



NOVEL - MATERIALS  
DISCOVERY LABORATORY



FRITZ-HABER-INSTITUT  
MAX-PLANCK-GESELLSCHAFT



# The NOMAD Artificial-Intelligence Toolkit: Web-Based FAIR-Data-Driven Materials Science

Luigi Sbailò

Humboldt –Universität zu Berlin

# A web based Artificial-Intelligence toolkit

The [NOMAD Archive](#) includes data from [AFLOW](#), Materials Project, OQMD, and more, and from numerous individual researchers.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, Materials Project, OQMD, and more, and from numerous individual researchers.

**Findable Accessible Interoperable Reusable (FAIR) data**

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, Materials Project, OQMD, and more, and from numerous individual researchers.

➡ **Findable AI Ready data**

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, Materials Project, OQMD, and more, and from numerous individual researchers.

→ **Findable AI Ready data**

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, Materials Project, OQMD, and more, and from numerous individual researchers.

→ **Findable AI Ready data**

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

# A web based Artificial-Intelligence toolkit

The **NOMAD Archive** includes data from **AFLOW**, Materials Project, OQMD, and more, and from numerous individual researchers.

→ **Findable AI Ready data**

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

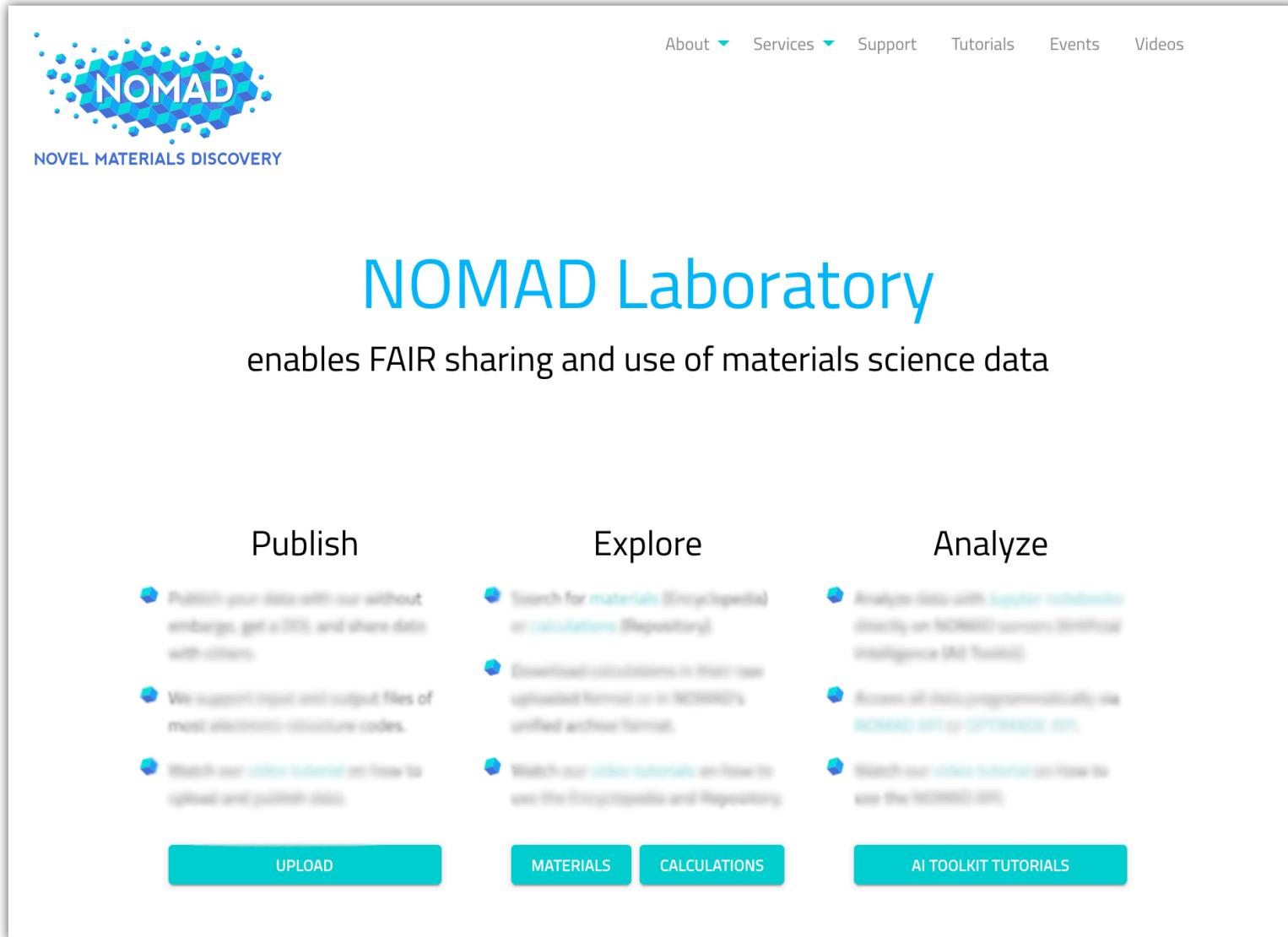
The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

In the *NOMAD Artificial-Intelligence Toolkit*, we maintain several notebooks demonstrating the latest applications of AI to materials science data.

# Technology

- Jupyter Notebook is used as environment for tutorials.
- Machine-learning packages, i.e. scikit-learn, tensorflow, ..., can be imported within the environment for immediate deployment.
- Servers deployed by the NOMAD AI toolkit are hosted by the Max Planck Computer & Data Facility in Garching.





The screenshot shows the homepage of the NOMAD Laboratory. At the top right is the NOMAD logo. The main title "NOMAD Laboratory" is in large blue font, followed by the subtitle "enables FAIR sharing and use of materials science data". Below this are three main sections: "Publish", "Explore", and "Analyze", each with a list of features and corresponding buttons.

## NOMAD Laboratory

enables FAIR sharing and use of materials science data

### Publish

- Publish your data with or without embargo, get a DOI, and share data with others.
- We support input and output files of most common simulation codes.
- Watch our [video tutorial](#) on how to upload and publish data.

[UPLOAD](#)

### Explore

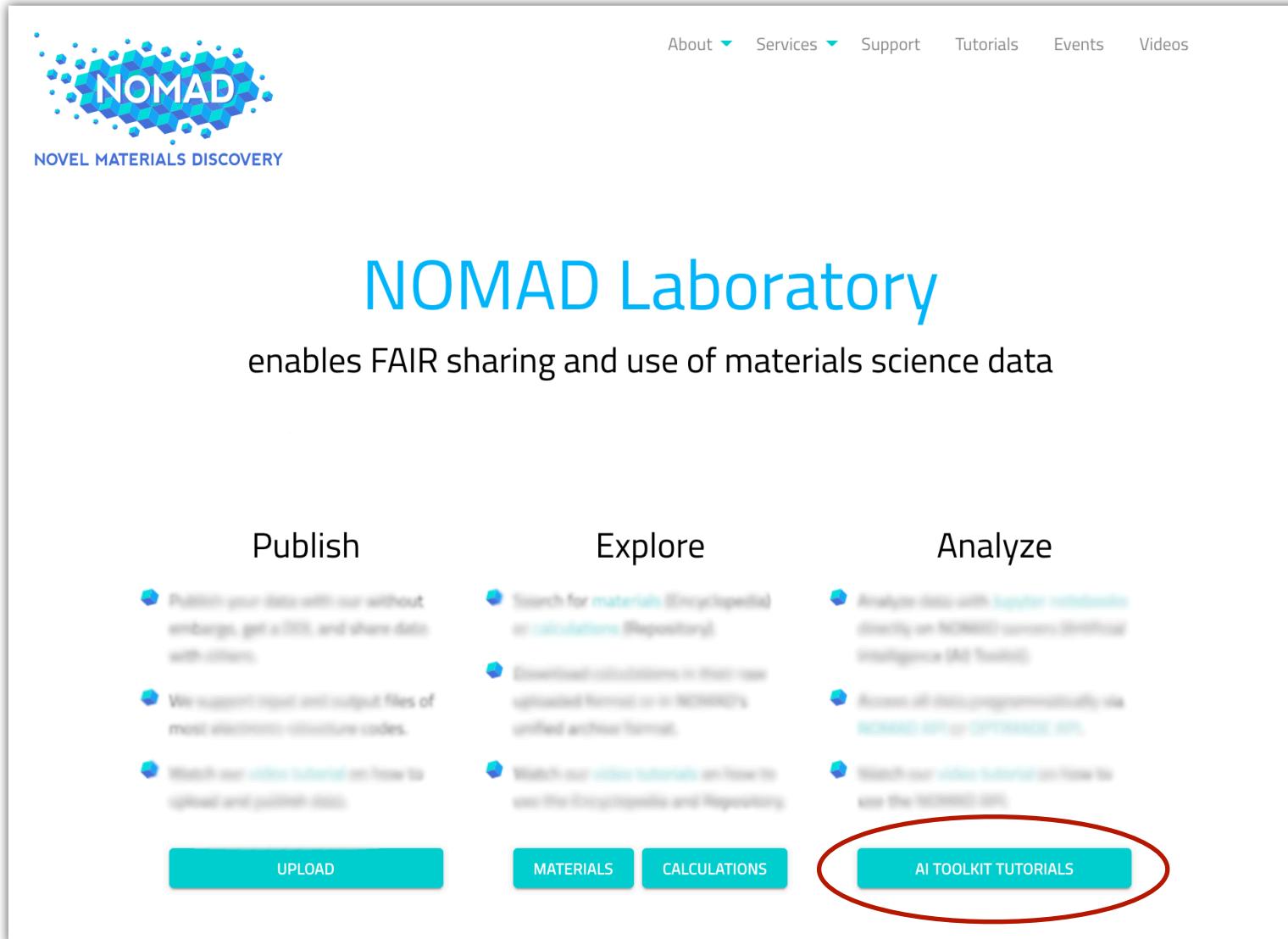
- Search for materials ([Encyclopedia](#)) or calculations ([Repository](#)).
- Download publications on most new deposited materials in NOMAD's unified archive format.
- Watch our [video tutorials](#) on how to use the Encyclopedia and Repository.

