Calphad Optimizer: SQL Insights on Refining, ML, and Quality Metrics

Bruno Reis, Florian Tang, Abdulmonem Obaied, Philipp Keuter, Moritz to Baben
GTT-Technologies

CALPHAD Conference, 28th of May 2024

Thermodynamic Software

Thermodynamic Databases

Consulting Services

Outline

- Introduction to Calphad Optimizer
- The SQLite results database
- Command line interface
- 1st Analysis: Optimization performance
- 2nd Analysis: Exploratory data analysis
- 3rd Analysis: Machine learning
- Conclusion

Introduction to Calphad Optimizer

Thermodynamic Software

Thermodynamic Databases

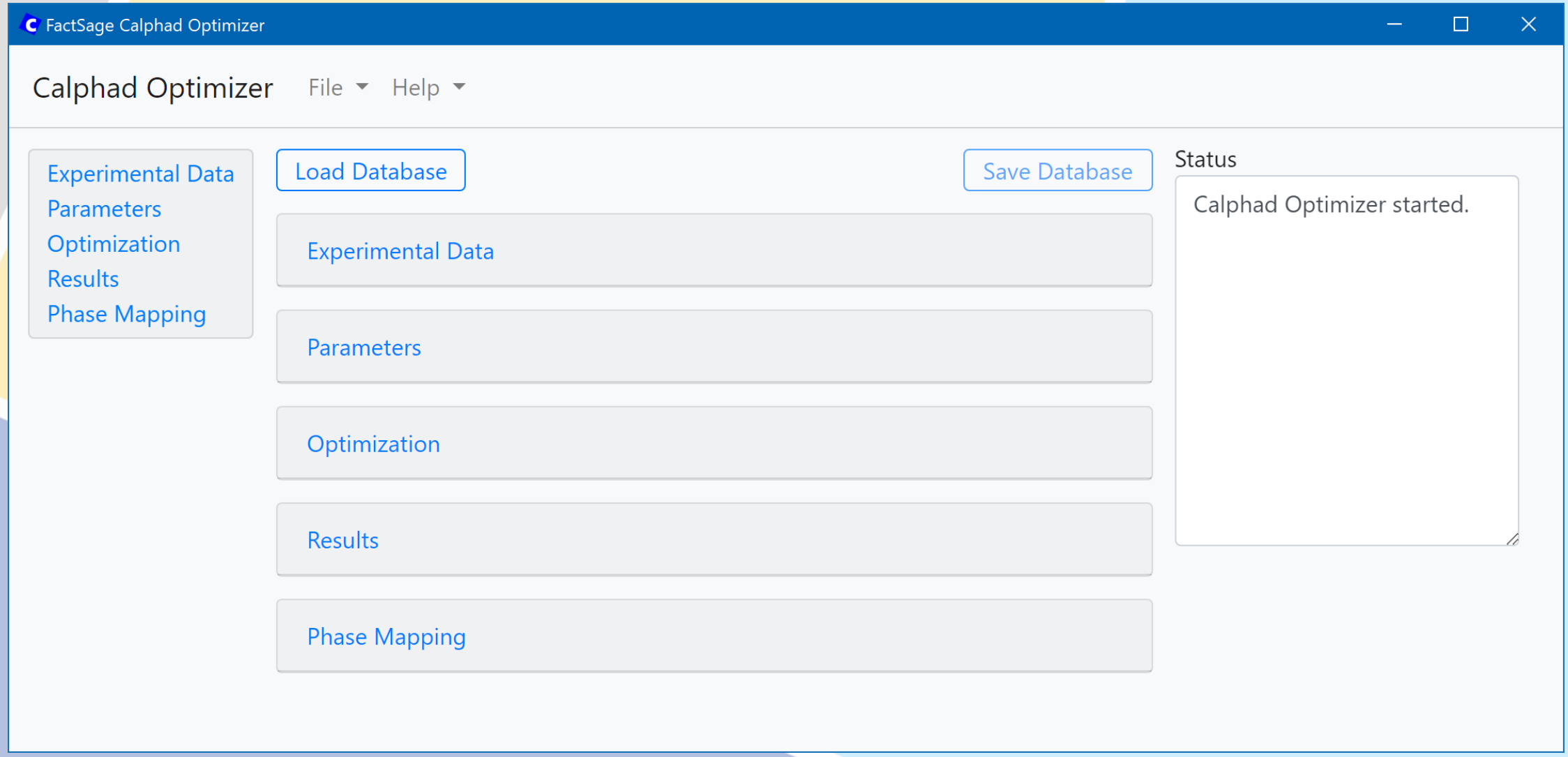
Consulting Services

Why does GTT develop Calphad Optimizer

- Calphad should be accessible to more users
  - A GUI tool that can be understood by any material scientists
  - Yet, no power user should be lost
- FAIR data! Responsibility of GTT:
  - I: Interoperability
    Provide data structures suited for open and accessible tools and workflows
  - R: Reproducibility
    Make the outcome of optimizations reliably deterministic
- Higher throughput than established tools
  - Minimize necessary interaction during optimization

Calphad Optimizer

Start-up screen



Calphad Optimizer

Optimization control

Optimization

Selected experiments All Clear

- [x] Hmix liquid Go To
- [x] Hmix bcc Go To
- [x] Hmix fcc Go To
- [x] Solidus fcc (Ni) Go To
- [x] Solidus bcc (Cr) Go To
- [x] Liquidus bcc (Cr) Go To
- [x] Liquidus fcc (Ni) Go To
- [x] Ni solubility in BCC (Cr) _ reversed Go To
- [x] Cr solubility in FCC (Ni) Go To
- [x] Activity 1873 K - Cr in liquid Go To
- [x] Activity 1273 K - Cr in bcc Go To

Selected parameters All Clear

	Variable	Identifier	Parameter	Value	Minimum	Maximum	Coupling (?)
☑	V1	LIQUID	$L^{0}_{Cr,Ni;}$ A	-1.23200e+4	-1.00000e+5	1.00000e+5	☐
☑	V2	LIQUID	$L^{0}_{Cr,Ni;}$ B	-1.12000e+0	-1.00000e+2	1.00000e+2	☐
☑	V3	LIQUID	$L^{1}_{Cr,Ni;}$ A	5.01000e+3	-1.00000e+5	1.00000e+5	☐
☑	V4	LIQUID	$L^{1}_{Cr,Ni;}$ B	2.01000e+0	-1.00000e+2	1.00000e+2	☐
☑	V5	FCC_A1	$L^{0}_{Cr,Ni:Va;}$ A	6.50000e+3	-1.00000e+5	1.00000e+5	☐
☑	V6	FCC_A1	$L^{0}_{Cr,Ni:Va;}$ B	-1.28200e+1	-1.00000e+2	1.00000e+2	☐
☑	V7	FCC_A1	$L^{1}_{Cr,Ni:Va;}$ A	2.61400e+4	-1.00000e+5	1.00000e+5	☐
☑	V8	FCC_A1	$L^{1}_{Cr,Ni:Va;}$ B	-9.29000e+0	-1.00000e+2	1.00000e+2	☐
☑	V9	BCC_A2	$L^{0}_{Cr,Ni:Va;}$ A	2.15700e+4	-1.00000e+5	1.00000e+5	☐
☑	V10	BCC_A2	$L^{0}_{Cr,Ni:Va;}$ B	-1.60400e+1	-1.00000e+2	1.00000e+2	☐
☑	V11	BCC_A2	$L^{1}_{Cr,Ni:Va;}$ A	2.43700e+4	-1.00000e+5	1.00000e+5	☐
☑	V12	BCC_A2	$L^{1}_{Cr,Ni:Va;}$ B	-4.92000e+0	-1.00000e+2	1.00000e+2	☐

Legend: Show

Optimizer settings

NOMAD

Hyperparameters

Optimization control

Optimize Test and plot

Calphad Optimizer

Comprehensive results in real-time

Results

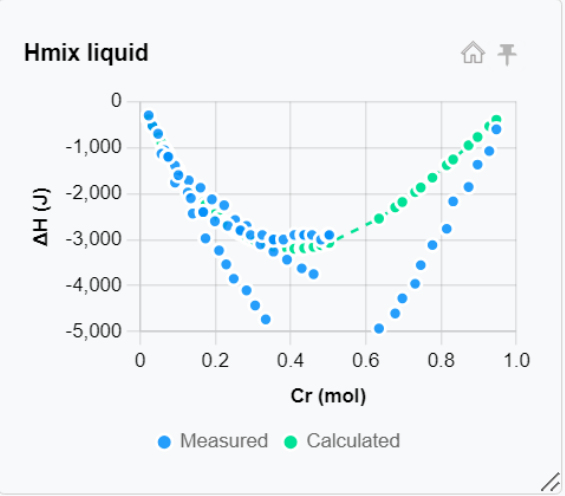
Summary

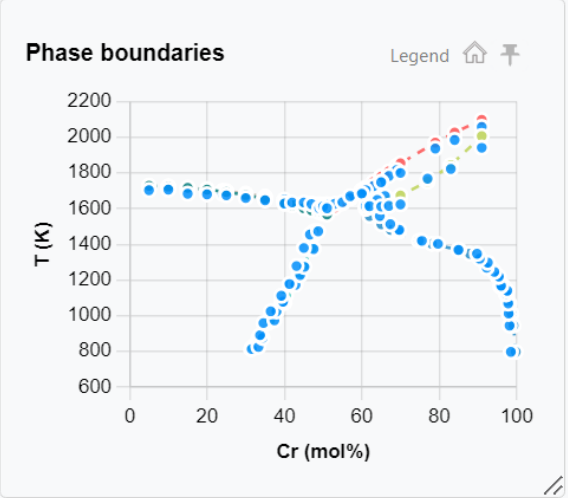
Data Group	No.	Error
★ Hmix liquid	Total: 62	2.08539e-2
★ Hmix bcc	Total: 11	1.21896e-1
★ Hmix fcc	Total: 15	2.04642e-1
★ Solidus fcc (Ni)	Total: 14	3.30316e-2
★ Solidus bcc (Cr)	Total: 10	1.55185e-2
★ Liquidus bcc (Cr)	Total: 16	2.49819e-2
★ Liquidus fcc (Ni)	Total: 13	1.60629e-2
★ Ni solubility in BCC (Cr) _ reversed	Total: 23	5.29259e-2
★ Cr solubility in FCC (Ni)	Total: 31	5.64063e-1
★ Activity 1873 K - Cr in liquid	Total: 53	3.39488e-2
★ Activity 1273 K - Cr in bcc	Total: 20	2.93893e-3

