pycalphad and ESPEI Workshop

ESPEI

CALPHAD XLVII

Why develop CALPHAD software tools?

- Better science
  - Develop new thermodynamic models
  - Improve and share methods as a community
  - Scripting: get more work done and be able to reproduce it
- Integration
  - Use standard tools in a huge Python ecosystem
  - Create calculation and data pipelines
  - Traverse length and time scales
  - Hierarchical models, machine learning
- Free
  - Freedom: share your work (or don't)
  - Cost: lower the barrier for entry

pycalphad

PHASES PSU PennState

Why develop CALPHAD software tools?

- I

- I

nad

Calculation software is NOT the whole picture!

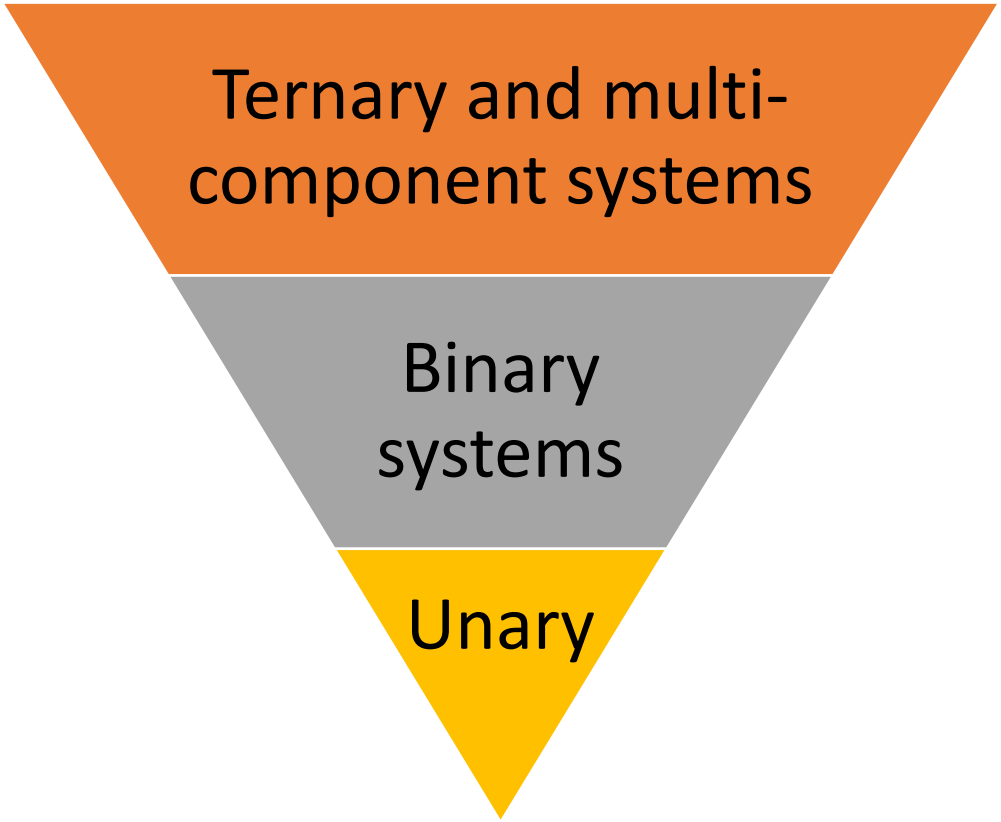
Databases and custom models come from somewhere

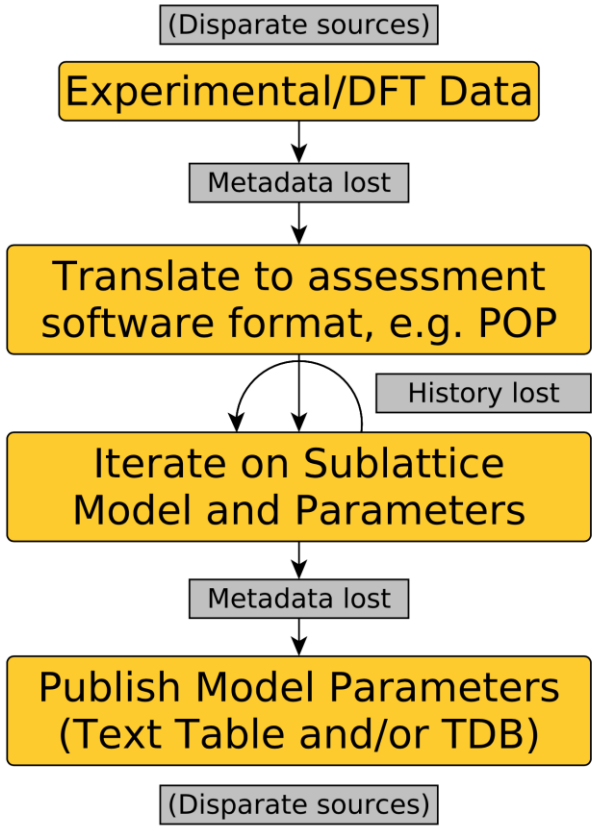
- Free
  - Freedom: share your work (or don't)
  - Cost: lower the barrier for entry

