approx \frac{\omega}{2\pi}\int_{-\pi/\omega}^{\pi/\omega} \tilde{F}\mathrm{d}t = F_0$$

Model parameter estimation and uncertainty quantification

The Bayesian framework is used to estimate the model parameters and their uncertainties based on

comparison with the experiments. This basically means that the parameters are not considered as scalars but as probability distributions (which can be characterised by mean and variance, for example). Namely, we use an MCMC method and an adaptive Metropolis algorithm to determine the statistical properties of the model parameters (Gelman et al., 2014).

The basis of the Bayesian approach is Bayes' rule

$$P(\theta|\mathbf{y}) = \frac{P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta})}{\int P(\boldsymbol{\theta})P(\mathbf{y}|\boldsymbol{\theta})d\boldsymbol{\theta}}$$

which uses a prior (known) distribution of the model parameters $\boldsymbol{\theta}$, $P(\boldsymbol{\theta})$, and a sampling distribution $P(\mathbf{y}|\boldsymbol{\theta})$ (being the probability of getting observation $\mathbf{y}$ given $\boldsymbol{\theta}$) to determine the posterior distribution $P(\boldsymbol{\theta}|\mathbf{y})$. The posterior distribution, being the distribution of the parameters $\boldsymbol{\theta}$ given the observations $\mathbf{y}$, is what we search. However, it is not possible to evaluate this expression analytically when having complex models with many parameters, and this is why Markov chain simulations are used. The residual $\mathbf{r}$ is the difference between the experimental results $\mathbf{y}$ and the model predictions $\mathbf{f}$ that depend on $\boldsymbol{\theta}$. If it is assumed that the residual vector is normally distributed with zero mean and zero autocorrelation, we have

$$P(\mathbf{y}|\theta) = \frac{1}{\sqrt{2\pi\hat{s}^2}}\exp\left(-\frac{RSS(\boldsymbol{\theta})}{2\hat{s}^2}\right)$$

where $\hat{s}$ is the residual standard deviation and $RSS(\boldsymbol{\theta}) = \mathbf{r}^{\mathrm{T}}\mathbf{r} = (\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))^{\mathrm{T}}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))$ is the residual sum of squares. In order to maximise the likelihood $P(\mathbf{y}|\boldsymbol{\theta})$, the residual sum of squares has to be minimised. Therefore, the MCMC procedure can be initialised by determining the prior distribution of $\boldsymbol{\theta}$ using an ordinary least squares (OLS) estimate. This is carried out with the lsqnonlin function in the MATLAB Optimization Toolbox that uses a trust-region-reflective algorithm. Next, the measurement standard deviation is estimated by

$$\hat{s}^2 \approx \frac{RSS(\boldsymbol{\theta}^*)}{n_m - n_p} \tag{25}$$

where $\boldsymbol{\theta}^*$ denotes the OLS estimate of $\boldsymbol{\theta}$, n_m is the number of measurements and n_p is the number of parameters. The covariance matrix of $\boldsymbol{\theta}^*$ is estimated by

$$\boldsymbol{\Sigma} \approx \hat{s}^2\left(\mathbf{J}(\boldsymbol{\theta}^*)^{\mathrm{T}}\mathbf{J}(\boldsymbol{\theta}^*)\right)^{-1} \tag{26}$$

where $\mathbf{J}$ is the Jacobian of $\mathbf{f}$ with respect to $\boldsymbol{\theta}$, which can be estimated by central difference.

The Metropolis algorithm is as follows (Gelman et al., 2014): initially, a random sample $\boldsymbol{\theta}_0$ is drawn from the distribution $N(\boldsymbol{\theta}^*, C^2\boldsymbol{\Sigma})$ being a multivariate normal distribution with mean $\boldsymbol{\theta}^*$ and covariance $C^2\boldsymbol{\Sigma}$, where $C \approx 2.4n_p^{-1/2}$ is the coverage factor. For every iteration i, another sample $\hat{\boldsymbol{\theta}}$ is drawn from the jumping distribution $N(\boldsymbol{\theta}_{i-1}, C^2\boldsymbol{\Sigma})$ and the likelihood ratio

$$R = \frac{P(\mathbf{y}|\hat{\boldsymbol{\theta}})}{P(\mathbf{y}|\boldsymbol{\theta}_{i-1})} = \exp\left(-\frac{RSS(\hat{\boldsymbol{\theta}}) - RSS(\boldsymbol{\theta}_{i-1})}{2\hat{s}^2}\right)$$

is calculated. This sample is only conditionally accepted

$$\boldsymbol{\theta}_i = \begin{cases} \hat{\boldsymbol{\theta}} & \text{with probability } \min(R, 1) \\ \boldsymbol{\theta}_{i-1} & \text{otherwise} \end{cases}$$

This means that the sample is always accepted when the likelihood has increased since the last sample, but it may also be accepted with a non-zero probability if the likelihood has decreased. Iterations continue until the sample distribution becomes stationary. Then the initial samples (the warm up) are rejected and the remaining samples as a whole represents the posterior distribution of $\boldsymbol{\theta}$.

In order to speed up convergence, a simple adaptation is added to the algorithm (Gelman et al., 2014; Haario et al., 2001): for every n_0th iteration i_0, where $n_0 \gg 1$, the covariance matrix of the jumping distribution is updated to be proportional to the posterior covariance estimated from the former simulations

$$C^2\boldsymbol{\Sigma} \leftarrow C^2\mathrm{cov}(\boldsymbol{\theta}_{i_0-1}, \boldsymbol{\theta}_{i_0-2}, \ldots, \boldsymbol{\theta}_0)$$

There are several criteria that should be fulfilled to ensure that the samples actually do represent the posterior distribution well (Gelman et al., 2014). The in-between and within variances of multiple independent Markov chains show whether the chains are similar and have converged. Also, the acceptance ratio should be around 0.23 when having more than five parameters (Gelman et al., 2014). If the acceptance ratio is too low, C^2 should be decreased and vice versa.

If the residual contains significant autocorrelation, that is, the number of actually independent measurements is reduced, both $\hat{s}^2$ and $\boldsymbol{\Sigma}$ become underestimated. To counteract this, equations (25) and (26) can be substituted accordingly (Enemark and Santos, 2015)

$$\tilde{s}^2 \approx \frac{\tilde{n}_m}{n_m}\frac{RSS(\boldsymbol{\theta}^*)}{\tilde{n}_m - n_p} \tag{27}$$

$$\tilde{\boldsymbol{\Sigma}} \approx \tilde{s}^2\frac{n_m}{\tilde{n}_m}\left(\mathbf{J}(\boldsymbol{\theta}^*)^{\mathrm{T}}\mathbf{J}(\boldsymbol{\theta}^*)\right)^{-1} \tag{28}$$

where $\tilde{n}_m$ is the number of independent measurements ensuring a residual with insignificant autocorrelation.

