OPEN ACCESS
IOP Publishing
Modelling and Simulation in Materials Science and Engineering
Modelling Simul. Mater. Sci. Eng. **27** (2019) 024001 (18pp) https://doi.org/10.1088/1361-651X/aaf0b8

MultOpt++: a fast regression-based model for the development of compositions with high robustness against scatter of element concentrations

Alexander Müller[1], Irina Roslyakova[2], Mario Sprenger[1], Paul Git[1], Ralf Rettig[3], Matthias Markl[1], Carolin Körner[1,4] and Robert F Singer[1,4]

[1] Chair of Materials Science and Engineering for Metals, Friedrich-Alexander-Universität Erlangen-Nürnberg, Martensstr. 5, D-91058 Erlangen, Germany
[2] ICAMS, Ruhr-Universität Bochum, Universitätstr.150, D-44801 Bochum, Germany
[3] Thermo-Calc Software AB, Råsundavägen 18, SE-169 67 Solna, Sweden
[4] Neue Materialien Fürth GmbH, Dr.-Mack-Straße 81, D-90762 Fürth, Germany

E-mail: alexander.mueller@fau.de

Received 28 May 2018, revised 4 November 2018
Accepted for publication 14 November 2018
Published 2 January 2019

Abstract

Alloys-by-design is a term used to describe new alloy development techniques based on numerical simulation. These approaches are extensively used for nickel-base superalloys to increase the chance of success in alloy development. During alloy production of numerically optimized compositions, unavoidable scattering of the element concentrations occurs. In the present paper, we investigate the effect of this scatter on the alloy properties. In particular, we describe routes to identify alloy compositions by numerical simulations that are more robust than other compositions. In our previously developed alloy development program package MultOpt, we introduced a sensitivity parameter that represents the influence of alloying variations on the final alloy properties in the post-optimization process, because the established sensitivity calculations require high computational effort. In this work, we derive a regression-based model for calculating the sensitivity that only requires one-time calculation of the regression coefficients. The model can be applied to any function with nearly linear behavior within the uncertainty range. The model is

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

then successfully applied to the computational alloys-by-design work flow to facilitate alloy selection using the sensitivity of a composition owing to the inaccuracies in the manufacturing process as an additional minimization goal.

Keywords: sensitivity, CALPHAD, regression analysis, alloys-by-design, superalloys

(Some figures may appear in colour only in the online journal)

1. Introduction

The conventional trial-and-error-based property testing method to develop new alloys demands large experimental effort. Especially for modern high performance alloys containing more than eight alloying elements, this approach is relatively expensive because of the high number of different alloy compositions [1]. The numerical alloy-by-design approach [2, 3], which is supported by the calculation of phase diagrams (CALPHAD) method [4], is the state of the art to predict new highly complex alloys.

We have successfully applied alloy-by-design methods and predicted the highly promising Ni-base superalloys ERBO/13 [5, 6] and ERBO/15 [7, 8]. Other researchers have also shown the capability of these design tools for various alloy classes [9–13].

During the manufacturing process of nickel-base superalloy components, various processes can lead to scattering of the element concentrations. In the course of the heat production of an alloy by vacuum inductive melting such a scattering occurs due to measurement inaccuracies of the scales, low purity of the master alloys, chemical reactions with the crucible or contaminated atmosphere [14]. In the state of the art process for casting of nickel-based superalloys called high rate solidification (HRS), the molten alloy is poured into a preheated mould under high vacuum and slowly withdrawn from the heating zone into the cooling zone [15]. Due to the liquid phase dwell time of several hours, volatile elements such as Al and Cr can evaporate [16]. The recently developed FCBC process is performed under a low pressure atmosphere and at higher withdrawal rates, preventing evaporation of these elements [17], but industrial application is still pending. Figure 1(a) shows an exemplary distribution of the element concentrations[5] along the withdrawal direction in a single crystal HRS casting process. In addition, the root mean squared error (RMSE) of the achieved element concentration from the target concentrations and the standard deviation of six independentantly produced castings of one experimental alloy with separately produced heats are shown. Besides a scattering of the Mo concentration and a systematic deviation of the Ti content, the decrease of the Al and Cr concentrations along the solidification path is visible. That means, that both the dispersion of the concentrations in the component (RMSE) and the reproducibility (standard deviation) of various heats and castings $\Delta\vec{c}$ must be considered regarding the sensitivity of the alloy properties.

The concept of sensitivity is shown schematically in figure 1(b). The slope of a plot of a property model function f against the element concentration indicates how strong a change in the concentration affects the spread of the respective property. Thus, a low sensitivity alloy exists in a concentration range where the slope is small. Therefore, it will only exhibit small variations in its properties for a given variation of the concentrations of the alloying elements.

[5] Measured by the optical emission spectrometry on Ametek Spectromaxx (SPECTRO Analytical Instruments GmbH, Kleve, Germany).

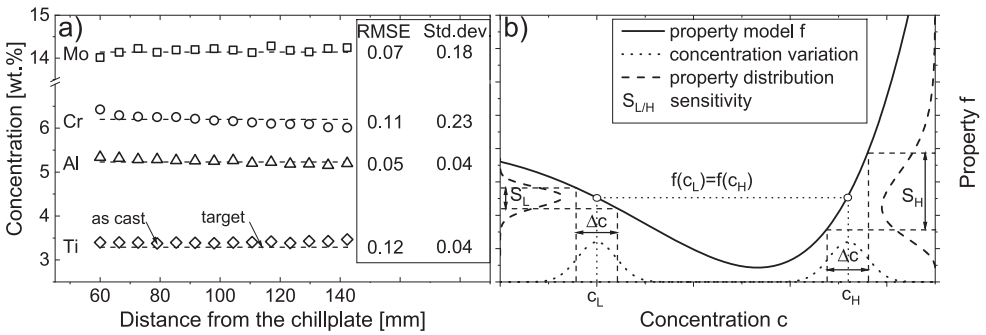

Figure 1. (a) Element concentration of an experimental nickel-base superalloy along a single crystal casting produced by the HRS process with the RMSE of the target composition. The standard deviation represents the reproducibility of target concentrations measured at six castings. These variations lead to the concentration deviation $\Delta\vec{c}$ from the nominal composition $\vec{c}$. (b) Schematic showing that a steeper function slope causes higher sensitivity S_H than S_L for two different compositions $\vec{c}_L$ and $\vec{c}_H$ with the same function value $f(\vec{c}_L) = f(\vec{c}_H)$ and the same inaccuracy $\Delta\vec{c}$. The examples for property models can be found in table 3.

In this case, the alloy $\vec{c}_L$ is the preferred alloy compared to the alloy $\vec{c}_H$, although both have the same property value $f(\vec{c}_L) = f(\vec{c}_H)$.

The sensitivity of the alloy properties is rarely been taken into account [18] during complex optimization processes because of the lack of fast models for the property robustness, which is defined as the inverse of the sensitivity. The general concept of multi-objective robust optimization was initially developed by Deb *et al* [19] from single-objective robust optimization [20, 21], and it is has been summarized in detail by Beyer *et al* [22]. Rettig *et al* [5] applied sensitivity analysis of the alloy properties in a post-optimization process using the alloy development program package MultOpt. An improvement of MultOpt called MultOpt++ has been briefly summarized by Markl *et al* [23].

In the current work, a new regression-based model for online sensitivity evaluation was derived. This will allow robust optimization of the alloy properties by including the property sensitivity as an additional minimization goal.

2. Methods

2.1. Alloy property prediction

Consider a multicomponent alloy consisting of n alloying elements. The chemical composition of this alloy can be represented as a n-dimensional vector: $\vec{c} = (c^{(1)},\ldots,c^{(n)})$, where $c^{(i)}$ $(i = 1, \ldots, n)$ is the concentration of ith alloying element. The typical design space for alloy development is given by a $2 \times n$ matrix containing the concentration range defined for a certain optimization problem $\mathbb{C}_{op} = [\vec{c}_{min}, \vec{c}_{max}]$, where $\vec{c}_{min} = (c_{min}^{(1)},\ldots,c_{min}^{(n)})$ and $\vec{c}_{max} = (c_{max}^{(1)},\ldots,c_{max}^{(n)})$ are n-dimensional vectors containing the minimum and maximum concentration values of the ith alloying elements, respectively. In some cases, the considered design space is extended by the temperature range $\mathbb{T}_r$, which is usually defined as $\mathbb{T}_r = [T_{min}, T_{max}]$, where T_{min} and T_{max} are the minimum and maximum temperatures relevant for alloy development.