PHASES

PennState

Model for traditional database development





Otis (2016).

PHASES

PennState

Established methods are effective, but have pain points

- Data formats can lead to ambiguity
  - *"Should the experiment be `T` or `X`?"*
  - How long does it take to understand a file someone else created?
- Critically assessed data is closely guarded
- Maintenance and re-optimization prevents new models from being tested and adopted => n^{th} generation of databases

pycalphad

PHASES

PSU

PennState

We need better modeling tools too

- DRY principle: Don’t Repeat Yourself
- We need better tools for rapid (re)assessment of new models
- Our tools should tell us
  - Where our models might make bad predictions
  - How we should spend our time doing experiments and calculations

PHASES PSU PennState

Solution: ESPEI

Extensible **S**elf-optimizing **P**hase **E**quilibria **I**nfrastructure

2 step approach:

1. Rapidly generate model parameters
2. Automatically optimize parameters simultaneously

→ Quantify uncertainty to make us better modelers

Parameter selection

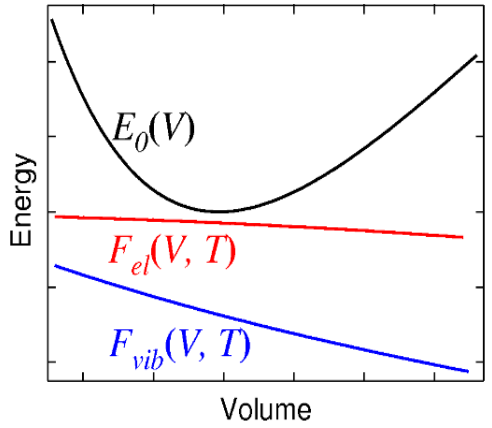
- Goal: Maximize model predictiveness, minimize model complexity
- Given a model: $\mathrm{a} + bT + cT\ln T + dT^{-1} + \sum_i e_i T^i$
- Question:

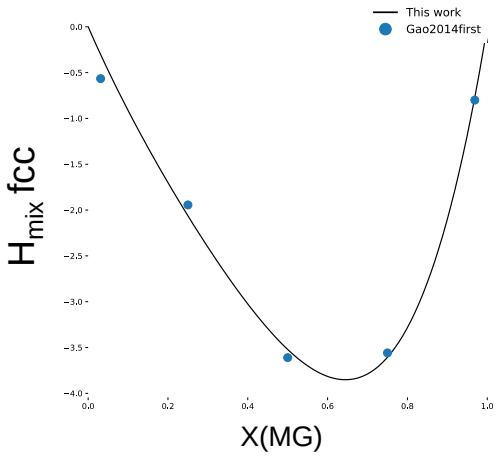
  How many parameters from this model should be used for each CALPHAD-type parameter?

- Traditional solution
  - Introduce parameters one at a time
  - Use domain knowledge
  - Rely on experimental information with increasing amounts from DFT
- ESPEI approach:
  - Rely heavily on DFT data
  - Use derivatives of Gibbs energy to simplify the problem
  - Apply corrected Akaike information criterion

