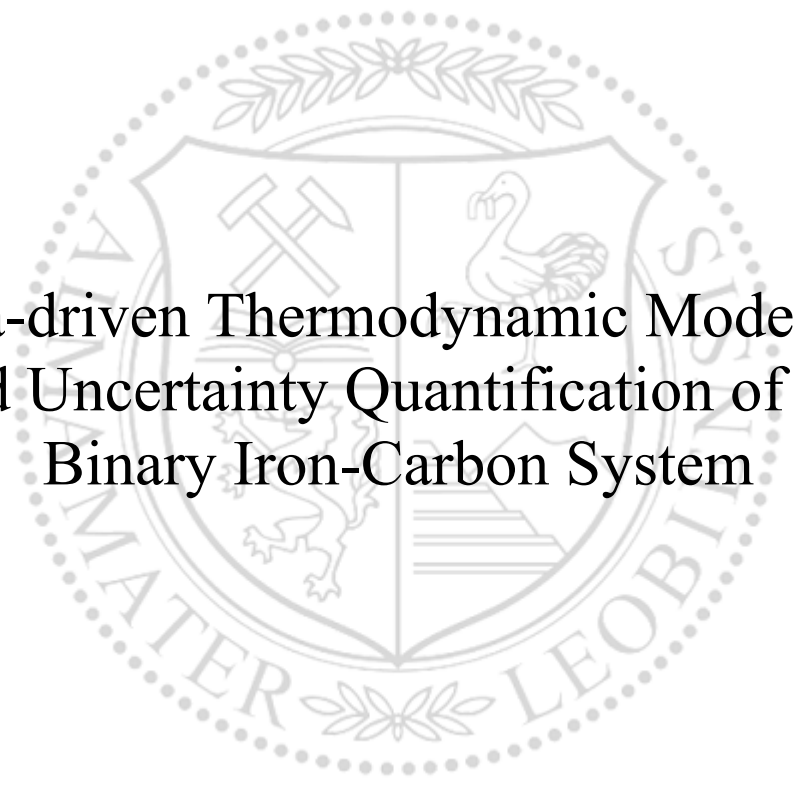




Chair of Physical Metallurgy and Metallic Materials

Master's Thesis



Data-driven Thermodynamic Modelling  
and Uncertainty Quantification of the  
Binary Iron-Carbon System

Johannes Ernst Bechter, BSc

September 2022



**EIDESSTÄTLICHE ERKLÄRUNG**

Ich erkläre an Eides statt, dass ich diese Arbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt, und mich auch sonst keiner unerlaubten Hilfsmittel bedient habe.

Ich erkläre, dass ich die Richtlinien des Senats der Montanuniversität Leoben zu "Gute wissenschaftliche Praxis" gelesen, verstanden und befolgt habe.

Weiters erkläre ich, dass die elektronische und gedruckte Version der eingereichten wissenschaftlichen Abschlussarbeit formal und inhaltlich identisch sind.

Datum 19.09.2022

---

Unterschrift Verfasser/in  
Johannes Ernst Bechter

# Acknowledgement

I want to thank my supervisor Univ.-Prof. Dr. Lorenz Romaner for giving me the opportunity to write my master's thesis at his chair and for his great support and guidance throughout this work.

Also a very big thank you to Dipl.-Ing. Tobias Spitaler who supported me with his helpful advice and expertise and also with his patience in helping me complete my thesis.

My special thanks to my friends, family and everybody who has believed in me and accompanied me on my way to concluding my studies. I am very grateful for having you by my side.

This thesis was composed under the scope of the COMET program within the K2 Center "Integrated Computational Material, Process and Product Engineering (IC-MPPE)" (Project No. 859480). This program is supported by the Austrian Federal Ministries for Transport, Innovation and Technology (BMVIT) and for Digital and Economic Affairs (BMDW), represented by the Austrian research funding association (FFG), and the federal states of Styria, Upper Austria and Tyrol.

# Abstract

The mostly accepted CALPHAD assessment of the binary iron-carbon system is from Gustafson (Scand. J. Metall. 14.5 (1985): 259-267). As is the case for most CALPHAD assessments, the proposed parametrization reports the chosen values of parameters without providing details about the procedure used to identify the optimal values and without information about their uncertainty or reliability. Therefore, the parametrization is not fully reproducible.

In this thesis, a database was created which contains the original thermodynamic data on phase boundaries, activities and formation enthalpies from experiments and ab-initio calculations along with the relevant meta-data specifying e.g. the original reference or experimental details. This python-based database allows adaption to user-specific requirements and easy reassessment at a later time, for example when new data is added.

The parameter optimization for the CALPHAD assessment was performed with ESPEI. The open-source software optimizes the parameters with Markov Chain Monte Carlo in the Bayesian framework and provides the associated probability distribution. This allows exploring propagation of parameter uncertainties and investigating the effect of choosing different sets of input data or model structure.

As a result, a new parametrization for the Fe-C system is presented which is fully reproducible and explains the thermodynamic data points with a higher probability compared to the parameter set proposed by Gustafson within the chosen error definition. The optimized phase diagram is presented along with the relating parameters in form of a TDB-file which considers the underlying data and estimates the uncertainties of the calculated phase boundaries. It is found that various data types help select reliable datasets and increase the accuracy of an assessment. Based on the calculated uncertainties it is pointed out that providing new data points in the eutectic and the high carbon-region can further improve the reliability of the assessment.

# Contents

|   |           |
|---|-----------|
| List of Abbreviations . . . . .   | VI        |
| <b>1 Introduction</b>   | <b>1</b>  |
| <b>2 Theory</b>   | <b>3</b>  |
| 2.1 Basic thermodynamics . . . . .  | 3         |
| 2.1.1 Laws of thermodynamics - enthalpy, heat capacity, entropy, Gibbs energy | 3         |
| 2.1.2 Solution phase thermodynamics . . . . .                                 | 4         |
| 2.1.3 Phase diagrams . . . . .  | 6         |
| 2.2 Binary iron-carbon system . . . . .                                       | 7         |
| 2.3 CALPHAD . . . . .   | 8         |
| 2.3.1 Introduction . . . . .  | 8         |
| 2.3.2 Model parametrization . . . . .   | 8         |
| 2.3.3 Software and databases . . . . .  | 11        |
| 2.3.4 Applications . . . . .  | 11        |
| <b>3 Computational methods</b>  | <b>13</b> |
| 3.1 Compound energy formalism . . . . .                                       | 13        |
| 3.1.1 Introduction . . . . .  | 13        |
| 3.1.2 Ordering and mixing effects . . . . .                                   | 14        |
| 3.1.3 Composition and temperature dependence . . . . .                        | 14        |
| 3.1.4 Phase descriptions . . . . .  | 15        |
| 3.2 ESPEI . . . . .   | 17        |
| 3.3 Bayesian inference and Markov Chain Monte Carlo . . . . .                 | 18        |
| 3.4 Priors, data weights and likelihood function . . . . .                    | 19        |
| 3.5 Error definition in ESPEI . . . . .                                       | 20        |
| 3.5.1 Phase boundary data . . . . .   | 20        |
| 3.5.2 Thermochemical data . . . . .   | 21        |
| 3.5.3 Activity data . . . . .   | 22        |
| <b>4 Results and discussion</b>   | <b>23</b> |
| 4.1 Database . . . . .  | 23        |
| 4.1.1 Motivation for a new database . . . . .                                 | 23        |
| 4.1.2 Covered data . . . . .  | 24        |

|          |   |           |
|----------|---|-----------|
| 4.2      | Results of ESPEI-optimization . . . . .                     | 33        |
| 4.2.1    | Probabilities, parameter evolutions, correlations . . . . . | 33        |
| 4.2.2    | Resulting phase diagram . . . . .                           | 36        |
| 4.2.3    | Modelled activities . . . . .                               | 42        |
| 4.2.4    | Modelled formation enthalpies . . . . .                     | 43        |
| 4.2.5    | Conflicting datasets . . . . .                              | 43        |
| 4.2.6    | Uncertainty quantification . . . . .                        | 45        |
| <b>5</b> | <b>Summary</b>  | <b>48</b> |
|          | <b>Appendix</b>   | <b>49</b> |
|          | <b>Bibliography</b>   | <b>52</b> |

## List of abbreviations

|         |  |
|---------|--|
| ACT     | Activity   |
| bcc     | Body-centered cubic  |
| CALPHAD | Calculation of phase diagrams                              |
| CEF     | Compound energy formalism                                  |
| DFT     | Density functional theory                                  |
| ESPEI   | Extensible Self-optimizing Phase Equilibria Infrastructure |
| fcc     | Face-centered cubic  |
| hcp     | hexagonal close-packed                                     |
| MCMC    | Markov Chain Monte Carlo                                   |
| SGTE    | Scientific Group Thermodata Europe                         |
| TDB     | Thermodynamic database                                     |
| ZPF     | Zero phase fraction  |

# 1 Introduction

Thermodynamics represents the basic foundation for the stability of phases, phase equilibria and transitions. CALPHAD (Calculation of Phase Diagrams) is a powerful tool within this fundamental framework which has gained large significance in materials science during the recent decades due to the increased computing power. The CALPHAD method models the Gibbs energies of the phases which help to understand and predict the thermodynamic behavior of the materials.

The foundation of the thermodynamic model is a database of experimental and theoretical data which contains the resulting models and model parameters of the Gibbs energies. This database can be used to calculate phase stabilities and material properties derived from the Gibbs energy models. The creation of a thermodynamic database is called an assessment and includes collecting experimental information and theoretical results, defining a model of the Gibbs energies of the different phases in the system and fitting the model to the available data. Common software for thermodynamic calculations which provide as well commercial databases are Thermo-Calc [1], FactSage [2] and MatCalc [3]. A recent open-source program is `pycalphad` [4]. After performing a thermodynamic assessment common databases only store a single value of the thermodynamic parameters without any further information about reliability, uncertainty as well as information on the original data input.

The open-source Python program ESPEI (Extensible Self-optimizing Phase Equilibria Infrastructure) [5] aims to perform thermodynamic parameter generation and optimization including uncertainty quantification. Using Bayesian inference and Markov Chain Monte Carlo (MCMC) algorithms, the probability distribution of the resulting parameters is calculated, based on the underlying input data for the assessment. The probability distribution of the parameters are then propagated to the quantities of interest and uncertainties can be calculated. Thereby, ESPEI is capable of quantifying the uncertainty which can be a crucial feature in materials design, so the boundaries in which the calculations can be trusted, are defined.

The goal of this work is to perform a thermodynamic assessment of the binary iron-carbon system with an uncertainty quantification, which will outline the reliability of the assessment. The databank contains thermodynamic data from experiments and ab-initio calculations along with meta-data such as the original article, experimental details or information about the reliability. The python-based database allows adaption to user-specific requirements and also easy reassessment at a later time, for example when new data is added.

After creating the thermodynamic database, the parameter optimization is performed with ESPEI, which also calculates the probability distribution of the parameters. The parameter uncertainties are then propagated to the quantities of interest, for example to calculate the



uncertainty of phase boundaries. Based on the resulting optimized phase diagram and calculated uncertainties, the effect of the underlying data can be investigated. The calculated uncertainties can guide future experiments to provide new data points in order to improve the accuracy of the assessment.

# 2 Theory

## 2.1 Basic thermodynamics

The basis of the CALPHAD method lies in the thermodynamics. Applying these fundamental principles allows the CALPHAD method to be used in a wide field of disciplines and applications. [6] The following chapters introduce the key concepts of thermodynamics which are needed to understand the CALPHAD method.

### 2.1.1 Laws of thermodynamics - enthalpy, heat capacity, entropy, Gibbs energy

The first law of thermodynamics (Equation 1) is known as the law of conservation of energy. [6] The change of the internal energy  $\Delta U$  can consist of heat  $q$  and work  $w$ . It can be described as the fact that energy cannot be created or destroyed in a closed system of constant volume, but the form of energy can be transformed from one to another. The term 'closed system' describes that there is no exchange of matter, energy or heat happening with its surroundings.

$$\Delta U = q - w \quad (1)$$

A quantity, as the internal energy  $U$ , is called a state function as it is only dependent on the state of the system, not on its history. For numerical calculations only differences of defined states are relevant as the internal energy  $U$  does not possess an absolute value. [7]

From this basic statement some important thermodynamic equations can be derived, introducing for example the enthalpy of a reaction. The enthalpy  $H$  is defined as  $q_p$ , the change in heat of the system with a reaction at constant pressure, and it can be obtained by transforming Equation 1 to

$$\Delta(U + w) = \Delta(U + pV) = q_p = \Delta H. \quad (2)$$

The heat capacity  $C_V$  of a system can be described according to Equation 3 as the amount of heat  $dq$  necessary in order to attain a temperature increase of  $dT$  at a constant volume  $V$ .

$$C = \frac{dq}{dT} \quad (3)$$

The entropy  $S$  is another state function which is a measure that reaches its maximum in the equilibrium state of the system. The entropy is defined by

$$S = \frac{dq}{T}. \quad (4)$$

In a closed system the entropy  $S$  can increase or stay constant, but never decrease. This means that the system cannot change its state from equilibrium as long as it is isolated from its environment. A change of state which increases the entropy is referred to as irreversible. [6] The second law of thermodynamics is relevant to decide if a reaction will spontaneously take place or not. This depends on the sign of the free energy  $G$  (also designated as Gibbs energy).  $G$  is the total energy of a system and it is minimal if the system reaches its equilibrium state. Therefore it can be derived that for a system with a constant composition, under isobaric and isothermal conditions, in equilibrium  $G$  becomes a minimum.

All these thermodynamic quantities are connected via the free energy  $G$  which is dependent on the internal energy  $U$ , volume  $V$  and pressure  $p$ , temperature  $T$  and entropy  $S$ .

$$G = U + pV - TS = H - TS \quad (5)$$

For changes of the system at constant pressure and temperature the free energy is defined as

$$dG = dq - T \cdot dS \quad (6)$$

This means that for a spontaneous reaction the change of free energy  $dG$  must have a negative sign. Consequently the equilibrium state of a system, at constant pressure and temperature, is the state with minimal free energy. [6]

## 2.1.2 Solution phase thermodynamics

CALPHAD calculations are often used for analyzing alloys consisting of two or more components which can form a solution under certain circumstances. The components can be represented by elements, ions and also molecules. The methods to describe the Gibbs energies of the solution phase is described in detail in Chapter 3.1. In this section the basic concepts of forming solutions and mixing effects are presented.

### Gibbs energy of binary solutions

The simplest case of a solution is a system of two components A and B occupying the sites in a crystal. Given  $N$  crystal sites in total and  $n_A$  sites occupied by component A, the mole fraction  $x_A$  of component A is

$$x_A = \frac{n_A}{N}. \quad (7)$$

Proceeding from Equation 5 the total Gibbs energy for such a binary system can be defined by

$$G = x_A G_A + x_B G_B + H_{mix} - T \Delta S_{mix}. \quad (8)$$

For an ideal solution there are no attractive or repulsive forces between both components and the Gibbs energy of the mixing reaction is

$$G_{mix}^{ideal} = RT(x_A \cdot \ln(x_A) + x_B \cdot \ln(x_B)). \quad (9)$$

However, the ideal case does not occur in reality, since attractive or repulsive interactions occur between the different components. For taking these interactions into account an excess term is added. This is implemented by introducing the regular solution energy parameter  $L_{A,B}$  which represents repulsive forces with a positive sign, and attractive forces with a negative sign.

$$G_{mix}^{exc} = x_A \cdot x_B \cdot L_{A,B} \quad (10)$$

Combining both equations results in the formulation of the regular solid solution:

$$G_{mix} = RT(x_A \ln(x_A) + x_B \ln(x_B)) + x_A x_B L_{A,B} \quad (11)$$

Given a system with a negative interaction parameter  $L_{A,B}$ , the resulting Gibbs energy function is a curve with one minimum. This means that over all compositions a solid solution of the components A and B is formed. For conditions where the temperature-dependent entropy term is not dominant, a positive parameter  $L_{A,B}$  results in a Gibbs energy curve that is shaped differently, exhibiting two minima though. The region between those two minima is characterized by the appearance of two phases in order to minimize the alloy's Gibbs energy at this composition and is designated as a miscibility gap. [6, 7]

### Activity in binary solutions

The effect of the different mixing behavior correlates with the activity of the components in the solution which is defined in Equation 12, where  $a_i$  is the activity of component  $i$  in the solution,  $p_i$  is the vapor pressure of  $i$  above the solution and  $p_i^0$  is the vapor pressure of the pure component  $i$ . [6] The activity also correlates with the chemical potentials  $\mu_i$  and  $\mu_i^0$  of the (pure) components and the temperature  $T$ .

$$a_i = \frac{p_i}{p_i^0} = \exp\left(\frac{\mu_i - \mu_i^0}{RT}\right) \quad (12)$$

In an ideal solution without any interactions of the present substances, the activity is the mole fraction. But for real solutions the interactions need to be taken into account and therefore the measured activity allows conclusions about the interacting mechanisms. [6] The chemical potential can also be derived from the Gibbs energy as

$$\mu_i = \left. \frac{\partial G}{\partial n_i} \right|_{n_{i+j} \text{ const.}} \quad (13)$$

### 2.1.3 Phase diagrams

A phase diagram allows indicating the different phases which are present at certain conditions as specified for example by composition and temperature. Thermodynamically, the calculation of a phase diagram is the result of a global Gibbs energy minimization process. [8]

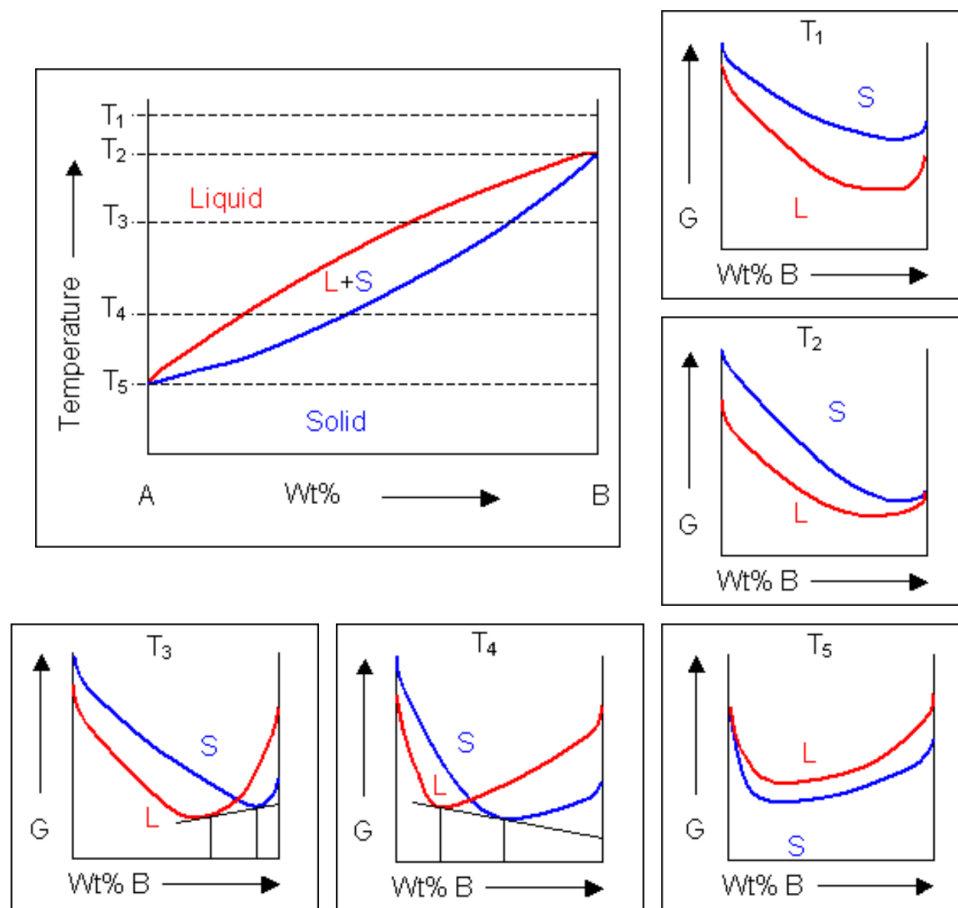


Figure 1: The stable phases are derived from the Gibbs energy curves at different temperatures. Graphics from [8].