Optimization progress

• Chart: All evaluations Add

Measured *vs* calculated





Status

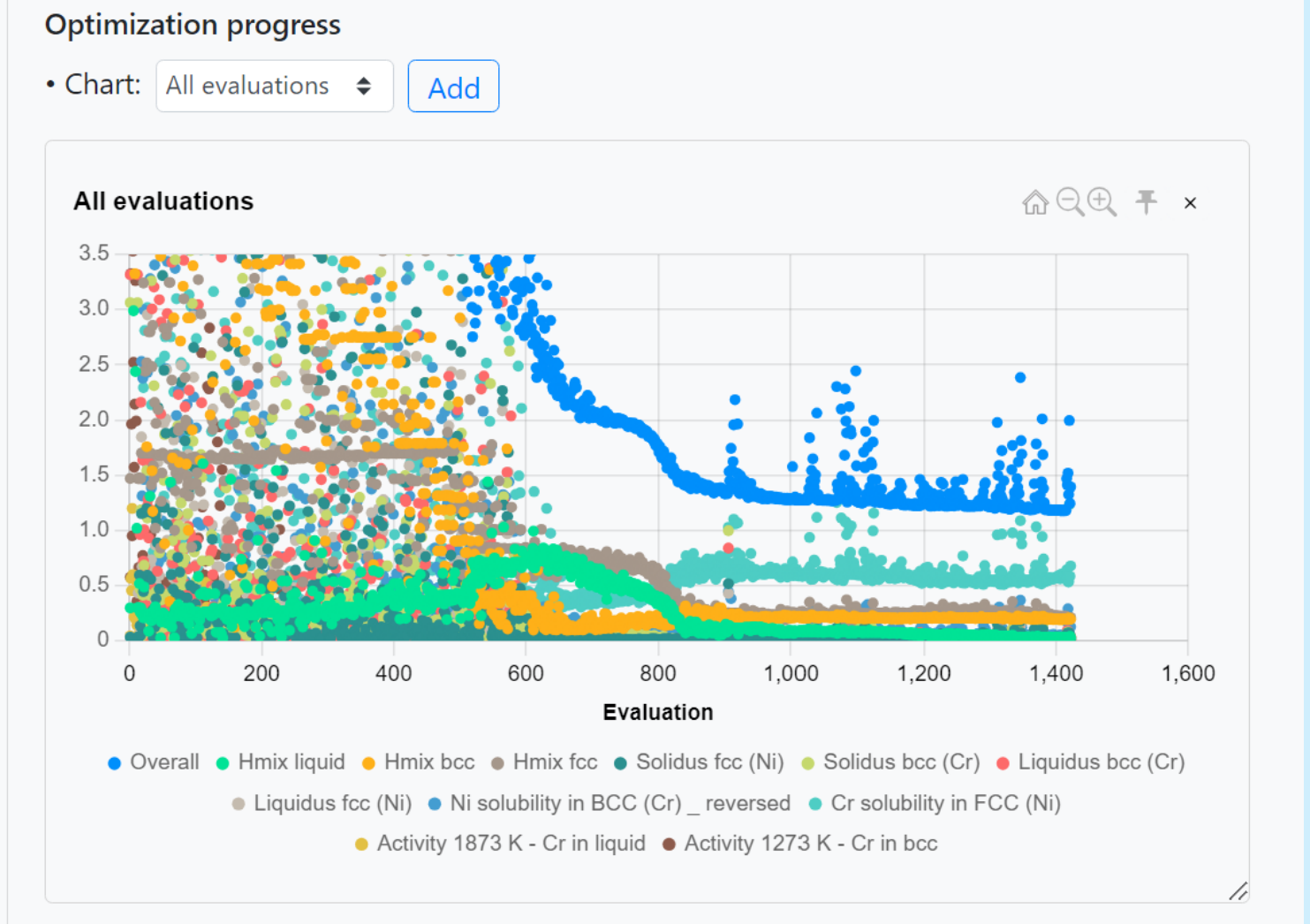
Testing experiments...
All successful.
Starting optimization...
Optimization saved as /history/2024-05-06_12-33-36-start.opt.
Optimization started.
Best evaluation 1: 7.06599e+1
Best evaluation 16: 3.38212e+1
Best evaluation 19: 1.77858e+1

Current evaluation: 26
Current error: 4.73527e+1
Elapsed time: 0:00:15
Successful evaluations: 26

Stop

Calphad Optimizer

Comprehensive results in real-time



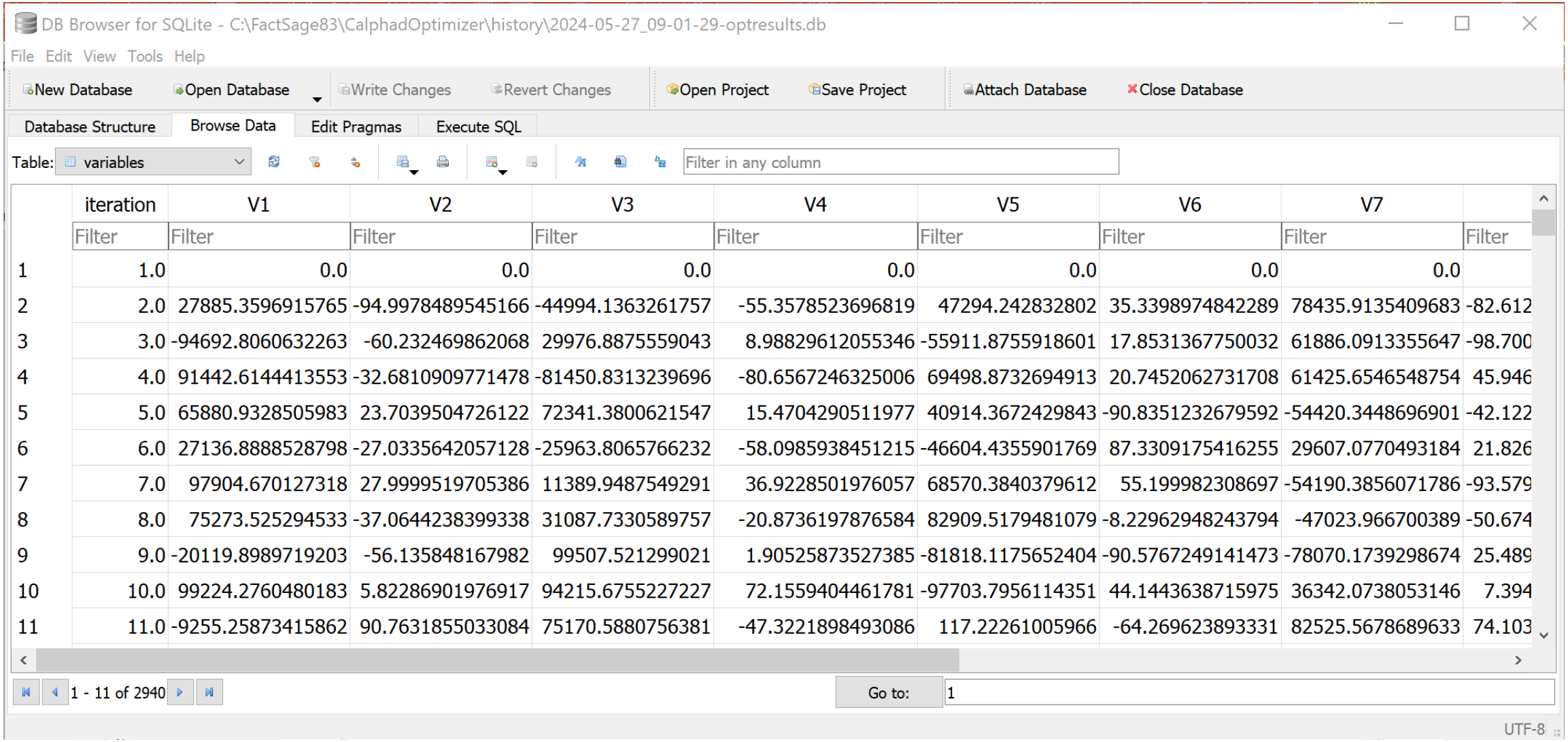
The SQLite results database

The SQLite results database

- What is SQLite?
  - A lightweight, file-based and relational database management system