Results and discussion

Experiments

With regard to spring dimensioning, we want to explore and utilise the non-linearities of the SMA

spring, when displacements are in the order of a few millimetres ($\Delta u \approx 2$ mm). This can be specified by $l = (\Delta u / \Delta \gamma(c)) = 2\,\text{mm}/1\%$ (equations (1)), because phase transformations are expected when $\Delta\gamma(c) \approx 1\%$. The maximum stiffness is required to be $k_A = 2$ N mm $^{-1}$. In order to determine the three parameters fully, it requires a third criterion. In this context, this criterion relates to the availability of specific wire diameters at the market (e.g. Memry Corporation, Amazon, Nitinol Devices & Components, Inc. and Johnson Matthey, Inc.).

The chosen spring dimensions are provided in Table 1. They are based on $G_A \approx 15$ GPa (Enemark and Santos, 2015), the values of l and k_A, equations (1) and (2) and the available spring diameters at the market.

The SMA springs are thermo-mechanically trained to get stabilised behaviour. The training consists of approximately 100 loading cycles using an amplitude of $A = 7.9$ mm, a frequency of $\omega = 0.5$ Hz and an ambient temperature of $T_\infty = 30$°C. Then around 100 cycles follow at $T_\infty = 70$°C. We found that it is important to explore the entire thermo-mechanical field during the training to avoid further training during the actual experiments. The choice of the minimum temperature (30°C) is based on the capabilities of the heat gun. At temperatures closer to the room temperature (25°C), the heat gun is not able to keep a constant temperature. The choice of the highest temperature (70°C) is made so that the spring is almost linearly elastic even for large deflections. A maximum amplitude of $A = 7.9$ mm is chosen because the spring easily breaks at higher amplitudes if combined with the high temperature. After the training, the spring has permanently elongated approximately 2.5 mm as a consequence of transformation-induced plastic strains.

The characterisation of the spring consists of experimentally obtained stable loading cycles at different ambient temperatures (T_∞ being 30°C, 50°C or 70°C), excitation amplitudes (A being 2.4, 3.3, 4.4, 5.8 and 6.9 mm) and frequencies (ω being 0.1, 0.3, 1.3, 4.0 and 10.0 Hz). The pre-tension length $x_0 = 6.5$ mm is kept constant. We found that the spring breaks after relatively few cycles (in the order of hundreds) if using $A > 5.8$ mm. For this reason, $A = 7.9$ mm is only used during training, and $A = 6.9$ mm is only used with slow speeds. Each experiment, of the $3 \cdot 5 \cdot 5 = 75$ different combinations, is performed a total of nine times using five different spring specimens to explore the uncertainties. Some experiments are discarded as outliers because of insufficient training.

Figure 7 shows 11 tests, which are used to calibrate the model. The size of the hysteresis loops and the tangential stiffness depends on both ambient temperature and oscillation frequency. The tangential stiffness during transformation is only slightly lower than during the thermo-elastic state. For SMA wires, the tangential stiffness during transformation is usually at least an order of magnitude lower than the pure austenitic or martensitic states (see, e.g., Enemark and Santos, 2015).

To get a better overview of the characteristic properties, the storage modulus K, the loss factor η and the pre-tension force F_0 are calculated from all the experimental results, and they are shown in Figure 8. They are represented by a mean value and an estimate of the 90% confidence interval based on the calculated

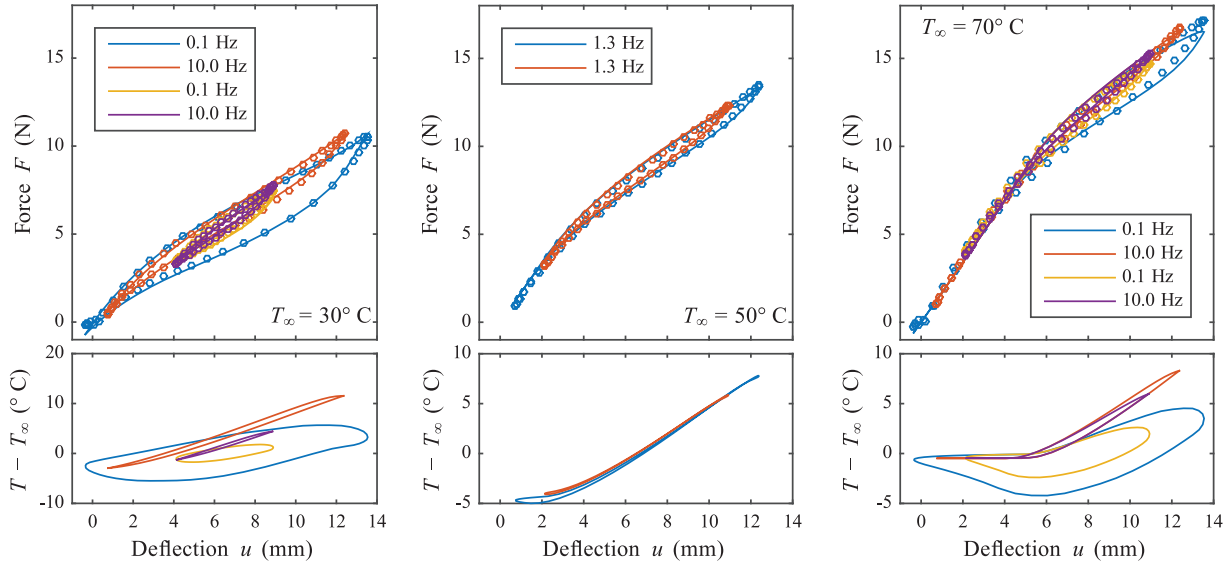

Figure 7. Cyclic tensile tests. Calibration data set (circles) and 2S model predictions (lines) of stable deflection cycles. For a given deflection, the force is measured and the force and temperature are predicted. The temperature loops are clockwise in the 0.1 and 1.3 Hz cases and counter clockwise in the 10.0 Hz cases.