Transformation from the design space $D = \mathbb{C}_{\mathrm{op}} \times \mathbb{T}_{\mathrm{op}}$ to the objective space $O = \vec{c} \times T$ requires the availability of property models:

$$(\vec{c} \in \mathbb{C}_{\mathrm{op}}, T \in \mathbb{T}_{\mathrm{op}}) \longmapsto f(\vec{c}, T).$$

Here, the property model is a functional description of the alloy properties, which can be obtained during a model development process. This model can be constructed based on known physical relationships or using statistical regression analysis with experimental data or empirical rules.

In this study, the applied models are based on CALPHAD calculations, combining both of the previously mentioned approaches. CALPHAD calculations are a very powerful tool for alloy development and allow prediction of several properties, even for complex compositions with more than eight alloying elements, such as Ni-base superalloys. The common thermodynamic calculations performed by the CALPHAD method during the optimization routine in this work are the calculation of transformation temperatures such as liquidus, solvus and solidus temperature and the amount of phases and their compositions. The databases used for such calculations are being continuously improved and show good agreement with experimental data [24–26] and less agreement for the prediction of topologically close packed (TCP) phases [27, 28].

2.2. Sensitivity

In general, the sensitivity $S_{f_k}(\vec{c}, \Delta\vec{c}, T)$, where k is the property index, describes the variation of the objective function $f_k(\vec{c}, T)$ with the variation or uncertainty of one $\Delta c^{(i)}$ or all $\Delta\vec{c}$ design variables with n elements. The widely used approaches for local sensitivity calculation, such as

$$S_{f_k}(\vec{c}, \Delta\vec{c}, T) = \left\| \Delta c^{(1)} \frac{\partial f_k(\vec{c}, T)}{\partial c^{(1)}}, \cdots, \Delta c^{(n)} \frac{\partial f_k(\vec{c}, T)}{\partial c^{(n)}} \right\| \tag{1}$$

as well as the standardized equation with distribution width $\sigma_{c^{(i)}}$ of the ith element within the range $\Delta c^{(i)}$ and the corresponding property distribution of $\sigma_{f_k^{(i)}(\vec{c},T)}$

$$S_{f_k}(\vec{c}, \Delta\vec{c}, T) = \left\| \frac{\sigma_{c^{(1)}} \partial f_k(\vec{c}, T)}{\sigma_{f_k^{(1)}(\vec{c},T)} \partial c^{(1)}}, \cdots, \frac{\sigma_{c^{(n)}} \partial f_k(\vec{c}, T)}{\sigma_{f_k^{(n)}(\vec{c},T)} \partial c^{(n)}} \right\| \tag{2}$$

are summarized in [29]. Equation (2) is recommended for sensitivity analysis, but it does not take into account the interactions between the design variables $c^{(i)}$ ($i = 1, \ldots, n$) and it is only valid for nearly linear functions [30]. To include these effects in sensitivity calculations, we focus on a variation-based method (3):

$$S_{f_k}(\vec{c}, \Delta\vec{c}, T) = \frac{\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)}{f_k^{\max}(\bar{\vec{c}}_{\mathrm{par}}) - f_k^{\min}(\bar{\vec{c}}_{\mathrm{par}})} \equiv S_{f_k}, \tag{3}$$

where $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)$ is the total distribution width (standard deviation) of the objective function f_k owing to the uncertainty in the input variables within the range $\vec{c} \pm \Delta\vec{c}$, in which $\vec{c}$ is the mean value (nominal composition) and $\Delta\vec{c}$ is double of the standard deviation of the concentration distribution, respectively. For simplification, we do not divide $S_{f_k}(\vec{c}, \Delta\vec{c}, T)$ into the single effect of each variable on the objective function f_k. The comparability of the sensitivities for different functions $f_k(\vec{c}, T)$ is ensured by the normalization value $f_k^{\max}(\bar{\vec{c}}_{\mathrm{par}}) - f_k^{\min}(\bar{\vec{c}}_{\mathrm{par}})$, which is ideally set during the optimization process as the objective function range from the set of the optimal Pareto-compositions $\bar{\vec{c}}_{\mathrm{par}}$.

The sensitivity of a multi-objective optimization problem (MOOP) has to combine the sensitivities of each objective function in a similar way to that proposed by Wang *et al* [31]:

$$S_{\mathrm{op}}(\vec{c},\,\Delta\vec{c},\,T)=\sqrt{\sum_{k=1}^{n_{\mathrm{goal}}}(S_{f_k})^2}=\sqrt{\sum_{k=1}^{n_{\mathrm{goal}}}\left(\frac{\sigma_{f_k}(\vec{c},\,\Delta\vec{c},\,T)}{f_k^{\max}(\bar{\vec{c}}_{\mathrm{par}})-f_k^{\min}(\bar{\vec{c}}_{\mathrm{par}})}\right)^2}\,. \tag{4}$$

2.3. Regression analysis of the property distribution

Equation (4) is based on the property distribution $\sigma_{f_k}(\vec{c},\,\Delta\vec{c},\,T)$, whose calculation during the optimization process is time consuming because of the large number of design variables. By ignoring the interactions between n variables, the required number of function calls can be reduced from 2^n in the simplest case of permutating all boundary values $\vec{c}_0 \pm \Delta\vec{c}$ to only 2n. This causes difficulties in locating good solutions for complex optimization problems, such as single-crystal Ni-base superalloy development. Depending on the number of optimization goals and input variables, each objective function has to be called for a typicall alloys-by-design routine in our approach up to 10^4–10^9 times. A simple optimization including interactions will then take $10^4 \times 2^8 \times 5$ ms ≈ 4 h using surrogate models [5]. The use of direct Thermo-Calc[6] (TC) calculations significantly increases the optimization time (in our alloy-by-design tool by a factor > 50).

In our approach, we first calculate the spatial distribution of $\sigma_{f_k}(\vec{c},\,\Delta\vec{c},\,T)$ within reasonable concentration and uncertainty ranges, as summarized in table 1. Additionally, some implicit and explicit constraints for development of single-crystal nickel-based superalloys for turbine blade production are defined to prevent calculation of impractical solutions. Extensive regression analysis of approximately 50 000 calculated points is then performed to obtain the free parameters of the model and select the most appropriate regression model from a set of considered models for each alloy property distribution. The number of required calculations is comparable with the number of function calls for one complex alloy development problem ($\sim 10^7$ calls), but they only have to be performed once and can be more easily parallelized because no communication between processes is required.

During statistical analysis, three different linear regression models (5)–(7) with and without stepwise variable selection [32] are applied and compared with each other to identify the most appropriate model (MAM) for prediction of the sensitivity of each alloy property:

$$\sigma_{f_k}^{\mathrm{FL/SL}}(\vec{x})=\underbrace{\sum_i a^{(i)}x^{(i)}}_{\text{pure elements}}+a^{(0)}, \tag{5}$$

$$\sigma_{f_k}^{\mathrm{FLI/SLI}}(\vec{x})=\sigma_{f_k}^{\mathrm{FL/SL}}(\vec{x})+\underbrace{\sum_{i,j,i>j} a^{(i,j)}x^{(i)}x^{(j)}}_{\text{interactions}}\,, \tag{6}$$

$$\sigma_{f_k}^{\mathrm{FLIQ/SLIQ}}(\vec{x})=\sigma_{f_k}^{\mathrm{FLI/SLI}}(\vec{x})+\underbrace{\sum_i a^{(i)}(x^{(i)})^2}_{\text{quadratic terms}}\,, \tag{7}$$

where $\vec{x}=(\vec{c},\,\Delta\vec{c},\,T)^T$ is the vector of the merged input parameters and $a^{(i)}$ $(i=1,\,\ldots,\,p)$ are unknown parameters to be defined from the precalculated data points. In the first place, the differences between these models lie in their complexity. The goal is to find a simple model,

[6] Thermo-Calc Software AB, Solna, Sweden.