The equilibrium of a system is characterized by a state where all the stable phases meet their internal constraints, for example composition or balance of the phase proportions, while also the external conditions, such as temperature and pressure, are satisfied.

Practically, the equilibrium states are derived from the Gibbs energy curve of each phase at different temperatures plotted against composition. For the simplest case of two components with full mutual solubility in the solid and liquid state, the occurring phases represent an equilibrium of the different phases, meaning that the state with minimal Gibbs energy at the given temperature is stable. In Figure 1 the Gibbs energies are plotted for different temperatures. At the highest temperature  $T_1$  the liquid phase is stable at all composition as it has the lowest

Gibbs energy. At  $T_2$  both curves meet at  $x_B=1$ , where both, the liquid and solid phase, are in equilibrium and so the melting point of the pure component B is reached. For lower temperatures the stable state in the region between the two curve minima is described by a common tangent. In this region both the solid and liquid phase are stable and the compositions of the minima identify the boundaries from the one-phase to the two-phase region. [6]

## 2.2 Binary iron-carbon system

The binary Fe-C system is a fundamental alloy system for a large number of practical applications. The phase diagram represents the relation between composition and temperature which is relevant for the formation of the various phases. In the Fe-C phase diagram two slightly different systems are combined and superimposed in most presentations: the metastable iron-cementite ( $Fe_3C$ ) and the stable iron-graphite system. [9]

These two systems differ by the type of the appearing carbon-rich phase. Graphite is formed if the cooling rates are extremely low, leading to the true equilibrium state which is usually very difficult to observe. The metastable cementite is technically more relevant, as in practical use this carbon phase is present at usual conditions. In common depictions only the iron-rich part up to a carbon-concentration of 25 at.%, which is 6.67 wt.%, is presented with the left boundary representing pure iron, and the right boundary representing pure cementite ( $Fe_3C$ ) in the case of the metastable system.

Iron can exist in two different crystal structures at atmospheric pressure: the body-centered cubic (bcc) and the face-centered cubic (fcc) crystallographic unit cell. Both types are able to dissolve carbon on interstitial sites depending on the iron's crystal structure and the temperature. Bcc-iron exists in two crystallographically identical species: the low-temperature  $\alpha$ -phase, being stable from room temperature up to about 911 °C, dissolving 0.022 wt.% carbon at the maximum, and the high-temperature phase  $\delta$ -iron, that is stable from 1392 °C up to the melting point and it is able to dissolve up to 0.1 wt.% of carbon. The solid solution of bcc-iron and carbon is referred to as ferrite. The fcc-iron, designated as  $\gamma$ -iron or austenite, can dissolve up to 2.06 wt.% of carbon. Fe-C alloys with carbon contents over circa 2 wt.% up to 4 wt.% are referred to as cast iron. [10, 9]

Concerning the values of maximum solubility and transformation temperatures it is noted that some of these values are not undisputed, but still represent a commonly accepted and widely used agreement for use in technology.

## 2.3 CALPHAD

### 2.3.1 Introduction

CALPHAD is a method based on the thermodynamics of the phases which allows gaining deeper understanding of the studied system and to make more accurate predictions on the appearance of phases in an alloy system. The goal is to generate a thermodynamic model from information on phase equilibria and thermochemical properties, to describe the Gibbs energy functions of the phases in a specific system. Building the Gibbs energy model starts with Gibbs energy functions of the pure elements, followed by the binary descriptions. With increasing complexity of the studied system, the Gibbs energy model is described by a rising number of parameters which are adjusted to fit the model to the covered data.

The CALPHAD method aims to obtain a comprehensive description of the Gibbs energies involved in the system in order to calculate and predict the phase diagram and other thermochemical properties in regions with little experimental information and to validate experimental findings.

### 2.3.2 Model parametrization

The basic steps of the CALPHAD method are shown in Figure 2. The first step is to assign the models for sublattices and Gibbs energies for the involved phases which are parameterized functions (Equation 14) of the state variables  $x_i$  (pressure, temperature, composition, etc.) and empirical parameters  $\theta_k$ . The empirical parameters can have individual forms for the various phases  $\varphi$ . [11]

$$G^\varphi = G^\varphi(x_1, \dots, x_i, \theta_1, \dots, \theta_k) \quad (14)$$

The Gibbs energy functions of the individual phases need to be defined and evaluated for a wide range of conditions. Contributions of the composition on the Gibbs energies regarding mixing and ordering effects are taken into account by using the Compound Energy Formalism (CEF) [5, 11]. This formalism is explained in more detail in Chapter 3.1.

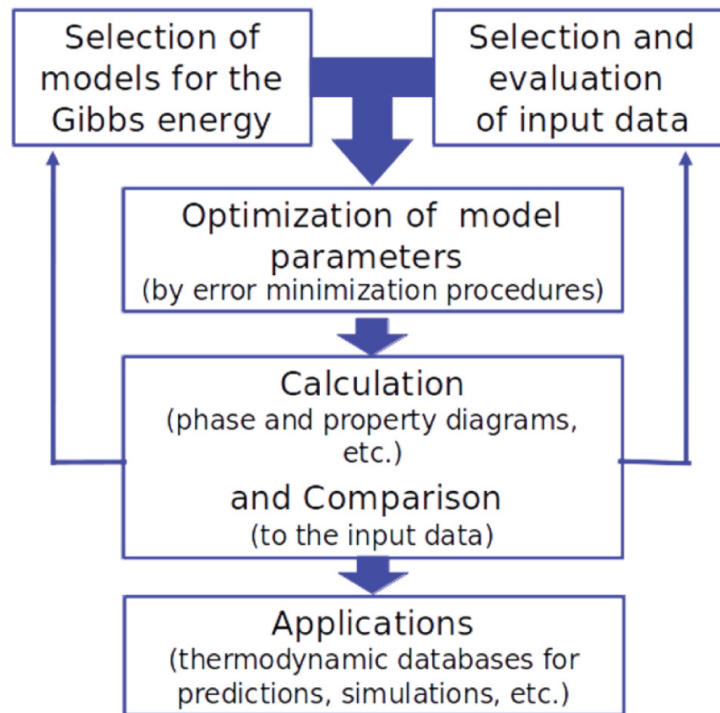


Figure 2: The iterative workflow of CALPHAD modelling: Continuous adjusting and re-evaluation of the model to the input data is performed to produce a thermodynamic database for sequential calculations. Graphics from [11].

Determining the model parameters is conventionally obtained by taking experimental data into account. In many use cases the main goal of establishing a thermodynamic model is to calculate phase diagrams and therefore the largest part of the collected data is phase boundary data. However it is crucial to also cover additional datatypes since phase equilibrium data alone is not sufficient to model the Gibbs energies of the different phases. Phase boundary data only gives information about the relation of the low-energy configuration of the phases for equal chemical potentials. [12] All available data based on the Gibbs energies of the various phases is critically evaluated in terms of reliability and a selection of usable datasets is made. In recent years also data from ab-initio methods gained importance for establishing a database as sometimes specific experiments cannot be conducted. Therefore also ab-initio thermochemical data is appropriate to contribute information for a comprehensive thermodynamic database. These additional datatypes can help to increase the database as well as evaluate which datasets should be used. Varying reliability of different datasets can be taken into account by assigning different weights. [13]

With larger numbers of different datatypes or components of the system the fitting algorithm requires to become more sophisticated and complex to be able to fit the model. Powerful mathematical methods, such as the least-squares method of Gauss [14], the Levenberg-Marquardt method [15] or Bayesian algorithm [16] are applied. Using such algorithms the model is refined



and optimized by adjusting the empirical parameters. Comparing and validating the calculated results with the input data revises the model in a continuous process. The resulting model should be capable of calculating the phase diagram and other thermochemical characteristics, not only for the compositions and temperatures given by the experimental data, but also for extrapolated regions. [13]

The CALPHAD method allows performing predictive calculations what makes this approach very interesting for multicomponent applications. This is because the thermodynamic descriptions of low-order subsystems can be extrapolated to more complex systems. It is also observed that the number of additional parameters for describing a high-order multicomponent system is reduced in comparison to a low-order system. This means that once the Gibbs energy models of the binary and ternary systems are determined, extrapolation to the quaternary or higher order system requires a decreasing number of additional parameters and also less data points. This characteristic can be made large advantage of for developing new materials which contain many components. [11, 12, 13]

However if there is no evidence for a stable phase in form of experimental data from low-order system, the phase stability in high-order system cannot be considered by CALPHAD. Consequently phase stability occurring exclusively in the high-order system cannot be estimated from the low-order Gibbs energy models. By introducing and using ab-initio calculations possible stable phases can be calculated and estimated and this drawback can be overcome. [11]

If in multicomponent phases there are contributions from further interactions, the Gibbs energy functions are extended by excess terms. These excess terms are added if they are necessary for a comprehensive description.

As this approach strongly bases on the information obtained from the low-order systems, changes in one of the sub-systems have strong effects on the relating low- and high-order systems. Practically this means that changing one of the unary descriptions affects the other unaries as well as the binary descriptions. This aspect leads to difficulties when models and endmember phase properties of large established databases should be amended. The Scientific Group Thermodata Europe (SGTE) developed a database of unary descriptions [17] which are the basis for the derived multicomponent descriptions. This complexity hinders updating the models with newer results. Hence being able to perform an assessment including the generation of a parameter set fitting for the available data can be a very attractive feature for recurring scientific applications. [5, 13]

### 2.3.3 Software and databases

Today there is large number of software for CALPHAD available, some of the most popular packages being FactSage [2], MatCalc [3], Pandat [18] or Thermo-Calc [19] as commercial products. During the last years there arose an increasing number of free open-source programs, for example pycalphad [4] and OpenCalphad [20]. The functions which are implemented in these software include basic calculations such as low-order phase diagrams but can also include more specific features for different use cases. In some of them (for example in MatCalc) modelling of thermodynamics and kinetics is combined to simulate precipitation kinetics, diffusion processes and phase transformations. [13]

The databases used in CALPHAD are generated by assessments of the corresponding subsystems. For technological applications where usually a fairly large number of elements are used in alloys, this approach results in extensive workloads for assessing all low-order subsystems. However, according to the dependencies explained earlier in Chapter 2.3.2, the number of necessarily determined subsystems is reduced in practical cases to the binaries and ternaries. In many cases quaternaries and higher-order subsystems are not essential to obtain a comprehensive assessment. [13, 17]

Once the model is fitted to the input data to satisfying extent, the resulting thermodynamic database (TDB) is ready to use in further applications. For example, phase equilibria in composition regions with little data may be predicted, or properties, such as microstructural, mechanical and other characteristics can be calculated. The TDB collects all the Gibbs energy functions of the phases of the system and can be used for thermodynamic calculations of multicomponent systems. [11, 13]

### 2.3.4 Applications

The range of applications of CALPHAD modelling was strongly driven by the increasing computing power in the recent decades. One of the most important fields since the beginning has been materials design. Therefore thermodynamic models are used to investigate possible alloys in terms of occurring phases or specific material properties. Thermodynamic databases are created or updated with new data and can be adapted to the individual use case. Then the resulting model may be used to estimate various material properties and predict promising alloy compositions. This comes along with drastically reduced experimental expenses in contrast to conventional alloy development where a lot of work goes into evaluating possible alloy compositions. However, the prediction of phases in high-order systems and also of certain intermetallic phases in multicomponent systems can still represent severe difficulties. This can

be encountered by taking ab-initio calculations into account. [13]

Besides the calculation of phase diagrams CALPHAD can also serve for modelling of other properties (for example diffusional behavior, molar volumes, elastic properties) where thermodynamic information from subsystems is extrapolated.

Many CALPHAD calculations are also aiming to gaining understanding of specific effects occurring in a material. Here the CALPHAD method can be a useful tool to comprehend certain mechanisms, concerning for example forming of precipitations or diffusion kinetics. [5, 13]

# 3 Computational methods

## 3.1 Compound energy formalism

### 3.1.1 Introduction

Selecting the expression used for describing the Gibbs energies in CALPHAD is a key decision for the whole approach. The model can serve predicting applications as well as gaining deeper understanding of the physical and chemical mechanisms underlying the resulting properties. Therefore the approach for selecting the appropriate model can vary from choosing a mathematical expression which leads to sufficient consensus to a model based on the causal physical mechanisms. Practical experiences showed that a model relating to physical principles is more capable of both, meeting the experimental observations and making specific predictions. [11]

The Compound Energy Formalism (CEF) represents a framework where the Gibbs energy models of the individual phases are described by the most appropriate expression respectively. Within the CEF the phases are characterized by different sublattices in which the constituents are placed. The constituents can be represented by atoms, molecules, ions and vacancies, which sit in defined sublattices and can mix according to the classical mixing theory. By introducing sublattices, the different interactions of the constituents on the different sublattice sites and interactions of the different species with each other can be taken into account. In the case of one sublattice and no allowed interactions of the different species, this formalism corresponds to ideal solution. For a sublattice being occupied by only one species, a stoichiometric phase is described. The Gibbs energies of the individual phases are determined by interaction parameters of the sublattices and the mole fractions of the constituents relative to the available sublattice sites. [11, 13]

$$G^\varphi = {}^{ref}G^\varphi + {}^{id}G^\varphi + {}^{ex}G^\varphi + {}^{phys}G^\varphi \quad (15)$$

The Gibbs energy term of the phase  $\varphi$  (Equation 15) can be written as the sum of the reference term  ${}^{ref}G^\varphi$ , the ideal mixing term  ${}^{id}G^\varphi$ , the excess term  ${}^{ex}G^\varphi$  and in some cases a physical term  ${}^{phys}G^\varphi$ , for example for magnetic, electronic, etc. contributions. The mixing and the excess term only appear for the case of two or more elements being mixed in the studied material. [11]

### 3.1.2 Ordering and mixing effects

If two or more species are present, different forms of ordering and mixing behavior can occur. There are some cases of mixing mechanisms to distinguish: One of the most simple cases is the substitutional random solution where all the constituents sit in one sublattice at random sites. This can be observed for example in liquids and solid solutions. The Gibbs energy consists of all terms given in Equation 15 being dependent on temperature and mole fraction. Dependencies on other variables, such as pressure, are mostly not considered in metallic alloys as ambient pressures are very low and usually this low pressures have negligible impact on the Gibbs energy. In stoichiometric phases there are multiple sublattices present with every species placed in one sublattice. As there is no compositional disorder within one sublattice, there is no mixing term to consider.

The third case is the ordered solution being an intermetallic compound within a finite composition range. For a binary system such a compound consists of two elements being placed in one sublattice each, but substitutional atoms of the other element are possible. Given two or more sublattices and multiple species being placed in the sublattices, the term of the Gibbs energy is dependent on the site fraction  $y_i^{(s)}$ , which is the mole fractions of each component  $i$  (atom, ion, molecule, vacancy) in the sublattice  $s$ . For each sublattice, the summation over the site fractions of this sublattice equals to 1, according to Equation 16. The site fractions are related to the overall composition according to Equation 17 with  $n^{(s)}$  being the stoichiometric coefficients in the sublattices and  $y_{VA}$  being site fractions occupied by vacancies. [11, 21]

$$\sum_i y_i^{(s)} = 1 \quad (16)$$

$$\frac{\sum_s n^{(s)} y_i^{(s)}}{\sum_s n^{(s)} (1 - y_{VA})} = x_i \quad (17)$$

### 3.1.3 Composition and temperature dependence

The regular solution energy parameter  $L_{A,B}$  that was introduced in Chapter 2.1.2 consists only of a single constant parameter. Parameters that are independent of temperature and composition are called strictly-regular. For a Gibbs energy model that should meet the real behavior, a strictly-regular model is not sufficient since changes in temperature and composition affect the observed interaction. Redlich and Kister defined an extension for the regular solution model which is known as the Redlich-Kister power series. [22] This extension is the most common as it is symmetrical and can be extrapolated to ternary and higher-order systems. Equation 18 describes the excess Gibbs energy of the phase  $\varphi$  including non-ideal mixing

behavior and interactions between the constituents.

$${}^{ex}G^\varphi = x_A \cdot x_B \sum_{v=0}^k (x_A - x_B)^v \cdot {}^vL_{A,B}^\varphi \quad (18)$$

The parameter  ${}^vL_{A,B}^\varphi$  can also be temperature-dependent in a linear or other form. If heat capacity data is available, the parameter  ${}^vL_{A,B}$  has contributions from the excess entropy  ${}^vS_{A,B}$  and the excess enthalpy  ${}^vh_{A,B}$ . [7]

$${}^vL_{A,B} = {}^vS_{A,B} + {}^vh_{A,B} \cdot T \quad (19)$$

### 3.1.4 Phase descriptions

For this work the framework of Gibbs energy descriptions as well as the excess parameters (one for the bcc phase, one for the fcc phase, five for the liquid phase) were adopted from Gustafson's assessment [23]. They serve as starting parameters for the optimization in order to fit to the literature data. The initial Redlich-Kister parameters of the different phases are listed in Table 1.

Table 1: List of the initial parameters for the optimization process, adopted from Gustafson [23]. The Gibbs energy descriptions of cementite and graphite were not changed.

| Phase  | excess model   | no. of sublattices | site ratios | constituents | excess parameters  |
|--------|----------------|--------------------|-------------|--------------|--|
| Bcc    | Redlich-Kister | 2                  | 1:3         | Fe:C,Va      | ${}^0L_{Fe:Va,C}^{bcc} = -190 T$   |
| Fcc    | Redlich-Kister | 2                  | 1:1         | Fe:C,Va      | ${}^0L_{Fe:Va,C}^{fcc} = -34671$   |
| Liquid | Redlich-Kister | 1                  | -           | Fe,C         | ${}^0L_{Fe,C}^{liq} = -124320 + 28.5T$<br>${}^1L_{Fe,C}^{liq} = +19300$<br>${}^2L_{Fe,C}^{liq} = +49260 - 19T$ |

### Ferrite and austenite phases

The Gibbs energy models of the different phases are based on Gustafson's descriptions [23] which follow the Compound Energy Formalism. The ferrite and austenite phases are described by two sublattices with Fe on one sublattice and interstitial C and vacancies placed on the other one. Regarding the denotation it is mentioned that components sitting in different sublattices are separated by a colon, while elements occupying the same sublattice are separated by a comma.