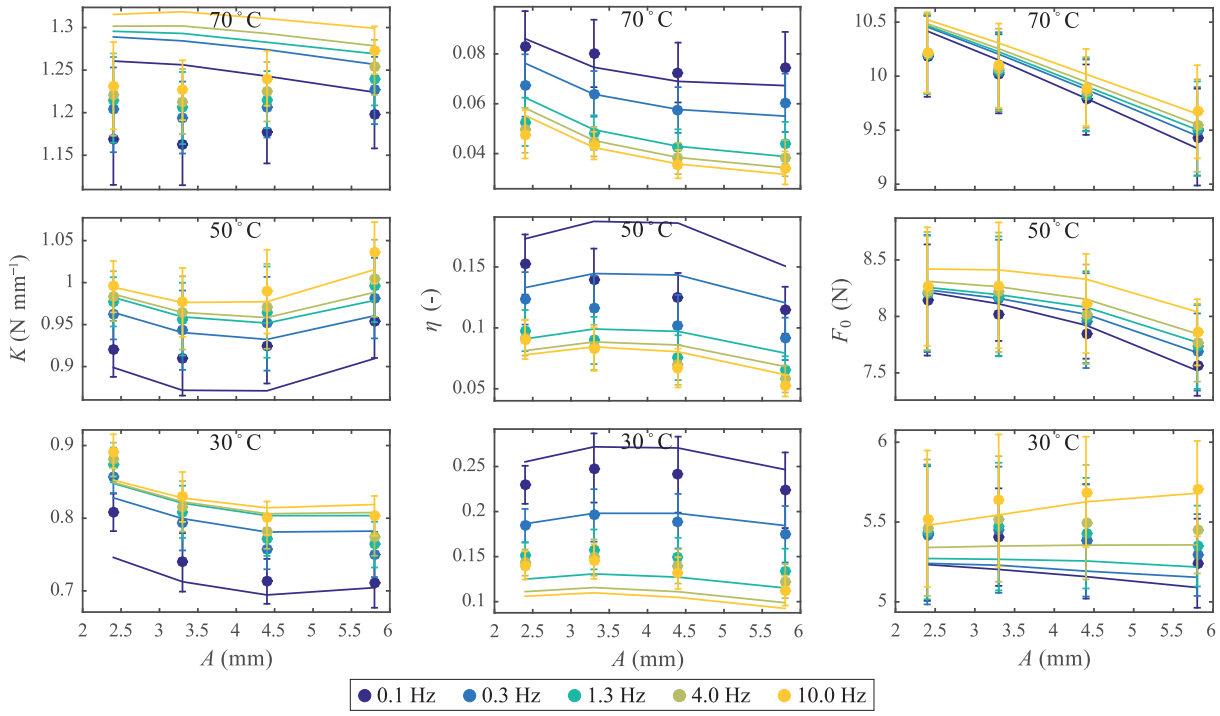

Figure 8. Storage modulus K, loss factor η and mean pre-tension force F_0 as function of excitation amplitude and frequency and ambient temperature. Experimental results are indicated by circles (mean) and error bars (estimated 90% confidence interval) from nine experiments. The lines are model predictions using the 2S model.

standard deviation and the Student's t-distribution using the appropriate number of degrees of freedom.

The storage modulus K depends on the ambient temperature to great extent, whereas the dependencies on frequency and amplitude are more moderate, even though the dependencies are clear, cf. Figure 8. The storage modulus increases with ambient temperature and frequency, whereas the dependency on amplitude is non-monotonic and coupled to the temperature.

The loss factor η also greatly depends on temperature and frequency, and it monotonically decreases with both. The dependency on amplitude is insignificant. The values of K and η seem to saturate with increasing frequency: The changes in their values are similar from 0.1 to 0.3 Hz and from 4.0 to 10 Hz.

The mean pre-tension force F_0 greatly depends on temperature, and the dependencies on frequency and amplitude are insignificant. Malecot et al. (2006) and He and Sun (2010) showed that the mean pre-tension force significantly increases with the excitation frequency caused by high temperature rises, which is a consequence of the latent heat of phase transformations and the lack of heat transfer to the surroundings. However, in this case, the level of forced convection is so high that the mean temperature and therefore also the mean force during a cycle do not increase significantly. Piedboeuf et al. (1998) and Gandhi and Wolons (1999) showed strong amplitude-dependent behaviour on SMA wires, whereas the amplitude dependencies of K, η and F_0 are insignificant in these experiments performed on springs. The reason is that the tangential stiffness during transformation is only slightly lower than in the thermo-elastic regions for the springs (maximum a factor of 4), which again is caused by the bending in addition to torsion of the spring wire and the otherwise complex stress and strain fields. Oppositely, the tangential stiffness during transformation for straight wires is usually considerably lower than in the thermo-elastic regions. In some conditions, the tangential stiffness is even 0 or slightly negative (Piedboeuf et al., 1998). Because there is a direct connection between the storage modulus and the tangential stiffness and because the loading cycles both include thermo-elastic and transformation regions, it results in a large amplitude dependency for the straight wires.

There are clear discrepancies between the experimental results highlighted by the confidence intervals. These discrepancies reflect the differences between the spring specimens, that is, material imperfections and geometrical inaccuracies. Also, small differences in training have an impact, and the characteristics change slightly during the lifetimes of the springs after the training process. This means that they never reach a perfectly stabilised state, but the changes are small and happen slowly.

Model fitting

The 2S model is fitted using OLS to the calibration data set highlighted in Figure 7, where the geometrical properties of the spring are fixed, and they are shown in

Table 2. Determined model parameters using OLS for both the 1S and 2S models, and the parameter distribution mean and confidence intervals using MCMC for the 2S model.

Model	1S	2S			
	OLS	OLS	Mean	5th Perc.	95th Perc.
H (mm m^{-1})	13.8	14.6	14.6	14.2	15.0
E_A (GPa)	38.5	36.8	36.8	36.5	37.0
E_M (GPa)	37.7	30.9	30.8	29.9	31.8
ν (–)	0.33	0.33		(Fixed)	
C_A (MPa K^{-1})	10.4	10.6	10.6	10.3	10.9
C_M (MPa K^{-1})	11.8	11.1	11.2	10.7	11.6
A_s (°C)	−6.0	−6.0		(Lower bound)	
A_f (°C)	35.0	35.0		(Upper bound)	
M_s (°C)	30.0	30.0		(Upper bound)	
M_f (°C)	−13.0	−13.0		(Lower bound)	
n_1^f (–)	0.58	0.71	0.70	0.65	0.76
n_2^f (–)	0.36	0.36		(Fixed)	
n_1^r (–)	0.74	0.72	0.72	0.68	0.76
n_2^r (–)	0.36	0.36		(Fixed)	
p (–)	0.99	0.99		(Upper bound)	
ρc_p (MJ m^{-3} K^{-1})	2.59	2.63	2.63	2.42	2.85
$\hat{h}$ (kW m^{-2} K^{-1})	0.34	0.37	0.37	0.33	0.42

OLS: ordinary least squares.
Confidence intervals are highlighted in terms of the 5th and 95th percentiles. The obtained residual standard deviations for the OLS fits are $\tilde{s}_{1S} = 0.26$ N and $\tilde{s}_{2S} = 0.24$ N, respectively (corrected for autocorrelation).