Table 1. Design variable range $\mathbb{C}_{\text{reg}} = [\vec{c}_{\min}, \vec{c}_{\max}] \times [\Delta\vec{c}_{\min} = \vec{0}, \Delta\vec{c}_{\max}] \times [T_{\min}, T_{\max}]$ and implicit and explicit constraints for development of single-crystal nickel-based blade superalloys for regression analysis. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, W_{ht}, p, $T_{\gamma'}$, T_{sol}, and T_{liq} are the solid solution strengthening index, density, misfit between the γ and γ' phases, γ' fraction, heat treatment window, cost, γ' solvus, and solidus and liquidus temperatures, respectively.

Design variables range: $\mathbb{C}_{\text{reg}}$										
	Al	Co	Cr	Mo	Re	Ta	Ti	W	Ru	T[°C]
$\vec{c}_{\min}$[wt%]	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	900
$\vec{c}_{\max}$[wt%]	7.0	20	15	12	10	15	7.0	15	10	1100
$\Delta\vec{c}_{\max}$[wt%]	0.3	0.4	0.3	0.4	0.5	0.5	0.3	0.5	0.4	—

Functions constraints			Concentrations constraints	
f_k	$f_k^{\min}$	$f_k^{\max}$	c_0	$c_0^{\max}$[at%]
ρ[kg m^{-3}]	7500	9200	$\sum_{i\neq\text{Co}}^{n} c^{(i)}$	50
γ'[mol%]	30	75	$\sum_{i=\text{Ta,Ti}}^{n} c^{(i)}$	$2 \cdot c^{(\text{Al})}$
T_{liq} [°C]	1250	1500	$\sum_{i=\text{Al,Ta,Ti}}^{n} c^{(i)}$	25
$\delta_{\gamma/\gamma'}$[%]	−1.0	0.5		
p[€/kg]	—	700		
I_{SSS} [at%]	—	50		
T_{sol}[°C]	1200	1500		
$T_{\gamma'}$[°C]	1000	1400		

which allows to describe the true function with as few coefficients as possible. Here, L, I, and Q correspond to the linear, interaction, and quadratic terms involved in the model construction, and S and F in the superscript denote regression models with and without stepwise selection of variables. For example, SLIQ describes a model with linear, interaction, and quadratic terms calculated by stepwise regression analysis.

By substituting the property distribution $\sigma_{f_k}(\vec{x})$ into equations (3) and (4) with the regression-based $\sigma_{f_k}^{\text{reg}}(\vec{x})$ (one of the previously determined (5)–(7)), we define fast models for prediction of the sensitivity of a single function as $S_{f_k}^{\text{p}}$ and of all optimized goals as S_{op}^{p}, respectively.

The value $\sigma_{f_k}^{\text{reg}}(\vec{c}, \Delta\vec{c}, T)$ can be also used to predict the property variation within a certain specification limit $\Delta\vec{c}_{\text{s}}$ by replacing the uncertainty $\Delta\vec{c}$ with $\Delta\vec{c}_{\text{s}}$.

2.4. Optimization

The computer-based development process for indentifying good alloy candidates is usually applied in the objective space (properties), where the algorithm attempts to achieve the best property value available in the whole design space (concentrations). Simultaneous optimization of several goals has to be able to handle so-called MOOPs, which also contain a certain number of constraints. The result of such a calculation is a Pareto-front describing the best possible combinations of non-dominated function values for a given optimization problem. The effectivity of evolutionary algorithms (EAs) for solving MOOPs has already been shown by many researchers in a wide range of approaches [21, 33–37].

The model for prediction of the sensitivity derived by regression analysis was applied to the optimization problem described in [5], which was slightly modified and extended by

Table 2. Definition of the optimization problem for the current work. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, W_{ht}, p, $T_{\gamma'}$, T_{sol}, and T_{liq} are the solid solution strengthening index, density, misfit between the γ and γ' phases, γ' fraction, heat treatment window, cost, γ' solvus, and solidus and liquidus temperatures, respectively. $\vec{c}_r$ represents the whole optimization range.

Optimization goals	$\underset{\vec{c}_r, T\,=\,1100\,^{\circ}\mathrm{C}}{\mathrm{Minimize}} \begin{bmatrix} -I_{SSS}(\vec{c}_r, T) \\ \rho(\vec{c}_r, T) \end{bmatrix}$
Constraints	
Design variables: $\mathbb{C}_{\mathrm{op}}$	Objective functions
$3.0[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Al})} \leqslant 7.0[\mathrm{wt\%}]$	$45[\mathrm{mol\%}] \leqslant \gamma'(\vec{c}_r, T) \leqslant 55[\mathrm{mol\%}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Ti})} \leqslant 5.0[\mathrm{wt\%}]$	$-0.6[\%] \leqslant \delta_{\gamma/\gamma'}(\vec{c}_r, T) \leqslant -0.4[\%]$
$5.0[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Cr})} \leqslant 15[\mathrm{wt\%}]$	$W_{ht}(\vec{c}_r, T) \geqslant 70[^{\circ}\mathrm{C}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Co})} \leqslant 15[\mathrm{wt\%}]$	$T_{\gamma'}(\vec{c}_r, T) \geqslant 1240[^{\circ}\mathrm{C}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Mo})} \leqslant 10[\mathrm{wt\%}]$	$T_{\mathrm{sol}}(\vec{c}_r, T) \geqslant 1330[^{\circ}\mathrm{C}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Ta})} \leqslant 10[\mathrm{wt\%}]$	$T_{\mathrm{liq}}(\vec{c}_r, T) \geqslant 1355[^{\circ}\mathrm{C}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{W})} \leqslant 12[\mathrm{wt\%}]$	$p(\vec{c}_r) \leqslant 230[€/\mathrm{kg}]$
$0.2[\mathrm{wt\%}] \leqslant c_r^{(\mathrm{Re})} \leqslant 5.0[\mathrm{wt\%}]$	$\rho(\vec{c}_r, T) \leqslant 8300[\mathrm{kg\ m^{-3}}]$

additional constraints (see table 2). It should be noted that the fast sensitivity calculation for the formation of third phases $S_{3\mathrm{rd}}^{\mathrm{p}}(\vec{c}, \Delta\vec{c}, T)$ is not performed in the current work and has to be applied in post-optimization for the whole Pareto-front or selected alloys only with direct and therefore extensive sensitivity calculations $S_{3\mathrm{rd}}(\vec{c}, \Delta\vec{c}, T)$ as given by equations (1) or (2), if required.

2.5. Implementation details

Regression analysis with stepwise variable selection and testing of the null hypothesis H_0 (an extended review of the method is given by Nickerson [38]) were performed with the built-in functions of MATLAB[7] software (version 2016a 64-bit), such as stepwise [32], swtest [39], lillietest [40, 41], adtest [42], and jbtest [43]. The in-house-developed alloy-by-design tool MultOpt++ (for details the readers are referred to [5] and [23]) was used for the CALPHAD calculations with the programming library TC-API version 2017b (64-bit) based on Thermo-Calc version 2017b[8]. All of the Thermo-Calc calculations were performed with the TTNi8[9] database and deactivated global minimization. The optimization kernel of MultOpt++ is the Geneva Ivrea–Via Arduino 1.6.1[10] optimization library [44] and it was modified to enable storage of a whole Pareto-front during the optimization process. For multi-criteria optimization, the EA was applied.