This sublattice model is also capable of considering magnetic contributions to the Gibbs energy. Therefore magnetic ordering effects are taken into account by the term  $G_m^{mo}$ . Since for the austenite magnetic effects are negligible, this contribution is only taken into account for ferrite. The Gibbs energy of one formula unit of the solid solution  $Fe_a(Va,C)_c$  is for the bcc phase:

$$G_m^{bcc} = y_{Va}^2 {}^0G_{Fe:Va}^h + y_C^2 {}^0G_{Fe:C}^h + RT [y_C^2 \ln(y_C^2) + y_{Va}^2 \ln(y_{Va}^2)] + y_{Va}^2 y_C^2 \cdot L_{Fe:Va,C}^{bcc} + G_m^{mo} \quad (20)$$

for the fcc phase:

$$G_m^{fcc} = y_{Va}^2 {}^0G_{Fe:Va}^h + y_C^2 {}^0G_{Fe:C}^h + RT [y_C^2 \ln(y_C^2) + y_{Va}^2 \ln(y_{Va}^2)] + y_{Va}^2 y_C^2 \cdot L_{Fe:Va,C}^{fcc} \quad (21)$$

$y_A^s$  is the site fraction of component  $A$  on sublattice  $s$  and  $L_{A:B}$  is the Redlich-Kister excess parameter describing the interaction between two components placed in two separate sublattices. The  ${}^0G$  values relate to the enthalpy of selected reference of each element at 298.15 K. The interaction parameter can be assumed to be composition dependent following a Redlich-Kister polynomial.  ${}^0G_{Fe:Va}^h$  represents the hypothetical Gibbs energy term of pure Fe without any magnetic contributions,  ${}^0G_{Fe:C}^h$  is a hypothetical term for the non-magnetic case of all interstitial sites being occupied with C, and  $G_m^{mo}$  adds the magnetic ordering effects.

For this work the magnetic contributions were not modified or optimized, so these terms were assumed from Gustafson who applied the model previously presented by Inden [24] and Hillert and Jarl [25]. This description of the magnetic contributions also allows to model the Curie temperature as a function of the contained C.

The Redlich-Kister parameters that were used as the starting point for the optimization are given in Table 1.

## Liquid phase

The liquid phase was described by the ordinary regular solution model whereas the excess model follows a Redlich-Kister polynomial. The Gibbs energy function follows:

$$G_m^{liq} = y_{Fe} {}^0G_{Fe} + y_C {}^0G_C + RT [y_{Fe} \ln(y_{Fe}) + y_C \ln(y_C)] + y_{Fe} \cdot y_C [{}^0L_{Fe,C}^{liq} + {}^1L_{Fe,C}^{liq} (1 - 2y_{Fe}) + {}^2L_{Fe,C}^{liq} (1 - 2y_{Fe})^2] \quad (22)$$

The Redlich-Kister excess-parameters for the liquid phase  $L_{Fe,C}^{liq}$  also were accepted from Gustafson [23] as the initial parameter values and the parameters are given in Table 1.

## Carbides

The assessment of the C-phases included cementite as stoichiometric compound of  $\text{Fe}_3\text{C}$  and solid carbon as graphite and diamond. For these phases the Gibbs energy descriptions were also accepted from Gustafson who adopted the model from Guillermet [26] and the optimization was not applied, hence these phase descriptions were not changed.

## 3.2 ESPEI

ESPEI (Extensible Self-optimizing Phase Equilibria Infrastructure) is an open-source, Python-based software tool using the CALPHAD framework including two steps of model parameter evaluation: The first step being parameter generation from a database of single-phase thermochemical data and as a second step parameter optimization by a Markov Chain Monte Carlo (MCMC) algorithm using the probability distributions to quantify uncertainties.

Additionally, the software-environment is capable of creating extensible databases of collected literature datasets which are used for the thermodynamic calculations.

ESPEI provides a software tool to use an existing thermodynamic description as well as create a new model of the Gibbs energies which can then be updated and optimized according to an underlying literature database. What is very attractive for practical applications is the ability to estimate the reliability of the obtained phase diagrams. This is achieved with Bayesian inference where the parameter estimation is performed autonomously to a great extent once the simulation is set up. This procedure aims to reduce the operator's influence on the fitting process and thus make the thermodynamic assessment more reproducible. Applying MCMC and Bayesian inference provides the probability distributions after the optimization which are used to calculate the model uncertainties. The concepts of MCMC and Bayesian statistics are described more closely in Chapter 3.3.

The process of parameter generation is not explained in this work as this step was not used for this thesis. The initial parameters were taken from a previous assessment and only parameter optimization was performed.

For the optimization of the model parameters the Gibbs energies are fitted to all data provided from covered literature data simultaneously with an iterative MCMC algorithm. The Gibbs energies are described by a temperature-dependent expression within the CEF using Redlich-Kister excess mixing parameters, see Equation 15. [11, 5, 17]

The parameter optimization is performed using the Bayesian inference approach evaluating the probability distributions of the obtained model parameters in order to minimize the residual sum



of squares between the determined parameters and the collected data. Besides the parameter set obtained in ESPEI, also parameters which stem from other assessments can be optimized to fit the created database. [12] As a result of the optimization within this work, a new, fully reproducible and comprehensible TDB of the binary Fe-C system based on a literature database was performed.

### 3.3 Bayesian inference and Markov Chain Monte Carlo

For finding a set of model parameters describing the input data best, Bayesian inference is applied. Bayes' statistics represent a framework for parameter consideration based on probability distributions. Bayes' theorem (Equation 23) defines the posterior probability distribution for a parameter vector  $\theta$ , conditioned on the data  $D$ . The prior  $p(\theta)$  is an estimate from the modeler of the probability distribution for the parameter vector, and  $p(D|\theta)$  is the probability of observing the data conditioned on a set of parameters which is the likelihood.  $p(D)$  is the probability of all possible data called evidence. This approach takes the likelihood of observing a set of data into account for updating the priors. [12]

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \quad (23)$$

Recently the Bayesian statistics approach has become widely accessible as computing power rapidly increased and software implementing Markov Chain Monte Carlo (MCMC) algorithms arose. MCMC is used for numerical evaluations of the posterior distributions which form a Markov chain. This means a distinct number of parameters are sampled and are used to determine the parameters which lead to maximum likelihood. This method for parameter optimization can be used for any type of system and also the number of degrees of freedom can lie in a wide range - restricted only by the computing power which accounts for reasonable computing times. Also systems with highly correlated parameters can exhibit bad convergence. In CALPHAD models the parameters correlate with each other as a changing parameter can be corrected by another parameter and the resulting Gibbs energy function stays the same. Hence ESPEI uses an ensemble sampler which introducing an ensemble of Markov chains to sample multiple parameters simultaneously. This sampler algorithm is implemented by using the emcee package [27] which generates multiple chains of Gaussian distributions. Based on these chains ESPEI determines the probability functions and likelihoods. The proposed parameters are evaluated by the Metropolis algorithm, given in Equation 24, which states that the proposed parameter set is accepted if the probability is increased and the parameter set is added to the Markov chain. But parameters which decrease the probability can be accepted with a probability

proportional to the decrease in posterior probability. [5]

$$p_{accept} = \min \left( \frac{p_{proposed}}{p_{current}}, 1 \right) \quad (24)$$

### 3.4 Priors, data weights and likelihood function

Via the prior distribution the operator is able to define a range of possible parameter values for the parameter optimization. ESPEI allows to specify the distribution functions and hyperparameters which give the centrality and the boundaries of the distribution function. The hyperparameters can be helpful if the operator wants to specify a range of parameter values or the probability is zero for certain values below and above the limit. The three basic distributions implemented in ESPEI are: uniform, triangular and normal distribution.

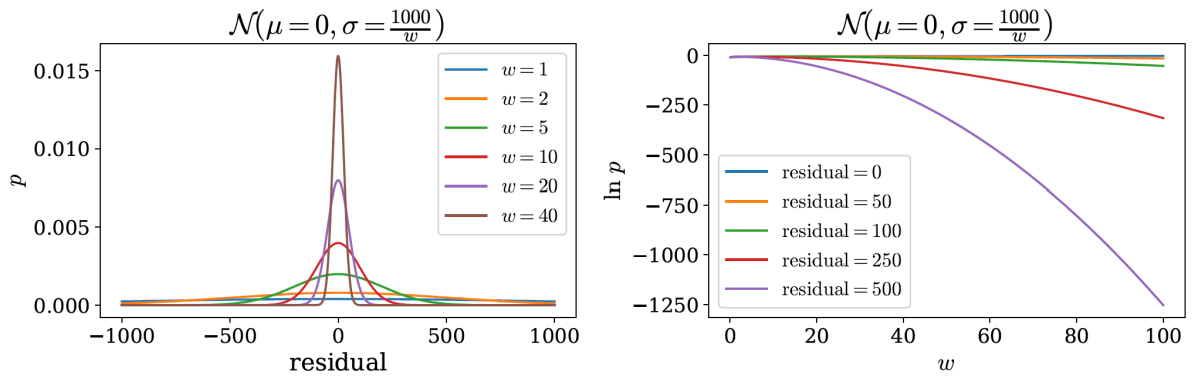


Figure 3: Dataset weights affect the probability distribution of the residuals (left). The log-probabilities (right) decrease with higher residuals. Figure from [12]

The uniform priors set a constant probability in a chosen range and are useful if the initial database parameter is not likely to be fitting. They can also be used if the user wants to ensure that for example the value of the enthalpy stays negative. A triangular distribution gives a sharp lower and upper boundary outside of which the probability is zero and the parameter must lie between these limits. Also the triangular function has a maximum in the probability density function describing a tendency to values lying at the center.

For this work the normal prior described by a Gaussian distribution was used for all optimized parameters. The prior was defined by a local relative of 1.0 and a scale relative of 0.4. This type of prior does not set any sharp limits on the parameter boundaries and the function was chosen to be centered on the initial parameter.

For evaluating the likelihood ESPEI relates to the standard deviations of the error of each data types. The likelihoods are normalized by the standard deviation of the residual which

is assumed as normal distributed and dependent on the specific datatype. This is done to maintain constant weighting of error from the various data.

For single-phase thermochemical data the following default values for standard deviations are assigned in ESPEI:  $500 \text{ J mol}^{-1}$  for enthalpies,  $0.2 \text{ J K}^{-1}$  for entropies and  $0.2 \text{ J K}^{-1} \text{ mol}^{-1}$  for heat capacities. For activity data the error is evaluated by calculating the chemical potentials (the reasons for that are explained in Chapter 3.5.3) and comparing the measured with the calculated values. The standard deviation of the chemical potential is  $500 \text{ J mol}^{-1}$ . [5]

The weight of each data type or dataset can also be changed and specifically chosen by the operator according to the reliability of the dataset. This is a crucial decision as the assigned weight of the datasets is directly linked to the calculated probability and determined uncertainty. Hence it is useful to define the individual dataset's weight following the reliability of the experimental data, representing for example measuring errors. As depicted in Figure 3 the higher weights affect the probability distribution to be narrower than for lower weights. For a narrower probability distribution, the likelihood of a certain residual is lower than for a broader curve.

In this work a weight factor of 20 was chosen for all datasets uniformly. All these parameters which are selected by the software user, such as dataset weights, type of distribution and the hyperparameters of the priors, are specified within a machine-readable input file. This basic information is necessary for ESPEI to set up the optimization calculations and is stored in JSON-files following a prescribed format scheme. [5]

## 3.5 Error definition in ESPEI

### 3.5.1 Phase boundary data

In most cases the main criterion of a thermodynamic assessment is to compare the resulting phase diagram with measured data. Accordingly phase equilibria data is important to consider if a new assessment is judged. Determining the residual for phase boundary data is possible in multiple ways. The direct approach compares the calculated phase diagram to the measured points by determining the differences in composition of the phase boundary or temperature. Within ESPEI, a different definition is applied which always gives a value for the residual from the model to the measured composition, even for the case when the relating phase boundary cannot be calculated. The approach for the error definition of phase boundaries is characterized by the concept of driving forces, where the measure for fitting the phase boundary data is the driving force necessary for a given phase to be stable at the measured composition.

For determining the driving forces of a measured phase equilibrium the target equilibrium

hyperplane serves as the reference. After conducting equilibrium calculations with all present phases are conducted at the measured temperatures and compositions, the target hyperplane is calculated as the mean of the chemical potentials as represented in Figure 4a. With each Gibbs energy function of the phases that are expected to be stable at this composition, equilibria are calculated at the projections of their compositions onto the Gibbs energy curves and the driving forces can then be determined as the residual between the target hyperplane and the calculated vertex Gibbs energy. [5, 12]

In order to obtain the probabilities of meeting the phase boundaries, the errors are assumed to follow a normal distribution around zero with a standard deviation of 1 kJ. [12]

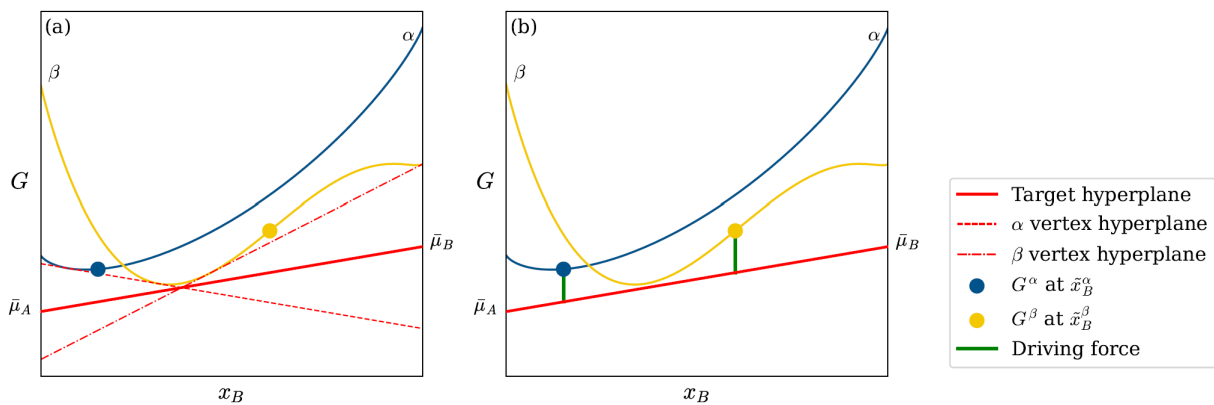


Figure 4: Example for determining the driving forces for phase boundary data: The blue and yellow points represent measured phase equilibrium compositions. The target hyperplane is constructed as the arithmetic mean of the calculated chemical potentials. The energy difference between the target hyperplane and the vertex energy is the driving force (b). Figure from [12].

### 3.5.2 Thermochemical data

The types of thermochemical data used in ESPEI includes two types: Thermochemical data with fixed internal degrees of freedom and equilibrium thermochemical data. For the thermochemical data with fixed internal degrees of freedom the thermochemical property can be obtained directly from the Gibbs energy model. Equilibrium data are typically measured for given values as temperature, pressure, composition. For these data, the thermochemical property is determined by an equilibrium calculation where various phases can occur. In this case the error is determined by calculating the thermochemical property according to the observed composition and phase fraction. Therefore the error for thermochemical data is defined as the difference from the predicted values to the experimental values. Also the standard deviations of the thermochemical data are assumed and implemented in ESPEI with  $500 \text{ J mol}^{-1}$  for enthalpies,  $0.2 \text{ J mol}^{-1} \text{ K}^{-1}$  for entropies,  $0.2 \text{ J mol}^{-1} \text{ K}^{-1}$  for heat capacities. [12]

### 3.5.3 Activity data

Due to the direct correlation with the chemical potential, activity data is very important to gain information on the Gibbs energies. Hence residuals of activity data are treated in a different way concerning error definition. As activity values range from 0 to 1, the variability of the observed values is usually rather small, causing small residuals. However, these small residuals can have a large effect on the calculated chemical potentials. But for determining the likelihood in terms of residuals, the chemical potentials should be preferred instead of activity data. Hence ESPEI uses the original activity data for deriving the Gibbs energy functions, but the calculated chemical potentials for likelihood estimation. Therefore the data provided by the user includes the observed activities along with the reference state in order for ESPEI to calculate the chemical potential according to Equation 25.  $\mu_A^{obs}$  and  $\mu_A^0$  being the observed chemical potential respectively the chemical potential of the reference state and  $a_A^{obs}$  the observed activity. [12]

$$\mu_A^{obs} = \mu_A^0 + RT \ln(a_A^{obs}) \quad (25)$$

# 4 Results and discussion

## 4.1 Database

### 4.1.1 Motivation for a new database

One of the goals of this work was to establish a literature database of the Fe-C binary phase diagram and to create a new thermodynamic database on this data collection within the ESPEI method. The motivation to collect these data and to create a new TDB-file was to enhance the information content of such databases and to reproduce the widely used thermodynamic assessment of Gustafson from the year 1985 [23]. Currently there are several established thermodynamic databases, which for the most part only report the optimized parameters without any further information about the acquisition of these data, such as origin of the data, experimental details or measurement errors. The connection to the original data is often unclear or missing what makes the process of parameter generation not fully reproducible. This lack of information can lead to difficulties and inexactness when, for example results from these databases should be interpreted and processed in subsequent steps. Also, there is only little information about the reliability of the thermodynamic models available from common databases which means that the uncertainty cannot be estimated from the underlying database. ESPEI provides a tool to create a new database from specifically selected literature data and to perform a full TDB-generation process from this database. This allows the operator to adapt literature database to his individual requirements and to generate a TDB which is used for the following CALPHAD simulations.

In the present work, this feature of customizing the database was used to also augment the content of information in the database, such as implementing a register of all used scientific articles, extended information about the measured data points and details about the applied experimental methods. The open-source Python-code of the ESPEI software allows adaption and re-evaluation of the database at a later point, when for example new datasets are available or certain characteristics are looked for in the data.

All these features provide new options to adapt the created TDB to the user's individual requirements and thus extending the possibilities of the CALPHAD method.

## 4.1.2 Covered data

To obtain a comprehensive experimental data collection of the binary Fe-C system, over 190 different data points from 13 scientific publications were collected and recorded.

The literature database only includes datasets which are completely based on their own measurements. This ensures to being able to capture as much information on the experiments itself and to exclude measurement errors from previous experiments.

As several types of measured data can serve as input data for ESPEI, the collected datasets include experimentally obtained values on phase equilibria, thermodynamic activities and ab-initio data on formation enthalpies from density functional theory (DFT) calculations.