PHASES PSU PennState

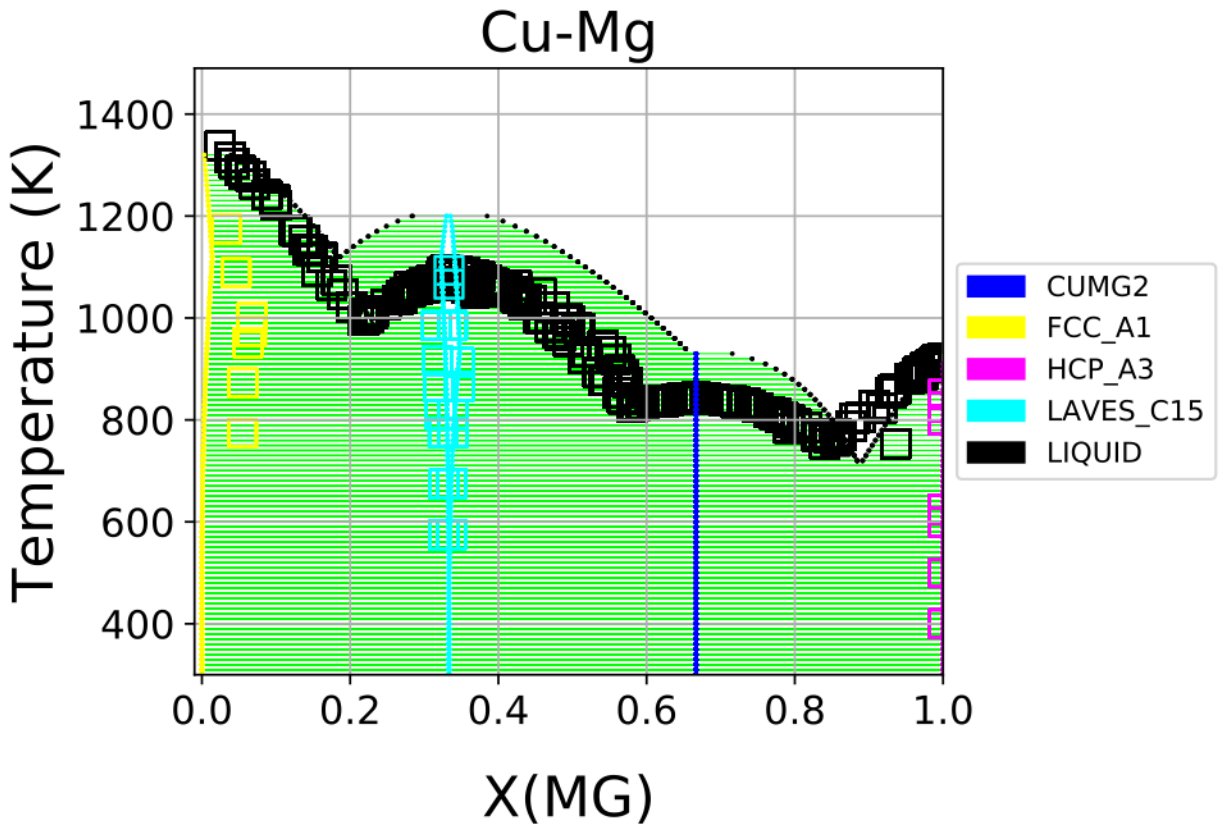
Parameter selection example: Cu-Mg phase diagram from single phase data





