

GEMSFITS: A coupled code for GEMS input data optimization

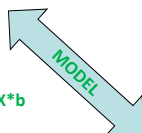
George D. Miron, Dmitrii A. Kulik, Svitlana V. Dmytrieva,
Thomas Wagner

Thermodynamic modeling

- **Insight into chemical processes:** chemical engineering, chemistry, materials research, and geochemistry
- **Simple well controlled experimental systems** → model more complex systems

Property I want to
know but cannot
measure

$$Y = a + X*b$$



Property I can
measure

Thermodynamic models

relations between thermodynamic properties,
calculation of measurable properties

+


Thermodynamic datasets

- actual values of thermodynamic properties
(compilation, derivation, consistency, errors, bookkeeping)

Simultaneous processing of data ensures that the derived properties are internally consistent
and accurately reproduce the experimental data (Anderson and Crerar, 1993)

What do we need?

GEMSFITS

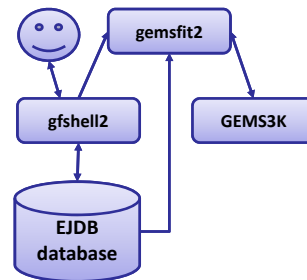
- Suite of appropriate models of mixing in phases → **TSolMod library**
- Efficient numerical method to perform thermodynamic computations → **GEMS3K**

- Optimization algorithms → **NLopt**
- Extensive, flexible collection of experimental data → **NoSQL (EJDB)**



GEMSfits

GEMSfits Code Package

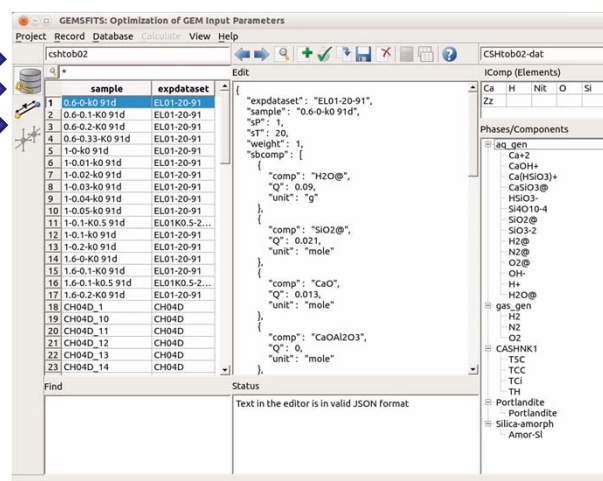
- Graphical user interface
- Experimental database management
- Setting up the fitting task
- Optimization algorithms
- Results and statistics




gfsell2: The GEMSfits GUI

- Database Mode
- Task Mode
- View Results Mode

(more in the live example)



```
{
  "_id": "0x321c209",
  "sample": "Ben_1",
  "ST": 101.5,
  "SP": 0,
  "expdataset": "Bhm_Benezeth_etal2001",
  "sbcomp": [
    {
      "comp": "AlO(OH)",
      "Q": 10,
      "unit": "g" },
    {
      "comp": "H2O",
      "Q": 400,
      "unit": "g" },
    {
      "comp": "HCl",
      "Q": 0.0888514,
      "unit": "g" } ],
  "expphases": [
    {
      "phase": "aq_gen",
      "phprop": [
        {
          "prop": "pH",
          "Q": 2.484 } ],
      "phIC": [
        {
          "IC": "Al",
          "Q": 3.68129E-4,
          "unit": "molal" } ] } ] }
```




GEMSfits database

<Key> : <Value>
"comp" : "AlO(OH)"


Spreadsheet → gfsHELL2 (csvtoejdb) → Database

- Flexible for various type of experiments
- Easy data selection for various optimization runs
- Select-Copy-Read-Update-Delete using the GUI in Database Mode
- Stores: experiments, tasks, results

JSON




Eidgenössische Technische Hochschule Zürich



PAUL SCHERRER INSTITUT

Fluids and Mineral Resources Group

GEMS Workshop
EMPA 7-8 May 2014



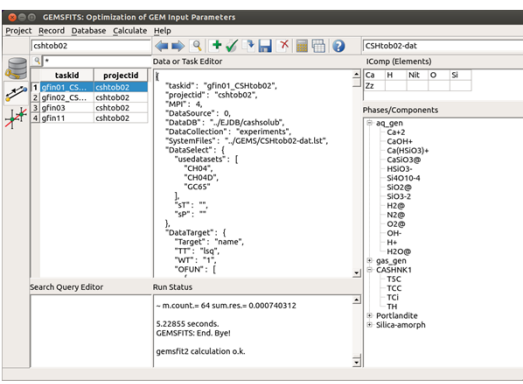
Setting up the fitting task


(in the task specification file)

- **What data to use:** select/exclude experiments (samples) from database
- **What to fit:** marking the adjustable parameters
- **What to compare:** solubility, solid-solution composition, pH, Eh, activities, etc.
- **How to calculate the residuals:** e.g. least squares, weighting


JSON

GEMSfits GUI in Task Mode with task specification editor





Eidgenössische Technische Hochschule Zürich



PAUL SCHERRER INSTITUT

Fluids and Mineral Resources Group

GEMS Workshop
EMPA 7-8 May 2014



What can you fit?