3. Results and discussion

3.1. Model selection

One of the main aims of statistical analysis was to identify the MAM of each alloy property to be used in MOOP. Usually, physical-based models are preferred to any formal mathematical

[7] MathWorks GmbH, Ismaning, Germany.
[8] Thermo-Calc Software AB, Solna, Sweden.
[9] Thermotech, Surrey, UK.
[10] Gemfony Scientific UG, Eggenstein-Leopoldshafen, Germany.

description, such as linear regression, and, ideally, the investigator who collects or generates the data should select such a model. However, if it is not possible to determine which model is the MAM based on expert knowledge, some quantitative statistical goodness-of-fit measure can be applied to compare competing models, such as the Akaike information criterion (AIC) (for detailed information refer to [32]), RMSE, or the coefficient of determination (R^2). Compared with the well-known RMSE and R^2 measures, application of the AIC for model comparison has the advantage that the number of model parameters is taken into account as a penalty term and it can thus avoid the so-called over-fitting problem. Taking into account all of the considered criteria, the MAM is the model with the smallest values of the AIC and RMSE statistics and the highest R^2 value. Thus, for all of the property distributions described with regression-based models using stepwise variable selection, we prefer the model that can achieve at least the same quality of prediction according to the best values of the considered statistical goodness-of-fit measures with full models but reduces the number of regression coefficients. Therefore, the $\sigma_{f_k}^{\mathrm{SLIQ}}(\vec{x})$ model from equation (7) was selected for further sensitivity predictions. Because of the large number of coefficients of the $\sigma_{f_k}^{\mathrm{SLIQ}}(\vec{x})$ model (up to 210), we focus in discussion on the $\sigma_{f_k}^{\mathrm{FL}}(\vec{x})$ or $\sigma_{f_k}^{\mathrm{SL}}(\vec{x})$ linear model.

3.2. Distribution calculation

The assumption of a normal distribution of objective function values $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T) \equiv \sigma_{f_k}$ owing to variation of the design variables is satisfied for more than 90% (see table 3) of the CALPHAD-based functions used in the current work by the null hypothesis H_0^{rej} test. This means that the function values of the remaining compositions have either a nonlinear trend close to the selected points within the $\Delta\vec{c}$ range or the distribution of randomly selected alloys is already a non-Gaussian distribution, for example, because of cutting off negative concentration values owing to limitation by the lower boundary.

For simulation of the uncertainty of a manufacturing process, the content of each element (i) is assumed to be normal distributed $\mathcal{N}(\mu, \sigma)$ with a mean value $\mu^{(i)}$ equal to the nominal alloy composition $c^{(i)}$. Here, the standard deviation $\sigma^{(i)}$ is defined as half of the respective process inaccuracy $\Delta c^{(i)}$, resulting in 95.45% of all of the alloys lying within the $\vec{c} \pm \Delta\vec{c}$ concentration range. The corresponding property inaccuracy $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T) = \sigma_{f_k}\left(\mathcal{N}\left(\vec{c}, \frac{1}{2}\Delta\vec{c}\right), T\right)$ is then calculated with 1000 randomly selected derivatives $\vec{c}_{\mathrm{rnd}}$ from a composition distribution defined by

$$\mathcal{N}\left(\vec{\mu} = \vec{c}, \vec{\sigma} = \frac{1}{2}\Delta\vec{c}\right) = \frac{1}{\frac{1}{2}\Delta\vec{c}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\vec{c} - \vec{c}_{\mathrm{rnd}}}{\frac{1}{2}\Delta\vec{c}}\right)^2\right). \tag{8}$$

In this case, the relative error in comparison with the distribution width calculated with 25 000 points is ≈2%. The inaccuracy defined in equation (8) can be adapted to a particular manufacturing process. So, e.g. a uniform distribution of elements can also be assumed within the inaccuracy range or the specification limits. Since such values of a process are usually known in weight percent, the concentration units in our model are also in weight percent. The variations in molar fraction can be recalculated by the CALPHAD method, if required. The property distribution calculation was performed with the TTNi8[11] database, which can differ for other databases but can still be used as a guidance value.

[11] Thermotech, Surrey, UK.

Table 3. Rejected fraction of null hypothesis H_0^{rej} tests of the composite normality (Matlab functions: *swtest* for the Shapiro–Wilk parametric hypothesis test [39], *lillietest* for the Lilliefors test [40, 41], *adtest* for the Anderson–Darling test [42], and *jbtest* for the Jarque–Bera test [43]) at the 0.05 [−] significance level for test alloys within the range defined in table 1. The mean rejected fraction $\langle H_0^{\text{rej}} \rangle$ represents the fraction of nonlinear property variations of the σ_{f_k} value within the $\Delta\vec{c}$ range. The analyzed property function $f_k(\vec{c}, T)$ distribution was calculated with 1000 derivatives randomly selected from the Gaussian distribution $\mathcal{N}\left(\mu = \vec{c}_0, \sigma^2 = \left(\frac{1}{2}\Delta\vec{c}\right)^2\right)$. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, p, $T_{\gamma'}$, T_{sol}, and T_{liq} are the solid solution strengthening index, density, misfit between the γ and γ' phases, γ' fraction, cost, γ' solvus, and solidus and liquidus temperatures, respectively.

H_0^{test} f_k	Swtest	Lillietest	Adtest	Jbtest	$\langle H_0^{\text{rej}} \rangle$ [−]
$\rho(\vec{c}, \Delta\vec{c}, T)$	0.058	0.040	0.038	0.044	0.045 ± 0.009
$\gamma'(\vec{c}, \Delta\vec{c}, T)$	0.091	0.059	0.069	0.066	0.071 ± 0.014
$T_{\text{liq}}(\vec{c}, \Delta\vec{c}, T)$	0.083	0.052	0.067	0.076	0.070 ± 0.013
$\delta_{\gamma/\gamma'}(\vec{c}, \Delta\vec{c}, T)$	0.065	0.047	0.060	0.028	0.050 ± 0.016
$p(\vec{c}, \Delta\vec{c}, T)$	0.065	0.049	0.061	0.039	0.053 ± 0.012
$I_{SSS}(\vec{c}, \Delta\vec{c}, T)$	0.078	0.048	0.055	0.078	0.065 ± 0.016
$T_{\text{sol}}(\vec{c}, \Delta\vec{c}, T)$	0.088	0.061	0.076	0.044	0.067 ± 0.019
$T_{\gamma'}(\vec{c}, \Delta\vec{c}, T)$	0.098	0.067	0.096	0.086	0.087 ± 0.014

3.3. Regression coefficients

In general, for a given linear function

$$f(\vec{x}) = \sum_{i=0}^{N} \alpha_i x^{(i)}, \tag{9}$$

where the regression coefficients α_i indicate the influence of a single design variable on the function variation. For further details, see [45]. The correctness of a linear model is correlated with the determination value R^2 and should ideally be close to 1 [29]. The mean R^2 value for all of the SLIQ property models calculated in the current work is 0.958 ± 0.013, while for FL it is 0.851 ± 0.051 (table 4). Nevertheless, as previously mentioned, the smaller number of coefficients in the FL model simplifies interpretation of the effect of each element and their uncertainty with respect to the whole property distribution $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)$. For better comparison of each predictor and thus its influence on the models in weight and atomic fractions, we used the standardized regression coefficients. Each input parameter $x^{(i)}$, as defined in equations (5)–(7), was standardized by the mean parameter value $\langle x^{(i)} \rangle$ and the corresponding mean values of the standard deviations within the whole design variables range defined in table 1 in weight $\sigma_{\text{wt}}^{(i)}$ and atomic $\sigma_{\text{at}}^{(i)}$ fractions:

$$x_j^{(i)} = \frac{2 \cdot (x_j^{(i)} - \langle x^{(i)} \rangle)}{\sigma_{\text{at}}^{(i)} + \sigma_{\text{wt}}^{(i)}}. \tag{10}$$

The coefficients for each function were normalized by the absolute coefficient value.