Table 1. A former differential scanning calorimetry (DSC) test showed that the phase transformation temperatures are $A_s = 4$°C, $A_f = 25$°C, $M_s = 20$°C and $M_f = -3$°C of the as-received material (Enemark and Santos, 2015). These temperatures are determined using tangential fits. However, when smooth hardening functions are used, the phase transformation temperatures used in the model can be altered to improve agreement between model predictions and experiments. This is exemplified by Lagoudas et al. (2012), where the phase transformation temperatures are changed up to 8°C to increase resemblance to a DSC curve. Also, the phase transformation temperatures may change as a consequence of the shape-setting heat treatment and the training process (Gloanec et al., 2013). Using these arguments, the phase transformation temperatures are allowed to change ±10°C from their DSC values when fitting the model. Since the experiments only contain partial transformations and not full transformations, it is not possible to uniquely identify all model parameters. The two parameters n_2^f and n_2^r determine the curvatures of the end of the forward transformation and the beginning of the reverse transformation, respectively, and they are not represented by the experiments. They are therefore fixed to be 0.36, making the curvature follow closely the cosine hardening function used by Brinson (1993), who has made a widely used constitutive model. Also, Poisson's ratio $\nu = 0.33$ is fixed, because it is not identifiable from the experiments. When fixing some parameters, the solution of other parameters might depend on the chosen fix points, so in this relation the solution is not unique.

In the OLS solution (Table 2), the phase transformation temperatures encounter their enforced boundaries. Also p encounters the upper boundary of its definition interval $(-1, 1)$, which is set to 0.99 in the numerical implementation. The model predictions and the calibration data set are compared in Figure 7 and the resemblance is high. The residual standard deviation is $\tilde{s} = 0.24$ N, where $n_m/\tilde{n}_m = 9$ is used to counteract the level of autocorrelation. Figure 7 also shows the spring temperature during the loading cycle, which forms a closed orbit confirming stabilised behaviour. The temperature primarily increases during loading and decreases during unloading, because the forward transformation is exothermic and the reverse transformation is endothermic. However, because the convection forces the temperature towards the ambient temperature, the temperature rate changes sign before the loading changes direction in each cycle. The maximum temperature depends on both frequency and amplitude. The mean temperature during a cycle is within ±5°C of the surroundings in all cases, which is because of the high level of convection.

To validate the model, model predictions and the remaining experiments are also compared. This is shown in terms of the complex modulus in Figure 8. The model predictions follow closely the experimental results both in terms of qualitative behaviour and low (quantitative) error. For K, the residual standard

Table 3. Some values of comparable model parameters found in the literature.

Parameter	OLS value	Ref. value	Reference
H (mm m^{-1})	14.6	33, 40	Lagoudas et al. (2012); Enemark and Santos (2015)
E_A (GPa)	36.8	33, 44	Lagoudas et al. (2012); Enemark and Santos (2015)
E_M (GPa)	30.9	23, 26	Lagoudas et al. (2012); Enemark and Santos (2015)
C_A (MPa K^{-1})	10.6	9	Enemark and Santos (2015)
C_M (MPa K^{-1})	11.1	12	Enemark and Santos (2015)
ρc_p (MJ m^{-3} K^{-1})	2.63	2.5 to 2.6, 3.2	Lagoudas et al. (2012); Zanotti et al. (2012) ($\rho \approx 6500$ kg m^{-3})
$\hat{h}$ (kW m^{-2} K^{-1})	0.37	0.1 to 0.4	Pathak et al. (2010) (straight wire, air velocity: 0–3.2 m s^{-1})

OLS: ordinary least squares.

deviation (between model and mean of experiments) is 0.043 N mm^{-1}, which can be compared with the mean standard deviation of the experiments of 0.023 N mm^{-1}. For η, the numbers are 0.021 compared with 0.012, and for F_0 it is 0.14 N compared with 0.24 N. The model errors are very acceptable, even though they cannot be explained by the measurement uncertainties entirely. The model errors are small compared to the full scale of the variables (K: 3.4%, η: 8.3% and F_0: 1.4%).

The literature values of some of the model parameters are provided in Table 3. The determined value of H is substantially lower than expected. The value is low because H determines not only the maximum transformation strain but also the tangential stiffness during transformation. A higher value of H would therefore result in too low stiffness during transformation. This can either be compensated by decreasing the values of M_f and A_s or by letting H be a function of the von Mises stress (Lagoudas et al., 2012). However, because only tests in the pseudoelastic region and the DSC test are available, it would not be possible to give a qualified estimate of the expression of H and it would not be reasonable to ignore the DSC results. The remaining parameters obtain reasonable values. It should be noted that the SMA wires used in this study are from the same company and have the same diameter as the wires used by Enemark and Santos (2015).

The OLS parameters of the 1S model (where only shear is taken into account) are also shown in Table 2. The same requirements regarding parameter boundaries are enforced as for the 2S model. Many of the model parameters are very similar. However, the values of E_M and E_A are almost equal, whereas they clearly differ in the 2S model. It is well-known that E_A is substantially higher than E_M as the 2S model predicts. The value of E_M has to attain a non-physically high value to counteract the deficiency regarding the geometrical non-linearities of the 1S model. These model parameters can therefore validate that the 2S model captures the physical behaviour of the spring to a higher extent compared with the 'standard' 1S model. On the other hand, the residual standard deviations of the two models are $\tilde{s}_{1S} = 0.26$ N and $\tilde{s}_{2S} = 0.24$ N, respectively, meaning that they perform almost equally well.

Parameter uncertainties

MCMC is used to determine the uncertainties of the parameters of the 2S model. The parameters that are fixed or have encountered their boundaries during OLS are considered fixed without uncertainties. This is necessary because otherwise the MCMC would find another optimum that does not satisfy the constraints. Four chains of each 15,000 samples are used, and the initial 7500 samples are removed before analysing the result. For every $n_0 = 500$ samples until the 7500th sample, the covariance matrix of the jumping distribution is updated. Only the calibration data set is used for determining the parameter uncertainties. The uncertainty analysis reflects the stability and the 'width' of the optimum determined by OLS, but not necessarily the observed discrepancies between the different springs used in the experiments. The resulting samples representing the parameter distributions are highlighted in terms of mean and 5th and 95th percentiles in Table 2, and they are shown as histograms and pairwise correlation plots in Figure 9. The parameters are well-defined, having uncertainties less than $\pm 12\%$ (5th and 95th percentiles compared with mean), and the means are either identical to or very close to the OLS estimates. The distributions are close to symmetric and also close to Gaussian. However, the tails of the distributions are non-Gaussian, and especially $\hat{h}$ is slightly skew. Therefore, the uncertainties found using the MCMC method differ from the initial linear estimate given in equation (28). Some parameters are highly correlated, for example, H with E_M, E_M with C_M and C_M with n_1^f. This means that they are difficult to isolate from each other with the given experiments. However, because of the non-linearities in the model with respect to the parameters, the uncertainties become limited. The parameters E_A, ρc_p and $\hat{h}$ have low correlations with all other parameters. Therefore, they are clearly identifiable. However, it should be highlighted that they might depend on the values of the fixed parameters.