The newly created literature database was implemented implemented in a machine-readable JSON-format. This enables the data to be directly used for the following steps within the ESPEI method. For providing the data to the database, the scientific articles were studied in terms of suitable data which then were digitized via JSON-files containing the experimental data and further metadata in a pre-specified format. The various JSON-files serve then as the input data for ESPEI.

### Phase equilibrium data

The largest number of data points in the database stems from phase equilibrium studies where the boundaries between two phase regions were investigated and determined. The phase boundaries covered by experimental data in the created literature database are marked in the Fe-C phase diagram in Figure 5.

The descriptions of these states consist of the measured carbon concentration, the temperature at this point and the phases which are present at both sides of the boundary. Since for the CALPHAD calculations the atmospheric pressure is also relevant, this information is also contained in the datasets. For those cases which had no information about the present atmospheric pressure available, normal pressure of 101 325 Pa was assumed.

Figure 5 exhibits all phase boundary datasets plotted in the Fe-C phase diagram. The red lines represent the phase boundaries proposed by Gustafson's assessment.

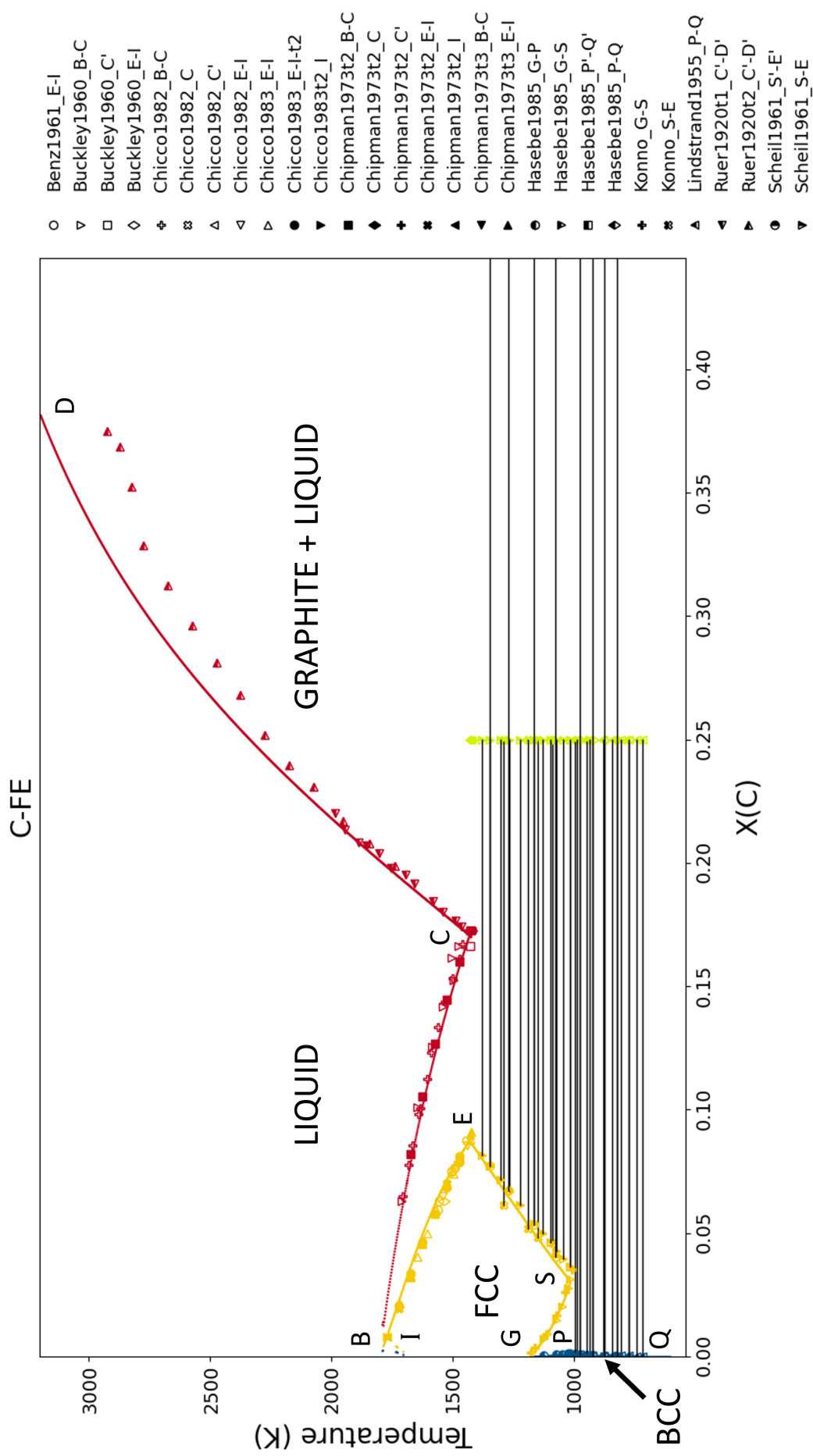


Figure 5: The experimental phase boundary data points of the stable and metastable system covered in the database. The red lines represent the phase boundaries from Gustafson's assessment [23]. For terms of distinction, all data on the stable system is marked with a prime symbol.



The created database on zero-phase-fraction (ZPF) data contains 153 data points from ten different scientific publications. An overview of the covered publications and the contained data is presented in Table 2.

Table 2: List of experimental data on phase boundaries covered in the literature database

| Author     | Year | Measured phase boundary   | No. of data points | Literature reference |
|------------|------|---|--------------------|----------------------|
| Benz       | 1961 | Austenite solidus (E-I)   | 8                  | [28]                 |
| Buckley    | 1960 | Austenite solidus (E-I) & liquidus (B-C), stable eutectic (C')    | 9                  | [29, 30]             |
| Chicco     | 1982 | Metastable eutectic (C), austenite liquidus (B-C) & solidus (E-I) | 31                 | [31]                 |
| Chicco     | 1983 | Austenite liquidus (A-B-C)  | 10                 | [32]                 |
| Chipman    | 1973 | Austenite solidus (E-I) & liquidus (B-C)                          | 14                 | [33]                 |
| Hasebe     | 1985 | Metastable & stable $\alpha$ -solvus (P-Q, P'-Q', G-P, G-S)       | 21                 | [34]                 |
| Konno      | -    | Metastable austenite solvus (G-S, S-E)                            | 10                 | [35]                 |
| Lindstrand | 1955 | Metastable $\alpha$ -solvus (P-Q)                                 | 8                  | [36]                 |
| Ruer       | 1920 | Stable liquidus (C'-D')   | 26                 | [37]                 |
| Scheil     | 1961 | Metastable & stable austenite solvus (S-E, S'-E')                 | 16                 | [38]                 |

The publication dates of the articles on ZPF-data range from 1920 [37] to 1985 [34]. It is noted that all experimental data that is registered in the literature database, was published with exact values in form of tables. No data was estimated from graphs or diagrams in order to prevent inaccuracies arising from graphic inspections.

The publications of Benz [28], Buckley [29, 30] and Chipman [33] were not accessible in the original form. But their experimental results were cited by Chicco 1982 [31] and Chicco 1983 [32] with the exact values in tabular form. Hence these data was also used for the literature collection.

As the Fe-C binary phase diagram as it is commonly known, in fact consists of the stable and metastable phase diagram superposed in one diagram, the goal of the present work was to create a thermodynamic description which is capable of describing both systems with one TDB. The datasets which are available, include specific measurements from the stable as well as from the metastable system, whereas some datasets can also be used for both systems.

To give an example, there are two different phase boundaries in Figure 6 marked in blue and orange respectively. The blue line represents the G-S phase boundary, which divides the bcc+fcc-phase region from the pure fcc-phase region, meaning that there is no carbide phase adjacent to either side of the phase boundary. In contrast, the orange line marks the S-E phase boundary, where the fcc-phase is in equilibrium with the fcc+Fe<sub>3</sub>C-phase region. As Fe<sub>3</sub>C occurs exclusively in the metastable phase diagram, the available measurements reporting data on this specific phase boundary were used only to describe the metastable system. Whereas available data on the G-S phase boundary (marked in blue) were used for both, the stable and

metastable system, as no system-specific information was contained. For terms of distinction, all data on the stable system is marked with a prime symbol in Table 2. On the following pages the covered phase equilibrium datasets are attached.

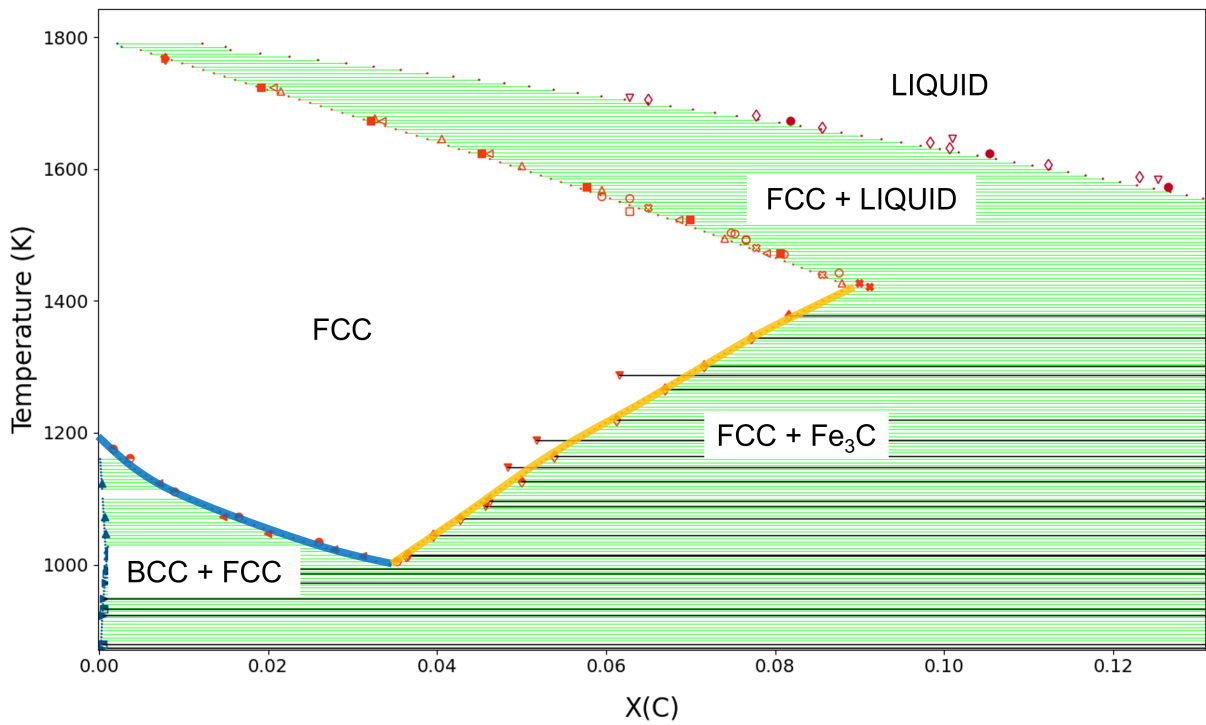


Figure 6: Data on the blue phase boundary are usable for the metastable and stable system, as there is no information on carbide phases contained. Whereas at the orange phase boundary the Fe<sub>3</sub>C phase is adjacent, which is a phase occurring exclusively in the metastable system, data on this type of phase boundary are usable only for the metastable system where the data is stemming from.

**Chicco 1982 - Table 1**

Exp. methods: Samples drawn from melt, measured by DTA, spectroscopy, direct combustion

| x_C    | Temperature |           |            |            |
|--------|-------------|-----------|------------|------------|
|        | C<br>[K]    | C'<br>[K] | E-I<br>[K] | B-C<br>[K] |
| 0,0649 |             |           | 1541,15    | 1706,15    |
| 0,0777 |             |           | 1481,15    | 1682,15    |
| 0,0855 |             |           | 1440,65    | 1663,15    |
| 0,0983 |             |           |            | 1640,65    |
| 0,1006 |             |           |            | 1631,65    |
| 0,1123 |             |           |            | 1606,15    |
| 0,1231 |             |           |            | 1588,15    |
| 0,1336 |             |           |            | 1559,15    |
| 0,1436 |             |           |            | 1531,15    |
| 0,1534 |             |           |            | 1501,15    |
| 0,1612 |             |           |            | 1472,65    |
| 0,1669 |             |           |            | 1458,65    |
| 0,1728 | 1419,15     | 1424,15   |            | 1439,15    |

**Chicco 1982 - Table 2**

(citing Buckley 1960)

Exp. methods: not available

| x_C    | Temperature |            |            |
|--------|-------------|------------|------------|
|        | C'<br>[K]   | E-I<br>[K] | B-C<br>[K] |
| 0,0628 |             | 1535,35    | 1708,65    |
| 0,1010 |             |            | 1645,55    |
| 0,1253 |             |            | 1583,45    |
| 0,1417 |             |            | 1540,55    |
| 0,1526 |             |            | 1493,55    |
| 0,1555 |             |            | 1504,75    |
| 0,1616 |             |            | 1474,35    |
| 0,1662 | 1426,75     |            |            |

**Chicco 1982 - Table 2**

(citing Benz 1961)

Exp. method: not available

| x_C    | Temp.      |
|--------|------------|
|        | E-I<br>[K] |
| 0,0594 | 1558,35    |
| 0,0628 | 1555,95    |
| 0,0748 | 1504,05    |
| 0,0752 | 1501,55    |
| 0,0765 | 1493,15    |
| 0,0765 | 1492,15    |
| 0,0810 | 1471,65    |
| 0,0875 | 1442,95    |

**Chicco 1982 - Table 3**

Exp. methods: Samples drawn from melt, DTA, spectroscopy, direct combustion

| Temp.<br>[K] | x_C    |        |
|--------------|--------|--------|
|              | E-I    | B-C    |
| 1420,15      | 0,0891 | 0,1770 |
| 1423,65      | 0,0883 | 0,1759 |
| 1428,15      | 0,0879 | 0,1752 |
| 1473,15      | 0,0794 | 0,1626 |
| 1523,15      | 0,0690 | 0,1461 |
| 1573,15      |        | 0,1280 |
| 1623,15      |        | 0,1061 |
| 1673,15      |        | 0,0818 |

**Chicco 1982 - Table 3**

(citing Chipman 1973)

Exp. method: not available

| Temp.<br>[K] | x_C    |        |
|--------------|--------|--------|
|              | E-I    | B-C    |
| 1421,15      | 0,0911 | 0,1728 |
| 1427,15      | 0,0899 | 0,1714 |
| 1473,15      | 0,0806 | 0,1598 |
| 1523,15      | 0,0699 | 0,1443 |
| 1573,15      | 0,0577 | 0,1265 |
| 1623,15      | 0,0453 | 0,1053 |
| 1673,15      | 0,0322 | 0,0818 |

**Chicco 1983 - Table 1**

Exp. methods: Samples drawn from melt, DTA, spectroscopy, direct combustion

| x_C       | Temp. E-I [K] |
|-----------|---------------|
| 0,0214841 | 1718,65       |
| 0,0326192 | 1677,15       |
| 0,0405147 | 1645,65       |
| 0,0500294 | 1604,65       |
| 0,0593984 | 1568,15       |
| 0,0740116 | 1495,15       |
| 0,0878745 | 1427,15       |

**Chicco 1983 - Table 2**

Exp. methods: Samples drawn from melt, DTA, spectroscopy, direct combustion

| x_C       | Temperature |         |         |         |        |
|-----------|-------------|---------|---------|---------|--------|
|           | A [K]       | I [K]   | E-I [K] | C [K]   | C' [K] |
| 0,0894844 |             |         |         | 1420,15 |        |
| 0,08868   |             |         |         |         | 1423,2 |
| 0,0789412 |             |         | 1473,2  |         |        |
| 0,0686251 |             |         | 1523,2  |         |        |
| 0,0577056 |             |         | 1573,15 |         |        |
| 0,0461548 |             |         | 1623,15 |         |        |
| 0,0335015 |             |         | 1673,15 |         |        |
| 0,0205846 |             |         | 1723,15 |         |        |
| 0,0083145 |             | 1767,15 |         |         |        |
| 0         | 1800,15     |         |         |         |        |

**Chicco 1983 - Table 2 (citing Chipman 1973)**

Exp. methods: not available

| x_C       | Temperature |         |         |         |         |         |
|-----------|-------------|---------|---------|---------|---------|---------|
|           | A [K]       | I [K]   | E-I [K] | C [K]   | C' [K]  | B-C [K] |
| 0         | 1800,15     |         |         |         |         |         |
| 0,0078554 |             | 1768,15 |         |         |         |         |
| 0,019233  |             |         | 1723,15 |         |         |         |
| 0,0321776 |             |         | 1673,15 |         |         |         |
| 0,0452904 |             |         | 1623,15 |         |         |         |
| 0,0577056 |             |         | 1573,15 |         |         |         |
| 0,0698724 |             |         | 1523,15 |         |         |         |
| 0,0805754 |             |         | 1473,15 |         |         |         |
| 0,0817982 |             |         |         |         |         | 1673,15 |
| 0,0898862 |             |         |         |         | 1427,15 |         |
| 0,0910899 |             |         |         | 1421,15 |         |         |

**Hasebe 1985 - Table 1**

Exp. method: Diffusion couple method

| Temp. [K] | x_C      |          |
|-----------|----------|----------|
|           | P'-Q'    | P-Q      |
| 773,15    |          | 6,51E-05 |
| 823,15    | 4,65E-05 | 1,07E-04 |
| 873,15    | 1,16E-04 | 2,14E-04 |
| 923,15    | 2,84E-04 | 4,09E-04 |
| 948,15    |          | 5,76E-04 |
| 973,15    | 6,27E-04 | 8,08E-04 |
| 993,15    |          | 1,00E-03 |

**Hasebe 1985 - Table 2**

Exp. method: Diffusion couple method

| Temp. [K] | x_C      |           |
|-----------|----------|-----------|
|           | G-P      | G-S       |
| 1013,15   | 1,02E-03 | 0,031205  |
| 1023,15   | 9,43E-04 | 0,0277442 |
| 1048,15   | 7,85E-04 | 0,0199543 |
| 1073,15   | 6,37E-04 | 0,0146612 |
| 1123,15   | 3,76E-04 | 0,0070282 |

**Konno - Table 2**

Exp. method: not available

| x_C       | Temp.<br>G-O-S-E<br>[K] |
|-----------|-------------------------|
| 0,0016252 | 1176,15                 |
| 0,0036625 | 1161,15                 |
| 0,0088649 | 1111,15                 |
| 0,0165211 | 1073,15                 |
| 0,025962  | 1035,15                 |
| 0,0352622 | 1005,15                 |
| 0,0457228 | 1088,15                 |
| 0,048310  | 1148,15                 |
| 0,0517435 | 1188,15                 |
| 0,0615077 | 1288,15                 |

**Lindstrand 1955 - Table 2**

Exp. method: C-solubility via elastic relaxation

| x_C      | Temp.<br>P-Q<br>[K] |
|----------|---------------------|
| 9,15E-04 | 986,15              |
| 6,04E-04 | 933,15              |
| 4,28E-04 | 879,15              |
| 3,16E-04 | 841,15              |
| 2,32E-04 | 807,15              |
| 1,67E-04 | 772,15              |
| 1,26E-04 | 741,15              |
| 9,30E-05 | 717,15              |