From a materials science point of view, the alloy designer can directly derive the effect of each element on the alloy sensitivity and its variation from the coefficient value, as shown for $\sigma_{f_k}^{\text{FL}}(\vec{c}, \Delta\vec{c}, T)$ in weight (figure 2(a)) and atomic (figure 2(c)) fractions. According to

Table 4. Comparison of the regression models. The best values are shown in *italics* and the selected model for further calculations is shown in **bold**. For better comparability of the RMSE, the standardized function values were used. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, W_{ht}, p, $T_{\gamma'}$, T_{sol}, and T_{liq} are solid solution strengthening index, density, misfit between the γ and γ' phases, γ' fraction, heat treatment window, cost, γ' solvus, and solidus and liquidus temperatures, respectively.

f_k		$\sigma_{f_k}^{\mathrm{FL}}$	$\sigma_{f_k}^{\mathrm{SL}}$	$\sigma_{f_k}^{\mathrm{FLI}}$	$\sigma_{f_k}^{\mathrm{SLI}}$	$\sigma_{f_k}^{\mathrm{FLIQ}}$	$\sigma_{f_k}^{SLIQ}$
ρ	AIC $\cdot 10^3$	54.08	54.07	31.83	31.76	−27.13	−27.21
	RMSE	0.322	0.322	0.286	0.286	0.210	0.210
	R^2	0.897	0.897	0.918	0.918	0.956	0.956
W_{ht}	AIC $\cdot 10^3$	79.16	79.16	−38.57	−38.62	−82.18	−82.23
	RMSE	0.367	0.367	0.197	0.197	0.157	0.157
	R^2	0.865	0.865	0.961	0.961	0.975	0.975
T_{liq}	AIC $\cdot 10^3$	72.96	72.96	20.55	20.53	−23.25	−23.27
	RMSE	0.355	0.355	0.269	0.269	0.214	0.214
	R^2	0.874	0.874	0.928	0.928	0.954	0.954
$\delta_{\gamma/\gamma'}$	AIC $\cdot 10^3$	125.0	125.0	37.12	37.08	1.675	1.644
	RMSE	0.467	0.467	0.294	0.294	0.244	0.244
	R^2	0.782	0.782	0.914	0.914	0.941	0.941
p	AIC $\cdot 10^3$	31.84	31.83	−9.737	−9.818	−38.32	−38.44
	RMSE	0.286	0.286	0.230	0.230	0.198	0.198
	R^2	0.918	0.918	0.947	0.947	0.961	0.961
T_{sol}	AIC $\cdot 10^3$	63.62	63.62	−17.76	−17.78	−71.03	−71.04
	RMSE	0.338	0.338	0.220	0.220	0.166	0.166
	R^2	0.886	0.886	0.952	0.952	0.972	0.972
I_{SSS}	AIC $\cdot 10^3$	132.7	132.7	37.71	37.65	12.59	12.55
	RMSE	0.486	0.486	0.295	0.295	0.258	0.258
	R^2	0.764	0.764	0.913	0.913	0.933	0.933
$T_{\gamma'}$	AIC $\cdot 10^3$	113.5	113.5	−2.463	−2.505	−66.43	−66.47
	RMSE	0.439	0.439	0.239	0.239	0.170	0.170
	R^2	0.807	0.807	0.943	0.943	0.971	0.971
γ'	AIC $\cdot 10^3$	75.34	75.34	−5.219	−5.271	−34.31	−34.37
	RMSE	0.360	0.360	0.235	0.235	0.202	0.202
	R^2	0.871	0.871	0.945	0.945	0.959	0.959

figure 2, the influence of each element on the property value $f_k(\vec{c}, T)$ (figures 2(b) and (d)) and the corresponding distribution $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)$ (figures 2(a) and (c)) can be easily determined.

Let us, for example, consider an optimization with the aim of a high amount of solid solution strengthening elements in the matrix given by the $I_{SSS}(\vec{c}, T)$ value. According to figure 2(b), the most effective elements for increasing the I_{SSS} index are Al and Ti, because they reduce the matrix phase fraction, and Re, Mo, and W, which act as solid solution strengthening elements. However, high amounts of these elements increase the sensitivity of the $\sigma_{I_{SSS}}(\vec{c}, \Delta\vec{c}, T)$ value, which can be seen in figure 2(a) (red squares). A slightly positive effect on the sensitivity can be achieved by increasing the amount of Cr (see figure 2(a)) (blue squares), which also has a negative effect on the solidus and liquidus temperatures (figure 2(b), blue squares).

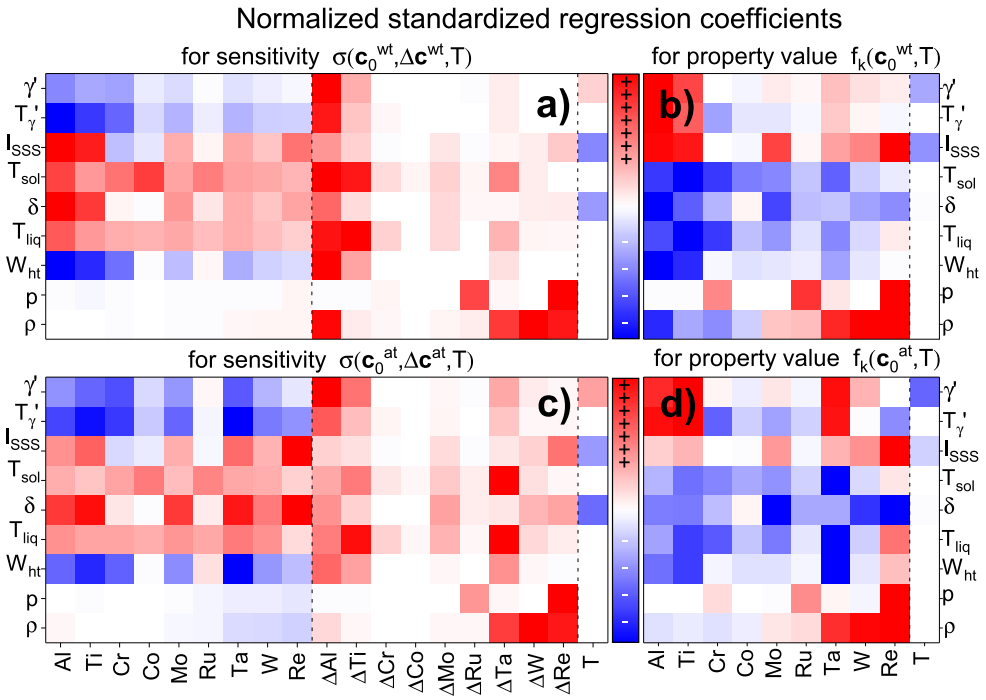

Figure 2. Normalized standardized regression coefficients of each element and their variations in wt% and at% for the (a) and (c) property variation $\sigma_{f_k}^{\mathrm{FL}}(\vec{c}, \Delta\vec{c}, T)$ and (b) and (d) property value within the FL regression model. The input parameters have been centered and divided by the mean distribution width of input parameters in weight and atomic fraction: $x_j^{(i)} = [2 \cdot (x_j^{(i)} - \langle x^{(i)} \rangle)] / [\sigma_{\mathrm{at}}^{(i)} + \sigma_{\mathrm{wt}}^{(i)}]$. Normalization was performed with the maximal absolute value of the coefficients for a certain function. The influence of each element on a given property is given by the color of the square: blue and red indicate increasing and reducing effect of the property distribution or property value, respectively. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, W_{ht}, p, $T_{\gamma'}$, T_{sol}, and T_{liq} are the solid solution strengthening index, density, misfit between the γ and γ' phases, γ' fraction, heat treatment window, cost, γ' solvus, and solidus and liquidus temperatures, respectively.

According to figure 2, the variations of Al and Ti cause the strongest property deviations for almost all of the properties, so their contents should be set as precisely as possible. Conversely, variation of cobalt has almost no effect because it exhibits similar characteristics to nickel. It should be noted that the variation of one element is balanced by the variation of the main element nickel, and therefore the coefficients describe the effects of both elements. In addition, the effect of ruthenium for CALPHAD-based calculations is insignificant, partly because of an incomplete description in the TTNi8 database, which has previously been shown by Matuszewski *et al* [46] and Ritter *et al* [26].

The evaluated coefficients for the SL model are given in table 5. These coefficients can be stored in thermodynamic databases for faster prediction of the distribution width of available properties. Direct calculation of distributions is already implemented in Thermo-Calc 2016b software, but it requires a high number of samples (several hundred for a complex Ni-based superalloy) to be calculated and cannot be included in the optimization routine.