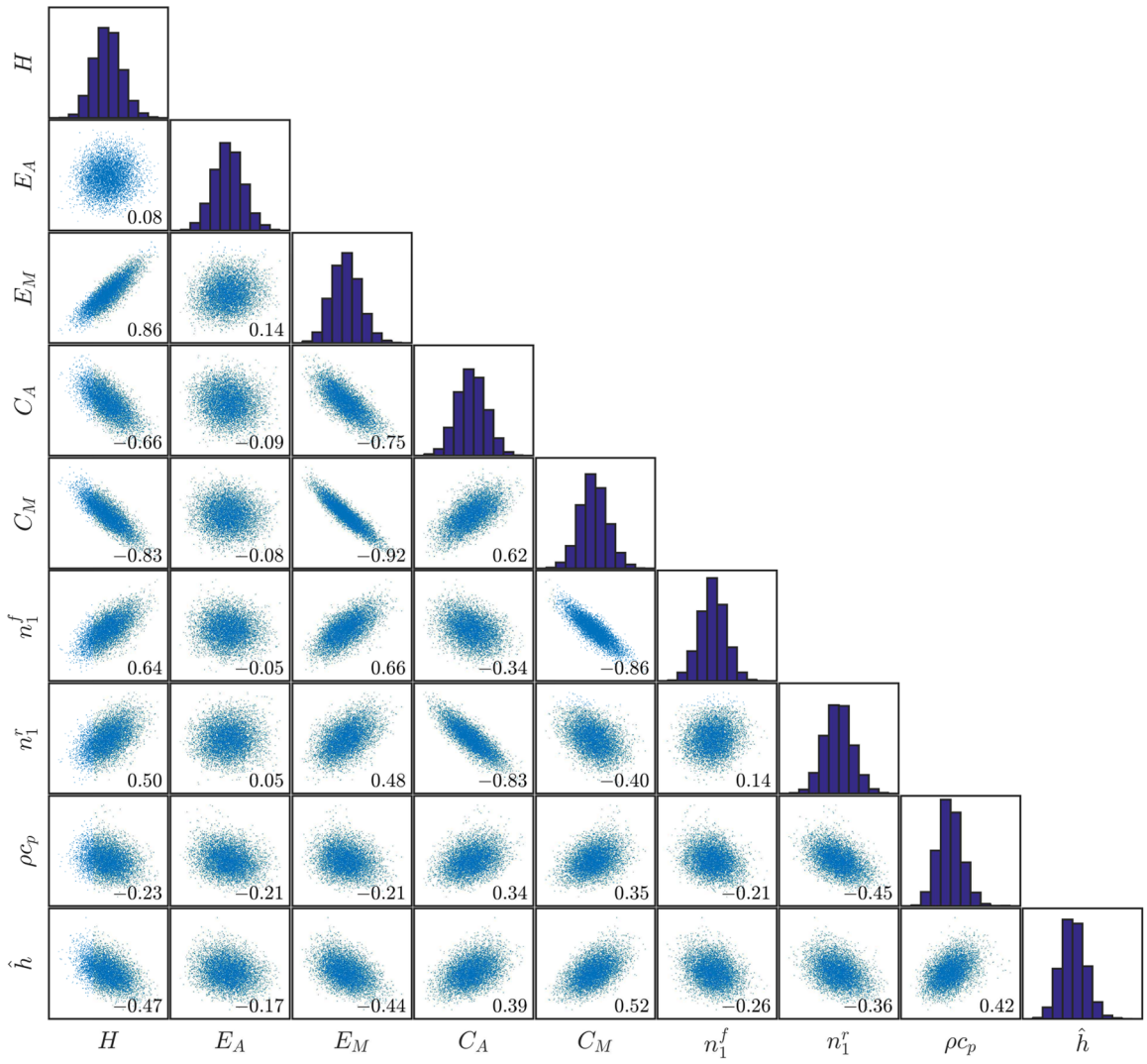

Figure 9. Accepted MCMC samples with the 2S model. Histograms are shown in the diagonal, and the off-diagonal plots are scatter plots showing the pairwise correlations. The linear correlation coefficients are highlighted in the corner of each plot.

Optimal damping conditions

Inspired by Piedboeuf et al. (1998) and He et al. (2010), who reported that the loss factor has a maximum with respect to frequency for constant convective conditions, additional experiments are carried out for frequencies lower than 0.1 Hz with another sixth spring. The results are shown in Figure 10. Compared to the experiments performed using springs 1–5 at 0.1 and 0.3 Hz, the results from spring 6 lie within the confidence interval. Indeed, the experimental results confirm that there exists an optimal frequency regarding the loss factor. It is at 0.08 Hz giving $\eta = 0.21$ for the given conditions for spring 6. It should be emphasised that the specific optimum depends on the convective conditions (He et al., 2010). At the same time, the storage modulus decreases significantly at low frequencies, while the pretension force is almost constant. The model is also able to predict this behaviour, even though it is only calibrated for frequencies higher than 0.1 Hz. The predicted optimal condition is $\eta = 0.25$ at $\omega = 0.9$ Hz. It is possible to explain why there exists an optimal frequency by the model. From the temperature loops (temperature as function of deflection) in Figure 11, the difference in spring temperature during loading and unloading is largest at 0.08 Hz. This means that the force increases during loading and decreases during unloading compared to the isothermal situation, which finally results in a hysteresis loop with a larger area. At lower frequencies (illustrated using 0.01 Hz), the spring temperature is almost constant due to the convection, which results in a narrow and flat temperature loop. At higher frequencies (illustrated using 0.30 Hz), the temperature loop becomes inclined and narrower again, because the convection is not strong enough to dissipate the thermal energy resulting from the exothermic loading process before the endothermic unloading process

Table 4. DSC results based on the mean of two experiments.

	A_s (°C)	A_f (°C)	M_s (°C)	M_f (°C)
Wire, batch 1	4	25	20	−3
Wire, batch 2	−2	24	18	−6
Spring	21	33	25	13

Uncertainty related to reproducibility is approximately ±1°C and uncertainty related to bias is approximately ±1°C.