[MATERIALS](#) [CALCULATIONS](#)

### Analyze

- Analyze data with [Machine Learning](#) (using our NOMAD's own Statistical Intelligence API Toolkit).
- Access all data programmatically via [API](#) or [COMMAND LINE](#).
- Watch our [video tutorial](#) on how to use the NOMAD API.

[AI TOOLKIT TUTORIALS](#)



The screenshot shows the NOMAD Laboratory homepage. At the top right is the NOMAD logo and the text "NOVEL MATERIALS DISCOVERY". The main title "NOMAD Laboratory" is in large blue text, followed by the subtitle "enables FAIR sharing and use of materials science data". Below this are three main sections: "Publish", "Explore", and "Analyze", each with a list of features and a red oval highlighting the "AI TOOLKIT TUTORIALS" button under "Analyze".

**Publish**

- Publish your data with or without embargo, get a DOI, and share data with others.
- We support input and output files of most common simulation codes.
- Watch our [video tutorial](#) on how to upload and publish data.

**Explore**

- Search for [materials](#) (Encyclopedia) or [calculations](#) (Repository).
- Download calculations in their raw deposited formats or in NOMAD's unified archive format.
- Watch our [video tutorials](#) on how to use the Encyclopedia and Repository.

**Analyze**

- Analyze data with [Tensor Networks](#) (using our NOMAD's open-source DeepChem AI Toolkit).
- Assess all documents automatically via [Machine Learning](#) or [Community](#).
- Watch our [video tutorial](#) on how to use the Tensor Network.

**AI TOOLKIT TUTORIALS**

# The NOMAD Artificial-Intelligence toolkit

## Powerful Artificial-Intelligence Tools for Materials Science

Find New Patterns and Information in Materials-Science Big Data



Query the archive →

View tutorials →

Reproduce results →

Get to work →

i

# The NOMAD Artificial-Intelligence toolkit

## Powerful Artificial-Intelligence Tools for Materials Science

Find New Patterns and Information in Materials-Science Big Data



Query the archive →

View tutorials →

Reproduce results → i

Get to work →

# Reproducing published results

## Identifying domains of applicability of machine-Learning models for materials science

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

## Discovery of new topological insulators in alloyed tetradymites

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

 ACCESS TUTORIAL

Authors: Luigi Sbailò | Thomas A. R. Purcell | Luca M. Ghiringhelli | Matthias Scheffler

### AI methods:

Supervised learning Classification Symbolic regression Features selection Atomic features

SISSO

### System:

Tetradymites Topological insulators

### Additional Resources:

[DOI](#) [PDF](#)

## ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

## Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides

Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler

# Reproducing published results

## Identifying domains of applicability of machine-Learning models for materials science

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

## Discovery of new topological insulators in alloyed tetradymites

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

 ACCESS TUTORIAL

## ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

## Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides

Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler

### AI methods:

Supervised learning | Classification | Symbolic regression | Features selection | Atomic features

Guohua Cao et al., Phys. Rev. Materials 4, 034204 (2020)

### Additional Resources:

[DOI](#) [PDF](#)

# Reproducing published results

## Identifying domains of applicability of machine-Learning models for materials science

Authors: Mohammad-Yasin Arif | Luigi Sbailò | Luca M. Ghiringhelli

## Discovery of new topological insulators in alloyed tetradymites

Learn how to find descriptive parameters (short formulas) that predict whether alloyed materials are topological or trivial insulators, using the example of tetradymites. This notebook is based on the algorithm 'sure independence screening and sparsifying operator' (SISSO) that enables to search for optimal descriptor by scanning huge feature spaces.

 ACCESS TUTORIAL

Authors: Luigi Sbailò | Thomas A. R. Purcell | Luca M. Ghiringhelli | Matthias Scheffler

### AI methods:

Supervised learning | Classification | Symbolic regression | Features selection | Atomic features

Guohua Cao et al., Phys. Rev. Materials 4, 034204 (2020)

### Additional Resources:

[DOI](#) [PDF](#)

## ARISE - Robust recognition and exploratory analysis of crystal structures via Bayesian deep learning

Authors: Andreas Leitherer | Angelo Ziletti | Luca M. Ghiringhelli

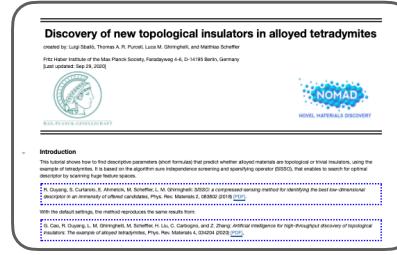
## Subgroup discovery of catalysts' genes for carbon-dioxide activation on semiconductor oxides

Authors: Aliaksei Mazheika | Luigi Sbailò | Luca M. Ghiringhelli | Sergey Levchenko | Matthias Scheffler

# Reproducing published results



## NOVEL MATERIALS DISCOVERY



— — — — —

# Discovery of new topological insulators in alloyed tetradymites

created by: Luigi Sbailò, Thomas A. R. Purcell, Luca M. Ghiringhelli, and Matthias Scheffler  
Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
[Last updated: Sep 29, 2020]



MAX-PLANCK-GESELLSCHAFT



NOMAD  
NOVEL MATERIALS DISCOVERY

Luigi Sbailò (Humboldt –Universität zu Berlin)

[luigi.sbailo@physik.hu-berlin.de](mailto:luigi.sbailo@physik.hu-berlin.de)

- NFDI workshop April 2022 -

# Reproducing published results

# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**

Georgios Cotsopoulos, Thomas A. R. Fennel, Luis M. Campos, and Thomas Schröder  
[Last updated: Sep 29, 2020]

**Introduction**  
This tutorial shows how to find descriptive parameters (descriptor formulas) that predict whether certain materials are topological or not. Materials, using the example of tetradymites. It is based on the algorithm semi-independent screening and searching (SISSO), that enables to search for optimal descriptor formulas.

**Description of the method**  
The SISSO algorithm is a semi-independent screening and searching method for identifying the best new descriptor. It is an improvement of the well-known recursive feature elimination (RFE) method for identifying the best few descriptors. With the default settings, the method reproduces the same results as RFE.

**Explanation of the method (click to expand/collapse)**  
The idea demonstrated in this tutorial is to start from simple physical quantities ("primary features"), whose properties of the constituent atoms allow one to calculate various correlations in the so-called "feature space". Then, SISSO is used to select only a few of these features that explain the data. By clicking directly on "Run", i.e., with the default selection, you can reproduce the 2D map as published in [\[PRM2020\]](#). You can also select primary features by yourself. In the "Operations" section, you can choose which operations are allowed for the feature selection. In the "Features" section, the number of features that are needed for each iteration of the SIS loop, and the max number of dimensions of the model. If a model is considered to be "good", it is saved in the "Settings" section. Finally, the "Default selection" section contains the parameters for the run.

**Operations:**

- $x + y$
- $x - y$
- $|x - y|$
- $x \cdot y$
- $x/y$
- $\exp(x)$
- $\exp(-x)$
- $1/x$
- $x^2$
- $x^3$
- $\sqrt{x}$
- $\sqrt[3]{x}$
- $\log(x)$
- $|x|$

**Features:**

- $Z_{\text{cations}}$
- $\chi_{\text{cations}}$
- $\lambda_{\text{cations}}$
- $Z_{\text{anions}}$
- $\chi_{\text{anions}}$
- $\lambda_{\text{anions}}$

**Settings:**

SISSO rung:  
**PRM2020**

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration:  
**50**

Maximum number of dimensions:  
**2**

**Default selection**

Run

Plot interactive map

**Number of features generated:** 15000

**1D model**  
 $\# \text{misclassified: } 25$   
 $\text{SVM dividing line: } c_0 + a_0 * [[[Z11A*Z11B]/\sqrt{X16E}]/[[Z11A/Z11C]+[Z11B/Z11D]]] = 0$   
 $c_0:8.968 \quad | \quad a_0:-0.01492$

**2D model**  
 $\# \text{misclassified: } 0$   
 $\text{SVM dividing line: } c_0 + a_0 * \text{abs}[[[X16C+X16D]*[Z11E/X16B]]-[[Z11C+Z11D]+\text{abs}[Z11C-Z11D]]] + a_1 * [[[Z11A+Z11B]*[Z11C+Z11D]]-\text{abs}[Z11A*Z11D]-[Z11B*Z11C]]] = 0$   
 $c_0:25.55 \quad | \quad a_0:0.1313 \quad | \quad a_1:-0.004255$

**Ticks**  
 To open the plot appearance tools.  
 Tick the box next to the cross symbols in order to choose which windows visualize the next structure selected in the map above.  
 Compound  
 Z, Ac, Sc  
 Display  
 Compound  
 Z, Ac, Sc  
 Display

**Chemical structures**

**Input parameters**

**Operations:**

- $x + y$
- $x - y$
- $|x - y|$
- $x \cdot y$
- $x/y$
- $\exp(x)$
- $\exp(-x)$
- $1/x$
- $x^2$
- $x^3$
- $\sqrt{x}$
- $\sqrt[3]{x}$
- $\log(x)$
- $|x|$