**Ruer 1920 - Table 1**

Exp. methods: Quenching of graphite-saturated liquid Fe. Combustion gravimetry

| Temp.<br>[K] | x_C<br>C'-D' |
|--------------|--------------|
| 1463,15      | 0,17420      |
| 1488,15      | 0,17662      |
| 1538,15      | 0,18006      |
| 1578,15      | 0,18449      |
| 1658,15      | 0,19159      |
| 1693,15      | 0,19527      |
| 1753,15      | 0,19793      |
| 1803,15      | 0,20387      |
| 1853,15      | 0,20714      |
| 1888,15      | 0,20811      |
| 1943,15      | 0,21330      |
| 1983,15      | 0,22002      |

**Ruer 1920 - Table 2**

Exp. methods: Quenching of graphite-saturated liquid Fe. Combustion gravimetry

| Temp.<br>[K] | x_C<br>C'-D' |
|--------------|--------------|
| 1738,15      | 0,19892      |
| 1843,15      | 0,20779      |
| 1953,15      | 0,21715      |
| 2073,15      | 0,23073      |
| 2173,15      | 0,23939      |
| 2273,15      | 0,25181      |
| 2373,15      | 0,26801      |
| 2473,15      | 0,28117      |
| 2573,15      | 0,29619      |
| 2673,15      | 0,31235      |
| 2773,15      | 0,32850      |
| 2823,15      | 0,35222      |
| 2873,15      | 0,36848      |
| 2923,15      | 0,37501      |

**Scheil 1961 - Table 3**

Exp. method: Gas-equilibration, combustion analysis

| Temp.<br>[K] | x_C<br>S-E |
|--------------|------------|
| 1014,15      | 0,0363     |
| 1015,15      | 0,0365     |
| 1044,15      | 0,0395     |
| 1070,15      | 0,0427     |
| 1097,15      | 0,0461     |
| 1126,15      | 0,0500     |
| 1164,65      | 0,0538     |
| 1220,15      | 0,0612     |
| 1266,15      | 0,0669     |
| 1301,65      | 0,0716     |
| 1344,15      | 0,0772     |
| 1377,65      | 0,0816     |

**Scheil 1961 - Table 5**

Exp. method: Gas-equilibration, combustion analysis

| Temp.<br>[K] | x_C<br>S'-E' |
|--------------|--------------|
| 1076,15      | 0,0405       |
| 1165,15      | 0,0535       |
| 1268,65      | 0,0674       |
| 1344,15      | 0,0773       |

## Thermodynamic activities

Experimental data on thermodynamic activities are included in the database for the stable system from Lobo [39] who made 37 measurements of the activity of graphite in  $\alpha$ -ferrite and from Smith [40] who measured 42 activities of graphite in austenite. The experimental technique that was used in both articles was equilibration of the steel samples with  $\text{CH}_4/\text{H}_2$  and  $\text{CO}/\text{CO}_2$  gas mixtures, followed by measuring the weight gains of the samples in order to determine the absorbed carbon. The produced samples contained carbon contents ranging from 0.006 at.% to 0.076 at.% at temperatures from 682 °C to 848 °C in Lobo's measurements and 1.8 at.% to 6.6 at.% measured at 800 °C and 1000 °C in Smith's measurements. The experimental activity datasets are given in Table 3 and 4.

Table 3: Data on activities of graphite in bcc-Fe from Lobo [39] covered in the literature database.

| Temp. [K] | C-content $x_C$ | C-activity $a_C$ | Temp. [K] | C-content $x_C$ | C-activity $a_C$ |
|-----------|-----------------|------------------|-----------|-----------------|------------------|
| 1121.15   | 0.00022778      | 0.084            | 1026.15   | 0.00019990      | 0.172            |
| 1121.15   | 0.00033468      | 0.127            | 1026.15   | 0.00024173      | 0.259            |
| 1086.15   | 0.00016736      | 0.088            | 1026.15   | 0.00047872      | 0.431            |
| 1086.15   | 0.00024638      | 0.131            | 1026.15   | 0.00051124      | 0.517            |
| 1086.15   | 0.00033468      | 0.175            | 1026.15   | 0.00062737      | 0.603            |
| 1086.15   | 0.00044155      | 0.219            | 1026.15   | 0.00072491      | 0.689            |
| 1086.15   | 0.00046013      | 0.244            | 1026.15   | 0.00076206      | 0.732            |
| 1086.15   | 0.00046478      | 0.263            | 1000.15   | 0.00017666      | 0.268            |
| 1070.15   | 0.00015341      | 0.1              | 1000.15   | 0.00041832      | 0.58             |
| 1070.15   | 0.00032538      | 0.2              | 1000.15   | 0.00041367      | 0.625            |
| 1070.15   | 0.00047872      | 0.3              | 1000.15   | 0.000455487     | 0.714            |
| 1056.15   | 0.00013017      | 0.087            | 1000.15   | 0.00060879      | 0.892            |
| 1056.15   | 0.00025102      | 0.174            | 975.15    | 0.00019525      | 0.4              |
| 1056.15   | 0.00036720      | 0.261            | 975.15    | 0.00028820      | 0.6              |
| 1056.15   | 0.00051124      | 0.348            | 975.15    | 0.00038114      | 0.8              |
| 1056.15   | 0.00063666      | 0.434            | 975.15    | 0.00046943      | 1                |
|           |                 |                  | 955.15    | 0.00006044      | 0.158            |
|           |                 |                  | 955.15    | 0.00017666      | 0.473            |
|           |                 |                  | 955.15    | 0.00021849      | 0.63             |
|           |                 |                  | 955.15    | 0.00028356      | 0.788            |
|           |                 |                  | 955.15    | 0.00034397      | 1                |

Table 4: Data on activities of graphite in fcc-Fe from Smith [40] covered in the literature database.

| Temp.<br>[K] | C-content $x_C$ | C-activity $a_C$ |       | Temp.<br>[K] | C-content $x_C$ | C-activity $a_C$ |       |
|--------------|-----------------|------------------|-------|--------------|-----------------|------------------|-------|
| 1073.15      | 0.013797        | 0.290            | 0.280 | 1273.15      | 0.004633        | 0.050            | 0.041 |
| 1073.15      | 0.018330        | 0.397            | 0.385 | 1273.15      | 0.009232        | 0.097            | 0.081 |
| 1073.15      | 0.022831        | 0.509            | 0.496 | 1273.15      | 0.013797        | 0.144            | 0.127 |
| 1073.15      | 0.027230        | 0.630            | 0.614 | 1273.15      | 0.018330        | 0.195            | 0.173 |
| 1073.15      | 0.031736        | 0.753            | 0.742 | 1273.15      | 0.022831        | 0.247            | 0.225 |
| 1073.15      | 0.036141        | 0.888            | 0.882 | 1273.15      | 0.027230        | 0.301            | 0.282 |
| 1073.15      | 0.039642        | 1.000            | 1.000 | 1273.15      | 0.031736        | 0.362            | 0.341 |
|              |                 |                  |       | 1273.15      | 0.036140        | 0.424            | 0.406 |
|              |                 |                  |       | 1273.15      | 0.040515        | 0.491            | 0.447 |
|              |                 |                  |       | 1273.15      | 0.044858        | 0.563            | 0.547 |
|              |                 |                  |       | 1273.15      | 0.049170        | 0.639            | 0.626 |
|              |                 |                  |       | 1273.15      | 0.053453        | 0.720            | 0.709 |
|              |                 |                  |       | 1273.15      | 0.057706        | 0.806            | 0.800 |
|              |                 |                  |       | 1273.15      | 0.061929        | 0.899            | 0.896 |
|              |                 |                  |       | 1273.15      | 0.066123        | 1.000            | 1.000 |

### Formation enthalpies

The literature database also contains formation enthalpies from DFT-calculations from three different publications [41, 42, 43] which calculated solution enthalpies of C in  $\alpha$ - and  $\gamma$ -Fe with DFT. Table 5 gives the calculated enthalpies, considered phases and reference states of the calculations.

A comment regarding the dataset of Ponomareva [43]: The reference state of Fe for this DFT-calculation was reported as fcc-Fe. As ESPEI automatically assigns reference states for the calculations and uses bcc-Fe as the default, the reported formation enthalpy from the literature had to be corrected for usage in ESPEI. The corrected formation enthalpy was determined to be  $8077 \text{ J mol}^{-1}$  instead of  $177.01 \text{ J mol}^{-1}$ .

Table 5: Solution enthalpies of C in Fe from DFT-calculations covered in the database. The formation enthalpy of Ponomareva had to be corrected in terms of the considered Fe-phase and is marked with \* for this reason.

| Author     | Literature reference | Phase  | Reference state of C | $x_C$    | $H_{\text{form}}$ [J/mol]       |
|------------|----------------------|--------|----------------------|----------|---------------------------------|
| Hristova   | [41]                 | bcc-Fe | Diamond              | 0.018182 | 1262.93                         |
| Jiang      | [42]                 | bcc-Fe | Graphite             | 0.007752 | 553.41                          |
| Ponomareva | [43]                 | fcc-Fe | hcp-C                | 0.009174 | 177.01 *<br>(corrected to 8077) |

## 4.2 Results of ESPEI-optimization

### 4.2.1 Probabilities, parameter evolutions, correlations

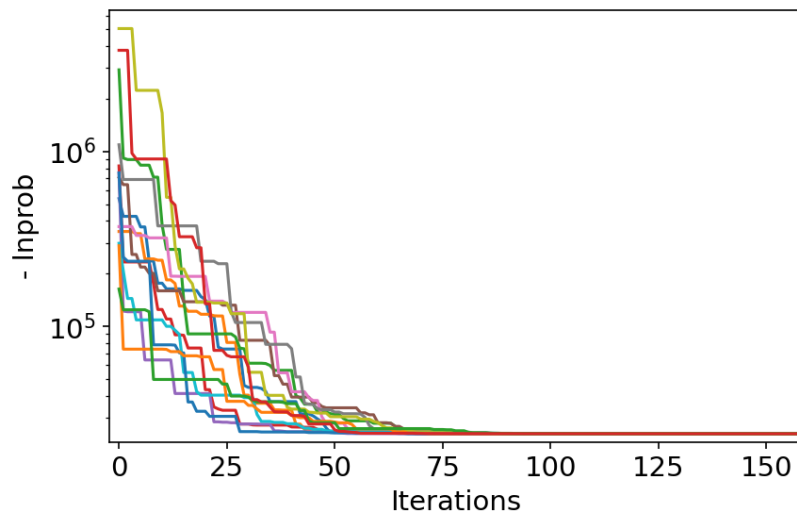


Figure 7: A detailed view on the probabilities of the various Markov chains, which represent the different parameter sets. Since a decrease of the negative logarithm corresponds with a growing probability, the likelihood significantly increases with growing number of iterations.

The literature database for the following results includes three data types: thermodynamic activities, phase boundary data and formation enthalpies from DFT-calculations. After performing an MCMC optimization within ESPEI, the convergence of the posterior probability distribution is evaluated and the final parameter set is reached when the MCMC process has converged. In Figure 7 the convergence of the negative logarithm of the probability function  $-\ln(\text{prob})$  is depicted. It is noted that in statistical methods often the logarithm of the probability is used



and that the falling values of  $-\ln(\text{prob})$  correspond to an increasing probability.

ESPEI allows to analyze the likelihoods resulting from the parameter sets: According to Figure 7 the probability of the resulting model increases significantly within the first 100 iteration steps of the optimization process. The log-probability of the initial parameter set proposed by Gustafson [23] was -30225 and after 3000 iterations, the parameters converged to an optimized parameter set with a log-probability increased to -23987. This verifies a significant improvement of the thermodynamical description as the experimental data are described more accurately. As the MCMC algorithm samples a set of starting parameters normally distributed around the initial parameters in order to overcome potential local minima, the  $-\ln(\text{prob})$  plotted in Figure 7 depicts drastically higher values for first iterations. These probabilities relate to the sampled starting parameters.

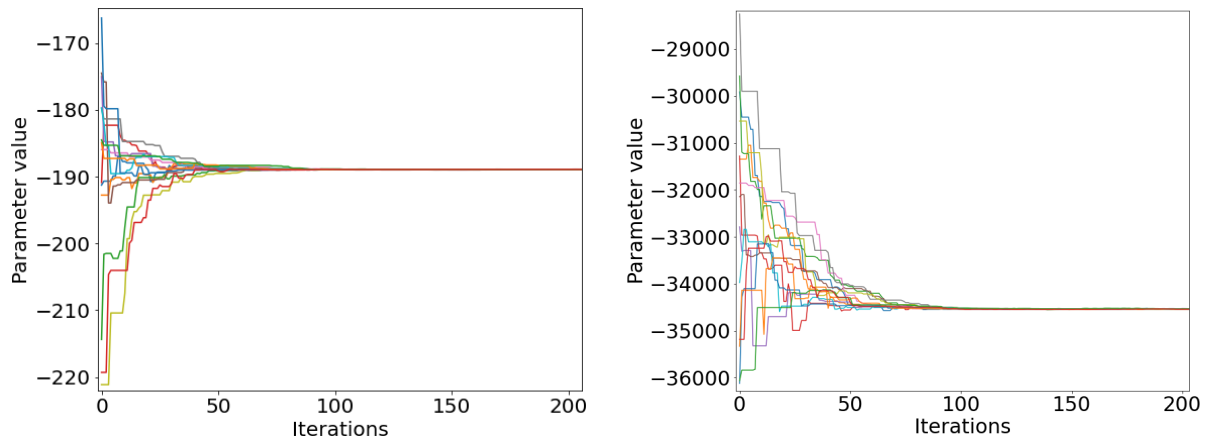


Figure 8: Evolution of the bcc- (left) and fcc-parameter chains (right image). After a so called burn-in phase, the MCMC algorithm proceeds with mostly small changes in the parameters.

The resulting Redlich-Kister parameters show two different trends during the optimization: The bcc- and fcc-parameters (Figure 8) both are relatively stable and do not differ greatly from the initial values after the first phase of roughly 100 iterations, where the different Markov chains consistently approach the optimal value.

In contrast, the parameters of the liquid phase (Figure 9) deviate significantly over a large part of the optimization. One liquid parameter even changes from a positive to a negative sign and they also vary in a much wider range from their initial values than the parameters of the bcc- and fcc-phase. Convergence was reached when the likelihood and the excess parameters attained a stable level for at least 500 iterations. The final set of parameters was obtained after 3000 iterations and the exact values are presented in Table 6.

The behavior of the liquid parameters, exhibiting pronounced spreading of the values during the optimization process, indicates strong correlations between the different parameters. The covariance is analyzed in a corner plot, shown in Figure 10, which presents the parameter distributions along the diagonal and the correlations between each parameter pair. The three

pairs marked in red stand out, showing elongated shapes of the graphs. This clearly implies correlations between the liquid parameters 1 with 2, 1 with 3 and 2 with 3.

This means that the increase of one liquid-parameter is corrected by the decrease of another one. Regarding the relatively large number of parameters and given the strong correlations, this is an indication that a reduced number of parameters could also be sufficient to describe the liquid phase.

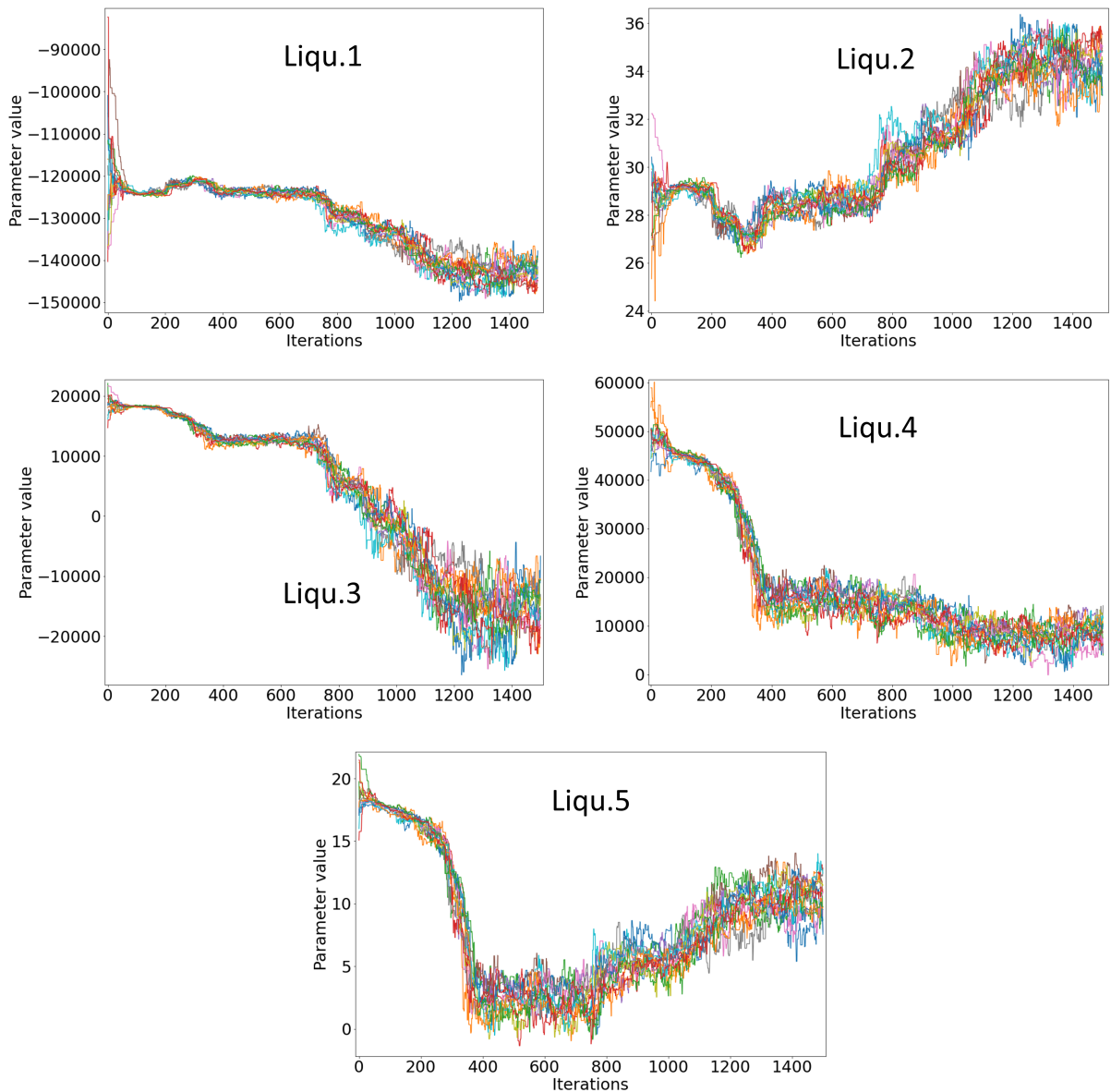


Figure 9: Evolution of the five liquid phase parameters over 1500 iterations. The excess parameters strongly deviate from their initial values in the beginning of the optimization process and oscillate in a wider range than the bcc- and fcc-parameters.