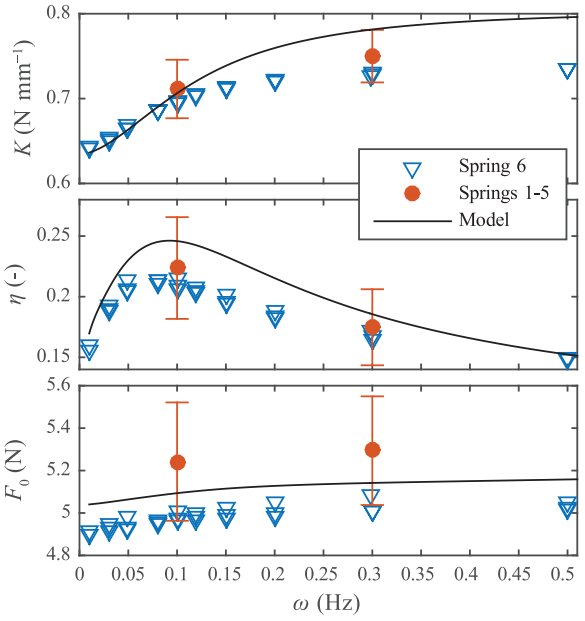

Figure 10. Investigation of the complex modulus for low frequencies using $A = 5.8$ mm and $T_\infty = 30$°C highlighting the optimal frequency regarding the loss factor. For springs 1–5, the 90% confidence interval is shown.

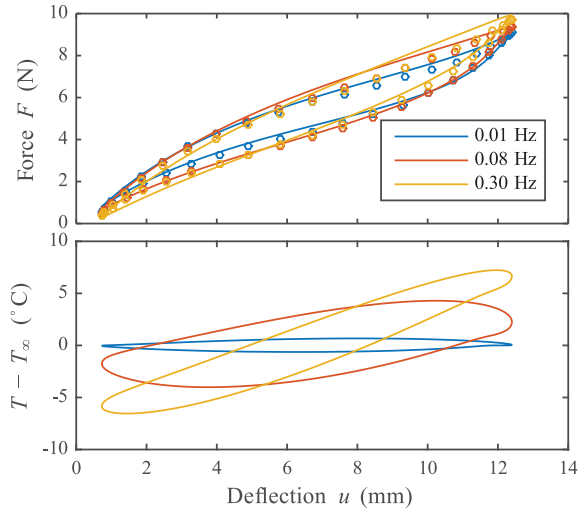

Figure 11. On top is the experimental (circles) and modelled (lines) force–deflection relationship using $A = 5.8$ mm and $T_\infty = 30$°C. Below is the predicted spring temperature as function of the deflection. The loops are clockwise in time.

occurs. At even higher frequencies, close to adiabatic conditions, the area of the temperature loop becomes 0, as may be seen in the 1.3 and 10.0 Hz cases in Figure 7.

Additional DSC tests

After characterisation, additional DSC analyses were carried out using a second batch of the as-received SMA wire and a piece of the SMA spring after the shape-setting treatment. The results are shown in Table 4 together with the reference test (wire, batch 1) that is used for characterisation. From Gloanec et al. (2013) (and others), it is known that heat treatments may significantly change the phase transformation temperature, which is the motivation for carrying out the additional tests. Also, it is interesting to determine whether the temperatures are stable across different batches.

From the results (Table 4), the transformation temperatures differ slightly (up to 6°C±2°C) between the wire batches, but it would not have a crucial impact if one or the other was used as the basis for the characterisation. However, the temperatures change considerably (up to 17°C±2°C relatively to wire, batch 1) as a consequence of the shape-setting heat treatment. If using these temperatures as a reference for the characterisation, it has a large impact on the rest of the model parameters, and equally important, the model fit notably worsens.

The methodology of the characterisation is based on information available in the literature and from manufacturers. For this reason, we chose to use the DSC results of the wire and not the spring, because these would be the values that would be available from the manufacturer. Again, it should be emphasised that the phase transformation temperatures may also change as a consequence of the training process that the springs are subjected to after the heat treatment. For example, Gloanec et al. showed that the phase transformation temperature values differed up to 15°C comparing an untested sample with a sample subjected to 3324 loading cycles.

Conclusion

Through justification of the modelling approach, it is shown that it is possible to use a single point (the star point) in the wire cross section to represent the global behaviour of the spring despite strong material non-linearities and complex stress–strain fields in the cross section. This is valid for moderate deflections. We also raise attention to the problem of choosing the maximum strain and stress in the cross section (at the circumference) to govern the SMA behaviour, because it may substantially misplace the transformation surfaces. At the same time, it is shown that normal strains and stresses are important if the global spring deformation is large, and therefore, these components should be included forming a 2D spring model.

The experiments show that the storage modulus K (stiffness), the loss modulus η (hysteresis) and the mean pre-tension force F_0 greatly depend on temperature. The high relative to low-temperature values (70°C relative to 30°C) are up to 170% for K, up to 40% for η, and up to 190% for F_0. To a smaller extent, K and η also depend on frequency, and less on the deflection amplitude. The largest changes in K and η are in the low-frequency range (below 1.0 Hz), whereas they saturate at higher frequencies (up to 10 Hz). There is no clear dependency of F_0 on either frequency or amplitude.

The 2D star point model with modified hardening and sub-loop behaviour is fitted to a calibration data set based on the available information about the transformation temperatures. The residual standard deviation is 0.24 N, which can be compared to the maximum obtained force of 17 N (i.e. 1.4%). All the model parameters attain physically sound values as a consequence of the model structure. Therefore, the model is able to explain the experimental behaviour. The proposed sub-loop functions have a sub-loop controlling parameter $p \in (-1, 1)$, which attain a value of 0.99 reflecting that the sub-loops are considerably wider and have a higher stiffness than predicted by the simple Duhem–Madelung sub-loop function corresponding to $p = 0$. The elastic stiffness during the mixed phase is equal to that of the pure austenitic phase. However, the value of the martensitic modulus is lower than the austenitic modulus when using the 2D model, which is physically sound. The appearing change in stiffness is caused by bending of the spring in addition to torsion resulting in a geometrical stiffening effect, which the model is able to capture. The 1D model does not capture this behaviour, but by choosing a (slightly unrealistic) high value of the martensitic modulus, the 1D and 2D models perform almost equally well.