- Benefits
  - Structure: well-defined schema organizes data into tables, rows, and columns
    - Tables: variables, errors, experiments results *vs* iteration

The SQLite results database



	iteration	V1	V2	V3	V4	V5	V6	V7	
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	2.0	27885.3596915765	-94.9978489545166	-44994.1363261757	-55.3578523696819	47294.242832802	35.3398974842289	78435.9135409683	-82.612
3	3.0	-94692.8060632263	-60.232469862068	29976.8875559043	8.98829612055346	-55911.8755918601	17.8531367750032	61886.0913355647	-98.700
4	4.0	91442.6144413553	-32.6810909771478	-81450.8313239696	-80.6567246325006	69498.8732694913	20.7452062731708	61425.6546548754	45.946
5	5.0	65880.9328505983	23.7039504726122	72341.3800621547	15.4704290511977	40914.3672429843	-90.8351232679592	-54420.3448696901	-42.122
6	6.0	27136.8888528798	-27.0335642057128	-25963.8065766232	-58.0985938451215	-46604.4355901769	87.3309175416255	29607.0770493184	21.826
7	7.0	97904.670127318	27.9999519705386	11389.9487549291	36.9228501976057	68570.3840379612	55.199982308697	-54190.3856071786	-93.579
8	8.0	75273.525294533	-37.0644238399338	31087.7330589757	-20.8736197876584	82909.5179481079	-8.22962948243794	-47023.966700389	-50.674
9	9.0	-20119.8989719203	-56.135848167982	99507.521299021	1.90525873527385	-81818.1175652404	-90.5767249141473	-78070.1739298674	25.489
10	10.0	99224.2760480183	5.82286901976917	94215.6755227227	72.1559404461781	-97703.7956114351	44.1443638715975	36342.0738053146	7.394
11	11.0	-9255.25873415862	90.7631855033084	75170.5880756381	-47.3221898493086	117.22261005966	-64.269623893331	82525.5678689633	74.103

The SQLite results database

- What is SQLite?
  - A lightweight, file-based and relational database management system

- Benefits
  - Structure: well-defined schema organizes data into tables, rows, and columns
    - Tables: variables, errors, experiments results *vs* iteration
  - Portability: self-contained, single file
    - Easy copying and sharing

The SQLite results database

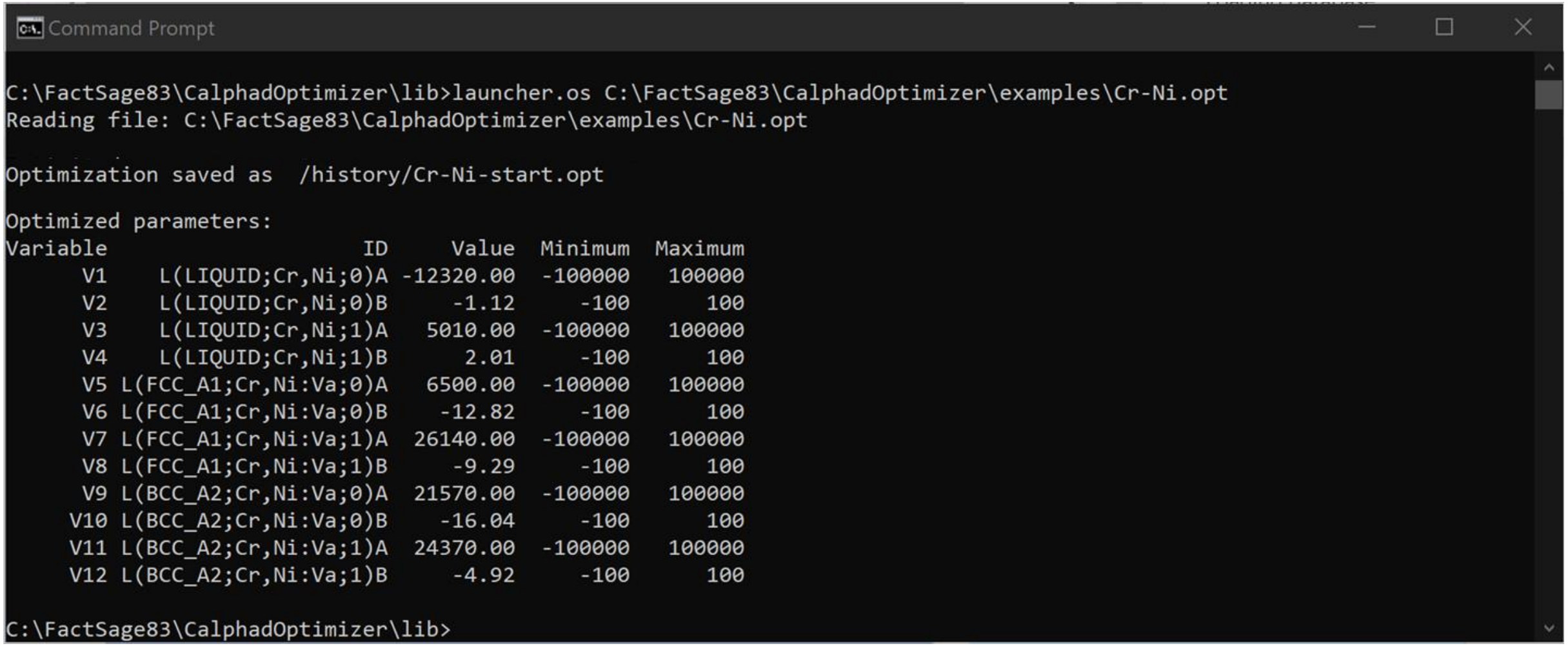
- What is SQLite?
  - A lightweight, file-based and relational database management system

- Benefits
  - Structure: well-defined schema organizes data into tables, rows, and columns
    - Tables: variables, errors, experiments results *vs* iteration
  - Portability: self-contained, single file
    - Easy copying and sharing
  - Simplicity: minimal configuration setup
    - Python example:

```
engine = sqlalchemy.create_engine('sqlite:///2024-05-06_14-30-17-optresults.db')
var_df = pandas.read_sql_table('variables', engine)
```

Command line interface

Command line interface
Command Prompt



Command line interface

python

```python
# Define the paths
lib_directory = r"C:\FactSage83\CalphadOptimizer\lib"
history_directory = r"C:\FactSage83\CalphadOptimizer\history"
file_name = "Cr-Ni"
opt_file = rf"C:\FactSage83\CalphadOptimizer\examples\{file_name}.opt"
co_executable = "launcher.os"

# Change the current working directory
os.chdir(lib_directory)

# Execute the optimization with CLI
subprocess.run(f'"{co_executable}" "{opt_file}"', check=True, shell=True)

# Change the current working directory
os.chdir(history_directory)

# Load SQLite results
engine = sqlalchemy.create_engine(f"sqlite:///{file_name}-optresults.db")
var_df = pd.read_sql_table("variables", engine)
```