ESPEI
Parameter Selection

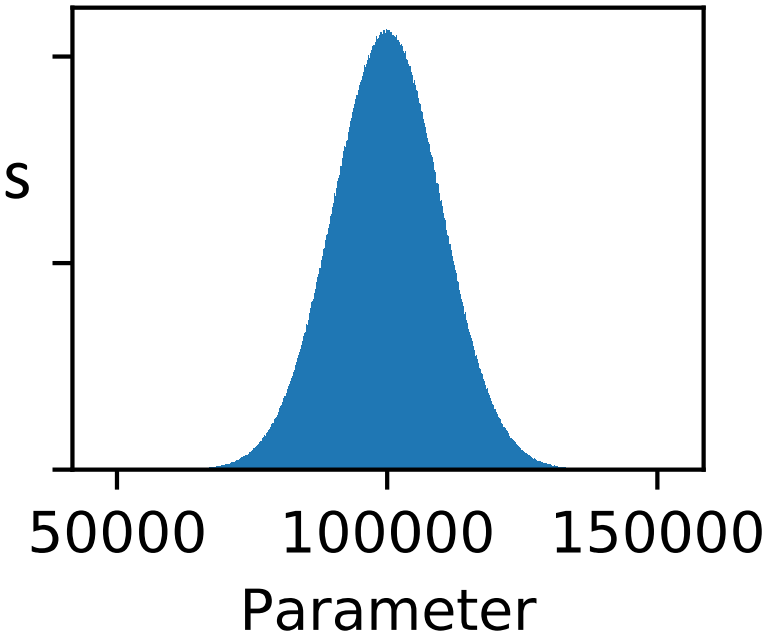
Run time: ≈ 3 seconds



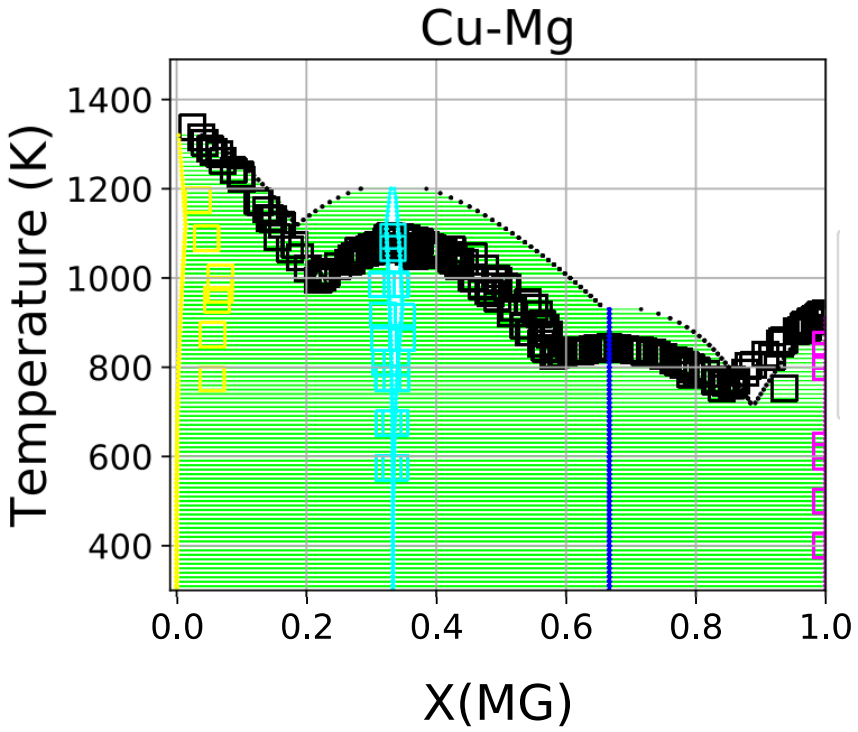
Ensemble Markov Chain Monte Carlo

- Goal of MCMC: find probability distribution for parameters
- Bayesian statistics method
- Start from a prior that is based on the current TDB
- Construct a Markov chain of possible parameter values