**Features:**

- $Z_{\text{cations}}$
- $\chi_{\text{cations}}$
- $\lambda_{\text{cations}}$
- $Z_{\text{anions}}$
- $\chi_{\text{anions}}$
- $\lambda_{\text{anions}}$

**Settings:**

SISSO rung:  
**PRM2020**

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration:  
**50**

Maximum number of dimensions:  
**2**

**Default selection**

Run

Plot interactive map

**Number of features generated:** 15000

**1D model**  
 $\# \text{misclassified: } 25$   
 $\text{SVM dividing line: } c_0 + a_0 * [[[Z11A*Z11B]/\sqrt{X16E}]/[[Z11A/Z11C]+[Z11B/Z11D]]] = 0$   
 $c_0:8.968 \quad | \quad a_0:-0.01492$

**2D model**  
 $\# \text{misclassified: } 0$   
 $\text{SVM dividing line: } c_0 + a_0 * \text{abs}[[[X16C+X16D]*[Z11E/X16B]]-[[Z11C+Z11D]+\text{abs}[Z11C-Z11D]]] + a_1 * [[[Z11A+Z11B]*[Z11C+Z11D]]-\text{abs}[Z11A*Z11D]-[Z11B*Z11C]]] = 0$   
 $c_0:25.55 \quad | \quad a_0:0.1313 \quad | \quad a_1:-0.004255$

# Reproducing published results

**Discovery of new topological insulators in alloyed tetradydimes**

Contributed by: Jürgen Roth, Thomas A. R. Pinen, Lucas M. Ohberg, and Matthias Scheidt  
Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany  
Last updated: Sep 29, 2020

  
NOVEL MATERIALS DISCOVERY

**Introduction**  
This tutorial shows how to find descriptive parameters which formulas to predict whether certain materials are topological or not. In addition, using the example of tetradydimes, it is based on the algorithm sure independence screening and sparsifying operator (SISSO), that enables us to optimally reduce the number of features.

**Methodology**  
SISSO is a machine learning method for identifying the best low-dimensional descriptor or an ensemble of related candidates. Phys. Rev. Materials 2, 083602 (2018) [CrossRef]

With the default settings, the method reproduces the same results from: J. M. M. Franca, J. C. L. Vasconcelos, and J. P. Z. P. de Souza, "Machine learning for high-throughput electronic structure calculations: The search of electronic bandgaps," Phys. Rev. Materials 2, 083602 (2018) [CrossRef]

**Description of the method code implementation**

The idea demonstrated in this tutorial is to start from simple physical quantities ("primary features"), here properties of the constituents like atoms such as Pauling electronegativity, atomic radius, and atomic mass. These are combined to form more complex features ("secondary features"), here combinations of primary features that are called "feature bases". Then, SISSO is used to select only a few of these formulas that explain the data.

By clicking directly on "Run", i.e., with the default selection, you can reproduce the SI2 map as published in [\[10\]](#). You can also select primary features and run the code. If you want to use a different set of primary features, you can do so by clicking on the checkboxes next to the primary features. The number of features are selected at each iteration of the SI2 steps, and the max number of dimensions of the model. The features considered here are the Pauling electronegativity ( $\chi_{\text{anions}}$ ), the atomic radius ( $Z_{\text{anions}}$ ), the atomic mass ( $Z_{\text{cations}}$ ), and the atomic volume ( $V$ ). The features  $\chi_{\text{cations}}$ ,  $Z_{\text{cations}}$ , and  $V$  have therefore projected the features to be selected into these four categories and areas. This means that by selecting, e.g., a primary feature  $\chi_{\text{cations}}$  and running the code, the code will automatically select the secondary features  $Z_{\text{cations}}$  and  $V$  and run the SISSO optimization. After the feature and other settings selected, press "Run". After the code has finished, the output is a 2D map showing the data points and the support vectors. If you click on "Plot interactive map", you can press "Plot interactive map" to reveal a map of tetrahedrae (topological insulators), the highest-dimensional model (d=2), the support vector machine (SVM) separating the two classes (topological insulators and non-topological insulators), and the decision boundary. The plot also shows the data points (magenta dots), inactive drop-down menu allow to assign axes, markers, and colors, to the descriptor components of choice.

With the selection of "PRM2020" (or default selection as SISSO rung), a special feature space is updated, which contains fewer features than the primary features. This means that the same result is obtained much faster. Note that the primary feature space contains thousands of the top-ranked features and SISSO finds the best 100 model.

The Python code for this notebook is by default hidden for easier reading. To toggle off/on the code, click [here](#).

**Operations:**

- $x + y$
- $x - y$
- $|x - y|$
- $x \cdot y$
- $x/y$
- $\exp(x)$
- $\exp(-x)$
- $1/x$
- $x^2$
- $x^3$
- $\sqrt{x}$
- $\sqrt[3]{x}$
- $\log(x)$
- $|x|$

**Features:**

- $Z_{\text{cations}}$
- $\chi_{\text{cations}}$
- $\lambda_{\text{cations}}$
- $Z_{\text{anions}}$
- $\chi_{\text{anions}}$
- $\lambda_{\text{anions}}$

**Settings:**

SISSO rung: PRM2020 ▾

To unfreeze the feature selection, please select any rung other than PRM2020.

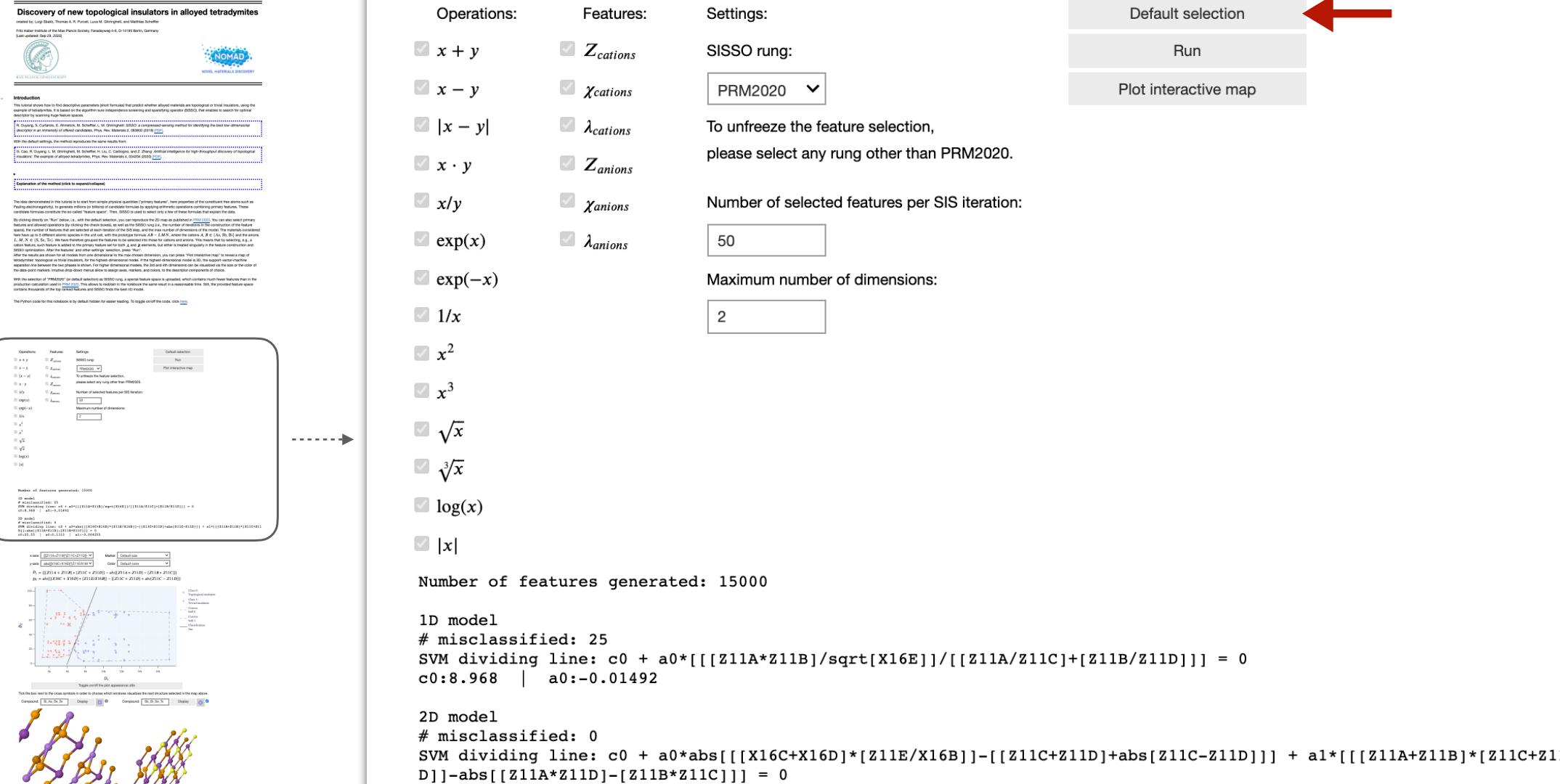
Number of selected features per SIS iteration: 50

Maximum number of dimensions: 2

**Default selection**

Run

Plot interactive map



Number of features generated: 15000

**1D model**  
 $\# \text{misclassified: } 25$   
 $\text{SVM dividing line: } c_0 + a_0 * [[Z11A*Z11B]/\sqrt{X16E}]/[[Z11A/Z11C]+[Z11B/Z11D]] = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

**2D model**  
 $\# \text{misclassified: } 0$   
 $\text{SVM dividing line: } c_0 + a_0 * \text{abs}[[X16C+X16D]*[Z11E/X16B]] - [[Z11C+Z11D]+\text{abs}[Z11C-Z11D]] + a_1 * [[Z11A+Z11B]*[Z11C+Z11D]] - \text{abs}[Z11A*Z11D] - [Z11B*Z11C]] = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$