1st Analysis: Optimization performance

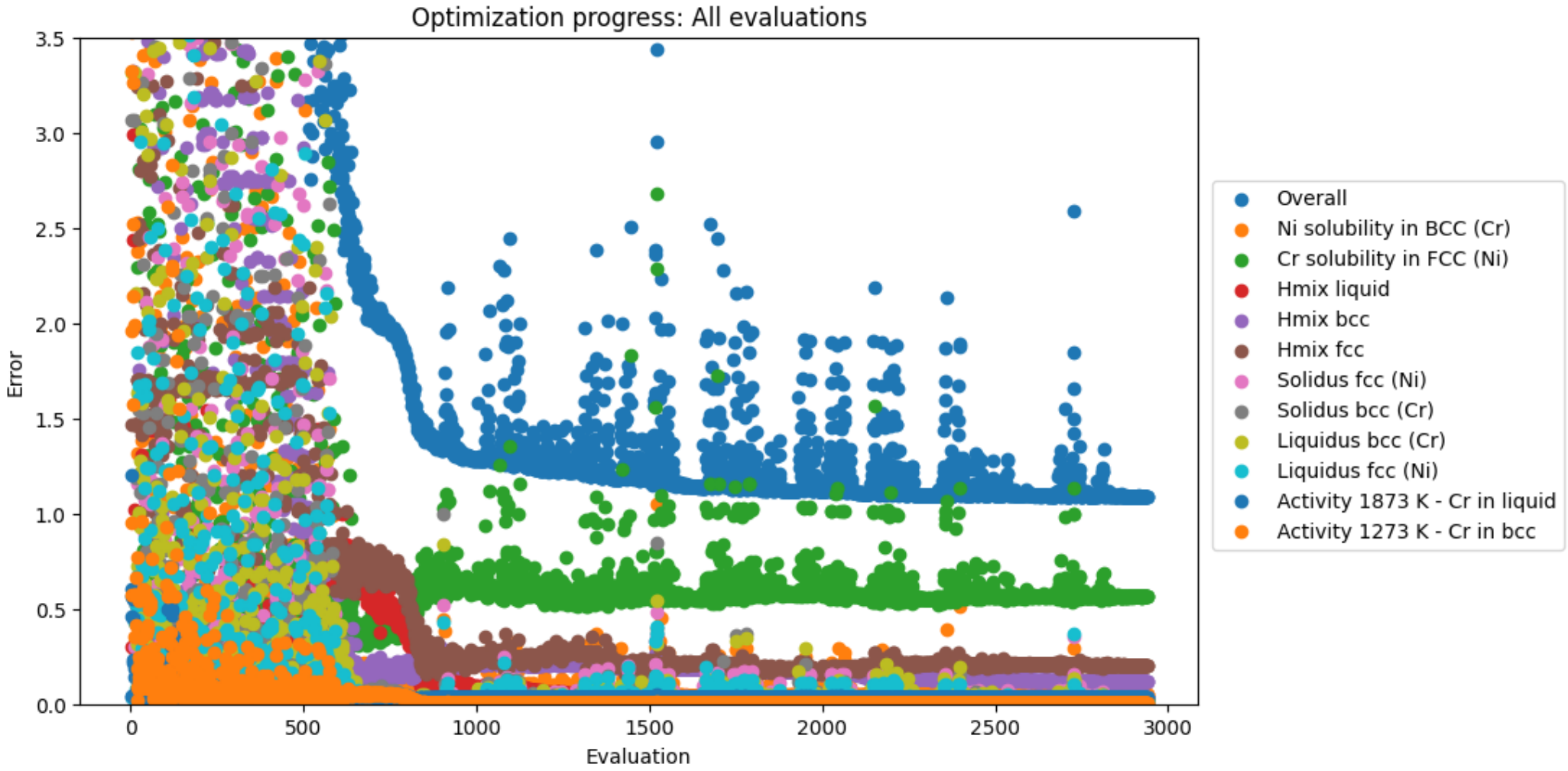
1st Analysis: Optimization performance

Error *vs* evaluation

```
# Load SQLite results
engine = sqlalchemy.create_engine(f"sqlite:///{file_name}-optresults.db")
df = pd.read_sql_table("errors", engine)

# Plotting
def plot_dispersion(df):
    y_col = df.columns[0]
    x_cols = df.columns[1:]

    plt.figure(figsize=(10, 6))

    for col in x_cols:
        plt.scatter(df[y_col], df[col], label=name_mapping.get(col, col))

    plt.xlabel('Evaluation')
    plt.ylabel('Error')
    plt.ylim(bottom=0, top=3.5)
    plt.title(f'Optimization progress: All evaluations')
    plt.legend(loc='upper center', bbox_to_anchor=(1.17, 0.8), ncol=1)
    plt.show()

plot_dispersion(df)
```