The observed frequency dependency is due to the latent heat of phase transformations that alters the spring temperature up to around ±10°C from the ambient temperature. This is captured by the model, because the constitutive equation is coupled to the energy equation taking into account the latent heat and the convective conditions. It is shown that the damping factor has an optimum with respect to frequency for the specific convective conditions. The optimum is a due to an interaction between the heat capacity, heat convection and the latent heat of the transformations that results in large temperature differences during loading and unloading. The model and the validation data sets are compared showing that the model errors are comparable in size with the experimental uncertainties evaluated from nine repetitions of the same experiments using five different spring specimens. Even though the experiments are relatively complex, it is possible to identify both the heat capacity and the convection coefficient independently of all other uncertain model parameters, which is based on their low correlation coefficients to other parameters. All parameters have relative uncertainties less than ±12%, which is due to the diversity of the experiments and because the model non-linearities limit the uncertainties, where parameter correlations are high.

A DSC analysis showed that the phase transformation temperatures change up to 17°C as a consequence of the shape-setting treatment. However, since the model calibration is based on the available information from the literature and suppliers and because the training process most likely would change the temperatures again, it was chosen to use the initial values based on the as-received wire. This reinforces the need of linking the constitutive models to uncertainty analysis to assure reliable predictions of SMA properties in the context of machine element design.

Acknowledgements

The authors would like to thank Mik Mathias Mønsted for designing the tensile test-bench and for carrying out experiments.

Declaration of Conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors would like to acknowledge the support of the Danish Ministry of Science, Innovation and Higher Education, for providing funding for the FTP Research project 12-127502 and the Brazilian Research Agencies CNPq, CAPES and FAPERJ and through the INCT-EIE (National Institute of Science and Technology – Smart Structures in Engineering), the CNPq and FAPEMIG. The Air Force Office of Scientific Research (AFOSR) is also acknowledged.

Notes

1. The transformation surfaces are not straight lines in the temperature–stress plane in the model by Lagoudas and co-workers, and therefore, the values of C_A and C_M are specific for zero stress.
2. The reason for the sign(ε) in the expression of σ is to ensure tension/compression symmetry, when ε_t is assumed to be a constant.

References

Aguiar RAA, Savi MA and Pacheco PMCL (2010) Experimental and numerical investigations of shape memory alloy helical springs. *Smart Materials and Structures* 19(2): 025008. DOI: 10.1088/0964-1726/19/2/025008.

An SM, Ryu J, Cho M, et al. (2012) Engineering design framework for a shape memory alloy coil spring actuator using a static two-state model. *Smart Materials and Structures* 21(5): 055009. DOI: 10.1088/0964-1726/21/5/055009.

Bo ZH and Lagoudas DC (1999) Thermomechanical modeling of polycrystalline SMAs under cyclic loading, part IV: modeling of minor hysteresis loops. *International Journal of Engineering Science* 37(9): 1205–1249. DOI: 10.1016/S0020-7225(98)00116-5.

Borges JM, Silva AA, de Araújo CJ, et al. (2013) Rotor-bearing vibration control system based on fuzzy controller and smart actuators. *International Journal of Multiphysics* 7(3): 197–205. DOI: 10.1260/1750-9548.7.3.207.

Brinson LC (1993) One-dimensional constitutive behavior of shape memory alloys: thermomechanical derivation with non-constant material functions and redefined martensite internal variable. *Journal of Intelligent Material Systems and Structures* 4(2): 229–242. DOI: 10.1177/1045389X9300400213.

Cho KJ, Hawkes E, Wood RJ, et al. (2008) Design, fabrication and analysis of a body-caudal fin propulsion system for a microrobotic fish. In: *IEEE international conference on robotics and automation*, Pasadena, CA, 19–23 May, pp. 706–711. New York: IEEE. DOI: 10.1109/ROBOT.2008.4543288.

Crews JH and Smith RC (2014) Quantification of parameter and model uncertainty for shape memory alloy bending actuators. *Journal of Intelligent Material Systems and Structures* 25(2): 229–245. DOI: 10.1177/1045389X13490842.

Crews JH, McMahan JA, Smith RC, et al. (2013) Quantification of parameter uncertainty for robust control of shape memory alloy bending actuators. *Smart Materials and Structures* 22(11): 115021. DOI: 10.1088/0964-1726/22/11/115021.

De Sousa VC and De Marqui C Jr (2014) Effect of pseudoelastic hysteresis of shape memory alloy springs on the aeroelastic behavior of a typical airfoil section. Epub ahead of print 19 December 2014. *Journal of Intelligent Material Systems and Structures*. DOI: 10.1177/1045389X14563862.

Elahinia MH, Hashemi M, Tabesh M, et al. (2012) Manufacturing and processing of NiTi implants: a review. *Progress in Materials Science* 57(5): 911–946. DOI: 10.1016/j.pmatsci.2011.11.001.

Enemark S and Santos IF (2015) Quasi-static characterisation of trained pseudoelastic shape memory alloy wire subjected to cyclic loading: transformation kinetics. *Journal of Intelligent Material Systems and Structures*. Epub ahead of print 8 June 2015. DOI: 10.1177/1045389X15585900.

Enemark S, Santos IF and Savi MA (2015) Shape memory alloys applied to improve rotor-bearing system dynamics – an experimental investigation. In: *The XVII international symposium on dynamic problems of mechanics* (ed V Steffen, D Rade and W Bessa), Natal, Brazil, 22–27 February. Rio de Janeiro: ABCM.

Enemark S, Savi MA and Santos IF (2014) Nonlinear dynamics of a pseudoelastic shape memory alloy system – theory and experiment. *Smart Materials and Structures* 23(8): 085018. DOI: 10.1088/0964-1726/23/8/085018.

Follador M, Cianchetti M, Arienti A, et al. (2012) A general method for the design and fabrication of shape memory alloy active spring actuators. *Smart Materials and Structures* 21(11): 115029. DOI: 10.1088/0964-1726/21/11/115029.

Gandhi F and Wolons D (1999) Characterization of the pseudoelastic damping behavior of shape memory alloy wires using complex modulus. *Smart Materials and Structures* 8(1): 49–56. DOI: 10.1088/0964-1726/8/1/005.

Gelman A, Carlin JB, Stern HS, et al. (2014) *Bayesian Data Analysis*. 3rd ed. Boca Raton, FL: CRC Press, Taylor & Francis Group.

Gloanec AL, Bilotta G and Gerland M (2013) Deformation mechanisms in a TiNi shape memory alloy during cyclic loading. *Materials Science and Engineering: A* 564: 351–358. DOI: 10.1016/j.msea.2012.11.051.

Haario H, Saksman E and Tamminen J (2001) An adaptive metropolis algorithm. *Bernoulli* 7(2): 223–242.