# Reproducing published results



Operations:	Features:	Settings:	Default selection
<input checked="" type="checkbox"/> $x + y$	<input checked="" type="checkbox"/> $Z_{cations}$	SISSO rung:	Run
<input checked="" type="checkbox"/> $x - y$	<input checked="" type="checkbox"/> $\chi_{cations}$	PRM2020	Plot interactive map
<input checked="" type="checkbox"/> $ x - y $	<input checked="" type="checkbox"/> $\lambda_{cations}$	To unfreeze the feature selection, please select any rung other than PRM2020.	
<input checked="" type="checkbox"/> $x \cdot y$	<input checked="" type="checkbox"/> $Z_{anions}$	Number of selected features per SIS iteration:	
<input checked="" type="checkbox"/> $x/y$	<input checked="" type="checkbox"/> $\chi_{anions}$	50	
<input checked="" type="checkbox"/> $\exp(x)$	<input checked="" type="checkbox"/> $\lambda_{anions}$	Maximum number of dimensions:	
<input checked="" type="checkbox"/> $\exp(-x)$		2	
<input checked="" type="checkbox"/> $1/x$			
<input checked="" type="checkbox"/> $x^2$			
<input checked="" type="checkbox"/> $x^3$			
<input checked="" type="checkbox"/> $\sqrt{x}$			
<input checked="" type="checkbox"/> $\sqrt[3]{x}$			
<input checked="" type="checkbox"/> $\log(x)$			
<input checked="" type="checkbox"/> $ x $			
<b>Output</b>			
<pre>Number of features generated: 15000  1D model # misclassified: 25 SVM dividing line: c0 + a0*[[Z11A*Z11B]/sqrt[X16E]]/[[Z11A/Z11C]+[Z11B/Z11D]] = 0 c0:8.968   a0:-0.01492  2D model # misclassified: 0 SVM dividing line: c0 + a0*abs[[X16C+X16D]*[Z11E/X16B]]-[[Z11C+Z11D]+abs[Z11C-Z11D]] + a1*[[Z11A+Z11B]*[Z11C+Z11D]]-abs[[Z11A*Z11D]-[Z11B*Z11C]] = 0 c0:25.55   a0:0.1313   a1:-0.004255</pre>			

# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**

Georgios Gouliatis, Thomas A. R. Fennel, Luis M. Campos, and Thomas Schröder  
[Last updated: Sep 29, 2020]

**Introduction**  
In this tutorial we show how to find descriptive parameters given formulas that predict whether certain materials are topological or not. Materials, using the example of tetradymites. It is based on the algorithm semi-independence screening and searching (SISSO), that enables to search for optimal descriptors in an extremely efficient way.

**Description of the method**  
The SISSO algorithm is a semi-independent screening and searching method for identifying the best few descriptors with the smallest number of dimensions. Phys. Rev. Mater. 2, 034202 (2018)

**With the default settings, the method reproduces the same results:**  
Phys. Rev. Mater. 2, 034202 (2018) and Phys. Rev. Mater. 2, 034203 (2018)

**Explanation of the method (click to expand/collapse)**

The idea demonstrated in this tutorial is to start from simple physical quantities ("primary features"), whose properties of the constituent atoms allow such an analysis. These primary features are combined to form "descriptor functions". These descriptor functions constitute the so-called "feature space". Then, SISSO is used to select only a few of these functions that explain the data. By clicking directly on "Run", i.e., with the default selection, you can reproduce the 2D map as published in [Phys. Rev. Mater.](#). You can also select primary features by clicking on the checkboxes in the "Operations" section. In this case, the "Run" button will be disabled. Instead, you will see a dropdown menu where you can specify, the number of features that are needed for each iteration of the SIS loop, and the max number of dimensions of the model. If no models constraint is selected, the maximum number of dimensions is set to 20. If a specific model is selected, the maximum number of dimensions is set to the dimension of the model. M, N, D, T, Sr, Tr, we have therefore grouped the features to be selected into three for cations and anions. This means that by selecting, e.g., a feature for cations, all features for cations will be selected. If you want to select a feature for anions, you have to click on the checkbox for anions. After selecting the features, you can press "Plot interactive map". To reveal a map of semiperiodic topology or misclassification, for the highest-dimensional model is 20, the aspect vector must be set to 1.0. If you want to change the aspect ratio, you can do so by clicking on the "Aspect" button. If you want to change the color of the data point markers, initialize drop-down menus above to assign pose, markers, and colors, to the descriptor components of choice.

**Python code for this notebook**  
This notebook contains the Python code used to generate the plots shown in the paper. When running the notebook, make sure that the environment variable `NOMAD_HOME` is set to the path of the directory containing the code. The code is contained in the `PRM2020` folder. This allows to maintain the location of the code in a reasonable time. Note, the present feature space contains thousands of the generated features and should not be used in its raw state.

**Code**

**Operations**

- $x + y$
- $x - y$
- $|x - y|$
- $x \cdot y$
- $x/y$
- $\exp(x)$
- $\exp(-x)$
- $1/x$
- $x^2$
- $x^3$
- $\sqrt{x}$
- $\sqrt[3]{x}$
- $\log(x)$
- $|x|$

**Features:**

- $Z_{\text{cations}}$
- $\chi_{\text{cations}}$
- $\lambda_{\text{cations}}$
- $Z_{\text{anions}}$
- $\chi_{\text{anions}}$
- $\lambda_{\text{anions}}$

**Settings:**

SISSO rung: PRM2020

**Default selection**

Run

Plot interactive map

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration: 50

Maximum number of dimensions: 2

**Number of features generated: 15000**

**1D model**  
 $\# \text{misclassified: } 25$   
 $\text{SVM dividing line: } c_0 + a_0 * [[Z11A*Z11B]/\sqrt{X16E}]/[[Z11A/Z11C]+[Z11B/Z11D]] = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

**2D model**  
 $\# \text{misclassified: } 0$   
 $\text{SVM dividing line: } c_0 + a_0 * \text{abs}[[X16C+X16D]*[Z11E/X16B]] - [[Z11C+Z11D]+\text{abs}[Z11C-Z11D]] + a_1 * [[Z11A+Z11B]*[Z11C+Z11D]] - \text{abs}[Z11A*Z11D] - [Z11B*Z11C]] = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$

**Ticks**  
 Toggling off the plot appearance tools.  
 Compound   
 Display

**Chemical structures**

**Operations:**

- $x + y$
- $x - y$
- $|x - y|$
- $x \cdot y$
- $x/y$
- $\exp(x)$
- $\exp(-x)$
- $1/x$
- $x^2$
- $x^3$
- $\sqrt{x}$
- $\sqrt[3]{x}$
- $\log(x)$
- $|x|$

**Features:**

- $Z_{\text{cations}}$
- $\chi_{\text{cations}}$
- $\lambda_{\text{cations}}$
- $Z_{\text{anions}}$
- $\chi_{\text{anions}}$
- $\lambda_{\text{anions}}$

**Settings:**

SISSO rung: PRM2020

**Default selection**

Run

Plot interactive map

To unfreeze the feature selection, please select any rung other than PRM2020.

Number of selected features per SIS iteration: 50

Maximum number of dimensions: 2

**Number of features generated: 15000**

**1D model**  
 $\# \text{misclassified: } 25$   
 $\text{SVM dividing line: } c_0 + a_0 * [[Z11A*Z11B]/\sqrt{X16E}]/[[Z11A/Z11C]+[Z11B/Z11D]] = 0$   
 $c_0: 8.968 \quad | \quad a_0: -0.01492$

**2D model**  
 $\# \text{misclassified: } 0$   
 $\text{SVM dividing line: } c_0 + a_0 * \text{abs}[[X16C+X16D]*[Z11E/X16B]] - [[Z11C+Z11D]+\text{abs}[Z11C-Z11D]] + a_1 * [[Z11A+Z11B]*[Z11C+Z11D]] - \text{abs}[Z11A*Z11D] - [Z11B*Z11C]] = 0$   
 $c_0: 25.55 \quad | \quad a_0: 0.1313 \quad | \quad a_1: -0.004255$

# Reproducing published results

**Discovery of new topological insulators in alloyed tetradymites**

Georgios G. Dotsikas, Thomas A. R. Fennel, Luis M. Campos, and Thomas Schröder  
Preprint number: NOMAD-2020-004, DOI: 10.1101/2020.09.29.203525  
[Last updated: Sep 29, 2020]

**NOMAD**  
NOVEL MATERIALS DISCOVERY

**Introduction**  
In this tutorial we will show how to determine parameters given formulas that predict whether certain materials are topological or trivial insulators, using the example of tetradymites. It is based on the algorithm semi-independence screening and searching criterion (SISCS), that enables to search for optimal compositions of elements in a material system.