Table 6: Initial [23] and optimal Redlich-Kister-parameters after 3000 iterations. The bcc- and fcc-parameters exhibit relatively small changes, whereas the liquid parameters varied in a wider range.

|           | BCC      | FCC       | Liquid    |       |           |          |       |
|-----------|----------|-----------|-----------|-------|-----------|----------|-------|
| Initial   | -190     | -3.47e+04 | -1.24e+05 | 28.5  | 1.93e+04  | 4.93e+04 | 19    |
| Optimized | -188.91  | -3.45e+04 | -1.45e+05 | 34.93 | -1.79e+04 | 7.88e+03 | 11.11 |
| Std. dev. | 1.25e-02 | 4.12      | 2.38e+03  | 0.74  | 3.77e+03  | 2.45e+03 | 1.21  |

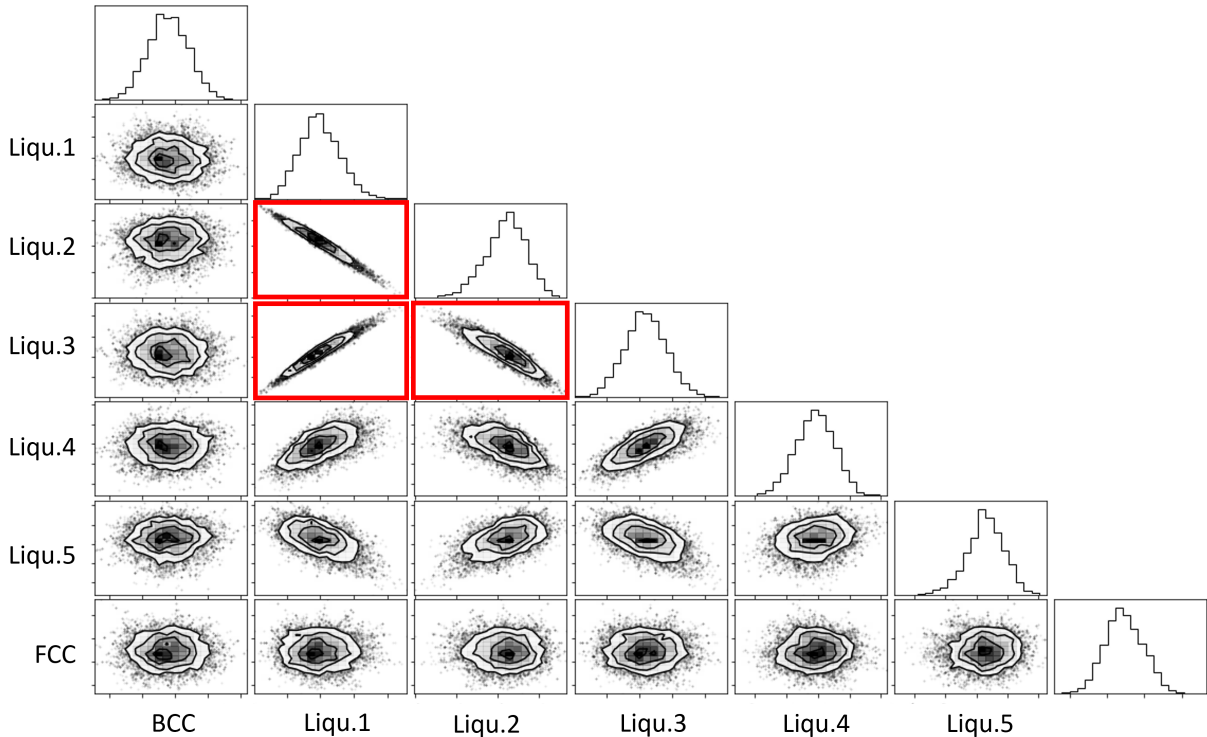


Figure 10: The corner plot represents the parameter distributions and the covariances. The three plots marked in red are standing out as their shapes are clearly elongated. This applies to strong correlations of the liquid parameters 1 with 2, 1 with 3 and 2 with 3.

## 4.2.2 Resulting phase diagram

One of the main goals of this thesis was to use ESPEI to generate a Gibbs energy model for the binary Fe-C system from the newly established literature database and to compare these results to Gustafson's assessment from 1985 [23]. For the model optimization with ESPEI the suggested model parameters from Gustafson were used as the initial parameters followed by an optimization of parameters via the MCMC algorithm.

The underlying experimental data includes phase boundary data for the stable and metastable system, thermodynamic activities of the stable system and solution enthalpies from DFT-calculations.

One aspect of the ESPEI method is that driving forces are defined in order to quantify the deviations of the calculated phase boundaries from the experimental phase equilibria. For each experimental data point the driving force is calculated and serves as a measure for the error between the obtained model and the literature data. The following plots (Figures 11 and 12) present the resulting optimized phase diagrams after 3000 iterations including the underlying experimental data points. Each considered data point is represented by a dot in the diagram with the color of the dot defining the determined driving force: Light yellow stands for low driving forces, meaning experimental and modelled phase boundaries are corresponding; dark red dots represent high driving forces and larger deviations.

As it is shown in Figures 11 and 12 the resulting phase diagrams are in good agreement with the experimental data: The stable Fe-C system (Figure 11) was calculated correctly, as the cementite phase does not appear. It is mentioned that for the calculations cementite was not excluded from the possible phases, but the thermodynamic model correctly predicted only graphite as an appearing carbide-phase in the stable phase diagram. The largest number of literature data points show low driving force. A region of higher driving forces is the eutectic point where a large number of data points is available. Small deviations can have a relatively strong effect on the calculated driving forces. The data points describing the liquidus line at high C-contents also exhibit large driving forces.

For the metastable phase diagram (Figure 12) only bcc-Fe, fcc-Fe, cementite and liquid were considered, as graphite is thermodynamically not a possible phase in the metastable system. The resulting phase diagram also agrees with the literature data, calculating slightly different phase boundaries than for the stable system. The driving forces in the region of the eutectic point are slightly higher than in other regions. For most experimental data points the driving forces are relatively equally distributed. Some scattered data points exhibit pronounced deviations.

The bcc-phase region is characterized by a very low solubility limit. According to Figures 13 and 14 the computed phase boundaries fit the experimental data without any significant differences considering the very small amounts of dissolved C.

In Figures 15 and 16 the phase diagrams of Gustafson compared to the results of this work and the underlying phase boundary data are depicted. In both, the stable and the metastable system, the differences between the ESPEI-optimized and Gustafson's version are not substantial.

In the stable phase diagram the experimental data points of the liquidus line at high C-contents clearly differ from Gustafson's version. The ESPEI-optimization leads to a differently curved liquidus line which fits the experimental data more accurate. It is noted that since the literature database used for this work is based on the literature sources used by Gustafson, the differences obviously are not large, but in some regions of the phase diagram there are noticeable improvements resulting from the MCMC-based optimization.

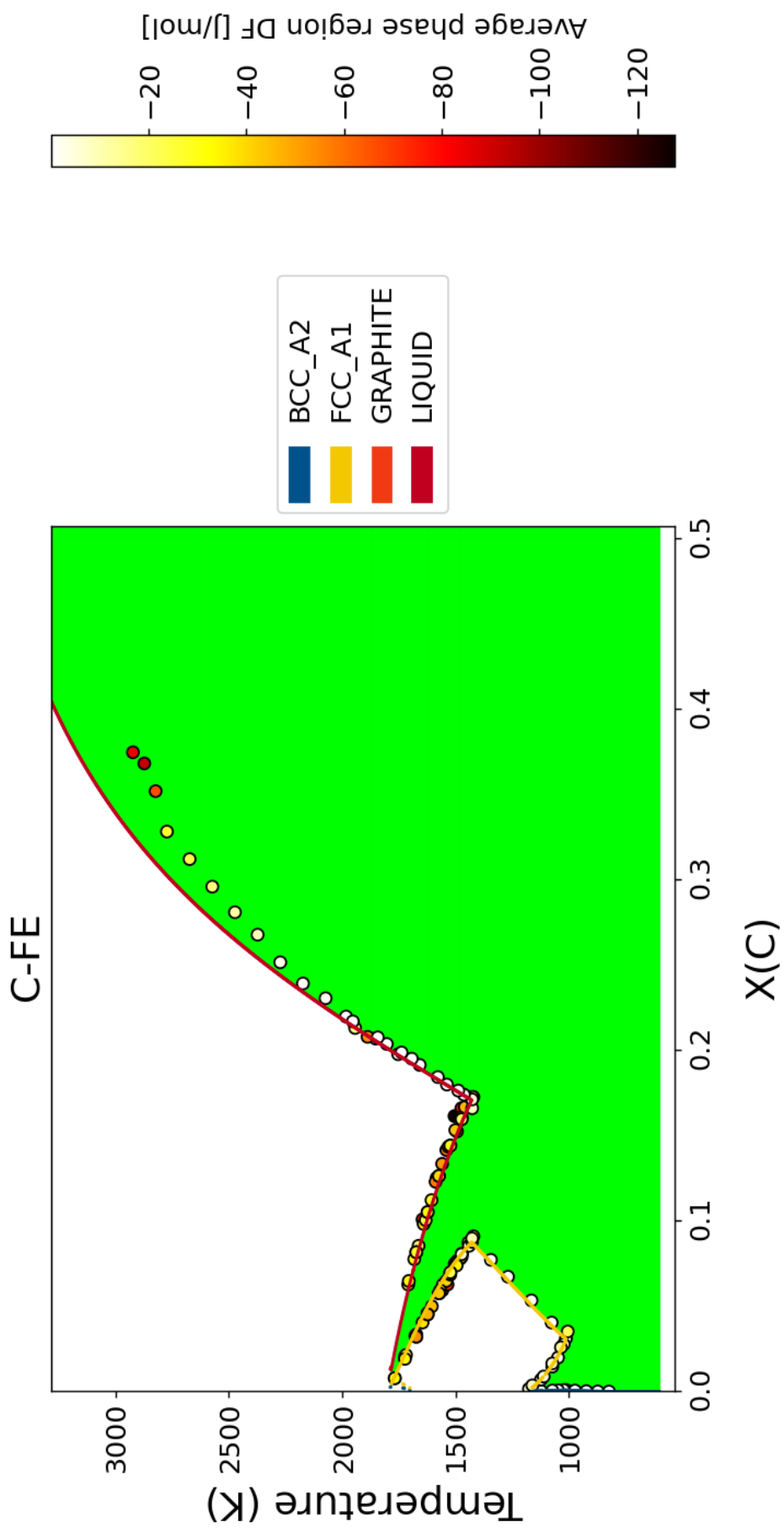


Figure 11: The stable Fe-C phase diagram: phase boundaries plotted as function of temperature and composition, resulting from the ESPEI-optimization fitting to activities, phase boundary data and formation enthalpies. Each dot represents an experimental data point, where the color of the dot defines the driving force.

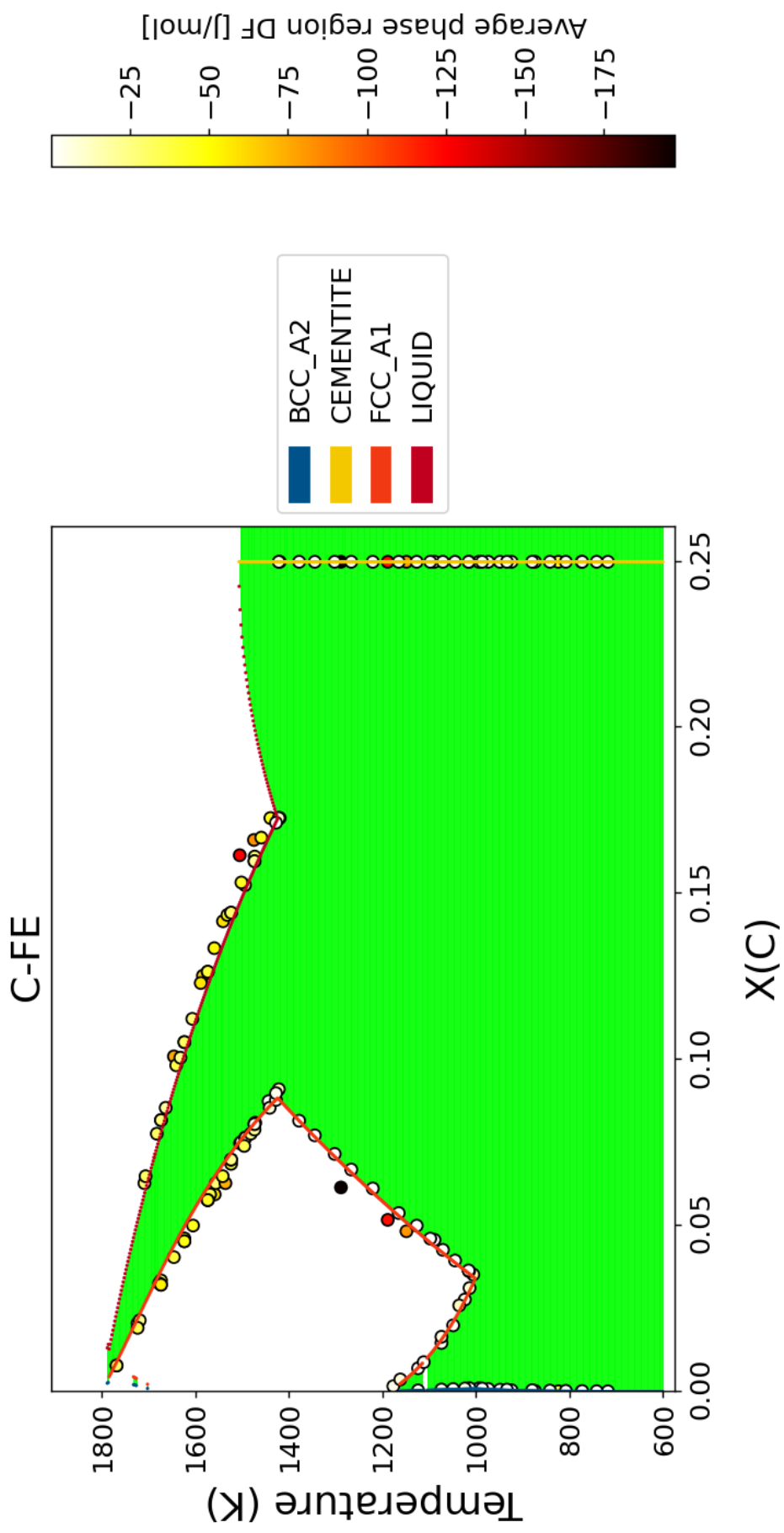


Figure 12: The metastable Fe-C phase diagram, resulting of the ESPEI-optimization fitting to activities, phase boundary data and formation enthalpies. Each dot represents an experimental data point, where the color of the dot defines the driving force.

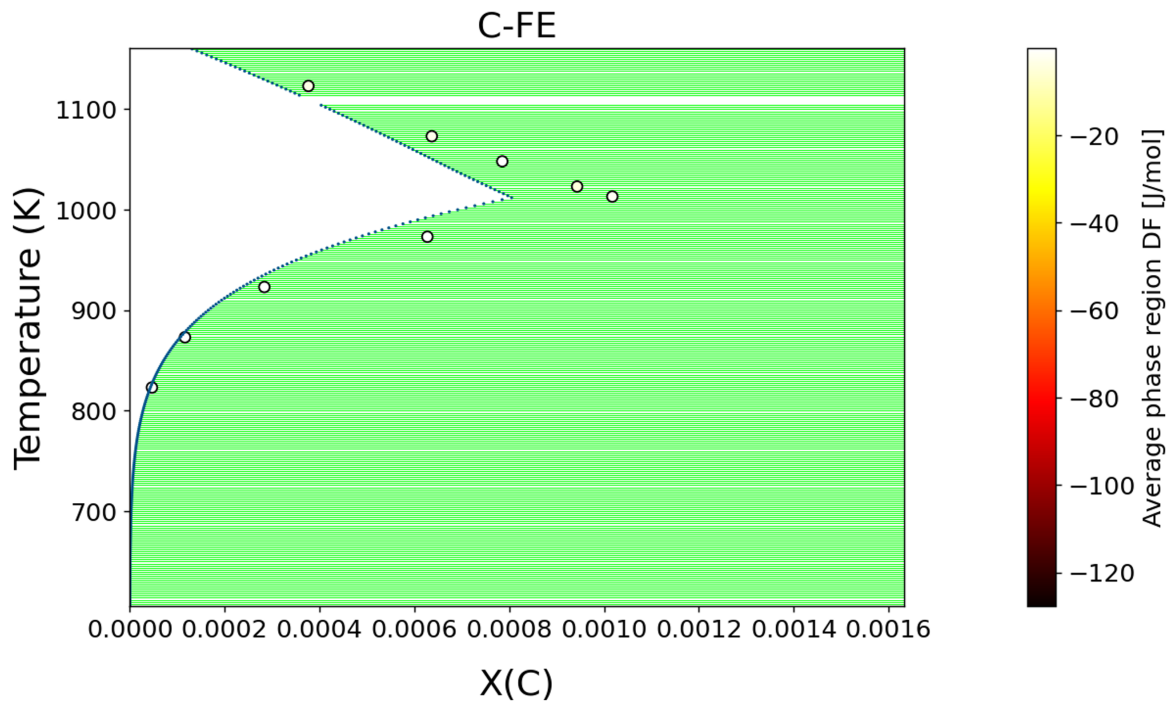


Figure 13: Zoom into the bcc-region of the stable phase diagram: The optimized bcc-solvus line corresponds with the experimental data points of Hasebe [34], resulting in small driving forces.