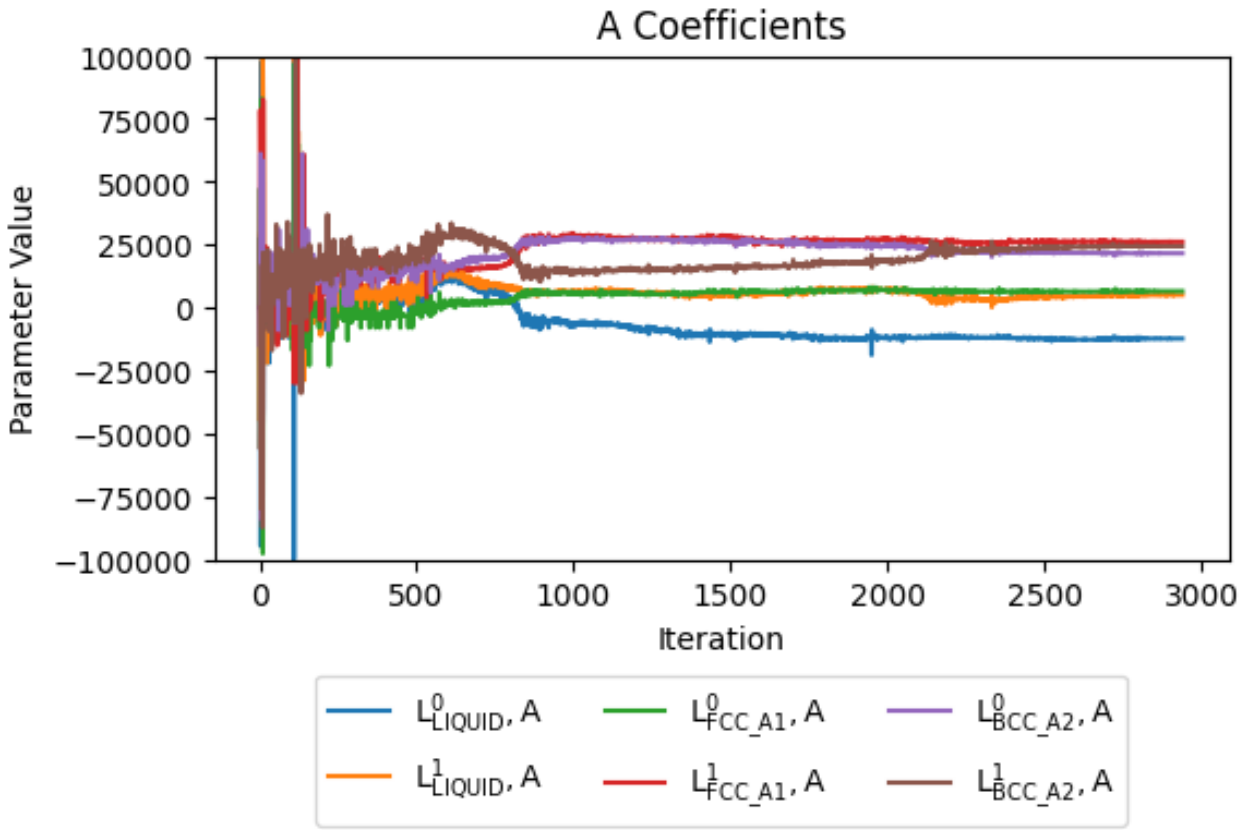
1st Analysis: Optimization performance

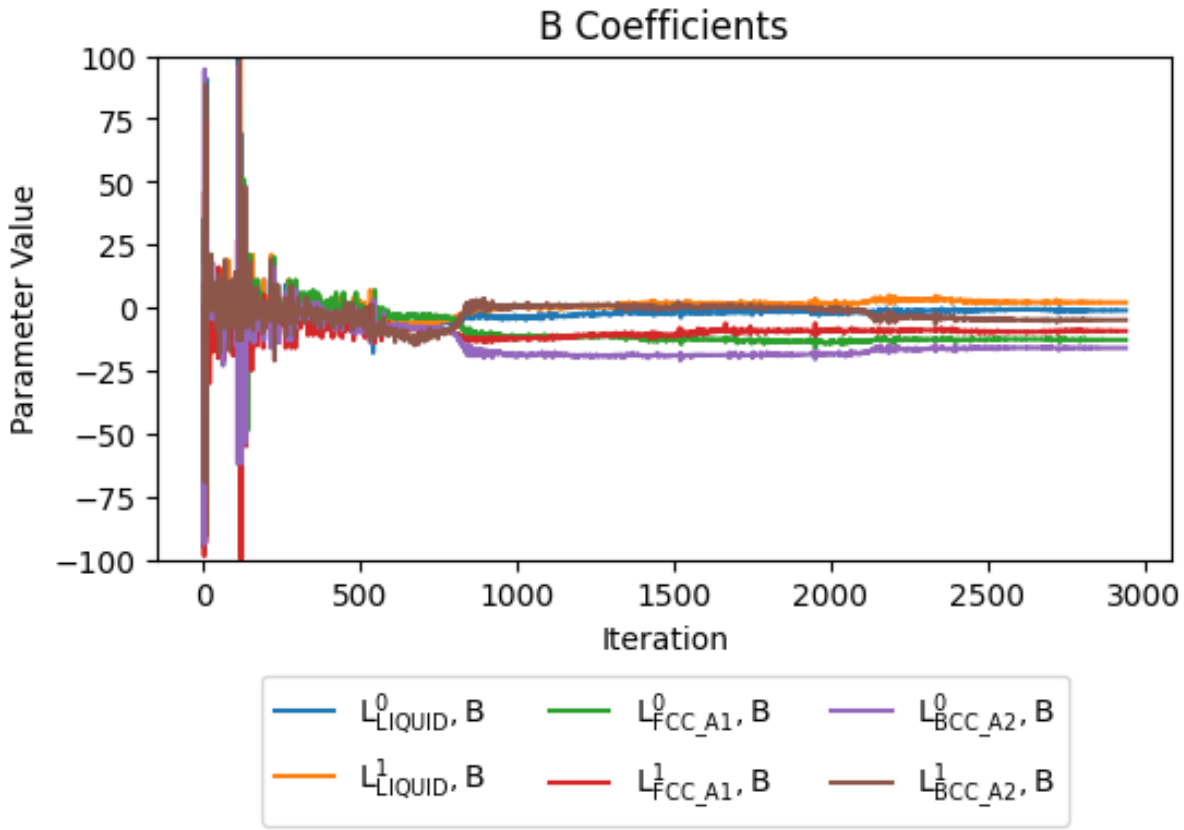
Error *vs* evaluation



1st Analysis: Optimization performance

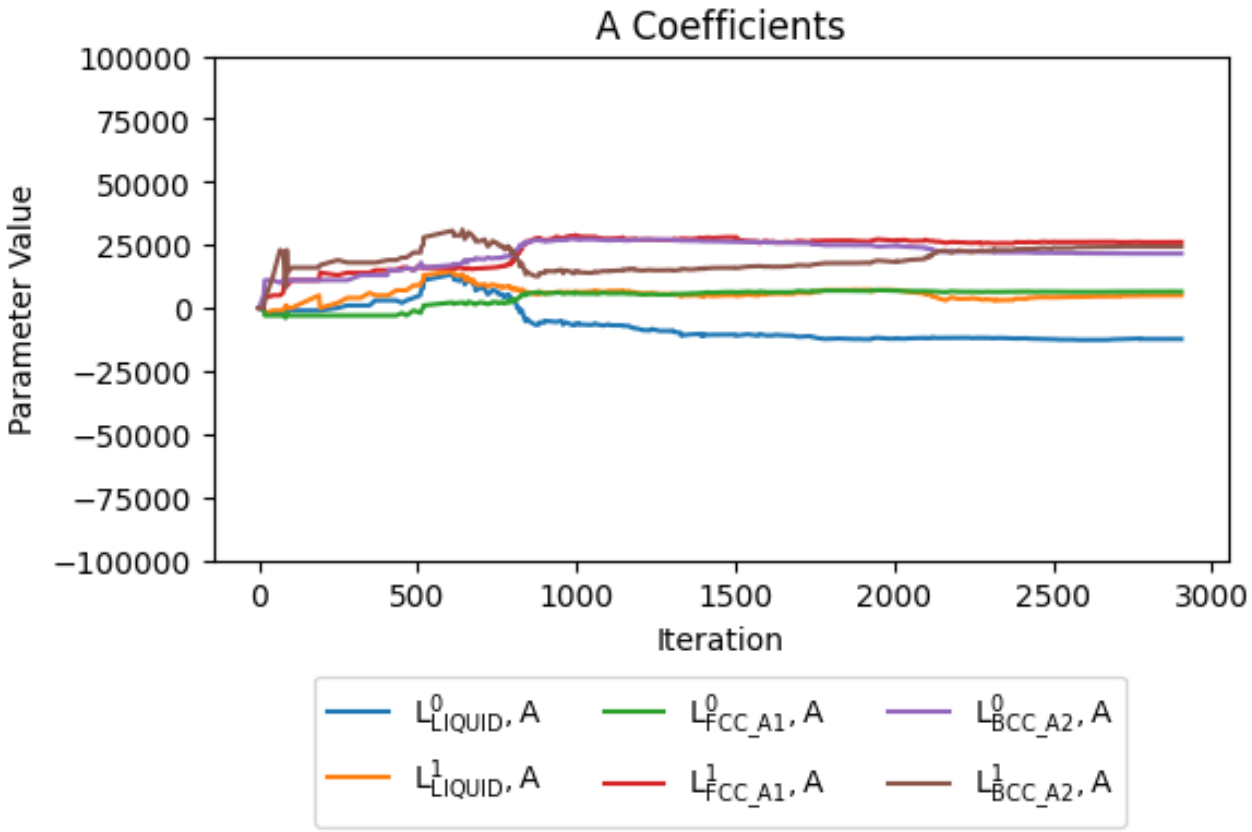
Parameter convergence

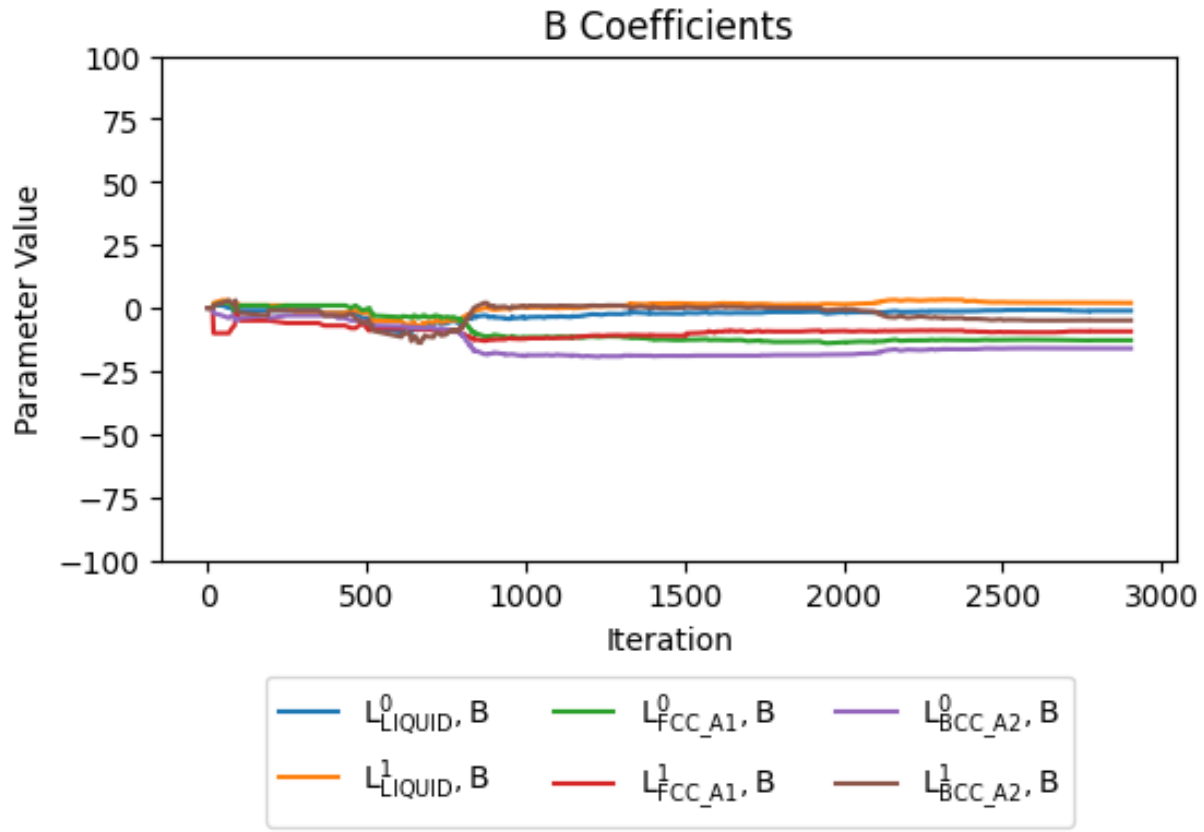




1st Analysis: Optimization performance

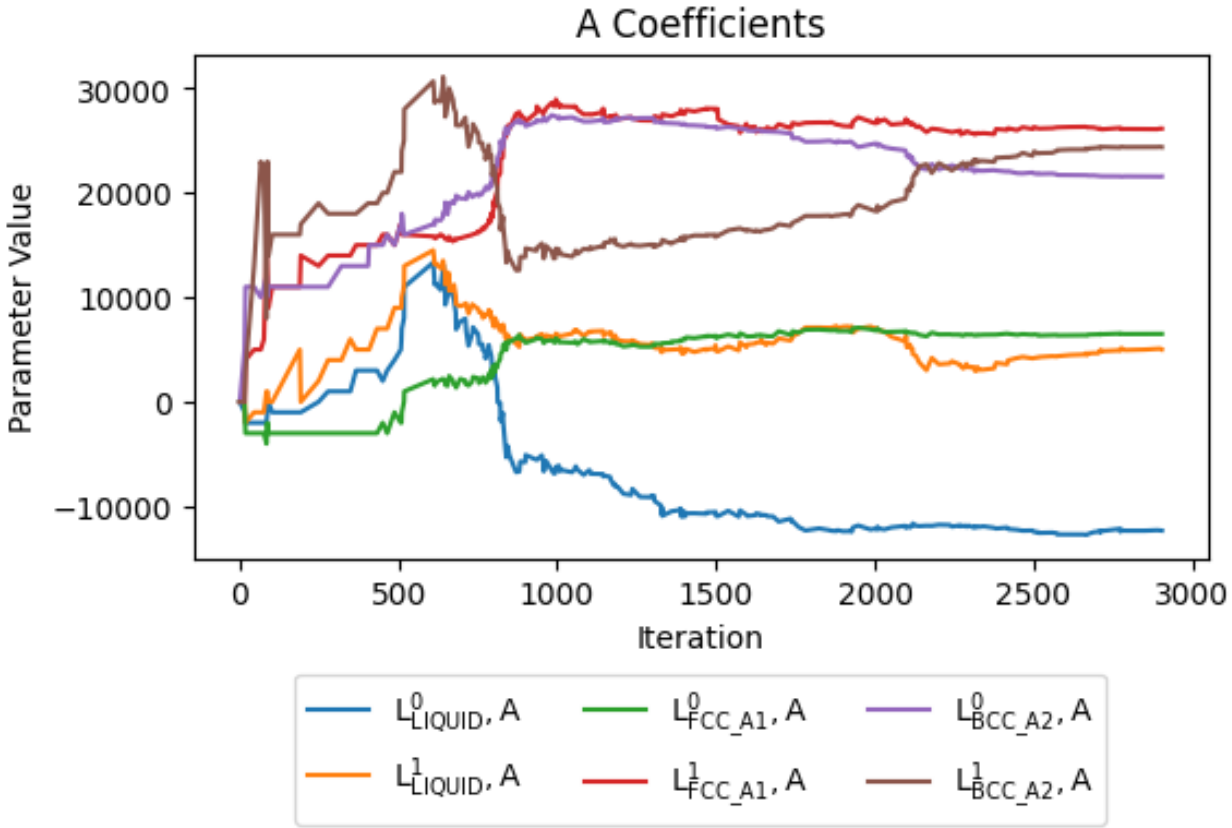
Parameter convergence – only improvements