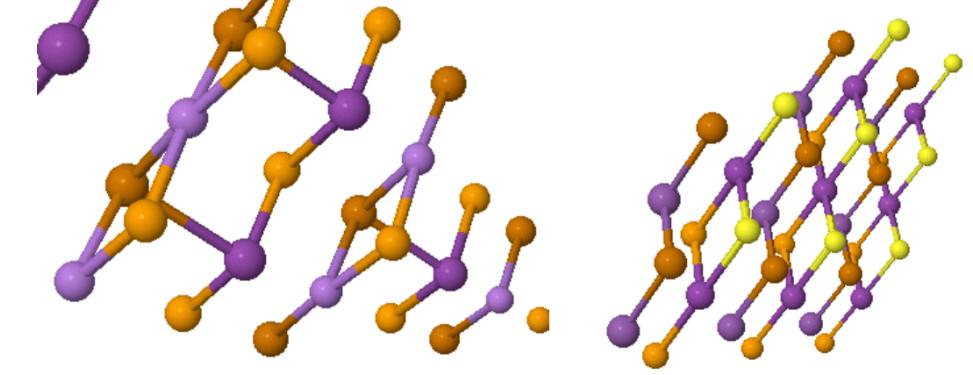
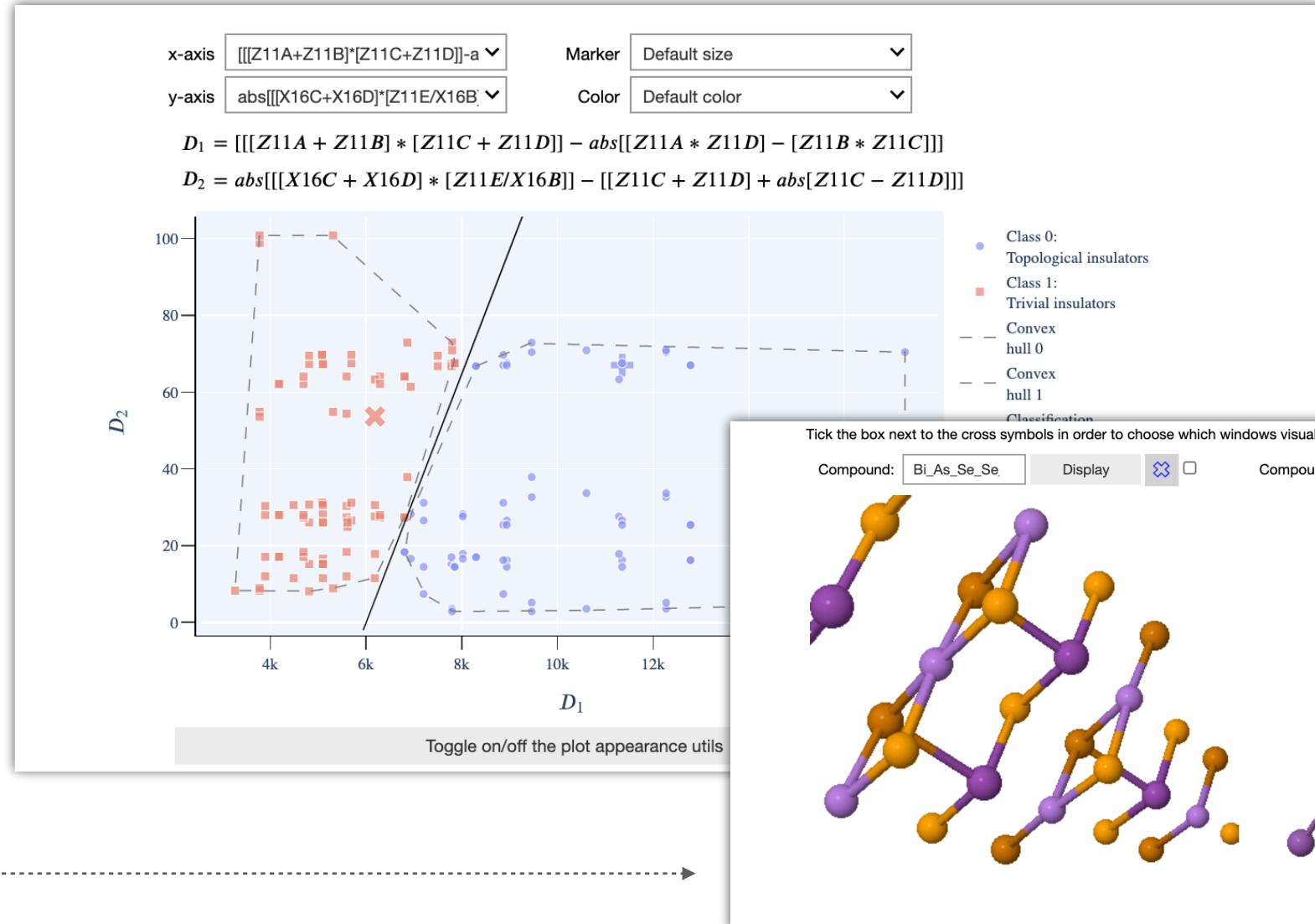
**Description of the method**  
The SISCS algorithm is a semi-independent screening and searching method for identifying the best low-dimensional topological insulators in an alloyed tetradymite. Phys. Rev. Materials 2, 034022 (2018) [\[link\]](#)  
With the default settings, the method reproduces the same results found in the paper.

