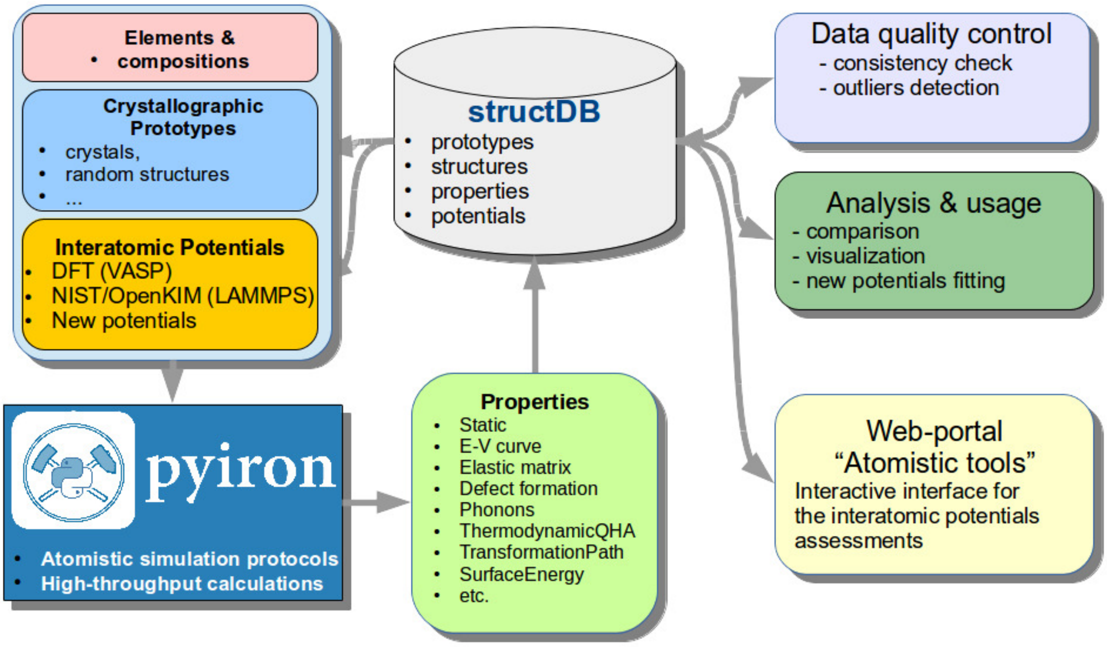
EMMC case study:

Data-driven Methods for Atomistic Simulations

Interview of Dr Yury Lysogorskiy, ICAMS, Ruhr-Universität Bochum

Writers: Alexandra Simperler and Gerhard Goldbeck



About Yury Lysogorskiy

Dr Yury Lysogorskiy (http://www.icams.de/content/people/icams-staff-members/?deta) is the leader of the research group "Data-driven methods for atomistic simulations" at the Interdisciplinary Centre for Advanced Materials Simulation (ICAMS) which is based at the Ruhr-Universität Bochum, Germany. Dr Lysogorskiy holds a PhD in Physics from the Kazan Federal University, Russia. His research interests are data-driven methods (machine learning, high-throughput calculations and data management), density functional theory and atomistic modelling and simulations.