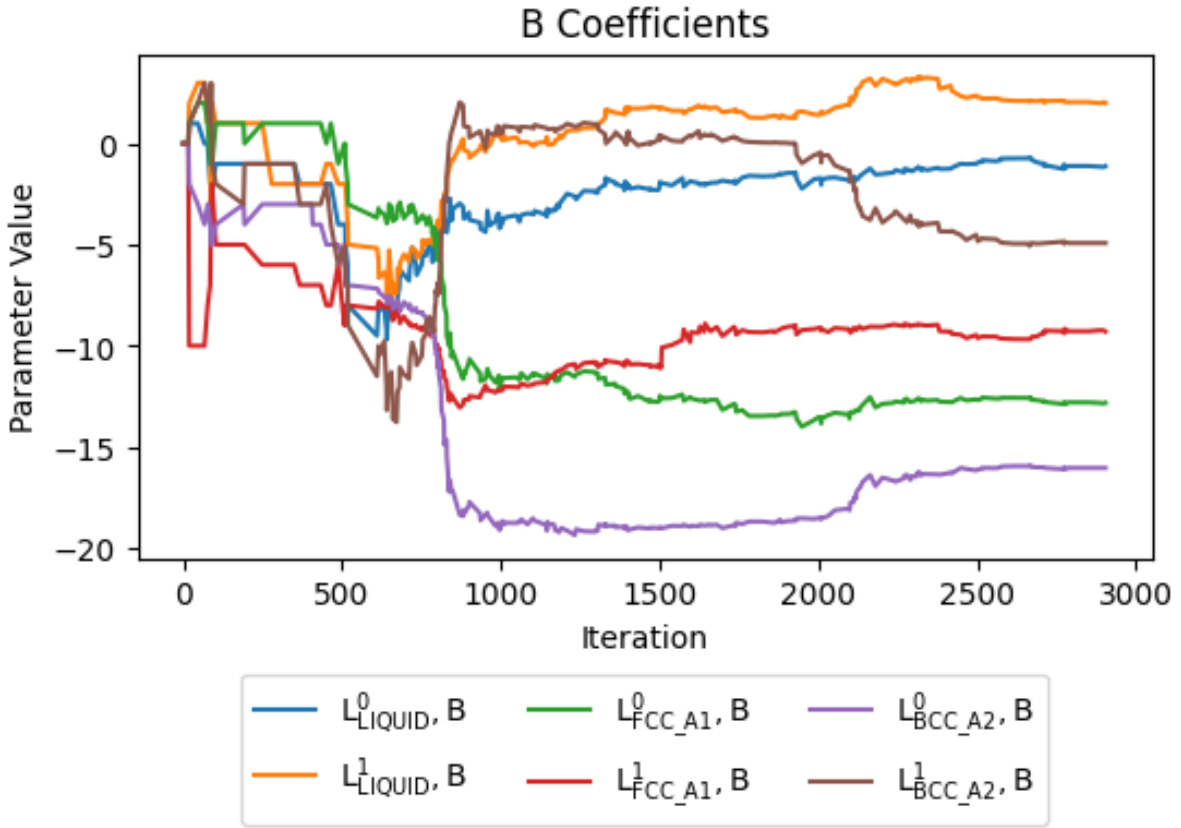




1st Analysis: Optimization performance

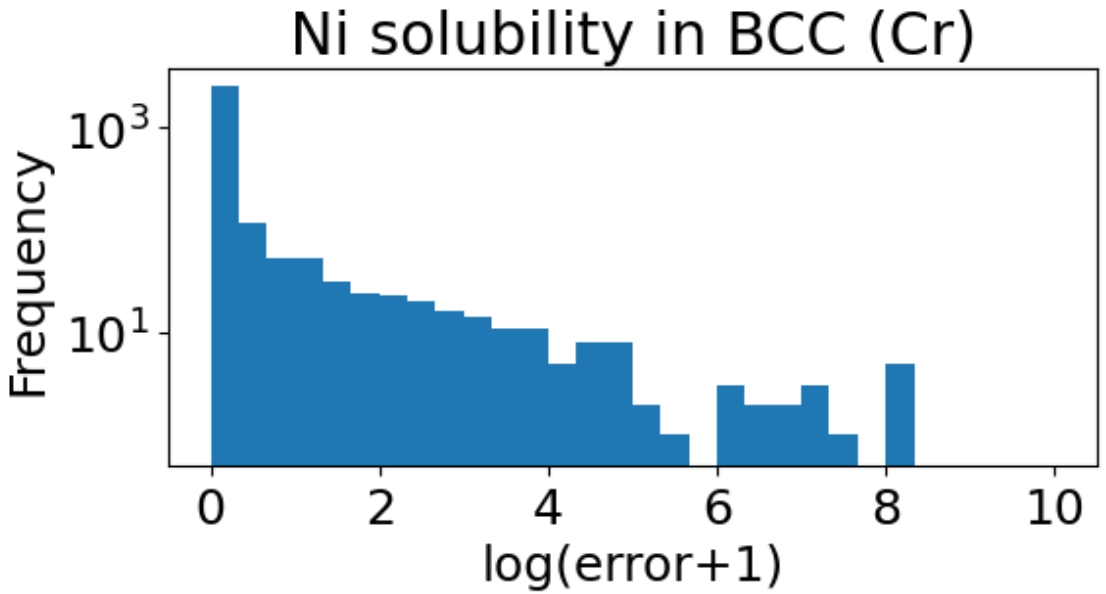
Parameter convergence – only improvements





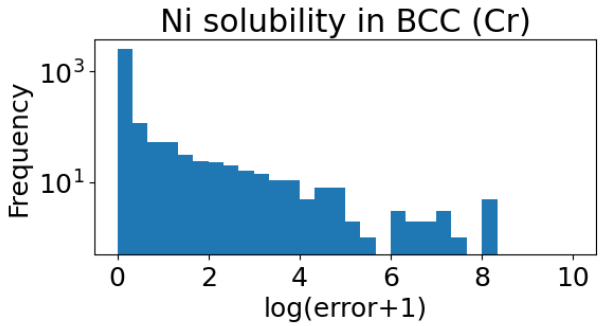
1st Analysis: Optimization performance

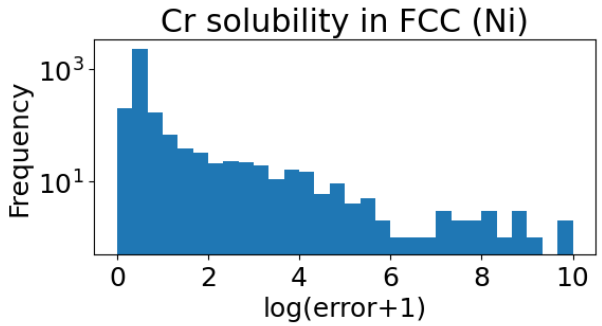
Error distribution for Ni solubility in BCC