**Comparison of the method with its implementation**  
The code demonstrated in this tutorial is to start from simple physical quantities ("primary features"), i.e., properties of the constituent atoms such as atomic radius, coordination number, and electronegativity, and to calculate the corresponding quantities in the so-called "feature space". Then, SISCS is used to select one or more of these variables that explain the data. By clicking directly on "Run" below, i.e., with no default selection, you can reproduce the 2D map as published in [\[Phys. Rev. Materials 2, 034022 \(2018\)\]](#). You can also select primary features by clicking on the checkboxes in the "Feature selection" section. The "Primary features" section contains the following checkboxes: "X", "Y", "Z", "X^2", "Y^2", "Z^2", "XY", "YZ", "ZX", "XYZ", "X^3", "Y^3", "Z^3", "XY^2", "YZ^2", "ZX^2", "XYZ^2", "X^4", "Y^4", "Z^4", "XY^3", "YZ^3", "ZX^3", "XYZ^3", "X^5", "Y^5", "Z^5", "XY^4", "YZ^4", "ZX^4", "XYZ^4", "X^6", "Y^6", "Z^6", "XY^5", "YZ^5", "ZX^5", "XYZ^5", "X^7", "Y^7", "Z^7", "XY^6", "YZ^6", "ZX^6", "XYZ^6", "X^8", "Y^8", "Z^8", "XY^7", "YZ^7", "ZX^7", "XYZ^7", "X^9", "Y^9", "Z^9", "XY^8", "YZ^8", "ZX^8", "XYZ^8", "X^10", "Y^10", "Z^10", "XY^9", "YZ^9", "ZX^9", "XYZ^9", "X^11", "Y^11", "Z^11", "XY^10", "YZ^10", "ZX^10", "XYZ^10", "X^12", "Y^12", "Z^12", "XY^11", "YZ^11", "ZX^11", "XYZ^11", "X^13", "Y^13", "Z^13", "XY^12", "YZ^12", "ZX^12", "XYZ^12", "X^14", "Y^14", "Z^14", "XY^13", "YZ^13", "ZX^13", "XYZ^13", "X^15", "Y^15", "Z^15", "XY^14", "YZ^14", "ZX^14", "XYZ^14", "X^16", "Y^16", "Z^16", "XY^15", "YZ^15", "ZX^15", "XYZ^15", "X^17", "Y^17", "Z^17", "XY^16", "YZ^16", "ZX^16", "XYZ^16", "X^18", "Y^18", "Z^18", "XY^17", "YZ^17", "ZX^17", "XYZ^17", "X^19", "Y^19", "Z^19", "XY^18", "YZ^18", "ZX^18", "XYZ^18", "X^20", "Y^20", "Z^20", "XY^19", "YZ^19", "ZX^19", "XYZ^19", "X^21", "Y^21", "Z^21", "XY^20", "YZ^20", "ZX^20", "XYZ^20", "X^22", "Y^22", "Z^22", "XY^21", "YZ^21", "ZX^21", "XYZ^21", "X^23", "Y^23", "Z^23", "XY^22", "YZ^22", "ZX^22", "XYZ^22", "X^24", "Y^24", "Z^24", "XY^23", "YZ^23", "ZX^23", "XYZ^23", "X^25", "Y^25", "Z^25", "XY^24", "YZ^24", "ZX^24", "XYZ^24", "X^26", "Y^26", "Z^26", "XY^25", "YZ^25", "ZX^25", "XYZ^25", "X^27", "Y^27", "Z^27", "XY^26", "YZ^26", "ZX^26", "XYZ^26", "X^28", "Y^28", "Z^28", "XY^27", "YZ^27", "ZX^27", "XYZ^27", "X^29", "Y^29", "Z^29", "XY^28", "YZ^28", "ZX^28", "XYZ^28", "X^30", "Y^30", "Z^30", "XY^29", "YZ^29", "ZX^29", "XYZ^29", "X^31", "Y^31", "Z^31", "XY^30", "YZ^30", "ZX^30", "XYZ^30", "X^32", "Y^32", "Z^32", "XY^31", "YZ^31", "ZX^31", "XYZ^31", "X^33", "Y^33", "Z^33", "XY^32", "YZ^32", "ZX^32", "XYZ^32", "X^34", "Y^34", "Z^34", "XY^33", "YZ^33", "ZX^33", "XYZ^33", "X^35", "Y^35", "Z^35", "XY^34", "YZ^34", "ZX^34", "XYZ^34", "X^36", "Y^36", "Z^36", "XY^35", "YZ^35", "ZX^35", "XYZ^35", "X^37", "Y^37", "Z^37", "XY^36", "YZ^36", "ZX^36", "XYZ^36", "X^38", "Y^38", "Z^38", "XY^37", "YZ^37", "ZX^37", "XYZ^37", "X^39", "Y^39", "Z^39", "XY^38", "YZ^38", "ZX^38", "XYZ^38", "X^40", "Y^40", "Z^40", "XY^39", "YZ^39", "ZX^39", "XYZ^39", "X^41", "Y^41", "Z^41", "XY^40", "YZ^40", "ZX^40", "XYZ^40", "X^42", "Y^42", "Z^42", "XY^41", "YZ^41", "ZX^41", "XYZ^41", "X^43", "Y^43", "Z^43", "XY^42", "YZ^42", "ZX^42", "XYZ^42", "X^44", "Y^44", "Z^44", "XY^43", "YZ^43", "ZX^43", "XYZ^43", "X^45", "Y^45", "Z^45", "XY^44", "YZ^44", "ZX^44", "XYZ^44", "X^46", "Y^46", "Z^46", "XY^45", "YZ^45", "ZX^45", "XYZ^45", "X^47", "Y^47", "Z^47", "XY^46", "YZ^46", "ZX^46", "XYZ^46", "X^48", "Y^48", "Z^48", "XY^47", "YZ^47", "ZX^47", "XYZ^47", "X^49", "Y^49", "Z^49", "XY^48", "YZ^48", "ZX^48", "XYZ^48", "X^50", "Y^50", "Z^50", "XY^49", "YZ^49", "ZX^49", "XYZ^49", "X^51", "Y^51", "Z^51", "XY^50", "YZ^50", "ZX^50", "XYZ^50", "X^52", "Y^52", "Z^52", "XY^51", "YZ^51", "ZX^51", "XYZ^51", "X^53", "Y^53", "Z^53", "XY^52", "YZ^52", "ZX^52", "XYZ^52", "X^54", "Y^54", "Z^54", "XY^53", "YZ^53", "ZX^53", "XYZ^53", "X^55", "Y^55", "Z^55", "XY^54", "YZ^54", "ZX^54", "XYZ^54", "X^56", "Y^56", "Z^56", "XY^55", "YZ^55", "ZX^55", "XYZ^55", "X^57", "Y^57", "Z^57", "XY^56", "YZ^56", "ZX^56", "XYZ^56", "X^58", "Y^58", "Z^58", "XY^57", "YZ^57", "ZX^57", "XYZ^57", "X^59", "Y^59", "Z^59", "XY^58", "YZ^58", "ZX^58", "XYZ^58", "X^60", "Y^60", "Z^60", "XY^59", "YZ^59", "ZX^59", "XYZ^59", "X^61", "Y^61", "Z^61", "XY^60", "YZ^60", "ZX^60", "XYZ^60", "X^62", "Y^62", "Z^62", "XY^61", "YZ^61", "ZX^61", "XYZ^61", "X^63", "Y^63", "Z^63", "XY^62", "YZ^62", "ZX^62", "XYZ^62", "X^64", "Y^64", "Z^64", "XY^63", "YZ^63", "ZX^63", "XYZ^63", "X^65", "Y^65", "Z^65", "XY^64", "YZ^64", "ZX^64", "XYZ^64", "X^66", "Y^66", "Z^66", "XY^65", "YZ^65", "ZX^65", "XYZ^65", "X^67", "Y^67", "Z^67", "XY^66", "YZ^66", "ZX^66", "XYZ^66", "X^68", "Y^68", "Z^68", "XY^67", "YZ^67", "ZX^67", "XYZ^67", "X^69", "Y^69", "Z^69", "XY^68", "YZ^68", "ZX^68", "XYZ^68", "X^70", "Y^70", "Z^70", "XY^69", "YZ^69", "ZX^69", "XYZ^69", "X^71", "Y^71", "Z^71", "XY^70", "YZ^70", "ZX^70", "XYZ^70", "X^72", "Y^72", "Z^72", "XY^71", "YZ^71", "ZX^71", "XYZ^71", "X^73", "Y^73", "Z^73", "XY^72", "YZ^72", "ZX^72", "XYZ^72", "X^74", "Y^74", "Z^74", "XY^73", "YZ^73", "ZX^73", "XYZ^73", "X^75", "Y^75", "Z^75", "XY^74", "YZ^74", "ZX^74", "XYZ^74", "X^76", "Y^76", "Z^76", "XY^75", "YZ^75", "ZX^75", "XYZ^75", "X^77", "Y^77", "Z^77", "XY^76", "YZ^76", "ZX^76", "XYZ^76", "X^78", "Y^78", "Z^78", "XY^77", "YZ^77", "ZX^77", "XYZ^77", "X^79", "Y^79", "Z^79", "XY^78", "YZ^78", "ZX^78", "XYZ^78", "X^80", "Y^80", "Z^80", "XY^79", "YZ^79", "ZX^79", "XYZ^79", "X^81", "Y^81", "Z^81", "XY^80", "YZ^80", "ZX^80", "XYZ^80", "X^82", "Y^82", "Z^82", "XY^81", "YZ^81", "ZX^81", "XYZ^81", "X^83", "Y^83", "Z^83", "XY^82", "YZ^82", "ZX^82", "XYZ^82", "X^84", "Y^84", "Z^84", "XY^83", "YZ^83", "ZX^83", "XYZ^83", "X^85", "Y^85", "Z^85", "XY^84", "YZ^84", "ZX^84", "XYZ^84", "X^86", "Y^86", "Z^86", "XY^85", "YZ^85", "ZX^85", "XYZ^85", "X^87", "Y^87", "Z^87", "XY^86", "YZ^86", "ZX^86", "XYZ^86", "X^88", "Y^88", "Z^88", "XY^87", "YZ^87", "ZX^87", "XYZ^87", "X^89", "Y^89", "Z^89", "XY^88", "YZ^88", "ZX^88", "XYZ^88", "X^90", "Y^90", "Z^90", "XY^89", "YZ^89", "ZX^89", "XYZ^89", "X^91", "Y^91", "Z^91", "XY^90", "YZ^90", "ZX^90", "XYZ^91", "X^92", "Y^92", "Z^92", "XY^91", "YZ^91", "ZX^91", "XYZ^92", "X^93", "Y^93", "Z^93", "XY^92", "YZ^92", "ZX^92", "XYZ^93", "X^94", "Y^94", "Z^94", "XY^93", "YZ^93", "ZX^93", "XYZ^94", "X^95", "Y^95", "Z^95", "XY^94", "YZ^94", "ZX^94", "XYZ^95", "X^96", "Y^96", "Z^96", "XY^95", "YZ^95", "ZX^95", "XYZ^96", "X^97", "Y^97", "Z^97", "XY^96", "YZ^96", "ZX^96", "XYZ^97", "X^98", "Y^98", "Z^98", "XY^97", "YZ^97", "ZX^97", "XYZ^98", "X^99", "Y^99", "Z^99", "XY^98", "YZ^98", "ZX^98", "XYZ^99", "X^100", "Y^100", "Z^100", "XY^99", "YZ^99", "ZX^99", "XYZ^99", "X^101", "Y^101", "Z^101", "XY^100", "YZ^100", "ZX^100", "XYZ^100", "X^102", "Y^102", "Z^102", "XY^101", "YZ^101", "ZX^101", "XYZ^102", "X^103", "Y^103", "Z^103", "XY^102", "YZ^102", "ZX^102", "XYZ^103", "X^104", "Y^104", "Z^104", "XY^103", "YZ^103", "ZX^103", "XYZ^104", "X^105", "Y^105", "Z^105", "XY^104", "YZ^104", "ZX^104", "XYZ^105", "X^106", "Y^106", "Z^106", "XY^105", "YZ^105", "ZX^105", "XYZ^106", "X^107", "Y^107", "Z^107", "XY^106", "YZ^106", "ZX^106", "XYZ^107", "X^108", "Y^108", "Z^108", "XY^107", "YZ^107", "ZX^107", "XYZ^108", "X^109", "Y^109", "Z^109", "XY^108", "YZ^108", "ZX^108", "XYZ^109", "X^110", "Y^110", "Z^110", "XY^109", "YZ^109", "ZX^109", "XYZ^109", "X^111", "Y^111", "Z^111", "XY^110", "YZ^110", "ZX^110", "XYZ^111", "X^112", "Y^112", "Z^112", "XY^111", "YZ^111", "ZX^111", "XYZ^112", "X^113", "Y^113", "Z^113", "XY^112", "YZ^112", "ZX^112", "XYZ^113", "X^114", "Y^114", "Z^114", "XY^113", "YZ^113", "ZX^113", "XYZ^114", "X^115", "Y^115", "Z^115", "XY^114", "YZ^114", "ZX^114", "XYZ^115", "X^116", "Y^116", "Z^116", "XY^115", "YZ^115", "ZX^115", "XYZ^116", "X^117", "Y^117", "Z^117", "XY^116", "YZ^116", "ZX^116", "XYZ^117", "X^118", "Y^118", "Z^118", "XY^117", "YZ^117", "ZX^117", "XYZ^118", "X^119", "Y^119", "Z^119", "XY^118", "YZ^118", "ZX^118", "XYZ^119", "X^120", "Y^120", "Z^120", "XY^119", "YZ^119", "ZX^119", "XYZ^119", "X^121", "Y^121", "Z^121", "XY^120", "YZ^120", "ZX^120", "XYZ^121", "X^122", "Y^122", "Z^122", "XY^121", "YZ^121", "ZX^121", "XYZ^122", "X^123", "Y^123", "Z^123", "XY^122", "YZ^122", "ZX^122", "XYZ^123", "X^124", "Y^124", "Z^124", "XY^123", "YZ^123", "ZX^123", "XYZ^124", "X^125", "Y^125", "Z^125", "XY^124", "YZ^124", "ZX^124", "XYZ^125", "X^126", "Y^126", "Z^126", "XY^125", "YZ^125", "ZX^125", "XYZ^126", "X^127", "Y^127", "Z^127", "XY^126", "YZ^126", "ZX^126", "XYZ^127", "X^128", "Y^128", "Z^128", "XY^127", "YZ^127", "ZX^127", "XYZ^128", "X^129", "Y^129", "Z^129", "XY^128", "YZ^128", "ZX^128", "XYZ^129", "X^130", "Y^130", "Z^130", "XY^129", "YZ^129", "ZX^129", "XYZ^129", "X^131", "Y^131", "Z^131", "XY^130", "YZ^130", "ZX^130", "XYZ^131", "X^132", "Y^132", "Z^132", "XY^131", "YZ^131", "ZX^131", "XYZ^132", "X^133", "Y^133", "Z^133", "XY^132", "YZ^132", "ZX^132", "XYZ^133", "X^134", "Y^134", "Z^134", "XY^133", "YZ^133", "ZX^133", "XYZ^134", "X^135", "Y^135", "Z^135", "XY^134", "YZ^134", "ZX^134", "XYZ^135", "X^136", "Y^136", "Z^136", "XY^135", "YZ^135", "ZX^135", "XYZ^136", "X^137", "Y^137", "Z^137", "XY^136", "YZ^136", "ZX^136", "XYZ^137", "X^138", "Y^138", "Z^138", "XY^137", "YZ^137", "ZX^137", "XYZ^138", "X^139", "Y^139", "Z^139", "XY^138", "YZ^138", "ZX^138", "XYZ^139", "X^140", "Y^140", "Z^140", "XY^139", "YZ^139", "ZX^139", "XYZ^139", "X^141", "Y^141", "Z^141", "XY^140", "YZ^140", "ZX^140", "XYZ^141", "X^142", "Y^142", "Z^142", "XY^141", "YZ^141", "ZX^141", "XYZ^142", "X^143", "Y^143", "Z^143", "XY^142", "YZ^142", "ZX^142", "XYZ^143", "X^144", "Y^144", "Z^144", "XY^143", "YZ^143", "ZX^143", "XYZ^144", "X^145", "Y^145", "Z^145", "XY^144", "YZ^144", "ZX^144", "XYZ^145", "X^146", "Y^146", "Z^146", "XY^145", "YZ^145", "ZX^145", "XYZ^146", "X^147", "Y^147", "Z^147", "XY^146", "YZ^146", "ZX^146", "XYZ^147", "X^148", "Y^148", "Z^148", "XY^147", "YZ^147", "ZX^147", "XYZ^148", "X^149", "Y^149", "Z^149", "XY^148", "YZ^148", "ZX^148", "XYZ^149", "X^150", "Y^150", "Z^150", "XY^149", "YZ^149", "ZX^149", "XYZ^149", "X^151", "Y^151", "Z^151", "XY^150", "YZ^150", "ZX^150", "XYZ^151", "X^152", "Y^152", "Z^152", "XY^151", "YZ^151", "ZX^151", "XYZ^152", "X^153", "Y^153", "Z^153", "XY^152", "YZ^152", "ZX^152", "XYZ^153", "X^154", "Y^154", "Z^154", "XY^153", "YZ^153", "ZX^153", "XYZ^154", "X^155", "Y^155", "Z^155", "XY^154", "YZ^154", "ZX^154", "XYZ^155", "X^156", "Y^156", "Z^156", "XY^155", "YZ^155", "ZX^155", "XYZ^156", "X^157", "Y^157", "Z^157", "XY^156", "YZ^156", "ZX^156", "XYZ^157", "X^158", "Y^158", "Z^158", "XY^157", "YZ^157", "ZX^157", "XYZ^158", "X^159", "Y^159", "Z^159", "XY^158", "YZ^158", "ZX^158", "XYZ^159", "X^160", "Y^160", "Z^160", "XY^159", "YZ^159", "ZX^159", "XYZ^159", "X^161", "Y^161", "Z^161", "XY^160", "YZ^160", "ZX^160", "XYZ^161", "X^162", "Y^162", "Z^162", "XY^161", "YZ^161", "ZX^161", "XYZ^162", "X^163", "Y^163", "Z^163", "XY^162", "YZ^162", "ZX^162", "XYZ^163", "X^164", "Y^164", "Z^164", "XY^163", "YZ^163", "ZX^163", "XYZ^164", "X^165", "Y^165", "Z^165", "XY^164", "YZ^164", "ZX^164", "XYZ^165", "X^166", "Y^166", "Z^166", "XY^165", "YZ^165", "ZX^165", "XYZ^166", "X^167", "Y^167", "Z^167", "XY^166", "YZ^166", "ZX^166", "XYZ^167", "X^168", "Y^168", "Z^168", "XY^167", "YZ^167", "ZX^167", "XYZ^168", "X^169", "Y^169", "Z^169", "XY^168", "YZ^168", "ZX^168", "XYZ^169", "X^170", "Y^170", "Z^170", "XY^169", "YZ^169", "ZX^169", "XYZ^169", "X^171", "Y^171", "Z^171", "XY^170", "YZ^170", "ZX^170", "XYZ^171", "X^172", "Y^172", "Z^172", "XY^171", "YZ^171", "ZX^171", "XYZ^172", "X^173", "Y^173", "Z^173", "XY^172", "YZ^172", "ZX^172", "XYZ^173", "X^174", "Y^174", "Z^174", "XY^173", "YZ^173", "ZX^173", "XYZ^174", "X^175", "Y^175", "Z^175", "XY^174", "YZ^174", "ZX^174", "XYZ^175", "X^176", "Y^176", "Z^176", "XY^175", "YZ^175", "ZX^175", "XYZ^176", "X^177", "Y^177", "Z^177", "XY^176", "YZ^176", "ZX^176", "XYZ^177", "X^178", "Y^178", "Z^178", "XY^177", "YZ^177", "ZX^177", "XYZ^178", "X^179", "Y^179", "Z^179", "XY^178", "YZ^178", "ZX^178", "XYZ^179", "X^180", "Y^180", "Z^180", "XY^179", "YZ^179", "ZX^179", "XYZ^179", "X^181", "Y^181", "Z^181", "XY^180", "YZ^180", "ZX^180", "XYZ^181", "X^182", "Y^182", "Z^182", "XY^181", "YZ^181", "ZX^181", "XYZ^182", "X^183", "Y^183", "Z^183", "XY^182", "YZ^182", "ZX^182", "XYZ^183", "X^184", "Y^184", "Z^184", "XY^183", "YZ^183", "ZX^183", "XYZ^184", "X^185", "Y^185", "Z^185", "XY^184", "YZ^184", "ZX^184", "XYZ^185", "X^186", "Y^186", "Z^186", "XY^185", "YZ^185", "ZX^185", "XYZ^186", "X^187", "Y^187", "Z^187", "XY^186", "YZ^186", "ZX^186", "XYZ^187", "X^188", "Y^188", "Z^188", "XY^187", "YZ^187", "ZX^187", "XYZ^188", "X^189", "Y^189", "Z^189", "XY^188", "YZ^188", "ZX^188", "XYZ^189", "X^190", "Y^190", "Z^190", "XY^189", "YZ^189", "ZX^189", "XYZ^189", "X^191", "Y^191", "Z^191", "XY^190", "YZ^190", "ZX^190", "XYZ^191", "X^192", "Y^192", "Z^192", "XY^191", "YZ^191", "ZX^191", "XYZ^192", "X^193", "Y^193", "Z^193", "XY^192", "YZ^192", "ZX^192", "XYZ^193", "X^194", "Y^194", "Z^194", "XY^193", "YZ^193", "ZX^193", "XYZ^194", "X^195", "Y^195", "Z^195", "XY^194", "YZ^194", "ZX^194", "XYZ^195", "X^196", "Y^196", "Z^196", "XY^195", "YZ^195", "ZX^195", "XYZ^196", "X^197", "Y^197", "Z^197", "XY^196", "YZ^196", "ZX^196", "XYZ^197", "X^198", "Y^198", "Z^198", "XY^197", "YZ^197", "ZX^197", "XYZ^198", "X^199", "Y^199", "Z^199", "XY^198", "YZ^198", "ZX^198", "XYZ^199", "X^200", "Y^200", "Z^200", "XY^199", "YZ^199", "ZX^199", "XYZ^199", "X^201", "Y^201", "Z^201", "XY^200", "YZ^200", "ZX^200", "XYZ^201", "X^202", "Y^202", "Z^202", "XY^201", "YZ^201", "ZX^201", "XYZ^202", "X^203", "Y^203", "Z^203", "XY^202", "YZ^202", "ZX^202", "XYZ^203", "X^204", "Y^204", "Z^204", "XY^203", "YZ^203", "ZX^203", "XYZ^204", "X^205", "Y^205", "Z^205", "XY^204", "YZ^204", "ZX^204", "XYZ^205", "X^206", "Y^206", "Z^206", "XY^205", "YZ^205", "ZX^205", "XYZ^206", "X^207", "Y^207", "Z^207", "XY^206", "YZ^206", "ZX^206", "XYZ^207", "X^208", "Y^208", "Z^208", "XY^207", "YZ^207", "ZX^207", "XYZ^208", "X^209", "Y^209", "Z^209", "XY^208", "YZ^208", "ZX^208", "XYZ^209", "X^210", "Y^210", "Z^210", "XY^209", "YZ^209", "ZX^209", "XYZ^209", "X^211", "Y^211", "Z^211", "XY^210", "YZ^210", "ZX^210", "XYZ^211", "X^212", "Y^212", "Z^212", "XY^211", "YZ^211", "ZX^211", "XYZ^212", "X^213", "Y^213", "Z^213", "XY^212", "YZ^212", "ZX^212", "XYZ^213", "X^214", "Y^214", "Z^214", "XY^213", "YZ^213", "ZX^213", "XYZ^214", "X^215", "Y^215", "Z^215", "XY^214", "YZ^214", "ZX^214", "XYZ^215", "X^216", "Y^216", "Z^216", "XY^215", "YZ^215", "ZX^215", "XYZ^216", "X^217", "Y^217", "Z^217", "XY^216", "YZ^216", "ZX^216", "XYZ^217", "X^218", "Y^218", "Z^218", "XY^217", "YZ^217", "ZX^217", "XYZ^218", "X^219", "Y^219", "Z^219", "XY^218", "YZ^218", "ZX^218", "XYZ^219", "X^220", "Y^220", "Z^220", "XY^219", "YZ^219", "ZX^219", "XYZ^219", "X^221", "Y^221", "Z^221", "XY^220", "YZ^220", "ZX^220", "XYZ^221", "X^222", "Y^222", "Z^222", "XY^221", "YZ^221", "ZX^221", "XYZ^222", "X^223", "Y^223", "Z^223", "XY^222", "YZ^222", "ZX^222", "XYZ^223", "X^224", "Y^224", "Z^224", "XY^223", "YZ^223", "ZX^223", "XYZ^224", "X^225", "Y^225", "Z^225", "XY^224", "YZ^224", "ZX^224", "XYZ^225", "X^226", "Y^226", "Z^226", "XY^225", "YZ^225", "ZX^225", "XYZ^226", "X^227", "Y^227", "Z^227", "XY^226", "YZ^226", "ZX^226", "XYZ^227", "X^228", "Y^228", "Z^228", "XY^227", "YZ^227", "ZX^227", "XYZ^228", "X^229", "Y^229", "Z^229", "XY^228", "YZ^228", "ZX^228", "XYZ^229", "X^230", "Y^230", "Z^230", "XY^229", "YZ^229", "ZX^229", "XYZ^229", "X^231