- Non-ideal solution **activity model parameters**
- **G° values** of dependent components (e.g. aqueous species, solid solution end members, pure condensed substances)
- **Temperature** and/or **pressure, bulk chemical composition** for inverse modeling tasks

Constraint options

- Upper and lower bounds on freely-adjustable parameters
- Parameters can be declared as freely adjustable 'F'; linearly constrained 'L'; reaction-constrained 'R'

Fitting parameter markup in JSON

```
<PMC>, <G0>, <bIC>,  
<TK>, <P>  
...  
F{  
  "IV" : -10300,  
  "LB" : -16000,  
  "UB" : -6000  
} ...  
...  
F-10300  
...
```

```
<G0>  
...  
R{  
  "IV" : -696322,  
  "Ref" : "We93",  
  "order" : "1",  
  "nC" : 4,  
  "xDC" : "AlOH+2",  
  "RC" : [  
    "Al+3",  
    "H2O@",  
    "H+",  
    "AlOH+2"  
  ],  
  "Rcoef" : [  
    -1,  
    -1,  
    1,  
    1  
  ]  
}  
...
```

```
<bIC>  
...  
L{  
  "LE" : "H",  
  "IV" :  
44.5761406330092,  
  "LEs" : [  
    "Na",  
    "Cl"  
  ],  
  "Lcoef" : [  
    1,  
    1  
  ]  
}  
...
```



Objective function

- Determines what experimental data is compared with its computed counterpart
- Many objective function terms → different type of data can be used simultaneously

Nested target function

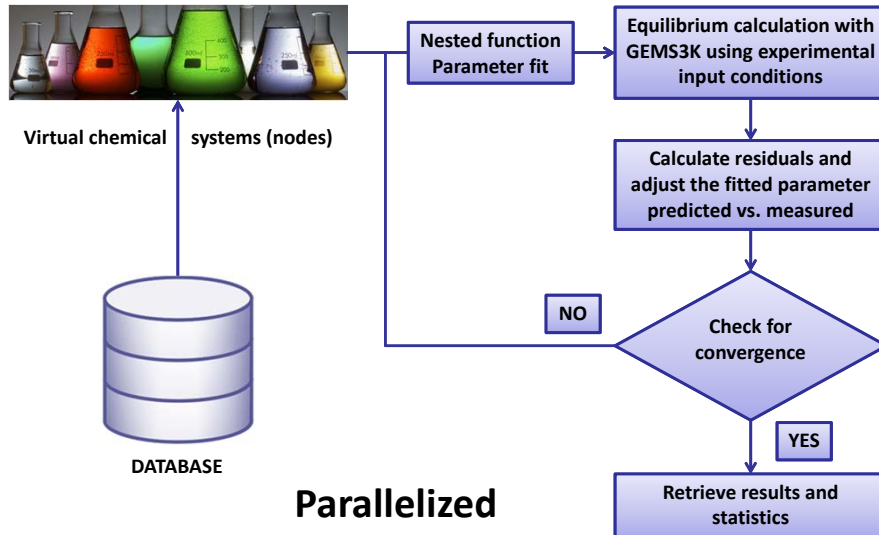
- Adjust parameters specific to each experiment after each optimization step (e.g. adjust pH)
- Used for inverse modeling (e.g. adjusts T and P to a given phase composition for each sample)

Objective and nested functions in JSON

```
{
  "Target" : "name",
  "TT" : "lsq",
  "WT" : "1",
  "OFUN" : [
    {
      "EPH" : "aq_gen",
      "CT" : "IC",
      "CN" : "Al",
      "unit" : "log_molal"
    }
  ],
  "NFUN" : [
    {
      "Ptype" : "bIC",
      "EPH" : "aq_gen",
      "CT" : "prop",
      "CN" : "pH",
      "unit" : "-loga"
    }
  ]
}
```

```
{
  "Target" : "name",
  "TT" : "lsq",
  "WT" : "1",
  "OFUN" : [
    {
      "EPH" : "aq_gen",
      "CT" : "IC",
      "CN" : "Si",
      "unit" : "log_molal"
    },
    {
      "EPH" : "aq_gen",
      "CT" : "IC",
      "CN" : "Ca",
      "unit" : "log_molal"
    }
  ],
  "ADDOUT" : [
    {
      "SRC" : "meas",
      "EPH" : "CSH",
      "CT" : "prop",
      "CN" : "Q",
      "unit" : "mol"
    }
  ]
}
```

gemsfit2 code flowchart



Weighting & outliers

- Set in the database of experiments or in the task configuration
- Set by the user for each sample, for each objective function term (type of data)
- Using the measurement error, measured value, measurement average
- Give outliers zero or low weight
- Tukey's Biweight: uses a threshold value (C) and gives samples which residuals exceed it a zero weight, and the rest - a weight as function of the distance to C .