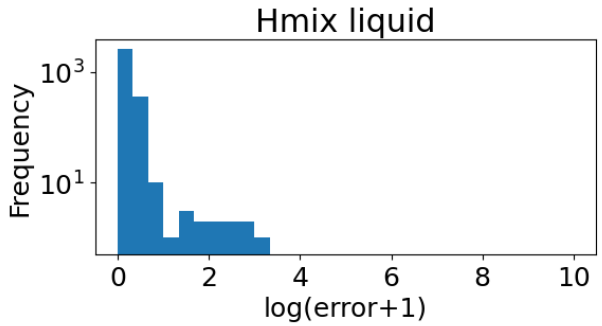


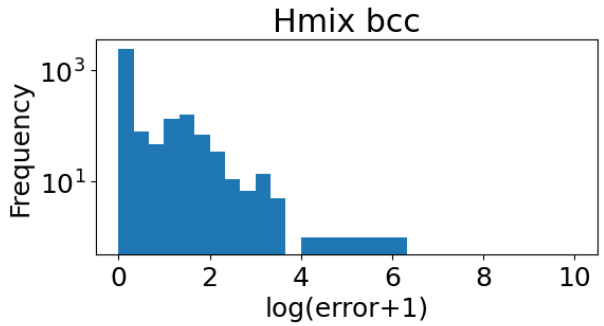
1st Analysis: Optimization performance

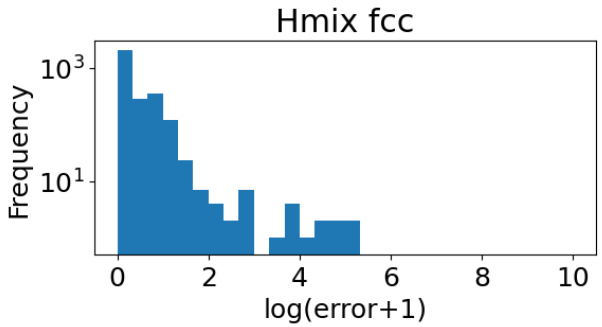
Error distribution per experiment

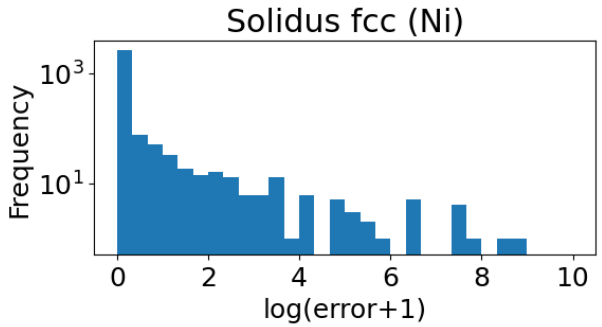


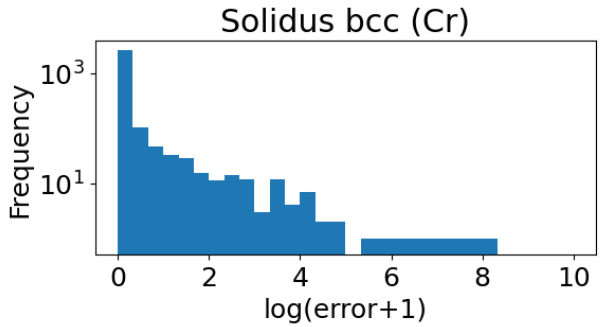


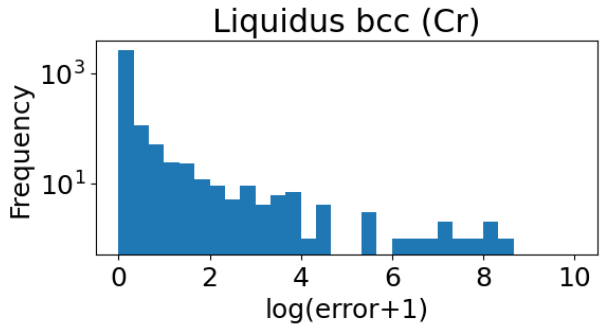


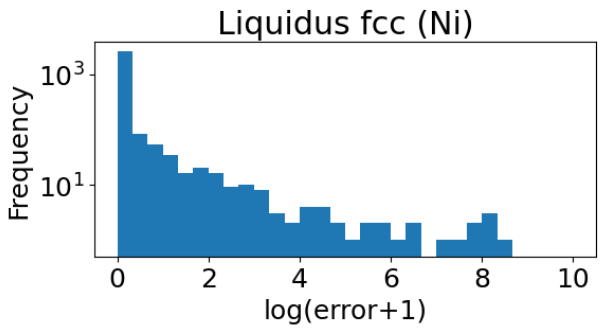


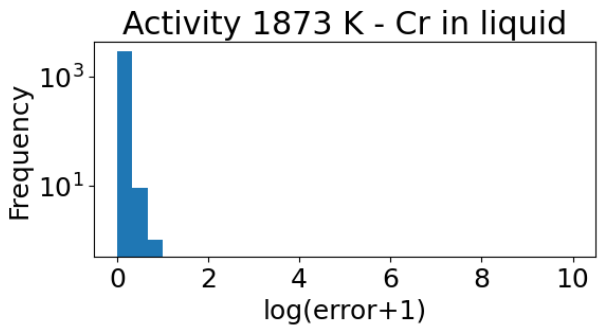


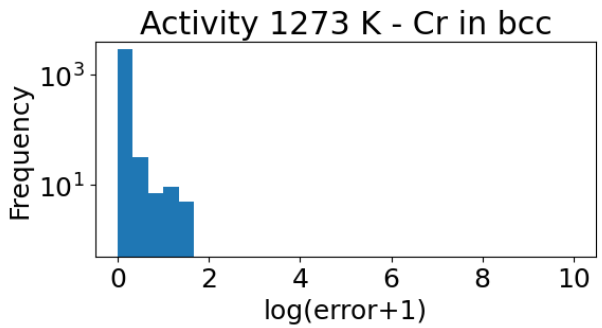






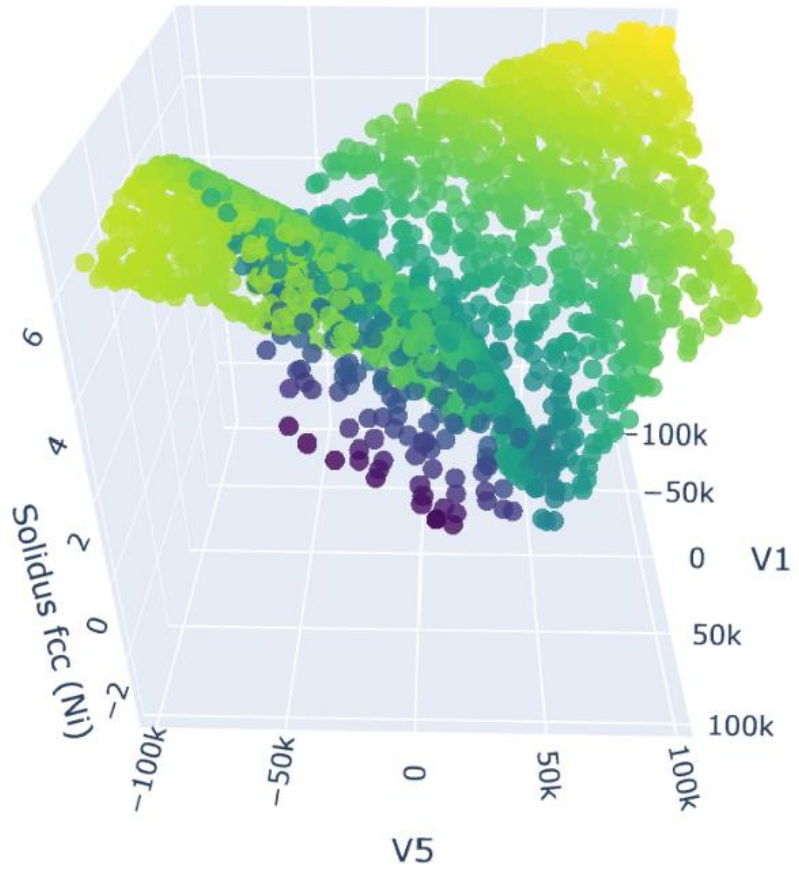


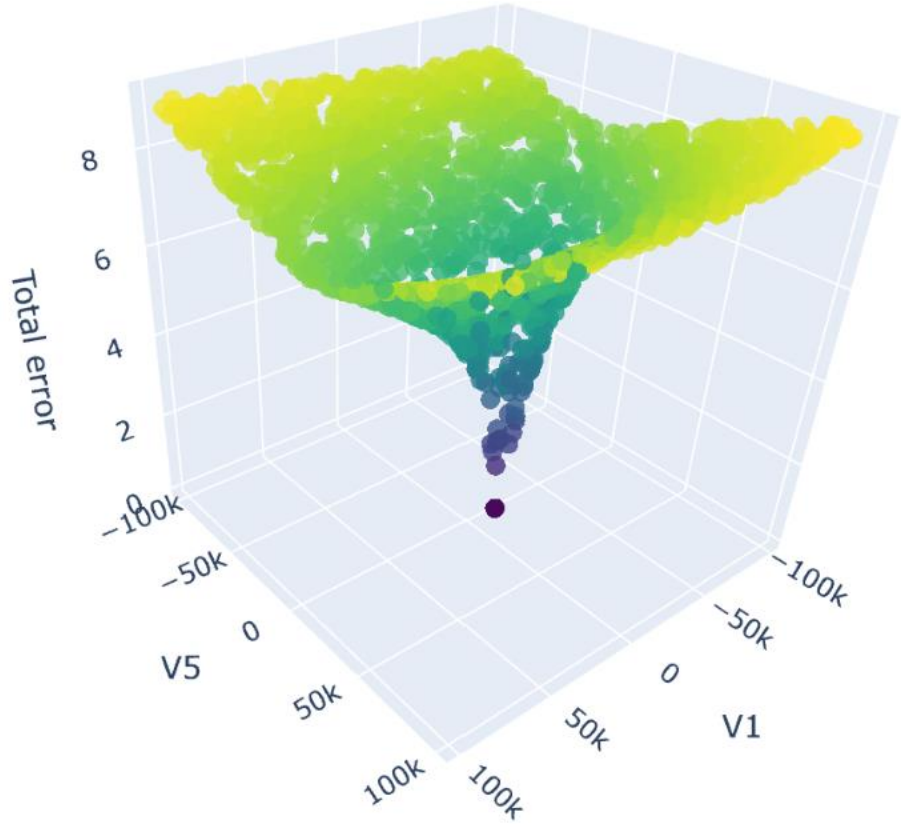




1st Analysis: Optimization performance

3D shape of the error function





2nd Analysis: Exploratory data analysis

2nd Analysis: Exploratory data analysis

Cr-Ni experimental data	Sample Size
Activity of Cr in BCC_A2	20
Activity of Cr in liquid	53
Heat of mixing of liquid	62
Heat of mixing of BCC_A2	11
Heat of mixing of FCC_A1	15
Solidus (FCC_A1)	14
Solidus (BCC_A2)	10
Liquidus (FCC_A1	13
Liquidus (BCC_A2)	16
Solvus (BCC_A2)	23
Solvus (FCC_A1)	31
Total	268