- For more on this come to our demo or Brandon's talk

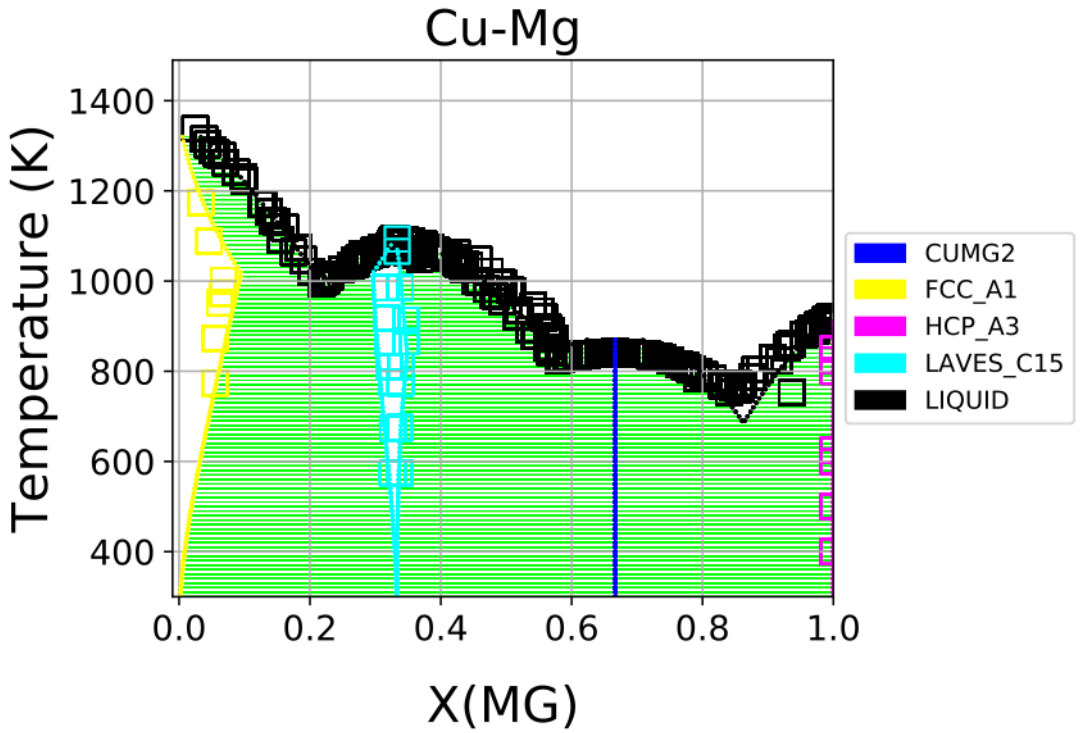


MCMC example: Cu-Mg phase diagram



ESPEI
MCMC

Run time: ≈ 8 hours on a laptop



PHASES PSU PennState

ESPEI datasets

- Critically assessed data is key to obtaining good results
- Data is digitized in JSON, a machine and human readable format

Single phase data


```
{
  "components": ["CU", "MG", "VA"],
  "phases": ["LAVES_C15"],
  "solver": {
    "mode": "manual",
    "sublattice_site_ratios": [2, 1],
      "sublattice_configurations": [["CU", "MG"],
                                    ["MG", "CU"],
                                    ["MG", "MG"],
                                    ["CU", "CU"]]
  },
  "conditions": {
    "P": 101325,
    "T": 298.15
  },
  "output": "HM_FORM",
    "values":   [[[-15720, 34720, 7000, 15500]]],
  "reference": "Zhou2007",
  "comment": "DFT reported in S. H. Zhou et al., J. Phase Equili
}
```


ZPF data


```
{
  "components": ["CU", "MG", "VA"],
  "phases": ["LIQUID", "FCC_A1", "LAVES_C15"],
  "conditions": {
      "P": 101325,
      "T": [773.85, 872.75, 951.15, 973.15, 1001.35, 1088.45, 1174.25]
  },
  "broadcast_conditions": false,
  "output": "ZPF",
  "values":   [
               [["FCC_A1", ["MG"], [0.0545706]], ["LAVES_C15", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0551807]], ["LAVES_C15", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0621264]], ["LAVES_C15", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0666104]], ["LAVES_C15", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0679314]], ["LIQUID", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0459052]], ["LIQUID", ["MG"], [null]]],
               [["FCC_A1", ["MG"], [0.0329389]], ["LIQUID", ["MG"], [null]]]
              ],
    "reference": "Jones1931",
    "comment": "solubility lines from Fig A, converted to atomic percent"
}
```


ate

ESPEI datasets

- Some digitized data freely available at
  - https://github.com/PhasesResearchLab/ESPEI-datasets
  - Contributions, fixes and suggestions are welcome
- Specification and guide for writing ESPEI datasets
  - http://espei.org/en/latest/input_data.html
- Prototype digitization of POP files to Python dictionaries
  - Used at Citrine Informatics
  - Conversion to ESPEI JSON coming soon
  - https://github.com/PhasesResearchLab/popparsing
  - http://popparsing.readthedocs.io/
  - We'd love suggestions and for you to report failures

PHASES PSU PennState

Description of the phase models

```
{
  "components": ["CU", "MG", "VA"],
  "refdata": "SGTE91",
  "phases": {
            "LIQUID" : {
                      "sublattice_model": [["CU", "MG"]],
                      "sublattice_site_ratios": [1]
            },
            "CUMG2": {
                      "sublattice_model": [["CU"], ["MG"]],
                      "sublattice_site_ratios": [1, 2]
            },
            "FCC_A1": {
                   "sublattice_model": [["CU", "MG"], ["VA"]],
                      "sublattice_site_ratios": [1, 1]
            },
            "HCP_A3": {
                      "sublattice_site_ratios": [1, 0.5],
                      "sublattice_model": [["CU", "MG"], ["VA"]]
            },
            "LAVES_C15": {
                      "sublattice_site_ratios": [2, 1],
                      "sublattice_model": [["CU", "MG"], ["CU", "MG"]]
            }
    }
}
```