Optimization algorithms

- Various global and local algorithms implemented in Nlopt library (<http://ab-initio.mit.edu/nlopt/>)
- Global algorithms – large search space, slower, converges to the “true” minimum value
- Local algorithms – small search space, much faster, can converge to local minimum values
- Start global, continue with local for higher precision

Statistics

- Goodness of fit, sensitivity analysis, correlation coefficients
- Parameter confidence intervals using
Monte Carlo simulations
 - Uses the experimental data scatter to create many random simulated measured data
 - Finds the best fit value of parameters for each set
 - Confidence intervals computed from the standard deviation of parameters obtained from each data set

Output

- Output files in csv format read in by the GUI to View Results window tabs
- Can be saved in the database

sample	expdataset	vt	sp	aq_gen.Si.meas	aq_gen.Si.calc	residual	aq_gen.Ca.meas	aq_gen.Ca.calc
0	GC65_1	GC65	25	1	0.00231	0.0031760726	-0.0008660726	0.00074
1	GC65_2	GC65	25	1	0.0026	0.0031751197	-0.0005751197	0.00093
2	GC65_3	GC65	25	1	0.0034	0.0031763737	-0.0002236269	0.00147
3	GC65_4	GC65	25	1	0.0043	0.0031760255	-0.0011539745	0.00168
4	GC65_5	GC65	25	1	0.00407	0.0031754012	-0.0008945985	0.0019
5	GC65_6	GC65	25	1	0.0039	0.0031758885	-0.0007241115	0.00195
6	GC65_7	GC65	25	1	0.00387	0.0031758251	-0.0006941749	0.00197
7	GC65_8	GC65	25	1	0.00377	0.00056415892	-0.002058411	0.002
8	GC65_9	GC65	25	1	0.00147	0.00025016495	-0.0012198351	0.00175
9	GC65_10	GC65	25	1	0.00025	0.00019899154	-0.01084626-05	0.00213
10	GC65_11	GC65	25	1	9e-05	0.00017569272	-8.5692715e-05	0.0039
11	GC65_12	GC65	25	1	6e-05	0.0001638502	-0.0001038502	0.00528

GEMSFTS GUI View Results Mode

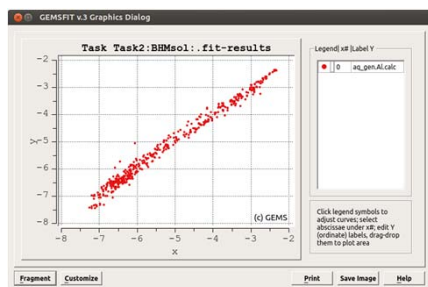
parameter	name	init.value	fitted.value	mc.mean	mc.stdev	conf99	conf95	conf90	CSS.sensitivity	correl.coef.2	correl.coef.3	correl.coef.4
1	PMc	3	-0.293	-0.29459141	-0.18897225	0.18315692	0.50485115	0.37459796	0.31120687	1.1565512	-0.622	1
2	PMc	4	1.63	1.6305428	1.6340671	0.10355895	0.28544842	0.21180183	0.17595981	11.952515	-0.9478	0.4872
0	PMc	2	62300	62333.336	62515.099	1416.7596	3905.1361	2897.5987	2407.2545	36.622154	1	-0.622



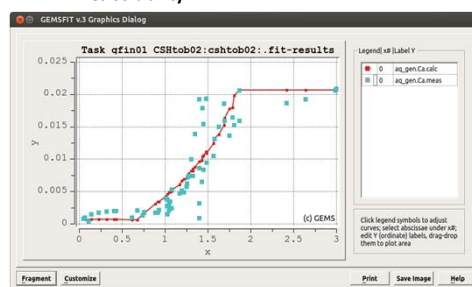
Output

- Values from the output tabs can be plotted in XY coordinates

measured Vs. calculated (Al solubility)



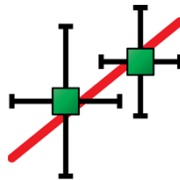
CSH amount Vs. measured & calculated Ca solubility



Outlook

- Plans to make GEMSFITS available for Windows (currently Linux only)
- Extensions to fit parameters of sorption models and mineral dissolution / precipitation kinetics models (TSorpMod and TKinMet libraries)
- Fit GEM-Selektor built-in equation-of-state parameters
- MongoDB server database (large data collection of various experiments) accessible to many scientists → optimizing TDBs and models for a large number of applications

Thank you!
And now let us try
GEMSFITS!



- This work is funded by ETH-19 11-2 grant at ETHZ IGP and is part of the GEMS TM project at LES PSI
- Thanks to Professor Ch.Heinrich (IGP) for support and discussions