Han H, Ang K, Wang Q, et al. (2006) Buckling enhancement of epoxy columns using embedded shape memory alloy spring actuators. *Composite Structures* 72(2): 200–211. DOI: 10.1016/j.comstruct.2004.11.015.

He YJ and Sun QP (2010) Frequency-dependent temperature evolution in NiTi shape memory alloy under cyclic loading. *Smart Materials and Structures* 19(11): 115014. DOI: 10.1088/0964-1726/19/11/115014.

He YJ, Yin H, Zhou R, et al. (2010) Ambient effect on damping peak of NiTi shape memory alloy. *Materials Letters* 64(13): 1483–1486. DOI: 10.1016/j.matlet.2010.03.068.

Holanda SA, Silva AA, de Araujo CJ, et al. (2014) Study of the complex stiffness of a vibratory mechanical system with shape memory alloy coil spring actuator. *Shock and Vibration* 2014: 162781. DOI: 10.1155/2014/162781.

Kim B, Lee M, Lee Y, et al. (2006) An earthworm-like micro robot using shape memory alloy actuator. *Sensors and Actuators A: Physical* 125(2): 429–437. DOI: 10.1016/j.sna.2005.05.004.

Lagoudas DC, Hartl D, Chemisky Y, et al. (2012) Constitutive model for the numerical analysis of phase transformation in polycrystalline shape memory alloys. *International Journal of Plasticity* 32–33: 155–183. DOI: 10.1016/j.ijplas.2011.10.009.

Liang C and Rogers CA (1993) Design of shape memory alloy springs with applications in vibration control. *Journal of Vibration and Acoustics* 115(1): 129–135. DOI: 10.1115/1.2930305.

Liu X, Wang Y, Yang D, et al. (2008) The effect of ageing treatment on shape-setting and superelasticity of a nitinol

stent. *Materials Characterization* 59(4): 402–406. DOI: 10.1016/j.matchar.2007.02.007.

Ma Y, Zhang Q, Zhang D, et al. (2014) A novel smart rotor support with shape memory alloy metal rubber for high temperatures and variable amplitude vibrations. *Smart Materials and Structures* 23(12): 125016. DOI: 10.1088/0964-1726/23/12/125016.

Malecot P, Lexcellent C, Foltete E, et al. (2006) Shape memory alloys cyclic behavior: experimental study and modeling. *Journal of Engineering Materials and Technology* 128(3): 335–345. DOI: 10.1115/1.2204947.

Mehrabi R, Kadkhodaei M and Elahinia M (2014) Constitutive modeling of tension-torsion coupling and tension-compression asymmetry in NiTi shape memory alloys. *Smart Materials and Structures* 23(7): 075021. DOI: 10.1088/0964-1726/23/7/075021.

Mirzaeifar R, DesRoches R and Yavari A (2011) A combined analytical, numerical, and experimental study of shape-memory-alloy helical springs. *International Journal of Solids and Structures* 48(3–4): 611–624. DOI: 10.1016/j.ijsolstr.2010.10.026.

Morin C, Moumni Z and Zaki W (2011) Thermomechanical coupling in shape memory alloys under cyclic loadings: experimental analysis and constitutive modeling. *International Journal of Plasticity* 27(12): 1959–1980. DOI: 10.1016/j.ijplas.2011.05.005.

Oehler SD, Hartl DJ, Lopez R, et al. (2012) Design optimization and uncertainty analysis of SMA morphing structures. *Smart Materials and Structures* 21(9): 094016. DOI: 10.1088/0964-1726/21/9/094016.

Oliveira SA, Savi MA and Santos IF (2014) Uncertainty analysis of a one-dimensional constitutive model for shape memory alloy thermomechanical description. *International Journal of Applied Mechanics* 6(6): 1450067. DOI: 10.1142/S1758825114500677.

Pathak A, Brei D and Luntz J (2010) Transformation strain based method for characterization of convective heat transfer from shape memory alloy wires. *Smart Materials and Structures* 19(3): 035005. DOI: 10.1088/0964-1726/19/3/035005.

Piedboeuf M, Gauvin R and Thomas M (1998) Damping behavior of shape memory alloys: strain amplitude, frequency and temperature effects. *Journal of Sound and Vibration* 214(5): 885–901. DOI: 10.1006/jsvi.1998.1578.

Sakuma T and Suzuki A (2007) Superelastic behavior under cyclic loading for coil spring of Ti-Ni shape memory alloy. *Materials Transactions* 48(3): 422–427. DOI: 10.2320/matertrans.48.422.

Saleeb AF, Dhakal B, Hosseini MS, et al. (2013) Large scale simulation of NiTi helical spring actuators under repeated thermomechanical cycles. *Smart Materials and Structures* 22(9): 094006. DOI: 10.1088/0964-1726/22/9/094006.

Sameallah S, Kadkhodaei M, Legrand V, et al. (2015) Direct numerical determination of stabilized dissipated energy of shape memory alloys under cyclic tensile loadings. *Journal of Intelligent Material Systems and Structures* 26(16): 2137–2150. DOI: 10.1177/1045389X14549869.

Savi MA, Pacheco PMCL, Garcia MS, et al. (2015) Nonlinear geometric influence on the mechanical behavior of shape memory alloy helical springs. *Smart Materials and Structures* 24(3): 035012. DOI: 10.1088/0964-1726/24/3/035012.

Shaw JA and Kyriakides S (1995) Thermomechanical aspects of NiTi. *Journal of the Mechanics and Physics of Solids* 43(8): 1243–1281. DOI: 10.1016/0022-5096(95)00024-D.

Speicher M, Hodgson DE, DesRoches R, et al. (2009) Shape memory alloy tension/compression device for seismic retrofit of buildings. *Journal of Materials Engineering and Performance* 18(5–6): 746–753. DOI: 10.1007/s11665-009-9433-7.

Tobushi H, Ohashi Y, Hori T, et al. (1992) Cyclic deformation of TiNi shape-memory alloy helical spring. *Experimental Mechanics* 32(4): 304–308. DOI: 10.1007/BF02325582.

Wahl AM (1944) *Mechanical Springs*. 1st ed. Cleveland, OH: Penton Publishing Co.

Wolons D, Gandhi F and Malovrh B (1998) Experimental investigation of the pseudoelastic hysteresis damping characteristics of shape memory alloy wires. *Journal of Intelligent Material Systems and Structures* 9(2): 116–126. DOI: 10.1177/1045389X9800900205.

Zanotti C, Giuliani P and Chrysanthou A (2012) Martensitic–Austenitic phase transformation of Ni–Ti SMAs: thermal properties. *Intermetallics* 24: 106–114. DOI: 10.1016/j.intermet.2012.01.026.