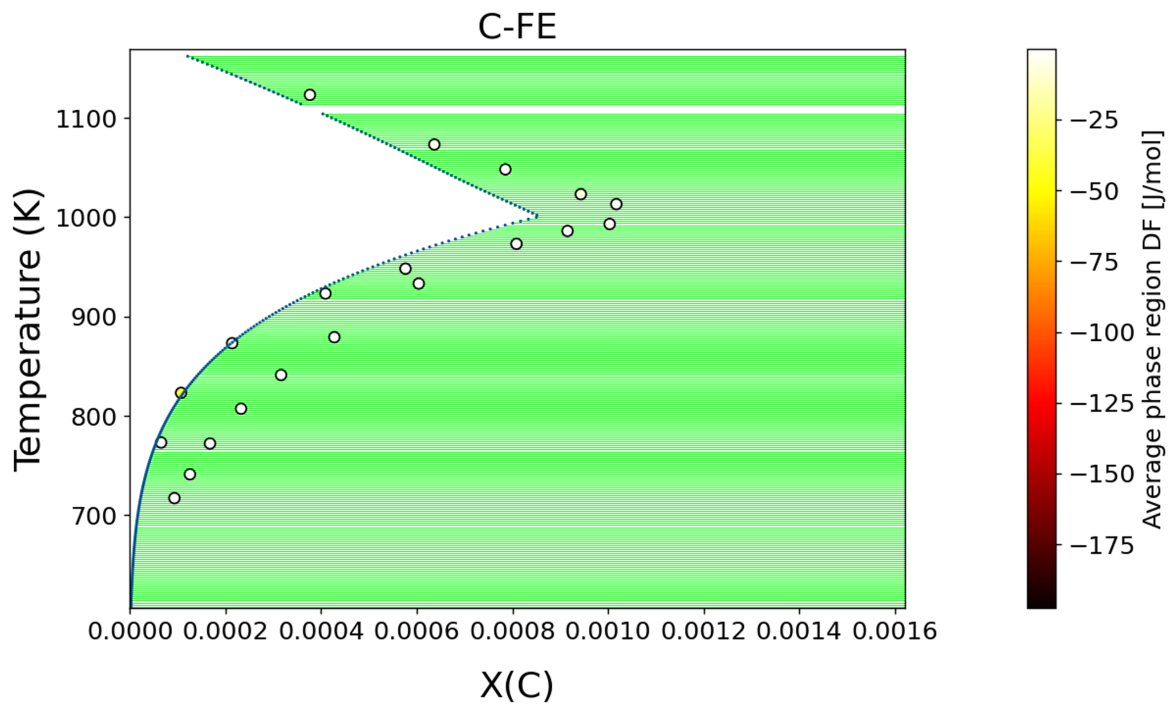


Figure 14: Zoom into the bcc-region of the metastable phase diagram: The datasets of Hasebe [34] and Lindstrand [36] both report similar maximum C-solubilities. The optimized solvus-line fits with the experimental data without causing large driving forces.

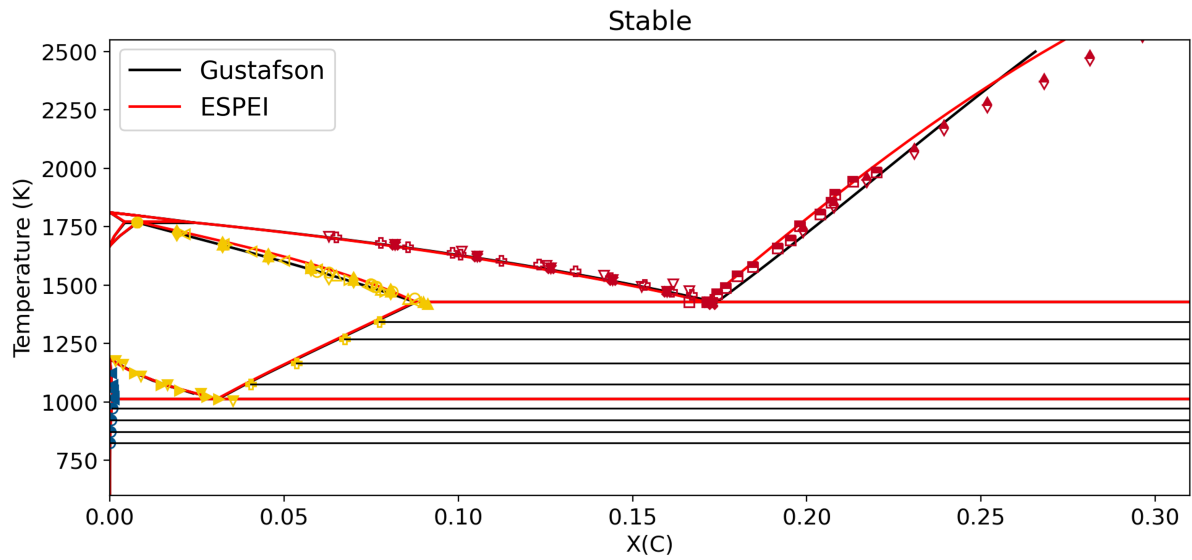


Figure 15: The optimized stable Fe-C phase diagram compared to Gustafson's version and the underlying phase boundary data. The largest deviations occur at the liquidus line at higher C-contents.

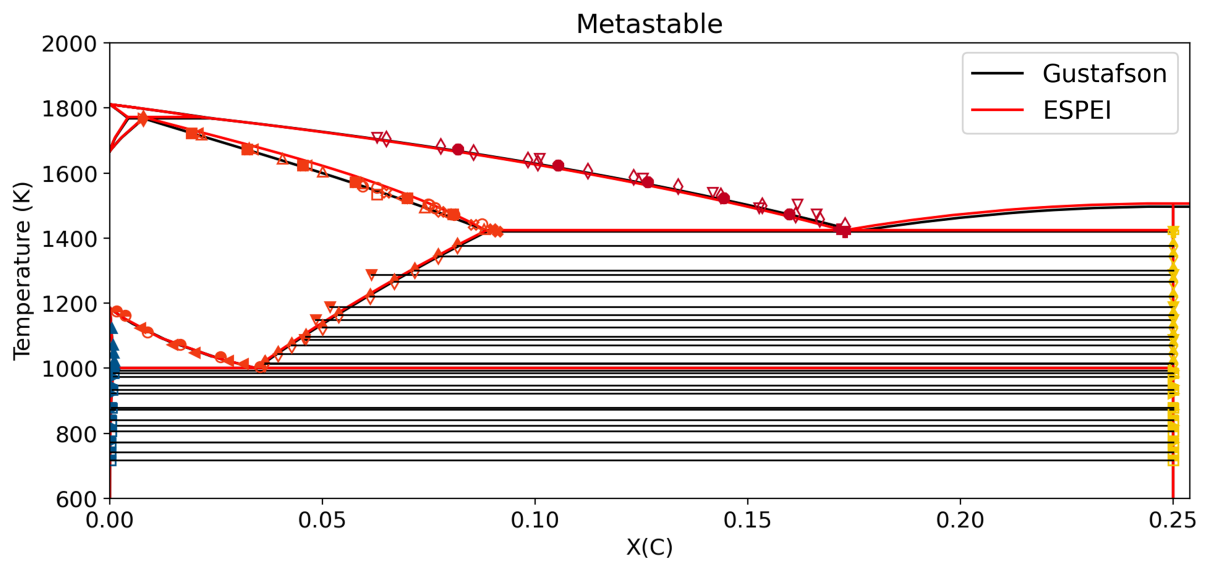


Figure 16: The optimized metastable Fe-C phase diagram compared to Gustafson's version and the underlying phase boundary data. The differences between the two models are relatively small.



### 4.2.3 Modelled activities

As another feature of the resulting Gibbs energy model, the thermodynamical activities were investigated as well. Therefore the thermodynamic activities of the various phases in the system are calculated from the Gibbs descriptions. The experimental measurements covered in the literature database agree with the activities of graphite in bcc- and fcc-Fe as well, as it is depicted in Figures 17 and 18.

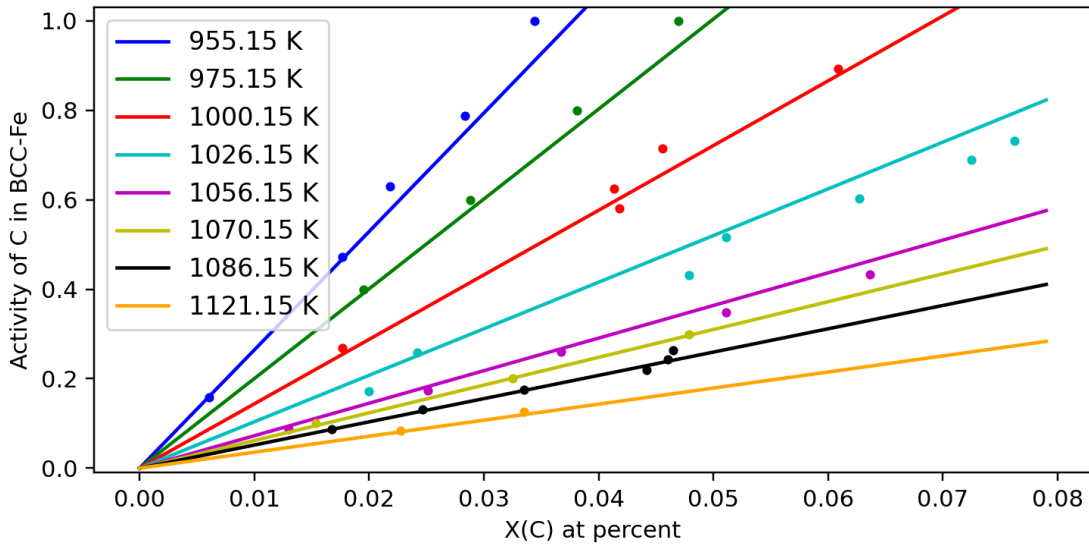


Figure 17: The thermodynamic model optimized with the MCMC algorithm also calculates the thermodynamic activities. After fitting and optimizing the model to all collected datatypes, the experimental points of graphite in bcc-Fe are corresponding with the computed activity model, represented by the lines.

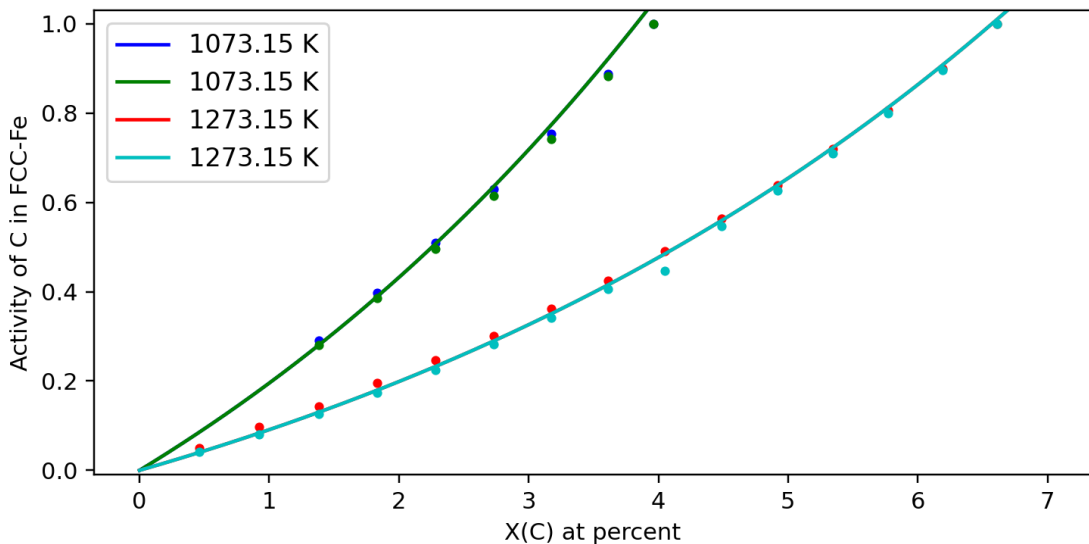


Figure 18: The model of the activities of graphite in fcc-Fe also exhibits agreement with the experimental data.

#### 4.2.4 Modelled formation enthalpies

The resulting thermodynamic database was also used to compare the formation enthalpies. The literature data which served as input data and the values obtained from the optimized model are presented in Table 7. The comparison shows that the model gives values higher than the literature data but still of a reasonable order of magnitude.

As mentioned in Chapter 4.1.2, for the calculations in ESPEI the bcc-phase is considered as the reference phase of Fe. As Ponomareva referred to fcc-Fe, the value was corrected to the bcc-Fe, which leads to a corrected formation enthalpy of  $8077 \text{ J mol}^{-1}$ .

Table 7: Comparison of the formation enthalpies from literature with the modelled values. As ESPEI considers the bcc-phase as the default reference phase of Fe, the enthalpy of Ponomareva had to be corrected.

| Author     | Literature reference | Phase  | Reference phase of C | $x_C$    | $H_{\text{form}}$ [J/mol] | $H_{\text{form}}^{\text{model}}$ [J/mol] |
|------------|----------------------|--------|----------------------|----------|---------------------------|--|
| Hristova   | [41]                 | bcc-Fe | Diamond              | 0.018182 | 1262.93                   | 1917.6                                   |
| Jiang      | [42]                 | bcc-Fe | Graphite             | 0.007752 | 553.41                    | 832.2                                    |
| Ponomareva | [43]                 | fcc-Fe | hcp-C                | 0.009174 | 8077 *<br>(corrected)     | 8294.6                                   |

#### 4.2.5 Conflicting datasets

An important aspect of creating a literature database is evaluating and selecting reliable and consistent data. During the process of gathering different data, there were also datasets available which reported contradictory data. This can be datasets which do not agree about the same trend in a property, for example on the bcc-solvus line. The bcc-phase region is characterized by very low carbon solubility which reaches its maximum at a temperature of about 1000 K. The literature search led to several publications with measurements on this phase boundary. The available literature sources on the bcc-solvus line in the stable system consisted of Hasebe [34] and Chipman [44].

Whereas Hasebe measured a maximum C-solubility of 0.06 at.%, Chipman reported 0.02 at.% and Gustafson's thermodynamic model gives a maximum C-solubility of about 0.09 at.%. Even though those differences are small in absolute numbers, the relative difference in this case is significant, as Hasebe reports a solubility limit almost three times as high as Chipman's. Considering the very low solubility limit of C in bcc-Fe, conducting measurements with high accuracy is difficult and the measurement errors are presumably relatively large in this region of the phase diagram.

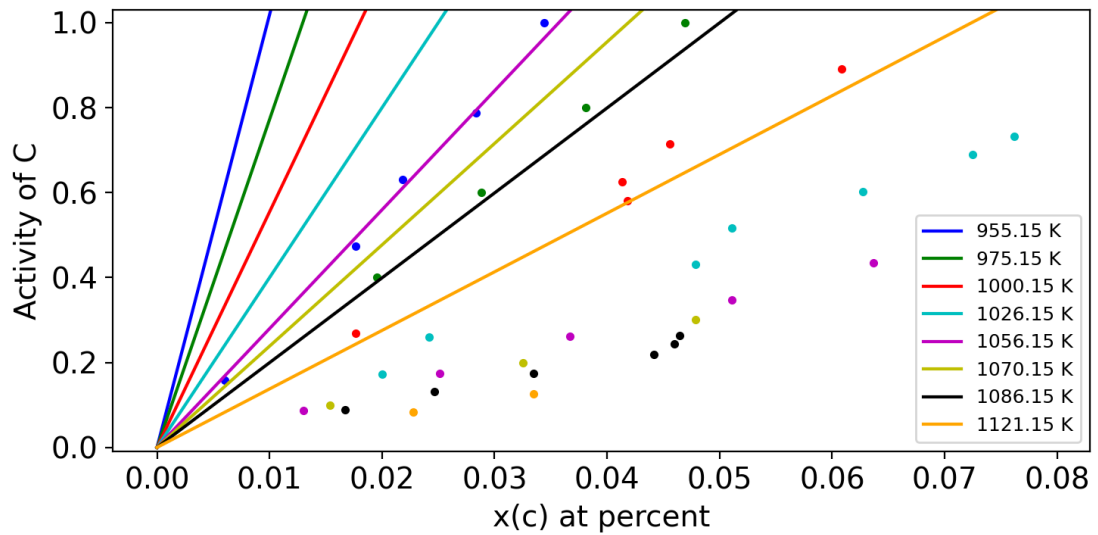


Figure 19: The model based on stable datasets and Chipman's data [44] with a lower C-solubility limit in bcc-Fe than Hasebe's [34] leads to significant deviations between the model and experimental activity data.

Including only one of the measurements for an MCMC-simulation, the algorithm is capable of fitting the phase boundaries to both datasets equally. Including both datasets for an optimization, the algorithm within ESPEI will always fit the phase boundary with lower C-contents due to the error-definition of the method. This is because the algorithm aims to minimize the driving forces, and therefore the phase boundary is calculated to lie at the lower C-concentration as this leads to minimal driving forces. Therefore, it is critical to select one of the two contradictory datasets.

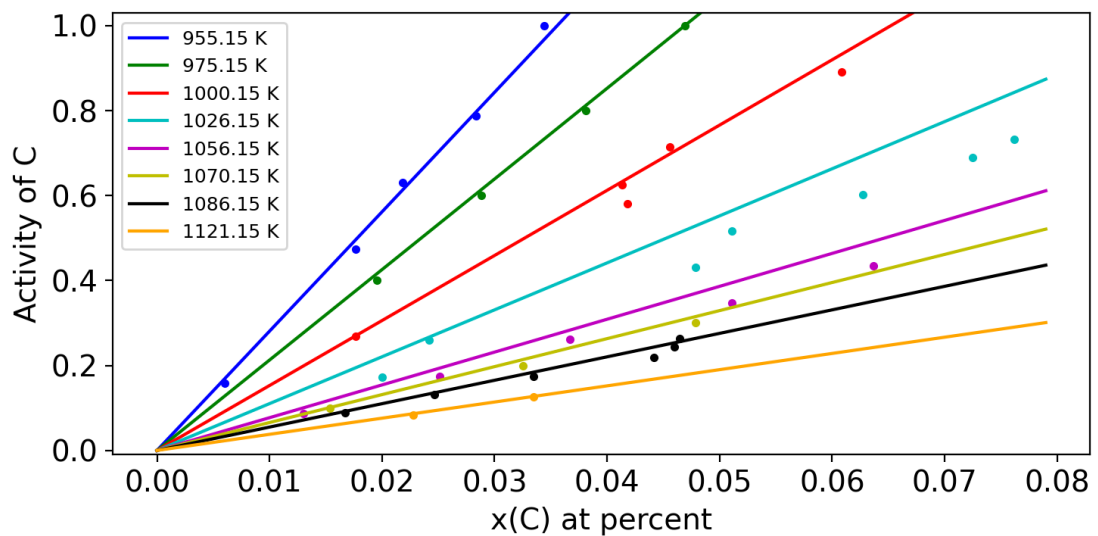


Figure 20: The obtained model of the stable system including Hasebe's dataset [34] is in good agreement with the measured activities.

While in the phase diagram it was not possible to differentiate the two datasets, the activity models preferred the dataset with the higher C-solubility limit: The modelled activities in Figure 19 were generated with a database including only stable datasets and only Chipman's data for the bcc-solvus line. The result exhibits significant deviations between the activity model and the measured data points, whereas the assessment of the stable system including only on Hasebe's bcc-solubility data (Figure 20) agrees with the experimental data. Based on these conclusions Chipman's data on the bcc-solubility limit was not included in the literature database, but Hasebe's dataset only.

## 4.2.6 Uncertainty quantification

After completing an MCMC simulation and obtaining a converged parameter set, the resulting probability distributions are propagated to the phase boundaries to give the uncertainties of the model and to evaluate the effects of the various datasets.

Figure 21 presents the calculated uncertainties of the stable phase diagram. In order to depict the uncertainties, the various phase boundaries relating to the different parameter chains are superimposed. The blurred lines represent the phase boundaries with larger uncertainties as the parameter chains propose several slightly different phase boundaries. The gaps occurring between some of the lines are due to numerical reasons and have no meaning.

For the stable phase diagram, most of the phase boundaries are characterized by small uncertainties. The region around the eutectic and the liquidus line at higher C-contents (Figure 22) are associated with the largest uncertainties. This correlates with the smaller number of available data for this part of the phase diagram. It can also be assumed that for high temperatures and high C-contents the measurement errors increase, resulting in lower reliability of the data points.