Essentials about data and modelling

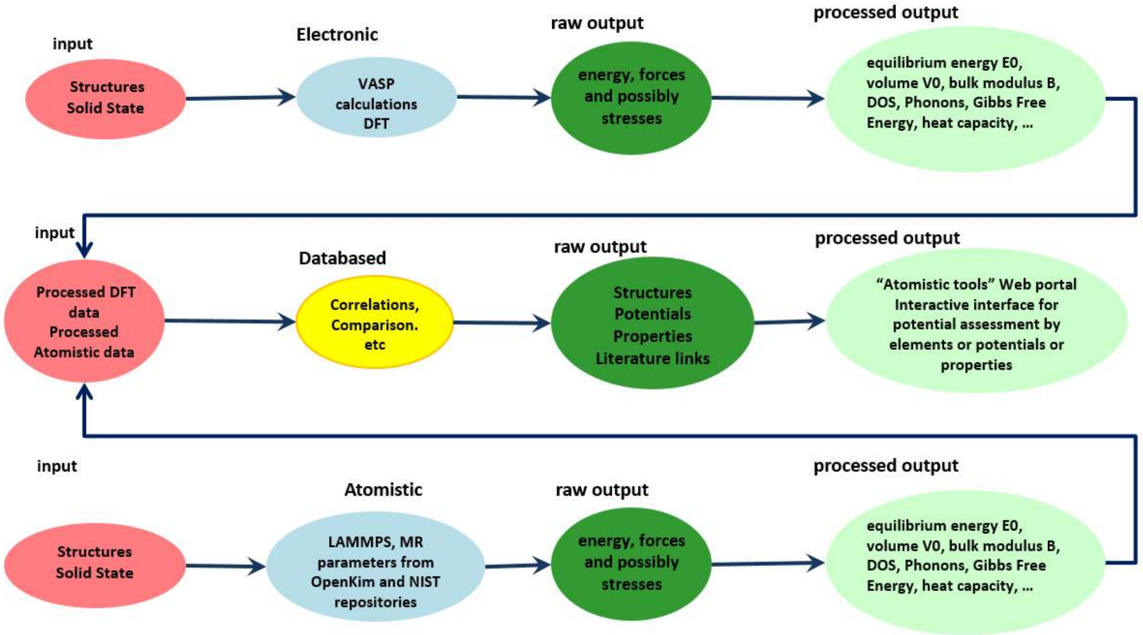
Dr Lysogorskiy started this project 2.5 years ago with the aim to validate and test interatomic potentials deposited in the NIST [1] and in the OpenKIM [2] repositories. He compares the predictions of the interatomic potentials to DFT reference calculations and, where available, to experimental data. The validation occurs on two levels, whereby level 1 is built on raw, unstructured reference data, which are typically energies, forces and stresses. Level 2 focusses on comparison with basic materials properties. The data used are calculated (DFT using VASP) and experimental (Pearson's Crystal Data, Landolt–Börnstein,

…). The data resulting from the actual potential validations are kept in a private repository at the moment. All collaborators can find data by elements, potentials and properties and to this community the data are FAIR. At some point in the future, this database will be accessible to the public. The database (PostgreSQL) is hosted by the university and JSON and HDF5 are used as auxiliary data formats. Python is used for coding.

When the data can be combined with electronic and atomistic models, potentials are extracted and links to the source (NIST, OpenKIM) and the original papers are provided. Also, the author of the validation data is stated. The access to data is controlled similar to UNIX file access, i.e. there are permissions if people can read and/or write into an entry.

About the Case Study

The case is based on Dr Lysogorskiy's papers "Transferability of interatomic potentials for molybdenum and silicon", Y. Lysogorskiy, T. Hammerschmidt, J. Janssen, J. Neugebauer and R. Drautz; Modelling Simul. Mater. Sci. Eng. 27 (2019) 025007 (DOI: https://doi.org/10.1088/1361-651X/aafd13) and "pyiron: An integrated development environment for computational materials science", J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, and J. Neugebauer; Computational Materials Science 163 (2019) 24-36 (DOI: https://doi.org/10.1016/j.commatsci.2018.07.043)


input
Structures Solid State
Electronic
VASP calculations DFT
raw output
energy, forces and possibly stresses
processed output
equilibrium energy E0, volume V0, bulk modulus B, DOS, Phonons, Gibbs Free Energy, heat capacity, …
input
Processed DFT data Processed Atomistic data
Databased
Correlations, Comparison. etc
raw output
Structures Potentials Properties Literature links
processed output
"Atomistic tools" Web portal Interactive interface for potential assessment by elements or potentials or properties
input
Structures Solid State
Atomistic
LAMMPS, MR parameters from OpenKim and NIST repositories
raw output
energy, forces and possibly stresses
processed output
equilibrium energy E0, volume V0, bulk modulus B, DOS, Phonons, Gibbs Free Energy, heat capacity, …

For this particular case, which were your objectives?