Time/iteration = **1.2 s**

2nd Analysis: Exploratory data analysis

Cr-Ni experimental data	Sample Size
Activity of Cr in BCC_A2	20
Activity of Cr in liquid	53
Heat of mixing of liquid	62
Heat of mixing of BCC_A2	11
Heat of mixing of FCC_A1	15
Solidus (FCC_A1)	14
Solidus (BCC_A2)	10
Liquidus (FCC_A1	13
Liquidus (BCC_A2)	16
Solvus (BCC_A2)	23
Solvus (FCC_A1)	31
Total	268

Time/iteration = **1.2 s**

EDA:

- Data Collection
- Data Cleaning
- Statistical Analysis
- Grouping
- Thresholding

2nd Analysis: Exploratory data analysis

Cr-Ni experimental data	Sample Size
Activity of Cr in BCC_A2	20
Activity of Cr in liquid	53
Heat of mixing of liquid	62
Heat of mixing of BCC_A2	11
Heat of mixing of FCC_A1	15
Solidus (FCC_A1)	14
Solidus (BCC_A2)	10
Liquidus (FCC_A1	13
Liquidus (BCC_A2)	16
Solvus (BCC_A2)	23
Solvus (FCC_A1)	31
Total	268

Time/iteration = **1.2 s**

EDA:

- Data Collection
- Data Cleaning
- Statistical Analysis
- Grouping
- Thresholding

Cr-Ni experimental data	Sample Size
Activity of Cr in BCC_A2	0
Activity of Cr in liquid	0
Heat of mixing of liquid	39
Heat of mixing of BCC_A2	3
Heat of mixing of FCC_A1	5
Solidus (FCC_A1)	7
Solidus (BCC_A2)	4
Liquidus (FCC_A1	6
Liquidus (BCC_A2)	6
Solvus (BCC_A2)	14
Solvus (FCC_A1)	15
Total	99

Time/iteration = **0.5 s**

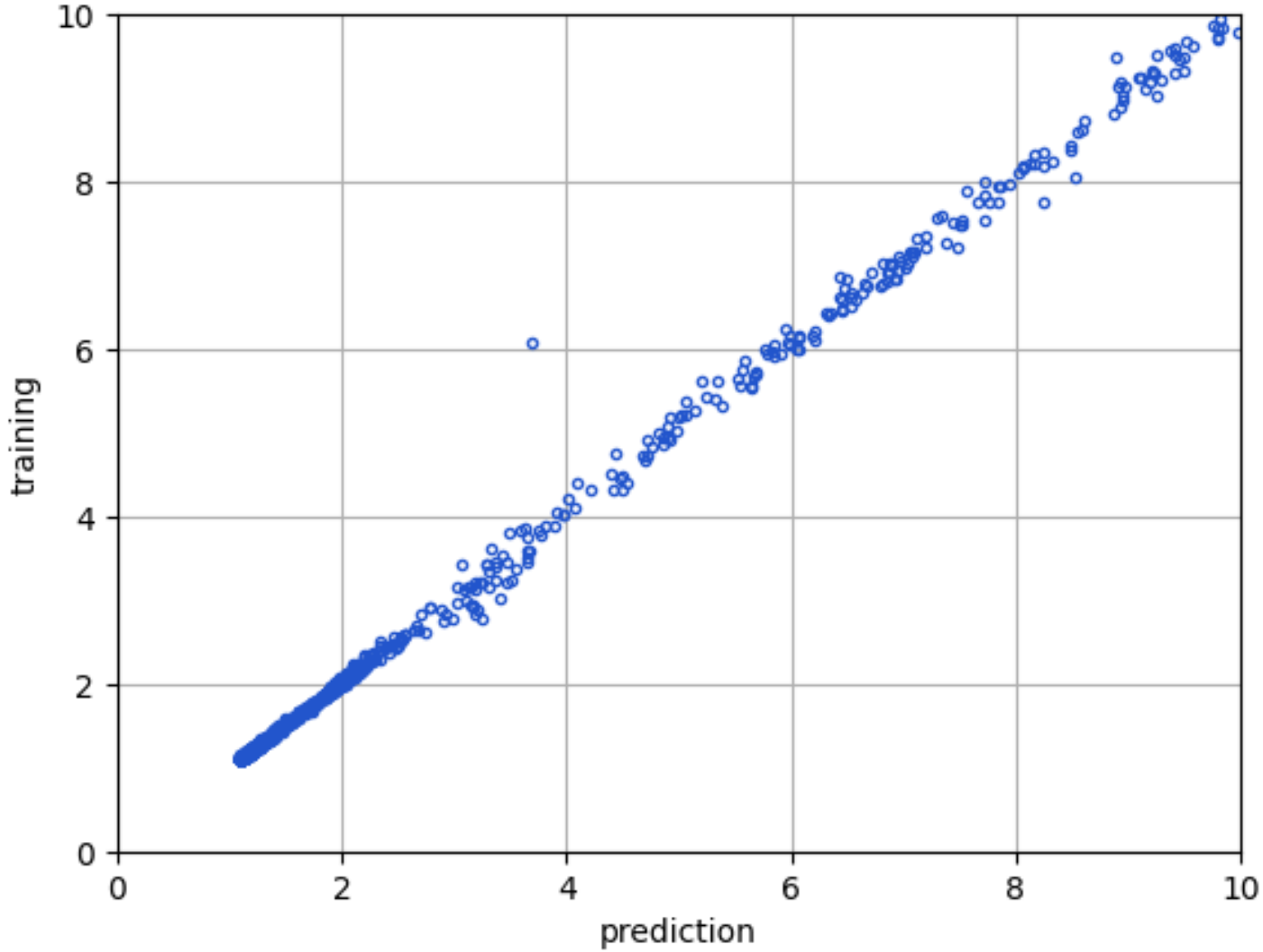
3rd Analysis: Machine learning

3rd Analysis: Machine learning

```
# read SQLite results
engine = sqlalchemy.create_engine(f"sqlite:///Cr-Ni-optresults.db")
y_df = pd.read_sql_query("SELECT error FROM errors", engine)
x_df = pd.read_sql_query("SELECT * FROM variables", engine)

# create neural network
model = Sequential()
model.add(Input(shape=(len(x_df.columns),), dtype='float32'))
normalization = Normalization(name="normalization")
normalization.adapt(x)
model.add(normalization)
nodes = 512
for _ in range(4)
    model.add(Dense(nodes, activation='relu'))
model.add(Dense(nodes, activation='linear'))
model.add(Dense(1, activation='linear'))

# compile
model.compile(optimizer=keras.optimizers.Adam(learning_rate=0.001), loss="mean_squared_error")

# fit
model.fit(x_df, y_df, epochs=300)
```

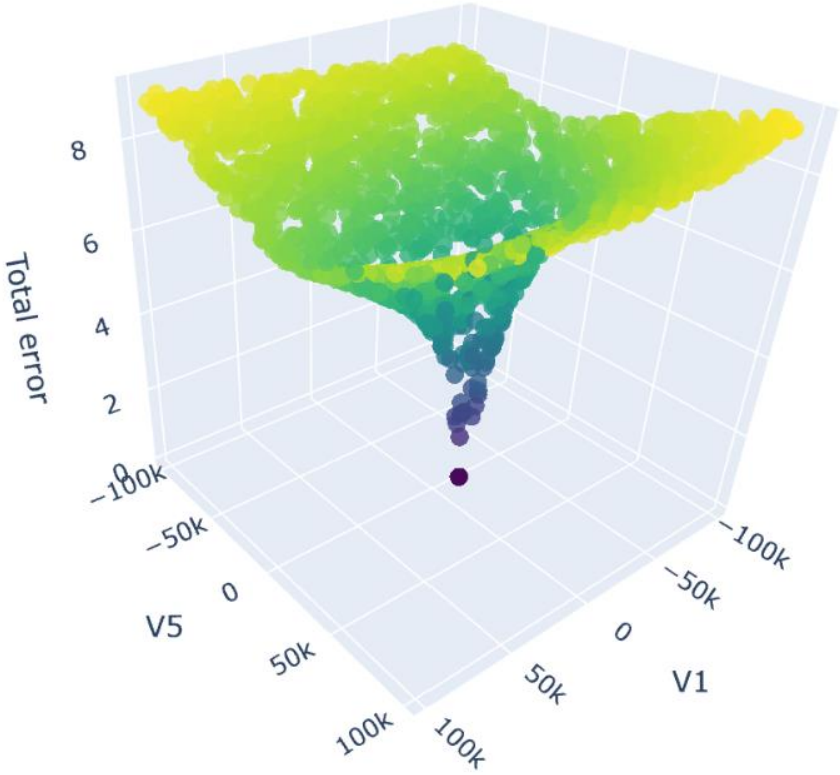
3rd Analysis: Machine learning

Fit of the error function using a deterministic neural network



Conclusion

- Calphad Optimizer
  - User-friendly GUI
  - Real-time responsiveness
  - Robust optimization algorithms
  - FAIR-compliant
    - Interoperable
    - Reusable
  - CLI-compatible
  - SQLite post-processing analysis
    - Optimization performance
    - Exploratory data analysis
    - Machine learning
    - And much more!



Stay tuned for our publication on demonstrating the performance of Calphad Optimizer in details!

Thank You!

Bruno Reis
br@gtt-technologies.de