PHASES

PennState

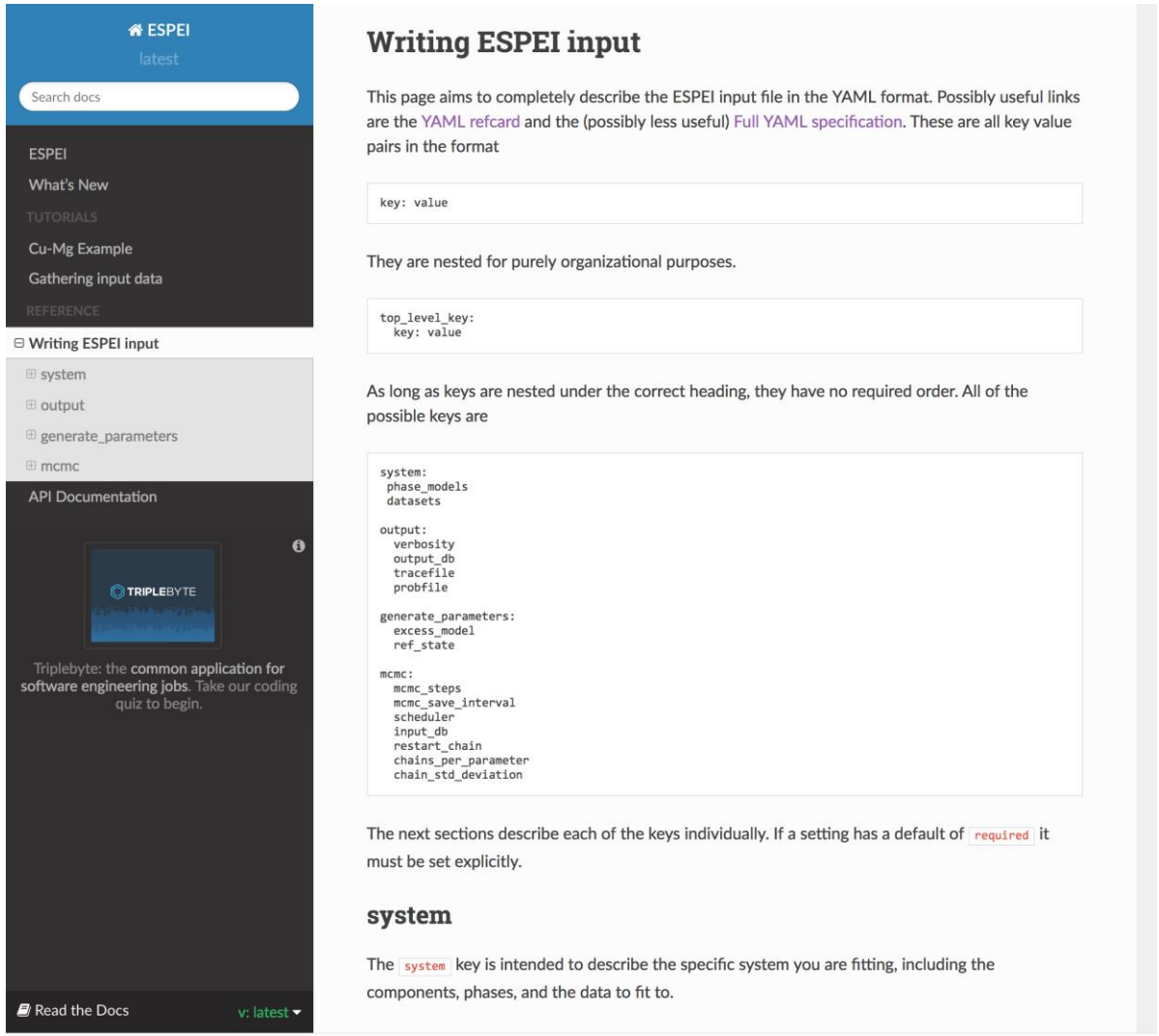
Running ESPEI: File based inputs

- Can do parameter generation, mcmc, or both
- Parameter generation
  - SGTE91 reference state provided
  - Easy to add other unary reference states
  - Choose an excess model (currently linear only)
- MCMC
  - Restart from previous run
  - Use MPI on a cluster
  - Many more customization options...

Example of minimal input for a full run

```
system:
  phase_models: phases.json
  datasets: my-input-data
generate_parameters:
  excess_model: linear
  ref_state: SGTE91
mcmc:
  mcmc_steps: 1000
```

Running ESPEI: File based inputs
Fully documented online



https://espei.org

PHASES PSU PennState