Our aim is to validate and test interatomic potentials deposited in the NIST and in the OpenKIM repositories. We want to provide an explicit and exhaustive comparison of different potentials with DFT over a wide range of materials properties and structures.

For this Case study did you create and/or apply a data-based model

We perform a postprocessing of certain calculations in order to extract more complicated physical properties. For example, we compute the bulk modulus B and its pressure derivative by a least-square fit of the E–V curves to fifth-order polynomials.

How did data play a key role in problem solving?

Experimental and DFT data are key to validate potentials. Often, new potentials are compared only to previously available potentials which is not sufficient enough for a sophisticated validation.

What methodologies have been applied?

As we had to handle a large number of combinations of interatomic potentials, properties and crystal structures, we were using the *pyiron* computational management framework [3]. Due to the integrated development environment aspect of *pyiron* we could implement and use therein protocols for the elastic matrix, phonons, vacancy formation energy, transformation paths, surface energy, thermodynamic properties, etc. and also utilized *pyiron*'s interfaces to VASP and LAMMPS.

What were the expected improvements by adding data to your modelling?

Generally, potentials have a limited transferability and therefore it is essential to assess the properties of a potential carefully before one applies it to a specific simulation. Our data-driven validation enables a practitioner to choose potentials in a more targeted way which leads to more reliable results.

For this particular case, did you have to invest a lot of work to make the data usable?

The data itself was not so much of a problem. However, we started from scratch to build the infrastructure around the data and it took us two years to perfection it. A lot of python code had to be written and we moved our validation routine from single core to the multi-core distributed system on a cluster.

For this particular case, what did you do with the data w. r. t. data-science?

We create new data and extract data, and to a certain extent we also reduce data. In order to get a more general idea of the predictive power of the potentials we consider correlations between different properties across different prototype crystal structures. We use Spearman's rank correlation coefficient to reduce the data space. For example, if the potentials reproduce a positive correlation between equilibrium energy and volume, it proves that they correctly capture the tendency of a compound to form closed-packed rather than open structures. Thus, we can better assess the transferability and consistency of interatomic potentials.

For this particular case, what did you do with the data w. r. t. materials applications?

Our validation method allows to identify the best potential for a structure and property of interest. In the future, we would like to enable that a practitioner can also validate their potentials and gain better knowledge about their performance.

For this particular case, what was the quantitative value of combining data with materials modelling?

Our validation method can aid users to reliably select the most appropriate forcefield for the task at hand. This will save person time and thus, will speed up their research.

What investments were made during the project?

We were awarded a grant which was used for salaries and the acquisition of a new server.

What sort of obstacles or barriers (if any) did you have to overcome to use data driven modelling?

We had to adapt the pyiron software and I converted from programming as a hobby to more serious coding.

Did using data improve your competitiveness/innovation power?

As I had to gain experience from scratch, I learned a lot about building an infrastructure and gained new competences additionally to my scientific curriculum.

What would you need the community to provide to enable data-driven materials modelling?

I would like 3rd party interatomic potential data to be machine accessible rather than needing human interference to access data. OpenKIM, for example, does have an API, that provides a standard for exchanging information between atomistic simulation codes and interatomic potentials. Standardization of repositories and curation are key, as one has to rely on that interatomic potentials survived their digitisation without errors/typos.

References

[1] NIST Interatomic Potential Repository: https://www.ctcms.nist.gov/potentials/

[2] OpenKIM, Knowledgebase of Interatomic Models: https://openkim.org/

[3] “pyiron: An integrated development environment for computational materials science”, J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, and J. Neugebauer; Computational Materials Science 163 (2019) 24-36 (DOI: https://doi.org/10.1016/j.commatsci.2018.07.043)