Table 5. Regression coefficients $a^{(i)}$ of the stepwise calculated linear model for the distribution width $\sigma^{\mathrm{SL}}_{f_k}(\vec{c}, \Delta\vec{c}, T)$ defined in equation (5) in dependence on element concentrations $c^{(i)}$ and their variations $\Delta c^{(i)}$ in weight fraction. I_{SSS}, ρ, γ', $\delta_{\gamma/\gamma'}$, p, $T_{\gamma'}$, T_{sol} and T_{liq} are solid solution strengthening index, density, misfit between γ and γ' phases, γ'-fraction, price, γ'-solvus, solidus and liquidus temperatures, respectively.

f_k / $a^{(i)}$	ρ	γ'	W_{ht}	T_{liq}	$\delta_{\gamma/\gamma'}$	p	T_{sol}	I_{SSS}	$T_{\gamma'}$
$a^{(0)}$	2.399	−0.010	20.15	−2.040	0.020	0.715	−3.613	0.592	18.66
$a^{(\mathrm{Al})}$	—	−21.49	−181.0	16.45	1.128	−0.786	23.46	13.17	−159.1
$a^{(\mathrm{Ti})}$	1.620	−13.89	−130.9	8.909	0.768	−1.204	10.83	10.28	−108.6
$a^{(\mathrm{Cr})}$	—	−8.122	−46.03	3.755	0.019	−0.211	7.931	−1.525	−43.75
$a^{(\mathrm{Co})}$	—	−0.996	−0.322	2.115	—	—	6.610	−0.330	−6.435
$a^{(\mathrm{Mo})}$	—	−4.709	−34.24	6.150	0.341	−0.261	8.324	3.007	−34.61
$a^{(\mathrm{Ru})}$	−1.094	−0.081	3.469	5.886	0.097	−0.585	14.35	0.487	−8.787
$a^{(\mathrm{Ta})}$	2.468	−4.488	−46.97	6.513	0.291	−0.238	9.694	3.680	−38.43
$a^{(\mathrm{W})}$	3.951	−2.374	−24.52	5.431	0.211	—	8.361	2.470	−24.35
$a^{(\mathrm{Re})}$	2.885	−1.202	−17.94	3.585	0.295	2.448	6.501	5.218	−20.10
$a^{(\Delta\mathrm{Al})} \times 10^{-2}$	10.94	5.594	21.45	2.838	0.081	−0.104	3.866	0.650	17.23
$a^{(\Delta\mathrm{Ti})} \times 10^{-2}$	1.142	2.388	10.60	4.178	0.026	−0.132	4.742	0.392	5.681
$a^{(\Delta\mathrm{Cr})} \times 10^{-2}$	0.333	0.041	0.122	0.750	0.001	0.719	0.798	—	0.874
$a^{(\Delta\mathrm{Co})} \times 10^{-2}$	0.199	—	—	0.051	—	0.077	0.190	—	—
$a^{(\Delta\mathrm{Mo})} \times 10^{-2}$	0.631	0.070	0.276	0.707	0.029	—	0.972	0.217	0.427
$a^{(\Delta\mathrm{Ru})} \times 10^{-2}$	0.955	—	0.098	0.063	0.004	8.641	0.284	−0.012	0.099
$a^{(\Delta\mathrm{Ta})} \times 10^{-2}$	12.07	0.443	3.584	1.275	0.005	0.314	2.526	0.098	1.455
$a^{(\Delta\mathrm{W})} \times 10^{-2}$	15.54	0.119	0.308	0.198	0.012	0.060	0.295	0.145	0.212
$a^{(\Delta\mathrm{Re})} \times 10^{-2}$	14.11	0.049	0.066	0.103	0.015	12.921	0.080	0.468	0.156
$a^{(T)} \times 10^{-3}$	—	2.052	—	—	−0.105	—	—	−1.462	—

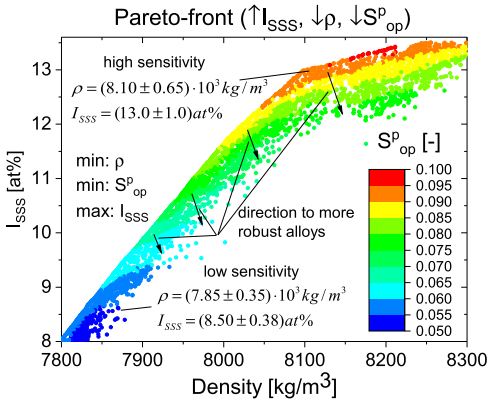

Figure 3. Pareto-front (↑ I_{SSS}, ↓ ρ, ↓ S^{p}_{op}) of optimization with predicted sensitivity $S^{p}_{op} = \sqrt{(S^{p}_{\rho})^2 + (S^{p}_{I_{SSS}})^2}$ as an additional minimization ↓ goal. The arrows indicate the direction to more robust solutions with inferior (I_{SSS}, ρ) values. Two alloys with high and low I_{SSS} sensitivity are selected as examples to indicate the different expected errors ($I^{high}_{SSS} = (13.0 \pm 1.0)$at% and $I^{low}_{SSS} = (8.50 \pm 0.38)$at%). The corresponding uncertainty value of each element is $\Delta c_i = 0.2$ wt%.

3.4. Optimization

The sensitivity of the optimization problem defined in equation (4) can be extended by any number of constraint functions n_{con}, as well as user-defined weighting factors w_k for each sensitivity depending on the requirements of the optimization goals

$$S_{op}(\vec{c}, \Delta\vec{c}, T) = \sqrt{\sum_{k=1}^{n_{goal}+n_{con}} w_k [S_{f_k}(\vec{c}, \Delta\vec{c}, T)]^2}\,. \tag{11}$$

Such an extension plays an important role if robust values of the constrained functions are also required (for example, a more stable cost of an alloy).

The result of the optimization problem (see table 2) by minimization (↓) of $S^{p}_{op} = \sqrt{(S^{p}_{\rho})^2 + (S^{p}_{I_{SSS}})^2}$ is shown in figure 3.

With this additional information, more robust alloys can be selected accepting somewhat inferior values of I_{SSS}, ρ or both. The density sensitivity only has a minor influence on the S^{p}_{op} value. In the case of equal weighting factors of $w_{I_{SSS}}$ and w_{ρ}, the preferred alloy will be the one with a low I_{SSS} value because this will lead to lower sensitivity. Because the S^{p}_{op} value is the squared sum of all of the single sensitivities, the sensitivity of each $S^{p}_{f_k}$ value can be recalculated in a post-optimization process, if required. Alternatively, each sensitivity S_{f_k} can be used as an additional optimization goal accepting a longer optimization time.

In the first instance, the different function slopes of the property functions f_k will have an effect on the property sensitivity S_{f_k}, as shown in figure 1(b). In addition, switching between interpolation functions within a CALPHAD database independently of the different temperature ranges, especially if temperature deviation ΔT is also considered, which is ignored in the current work, also influences S_{f_k}. Therefore, the sensitivity S_{f_k} can also be interpreted as the maximum relative error for a given input parameter distribution of optimized properties, giving the alloy designer the ability to select candidates with properties that are more stable or predicted more accurately by CALPHAD. This is indicated in figure 3 by the expected

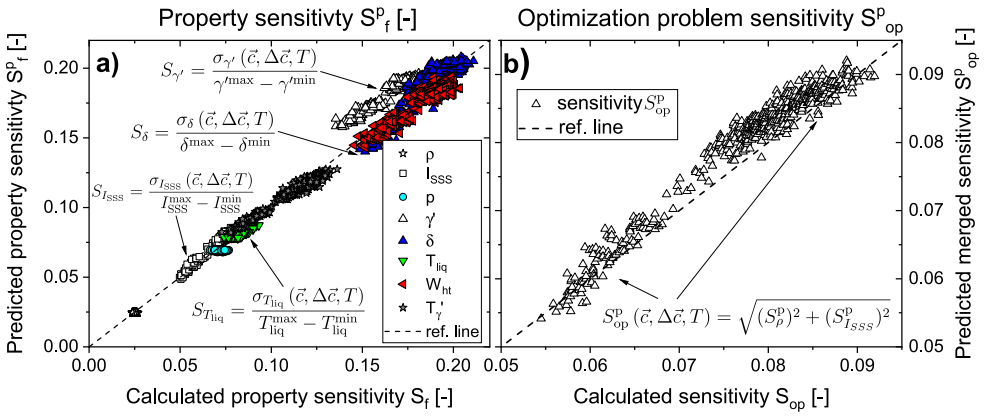

Figure 4. (a) Accuracy of the predicted sensitivity $S^{\mathrm{p}}_{f_k}$ for each property and (b) the combined sensitivity $S^{\mathrm{p}}_{\mathrm{op}} = \sqrt{S^2_{\rho} + S^2_{I_{SSS}}}$ compared with the recalculated S_{f_k} and S_{op} for optimal Pareto-alloys of the current optimization problem.

absolute errors for alloys with high and low sensitivity ($I^{\mathrm{high}}_{SSS} = (13.0 \pm 1.0)$at% compared with $I^{\mathrm{low}}_{SSS} = (8.50 \pm 0.38)$at%).