For the metastable system there is a relatively large number of data points available which causes high reliability of the modelled phase boundaries. As for the stable system, the region of higher C-contents show larger uncertainties as there is no experimental data available for C-contents higher than the eutectic point at 17.3 at.% which corresponds to 4.3 wt.%.

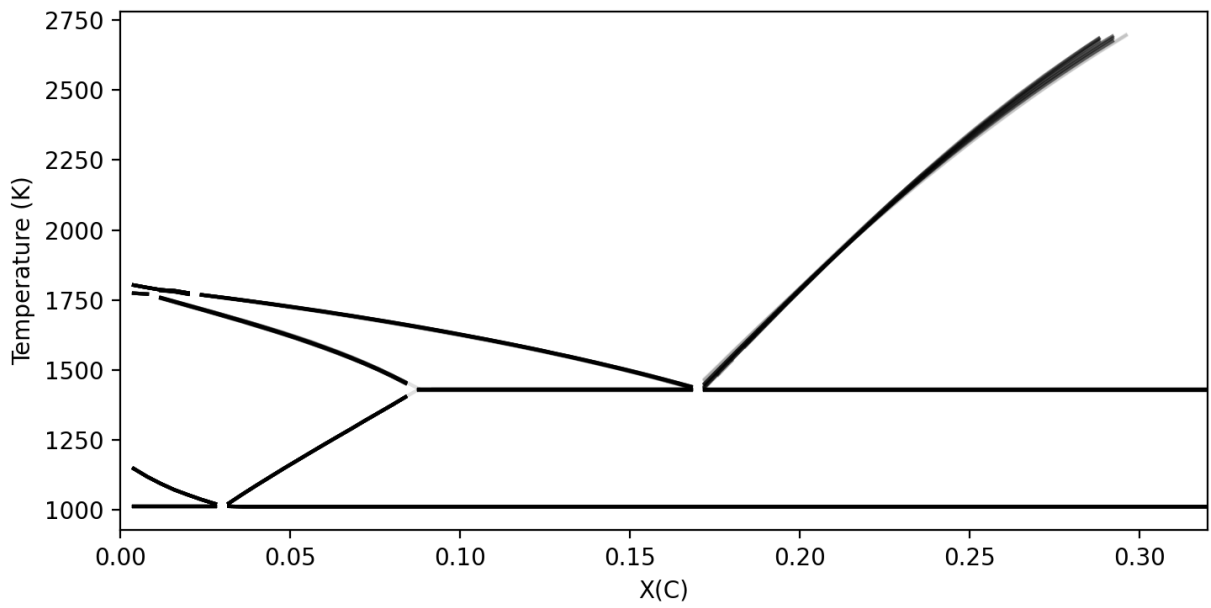


Figure 21: By superimposing the phase boundaries determined by the various parameter chains, the reliability of the modelled phase boundaries can be estimated. In regions with lower density of measured data the uncertainties are larger than in regions with more data points available.

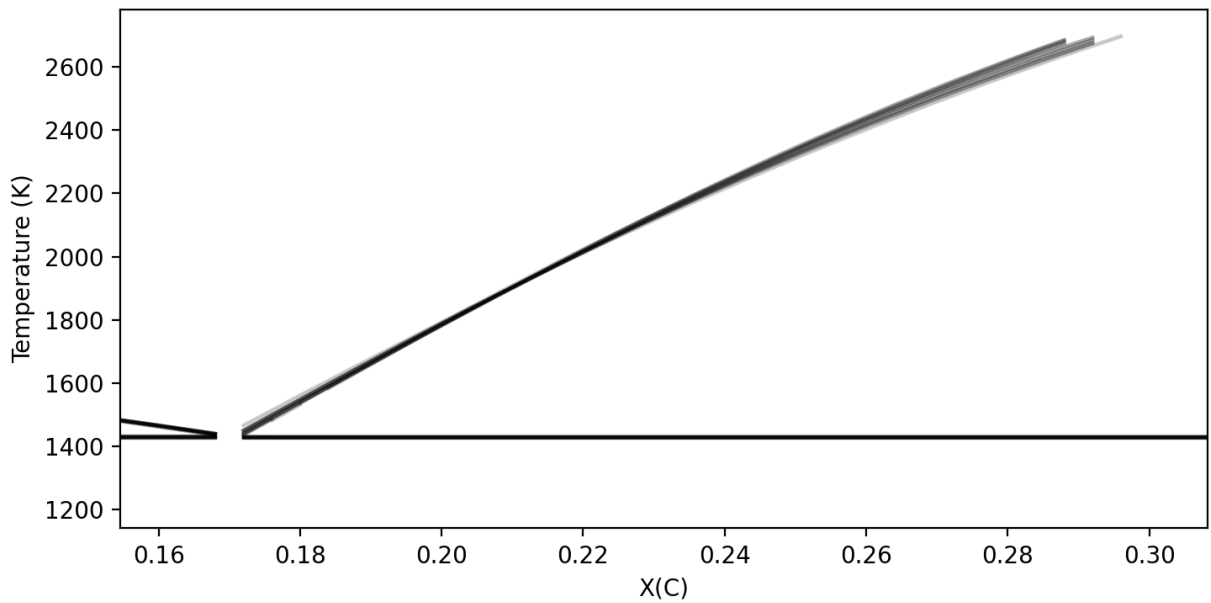


Figure 22: Zoom into the region of higher C-contents: less data points are available, which increases the uncertainty of the liquidus line. Also higher temperatures and high C-contents evoke larger measurement errors which lead to larger deviations as well.

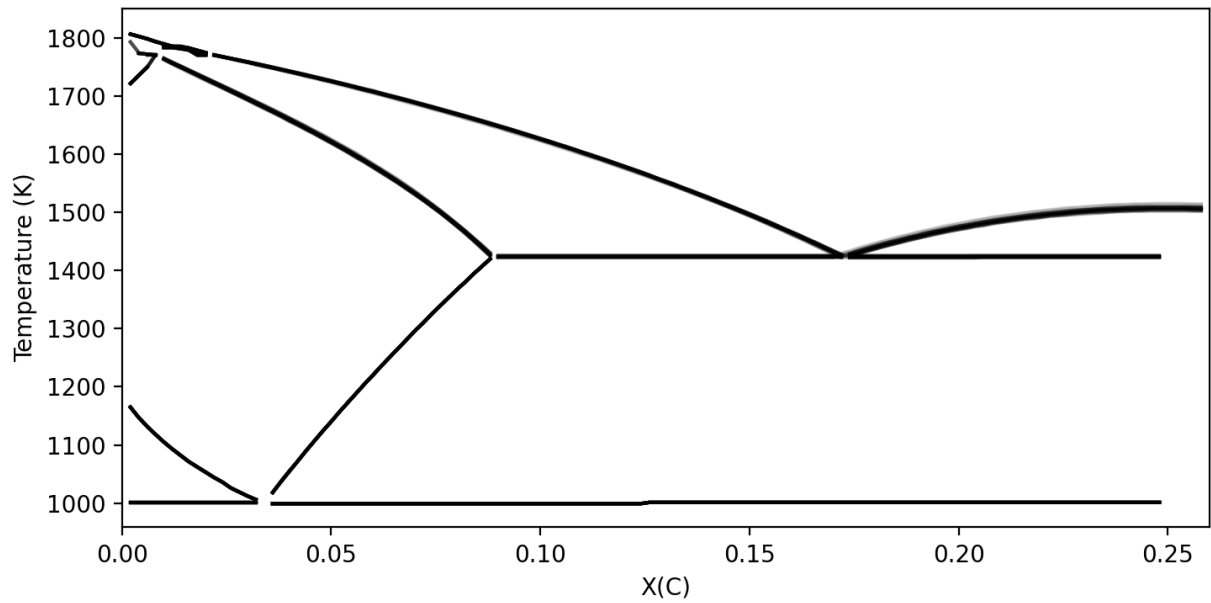


Figure 23: The large number of data points describing the metastable system lowers the uncertainties. Slightly larger uncertainties are observed for the liquidus line at higher C-contents.

Calculating the phase boundary uncertainties points out that providing additional data points in specific regions of the phase diagram can help to improve the accuracy of the thermodynamic description. While the available data for C-contents up to the eutectic point result in small uncertainties, new experiments at higher C-contents will have the largest benefit for the resulting thermodynamic model. This applies for both, the stable and the metastable system.

The calculated uncertainties of the modelled activities and formation enthalpies are very small. For the case of the activity, the deviations arising from uncertainty are smaller than the line thickness in Figure 17 and are therefore not relevant.

In the case of the formation enthalpies the determined uncertainties are negligible as well. These results correlate with the very small variances of the optimized bcc- and fcc-Redlich-Kister parameters (see Table 6 and Figure 10).

## 5 Summary

The goal of this work was to perform a thermodynamic assessment of the binary Fe-C system and to compare to the Gibbs energy model of Gustafson from 1985 [23] which is still the most accepted Fe-C-CALPHAD assessment. For this work the first step was a comprehensive literature search to collect thermodynamical data which served as the foundation of the assessment. The literature database contained phase boundary data and thermodynamic activities from experiments and formation enthalpies from ab-initio calculations. The Python-based database allows adaption to user-specific requirements and easy reassessment at a later time, for example when new data is added.

The parameter optimization for the CALPHAD assessment was performed with ESPEI with MCMC in the Bayesian framework also provides the probability distributions of the parameters. This allows the propagation of the parameter uncertainties and the effect of different input data on the quantities of interest that can be observed.

As a result, a new parametrization for the Fe-C binary system was obtained which is fully reproducible and has a higher probability to explain the thermodynamic data points compared to the parameter set proposed by Gustafson within the chosen error definition. The optimized phase diagram was presented along with the relating parameters in form of a TDB-file which considers the underlying data and estimates the uncertainties of the calculated phase boundaries.

It was found that the parameters of the liquid phase can be strongly altered and that a substantially different value should be chosen for the third parameter. Investigations revealed that including various data types help select reliable datasets and increase the accuracy of an assessment. Based on the calculated uncertainties it was pointed out that providing new phase boundary data points in the eutectic and the high carbon-region can further improve the reliability of the assessment. The calculated uncertainties of the activities were found to be very small. The formation enthalpies exhibited negligible uncertainties as well. This correlates with small variances of the optimized bcc- and fcc-Redlich-Kister parameters.

The ability to provide the resulting reliability of an assessment offers great potential for future investigations of multicomponent systems, as regions of the phase diagram with lower reliability can be estimated based on the uncertainty quantification and new experiments can be guided. This can greatly improve the accuracy of a thermodynamic assessment.









# Bibliography

- [1] Bo Sundman, Bo Jansson, and Jan-Olof Andersson. The thermo-calc databank system. *Calphad*, 9(2):153–190, apr 1985.
- [2] C.W. Bale, E. Bélisle, P. Chartrand, S.A. Deckerov, G. Eriksson, K. Hack, I.-H. Jung, Y.-B. Kang, J. Melançon, A.D. Pelton, C. Robelin, and S. Petersen. FactSage thermochemical software and databases — recent developments. *Calphad*, 33(2):295–311, jun 2009.
- [3] Ernst Kozeschnik and Bruno Buchmayr. Matcalc – A simulation tool for multicomponent thermodynamics, diffusion and phase transformation kinetics. In *Mathematical modelling of weld phenomena*, pages 349–361. 2001.
- [4] Richard Otis and Zi-Kui Liu. pycalphad: CALPHAD-based computational thermodynamics in python. *Journal of Open Research Software*, 5(1):1, jan 2017.
- [5] Brandon Bocklund, Richard Otis, Aleksei Egorov, Abdulmonem Obaied, Irina Roslyakova, and Zi-Kui Liu. ESPEI for efficient thermodynamic database development, modification, and uncertainty quantification: application to cu–mg. *MRS Communications*, 9(2):618–627, jun 2019.
- [6] N. Saunders and A.P. Miodownik. *CALPHAD: A Comprehensive Guide*. Elsevier Science Ltd, 1998.
- [7] Bo Sundman, H.L. Lukas, and S.G. Fries. *Computational thermodynamics: The Calphad Method*. Cambridge university press Cambridge, 2007.
- [8] Phase diagrams and solidification. <https://www.doitpoms.ac.uk/tlplib/phase-diagrams/phasediags1.php>, August 2022. DoITPoMS.
- [9] Werner Pepperhoff and Mehmet Acet. The structure of iron. In *Constitution and Magnetism of Iron and its Alloys*, pages 1–13. Springer Berlin Heidelberg, 2001.
- [10] H. Okamoto. The C-Fe (carbon-iron) system. *Journal of Phase Equilibria*, 13(5):543–565, oct 1992.
- [11] Gabriele Cacciamani. An introduction to the CALPHAD method and the compound energy formalism (CEF). *Tecnologia em Metalurgia Materiais e Mineração*, 13(1):16–24, 2016.
- [12] Brandon Bocklund. *Computational Design of Additively Manufactured Functionally Graded Materials by Thermodynamic Modeling with Uncertainty Quantification*. PhD thesis, The Pennsylvania State University, 2021.

- [13] Ursula R. Kattner. The CALPHAD method and its role in material and process development. *Tecnologia em Metalurgia Materiais e Mineração*, 13(1):3–15, 2016.
- [14] H.L. Lukas, E.Th. Henig, and B. Zimmermann. Optimization of phase diagrams by a least squares method using simultaneously different types of data. *Calphad*, 1(3):225–236, jan 1977.
- [15] Donald W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, jun 1963.
- [16] Erich Königsberger. Improvement of excess parameters from thermodynamic and phase diagram data by a sequential bayes algorithm. *Calphad*, 15(1):69–78, jan 1991.
- [17] A.T. Dinsdale. SGTE data for pure elements. *Calphad*, 15(4):317–425, oct 1991.
- [18] W. Cao, S.-L. Chen, F. Zhang, K. Wu, Y. Yang, Y.A. Chang, R. Schmid-Fetzer, and W.A. Oates. PANDAT software with PanEngine, PanOptimizer and PanPrecipitation for multi-component phase diagram calculation and materials property simulation. *Calphad*, 33(2):328–342, jun 2009.
- [19] Ping Fang Shi, Anders Engström, Bo Sundman, and John Ågren. Thermodynamic calculations and kinetic simulations of some advanced materials. *Materials Science Forum*, 675-677:961–974, feb 2011.
- [20] Bo Sundman, Ursula R Kattner, Mauro Palumbo, and Suzana G Fries. OpenCalphad - a free thermodynamic software. *Integrating Materials and Manufacturing Innovation*, 4(1):1–15, jan 2015.
- [21] Mats Hillert. The compound energy formalism. *Journal of Alloys and Compounds*, 320(2):161–176, 2001.
- [22] Otto Redlich and A. T. Kister. Algebraic representation of thermodynamic properties and the classification of solutions. *Industrial and Engineering Chemistry*, 40(2):345–348, feb 1948.
- [23] Per Gustafson. A thermodynamic evaluation of the fe–c system. *Scand. J. Metall.*, 14(5):259–267, 1985.
- [24] G. Inden. *Proceedings of CALPHAD V*, pages 1–13, 1976.
- [25] Mats Hillert and Magnus Jarl. A model for alloying in ferromagnetic metals. *Calphad*, 2(3):227–238, jan 1978.
- [26] AF Guillermet and P Gustafson. An assessment of the thermodynamic properties and pt phase diagram of iron trita-mac-0229. 1984.
- [27] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312, mar 2013.

- [28] MG Benz and JF Elliott. The austenite solidus and revised iron-carbon diagram. *Transactions of the Metallurgical Society of AIME*, 221(2):323–331, 1961.
- [29] RA Buckley and W Hume-Rothery. Liquidus and solidus relations in iron-rich iron-carbon alloys. *J. Iron Steel Inst.*, 196(12):403–406, 1960.
- [30] RA Buckley and W Hume-Rothery. *J. Iron Steel Inst.*, 200:142–143, 1962.
- [31] Bruno Chicco and Warren R. Thorpe. Experimental determination of the austenite and liquid phase boundaries of the Fe-C system. *Metallurgical Transactions A*, 13(7):1293–1297, jul 1982.
- [32] Bruno Chicco and Warren R. Thorpe. A further determination of the austenite solidus of the Fe-C system. *Metallurgical Transactions A*, 14(1):312–314, feb 1983.
- [33] J Chipman. *Metals Handbook, American Society for Metals*, 8(8):277, 1973.
- [34] M Hasebe, H Ohtani, and T Nishizawa. Effect of magnetic transition on solubility of carbon in bcc Fe and fcc Co-Ni alloys. *Metallurgical Transactions A*, 16(5):913–921, 1985.
- [35] Seibei Konno. A study of the A1 and A3 transformations in carbon steels by means of a differential dilatometer. *Report of the Research Institute for Iron, Steel and Other Metals*, 74.
- [36] E. Lindstrand. A method for the measurement of elastic relaxation, and its use for determination of the solubility of carbon in  $\alpha$ -iron. *Acta Metallurgica*, 3(5):431–435, sep 1955.
- [37] Rudolf Ruer and Julius Biren. Über die Löslichkeit des Graphits in geschmolzenem Eisen. *Zeitschrift für anorganische und allgemeine Chemie*, 113(1):98–112, oct 1920.
- [38] Erich Scheil, Theodor Schmidt, and Joachim Wüning. Ermittlung der Gleichgewichte von Kohlenoxyd-Kohlensäure-Gemischen mit dem Gamma-Mischkristall, mit Zementit und mit Graphit. *Archiv für das Eisenhüttenwesen*, 32(4):251–260, apr 1961.
- [39] Joseph A. Lobo and Gordon H. Geiger. Thermodynamics and solubility of carbon in ferrite and ferritic Fe-Mo alloys. *Metallurgical Transactions A*, 7(9):1347–1357, sep 1976.
- [40] Rodney P. Smith. Equilibrium of iron-carbon alloys with mixtures of CO-CO<sub>2</sub> and CH<sub>4</sub>-H<sub>2</sub>. *Journal of the American Chemical Society*, 68(7):1163–1175, jul 1946.
- [41] Elisaveta Hristova, Rebecca Janisch, Ralf Drautz, and Alexander Hartmaier. Solubility of carbon in  $\alpha$ -iron under volumetric strain and close to the  $\sigma_5$  (3 1 0)[0 0 1] grain boundary: Comparison of dft and empirical potential methods. *Computational Materials Science*, 50(3):1088–1096, 2011.
- [42] D. E. Jiang and Emily A. Carter. Carbon dissolution and diffusion in ferrite and austenite from first principles. *Physical Review B*, 67(21):214103, jun 2003.

- [43] A. V. Ponomareva, Yu. N. Gornostyrev, and I. A. Abrikosov. Ab initio calculation of the solution enthalpies of substitutional and interstitial impurities in paramagnetic fcc-Fe. *Physical Review B*, 90(1):014439, jul 2014.
- [44] John Chipman. Thermodynamics and phase diagram of the Fe-C system. *Metallurgical and Materials Transactions B*, 3(1):55–64, jan 1972.