3.5. Model accuracy

Because all of the proposed models are based on prediction of the property distributions, the accuracy of σ_{f_k} is important for the model correctness. The RMSE of the standardized function values and R^2 give a good indication of the model quality (table 4). The lowest calculated R^2 value for all of the models is 0.95 for the $\sigma_{\delta_{\gamma/\gamma'}}$ model, meaning that only approximately 5% cannot be described by the selected regression model. The recalculated property distribution σ_{f_k} of the randomly selected optimal compositions $\vec{c}_{\mathrm{op}}$ is compared with the predicted property distribution $\sigma^{\mathrm{p}}_{f_k}$ in figure 4(a). Because the absolute distribution value S_{f_k} depends on the Pareto-range $f^{\max}_k(\bar{\vec{c}}_{\mathrm{op}}) - f^{\min}_k(\bar{\vec{c}}_{\mathrm{op}})$ and therefore on the optimization problem itself, we determined the relative deviation values. They are between 3% and 5% for all of the sensitivities, except for $S_{\gamma'}$ with 8.5%. The overestimation of $S_{\gamma'}$ can be explained by the reduced optimization range $\bar{\bar{c}}_{\mathrm{par}} \in \mathbb{C}_{\mathrm{op}} \in \mathbb{C}_{\mathrm{reg}}$ compared with the whole regression analysis range $\mathbb{C}_{\mathrm{reg}} = [\vec{c}_{\min}, \vec{c}_{\max}]$ defined in table 1 or by stronger model deviations at the regression range boundaries (e.g. at the maximum temperature value of 1100 °C in the current case). Because we want to obtain a general model for the property distribution, we omit recalculation of $S_{\gamma'}$ within the reduced $\mathbb{C}_{\mathrm{op}}$ range. The accuracy of $S^{\mathrm{p}}_{\mathrm{op}} = \sqrt{(S^{\mathrm{p}}_{\rho})^2 + (S^{\mathrm{p}}_{I_{SSS}})^2}$ used as an optimized goal is shown in figure 4(b) with a relative error of 4.0%. With respect to the sensitivity of all of the constrained functions defined in equation (11), the predicted sensitivity error is equal to 13.0%.

3.6. Inaccuracy of the CALPHAD calculations

All of the CALPHAD calculations are assumed to be exact. However, scattering of the experimental data or deviations in the *ab initio* simulations result in uncertainty in the determined interaction parameters stored in the thermodynamic databases, and finally in

uncertainty in the CALPHAD calculations [24, 25, 47]. Especially the description of the TCP phases is quite poor [27, 28, 46]. Nevertheless, despite this insufficient accuracy, one has the advantage of separating potentially TCP-forming from potentially non-forming alloys. The predictive accuracy obtained is sufficient to test the stability of a large number of highly promising alloys determined by optimizing well-known properties and to reduce the post-processing effort in experimental validation significantly, taking into account the possible neglection of even better alloys.

For some alloy properties, the CALPHAD prediction inaccuracy $\sigma_{f_k}^{\mathrm{db}}$ with optional offset f_k^{off} of a certain measurement value $f_k(\vec{c}, T)$ can be estimated by comparing the experimental and calculated data. Here, the offset f_k^{off} describes the systematic over- or underestimation of a particular value $f_k(\vec{c}, T)$. A comparison between the liquidus, solidus, and γ' solvus temperatures calculated with Thermo-Calc using the TTNi7 database and the respective experimental values can be found in Rettig *et al* [24]. Replacing $\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)$ by a value $\sigma_{f_k}^{\mathrm{db}}$ that may not be concentration dependent and replacing $f_k(\vec{c}, T)$ by the offset-corrected function $\tilde{f}_k(\vec{c}, T) = f_k(\vec{c}, T) + f_k^{\mathrm{off}}$ in equation (3) results in a modified equation for the sensitivity $S_{\tilde{f}_k}^{\mathrm{db}}$. In this case, the $S_{\tilde{f}_k}^{\mathrm{db}}$ value predicts the property distribution owing to the inaccuracy of the thermodynamic database. The scattering of the element concentrations during the manufacturing process $\Delta\vec{c}$ and the estimated database inaccuracy $\sigma_{f_k}^{\mathrm{db}}$ can be connected in the extended distribution $\sigma_{f_k}^{\mathrm{ext}}(\vec{c}, \Delta\vec{c}, T) = \sqrt{(\sigma_{f_k}^{\mathrm{db}})^2 + (\sigma_{f_k}(\vec{c}, \Delta\vec{c}, T))^2}$. This allows more confident alloy selection because $\sigma_{f_k}^{\mathrm{ext}}(\vec{c}, \Delta\vec{c}, T) \geqslant \sigma_{f_k}(\vec{c}, \Delta\vec{c}, T)$.

4. Conclusions

A fast regression-based model for the property sensitivity has been derived and implemented in the optimization work flow of nickel-based superalloys, giving additional information about the distribution width of the optimized properties caused by inaccuracies in the input parameters. This novel contribution allows online sensitivity optimization with negligible increase of the computational time. Here, the model allows the selection of optimal compositions with direct consideration of possible deviations during a manufacturing process or of the property deviation within a certain specification limit. The derived method can be applied to almost any CALPHAD-based or semi-empirical model for alloy properties requiring only one-time pre-calculation of the regression coefficients. Furthermore, the model can be easily extended to take into account the imprecision of the thermodynamic databases.

By minimizing the sensitivity as an additional goal, the resulting Pareto-front indicates the direction to more stable alloys or alloys with smaller predicted property inaccuracy. Two alloys with low and high sensitivities have been discussed as examples.

The contribution of each element to the property distribution has also been calculated, which allows the most important elements that cause the greatest property uncertainty in the designed optimal alloy to be identified and therefore allows possible deviations to be controlled.

The average relative error of the predicted distribution width is ≈4%, and the resulting Pareto-front can be recalculated in a post-optimization process with a more exact method.

In the future, the model has to be extended by the sensitivity calculation for third and TCP phases, when better databases are available.

Acknowledgments

The authors acknowledge funding from the German Science Foundation (DFG) in the framework of the Collaborative Research Centre/Transregio 103 (projects B1, C6 and C7). We thank R Berlich and his team for their support regarding the Geneva optimization library. We thank Edanz Group[12] for editing a draft of this manuscript.

ORCID iDs

Alexander Müller https://orcid.org/0000-0003-3586-2965

References

[1] Xiong W and Olson G B 2016 Cybermaterials: materials by design and accelerated insertion of materials *npj Comput. Mater.* **2** 15009
[2] Rae C M F 2009 Alloys by design: modelling next generation superalloys *Mater. Sci. Technol.* **25** 479–87
[3] Reed R C, Tao T and Warnken N 2009 Alloys-by-design: application to nickel-based single crystal superalloys *Acta Mater.* **57** 5898–913
[4] Saunders N and Miodownik A P 1998 *CALPHAD: Calculation of Phase Diagrams: a Comprehensive Guide* (*Pergamon Materials Series*) (Oxford: Pergamon)
[5] Rettig R, Ritter N C, Helmer H E, Neumeier S and Singer R F 2015 Single-crystal nickel-based superalloys developed by numerical multi-criteria optimization techniques: design based on thermodynamic calculations and experimental validation *Modelling Simul. Mater. Sci. Eng.* **23** 035004
[6] Rettig R, Singer R, Neumeier S and Ritter N 2017 Creep-resistant, rhenium-free nickel base superalloy *US Patent* 2017/9580774 B2
[7] Rettig R, Matuszewski K, Müller A, Helmer H, Ritter N and Singer R 2016 Development of a low-density rhenium-free single crystal nickel-based superalloy by application of numerical multi-criteria optimization using thermodynamic calculations *13th Proc. Int. Symp. Superalloys* pp 35–44
[8] Göhler T, Rettig R, Singer R, Neumeier S and Ritter N 2017 Rhenium-free nickel base superalloy of low density *US Patent* 2017/0058383 A1
[9] Olson G B and Kuehmann C J 2014 Materials genomics: from CALPHAD to flight *Scr. Mater.* **70** 25–30
[10] Saunders N, Fahrmann M and Small C 2000 Application of CALPHAD calculations to Ni-based supperalloys *Superalloys* (TMS)
[11] Avraham S, Maoz Y and Bamberger M 2007 Application of the CALPHAD approach to Mg-alloys design *CALPHAD, Comput. Coupling Phase Diagr. Thermochem.* **31** 515–21
[12] Gheribi A E, Digabel S L, Audet C and Chartrand P 2013 Identifying optimal conditions for magnesium based alloy design using the mesh adaptive direct search algorithm *Thermochim. Acta* **559** 107–10
[13] Gheribi A E, Harvey J-P, Bélisle E, Robelin C, Chartrand P, Pelton A D, Bale C W and Digabel S Le 2016 Use of a biobjective direct search algorithm in the process design of material science applications *Optim. Eng.* **17** 27–45
[14] Choudhury A 1992 State of the art of superalloy production for aerospace and other application using VIM/VAR or VIM/ESR *ISIJ Int.* **32** 563–74
[15] Heckl A, Rettig R and Singer R 2009 Solidification characteristics and segregation behavior of nickel-base superalloys in dependence on different rhenium and ruthenium contents *Metall. Mater. Trans.* A **41** 202–11
[16] Alcock C B, Itkin V P and Horrigan M K 1984 Vapour pressure equations for the metallic elements: 298–2500 K *Can. Metall. Q.* **23** 309–13

[12] www.edanzediting.com/ac.

[17] Hofmeister M, Franke M M, Koerner C and Singer R F 2017 Single crystal casting with fluidized carbon bed cooling: a process innovation for quality improvement and cost reduction *Metall. Mater. Trans.* B **48** 3132–42
[18] Dulikravich G S and Egorov-Yegorov I N 2005 Robust optimization of concentrations of alloying elements in steel for maximum temperature, strength, time-to-rupture and minimum cost and weight *Int. Conf. on Computational Methods for Coupled Problems in Science and Engineering* ed E O Papadrakakis and B Schreer
[19] Deb K and Gupta H 2006 Introducing robustness in multi-objective optimization *Evol. Comput.* **14** 463–94
[20] Parmee I I 1996 The maintenance of search diversity for effective design space decomposition using cluster-oriented genetic algorithms (COGAs) and multi-agent strategies (GAANT) *Proc. of Adaptive Computing in Engineering Design and Control* ed I I Parmee and M J Denham
[21] Deb K and Gupta H 2005 Searching for robust Pareto-optimal solutions in multi-objective optimization *Evolutionary Multi-Criterion Optimization* ed C A Coello Coello, A Hernandez Aguirre and E Zitzler (Berlin: Springer) pp 150–64
[22] Beyer H-G and Sendhoff B 2007 Robust optimization—a comprehensive survey *Comput. Methods Appl. Mech. Eng.* **196** 3190–218
[23] Markl M *et al* 2018 Development of single-crystal Ni-base superalloys based on multi-criteria numerical optimization and efficient use of refractory elements *Metall. Mater. Trans.* A **49** 4134–45
[24] Rettig R, Heckl A, Neumeier S, Pyczak F, Göken M and Singer R 2009 Verification of a commercial CALPHAD database for Re and Ru containing nickel-base superalloys *Defect Diffus. Forum* **289–292** 101–8
[25] Ritter N, Schesler E, Müller A, Rettig R, Körner C and Singer R 2017 On the influence of Ta and Ti on heat-treatability and γ/γ'-partitioning of high W containing Re-free nickel-based superalloys *Adv. Eng. Mater.* **9** 1700150
[26] Ritter N C, Sowa R, Schauer J C, Gruber D, Goehler T, Rettig R, Povoden-Karadeniz E, Koerner C and Singer R F 2018 Effects of solid solution strengthening elements Mo, Re, Ru, and W on transition temperatures in nickel-based superalloys with high γ/γ'-volume fraction: comparison of experiment and CALPHAD calculations *Metall. Mater. Trans.* A **49** 3206–16
[27] Matuszewski K, Rettig R and Singer R F 2014 The effect of Ru on precipitation of topologically close packed phases in Re - containing N base superalloys: quantitative FIB-SEM investigation and 3D image modeling *MATEC Web Conf.* **14** 09001
[28] Zhao J-C and Henry M 2002 CALPHAD–is it ready for superalloy design? *Adv. Eng. Mater.* **4** 501–8
[29] Saltelli A, Tarantola S and Campolongo F 2000 Sensitivity analysis as an ingredient of modeling *Stat. Sci.* **15** 377–95
[30] Saltelli A 2008 *Global Sensitivity Analysis: The Primer* (Chichester: Wiley)
[31] Wang W, Caro S, Bennis F, Soto R and Crawford B 2014 Multi-objective robust optimization using a post-optimality sensitivity analysis technique: application to a wind turbine design *J. Mech. Des.* **137** 011403
[32] Burnham K P and Anderson D R 2003 *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (New York: Springer Science and Business Media) (https://doi.org/10.1007/b97636)
[33] Coello C A Coello 2002 Theoretical and numerical constraint-handling techniques used with evolutionary algorithms: a survey of the state of the art *Comput. Methods Appl. Mech. Eng.* **191** 1245–87
[34] Dulikravich G S, Jelisavcic N and Egorov I N 2006 Evolutionary optimization of chemistry of bulk metallic glasses *III European Conference on Computational Mechanics* ed C A Motasoares *et al* (Dordrecht: Springer) pp 413–413
[35] Deb K, Thiele L, Laumanns M and Zitzler E 2002 Scalable multi-objective optimization test problems *Proc. 2002 Congress on Evolutionary Computation, CEC* vol 1, pp 825–30
[36] Deb K 1999 Multi-objective genetic algorithms: problem difficulties and construction of test problems *Evol. Comput.* **7** 205–30
[37] Fleming P J, Purshouse R C and Lygoe R J 2005 Many-objective optimization: an engineering design perspective *Evolutionary Multi-Criterion Optimization* ed C A Coello Coello, A Hernández Aguirre and E Zitzler (Berlin: Springer) pp 14–32

[38] Nickerson R S 2000 Null hypothesis significance testing: a review of an old and continuing controversy *Psychol. Methods* **5** 241–301
[39] Royston P 1995 Algorithm as r94 *J. R. Stat. Soc.* C **44** 547–51
[40] Lilliefors H 1969 On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown *J. Am. Stat. Assoc.* **64** 387–9
[41] Lilliefors H 1967 On the Kolmogorov-Smirnov test for normality with mean and variance unknown *J. Am. Stat. Assoc.* **62** 399–402
[42] Anderson T W and Darling D A 1952 Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes *Ann. Math. Stat.* **23** 193–212
[43] Jarque C M and Bera A K 1987 A test for normality of observations and regression residuals *Int. Stat. Rev.* **55** 163–72
[44] Berlich R, Gabriel S and Garcia A 2011 Distributed parametric optimization with the Geneva Library *Data Driven e-Science* ed S Lin and E Yen (New York: Springer) (https://doi.org/10.1007/978-1-4419-8014-4_24)
[45] Helton J C 1993 Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal *Reliab. Eng. Syst. Saf.* **42** 327–67
[46] Matuszewski K, Müller A, Ritter N, Rettig R, Kurzydłowski K J and Singer R F 2015 On the thermodynamics and kinetics of tcp phase precipitation in Re− and Ru−containing Ni-base superalloys *Adv. Eng. Mater.* **17** 1127–33
[47] Copland E H, Jacobson N S and Ritzert F J 2001 Computational thermodynamic study to predict complex phase equilibria in the nickel-base superalloy René N6 *NASA Report* NASA/TM-2001-210897 pp 1–42