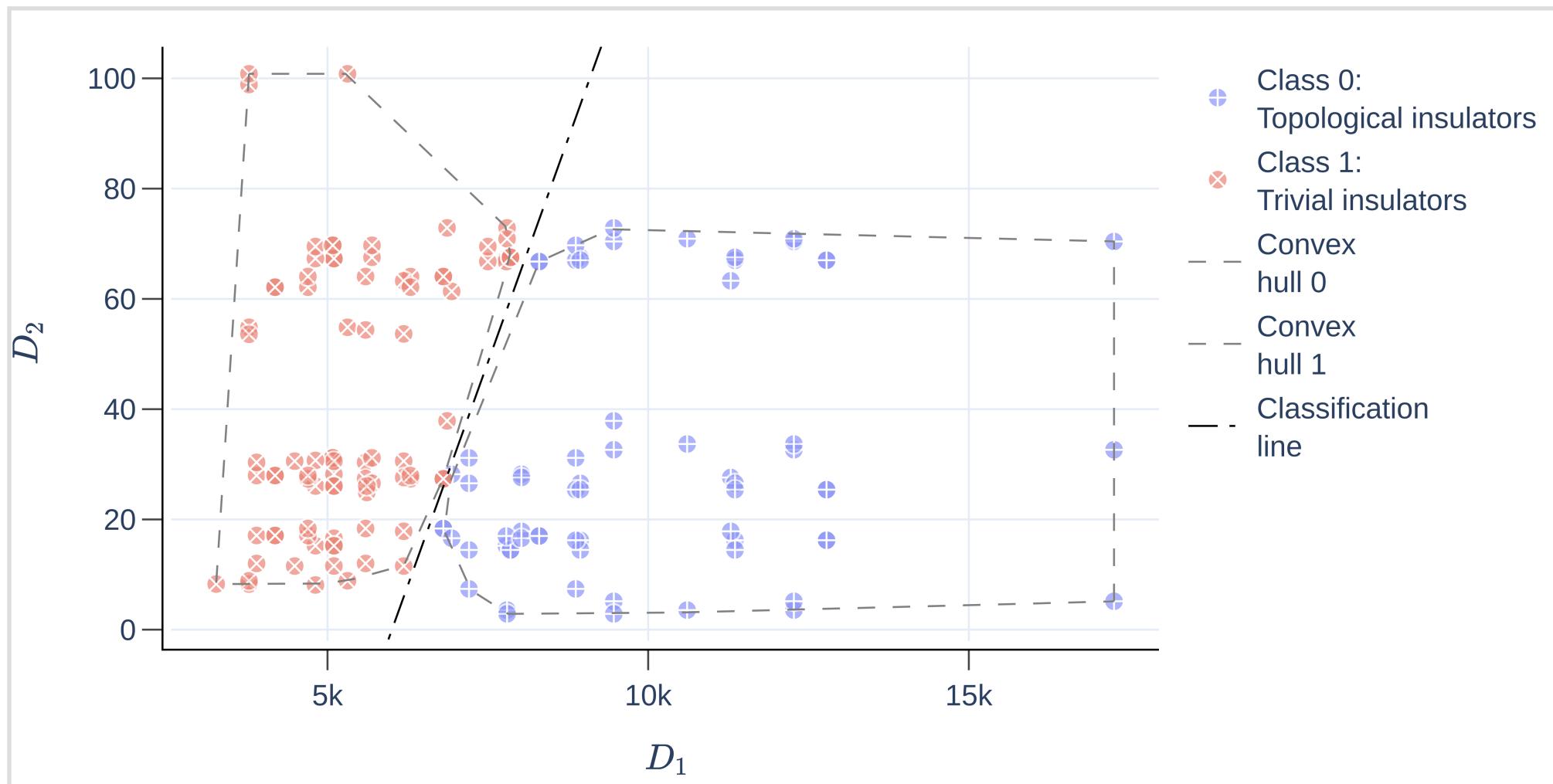
# Reproducing published results



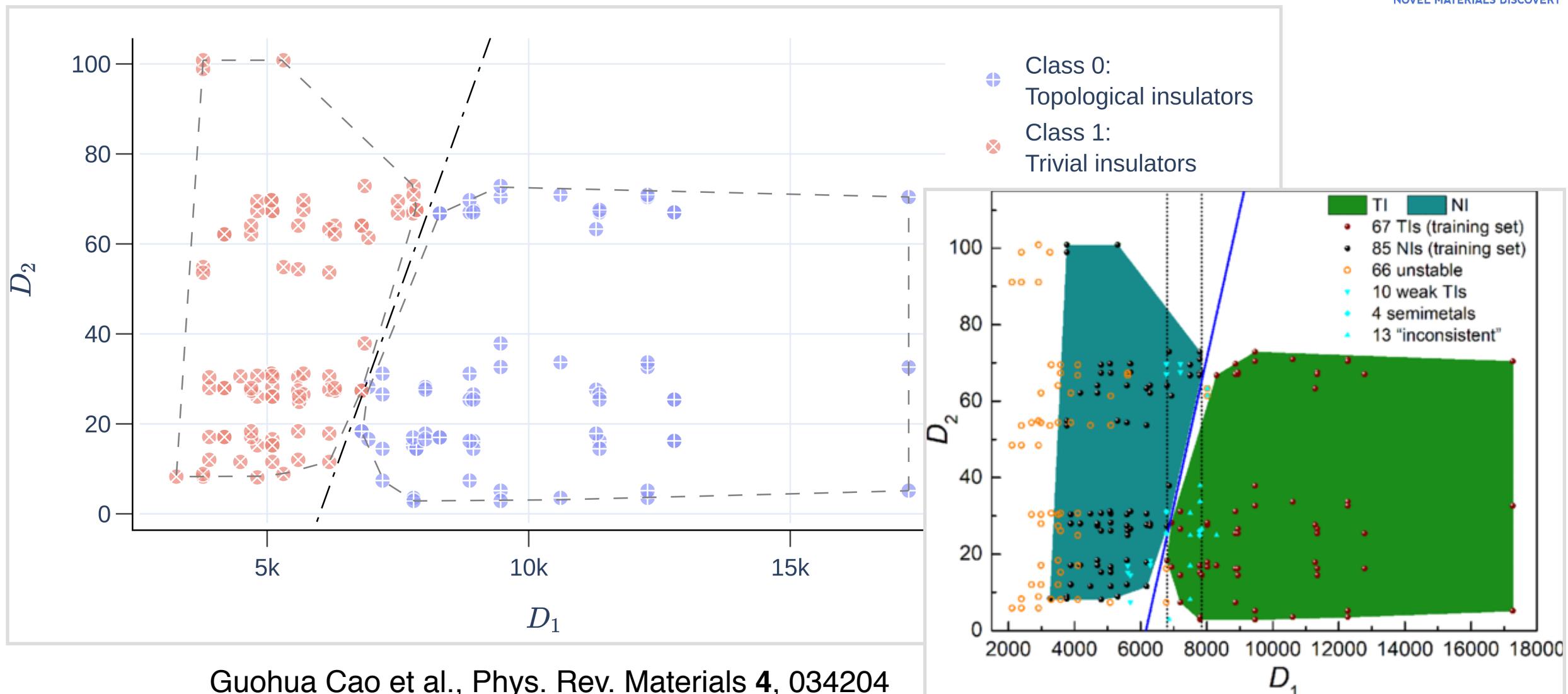
## NOVEL MATERIALS DISCOVERY



# Reproducing published results



# Reproducing published results



Guohua Cao et al., Phys. Rev. Materials 4, 034204

## Publish

- Publish your data with our without embargo, get a DOI, and share data with others.
- We support input and output files of most electronic-structure codes.
- Watch our [video tutorial](#) on how to upload and publish data.

[UPLOAD](#)

## Explore

- Search for [materials](#) (Encyclopedia) or [calculations](#) (Repository).
- All raw and processed data can be downloaded and used under the [CC BY 4.0](#).
- Watch our [video tutorials](#) on how to use the Encyclopedia and Repository.

[MATERIALS](#) [CALCULATIONS](#)

## Analyze

- Analyze data with [Jupyter notebooks](#) directly on NOMAD servers (Artificial Intelligence (AI) Toolkit).
- Access all data programmatically via [NOMAD API](#) or [OPTIMADE API](#).
- Watch our [video tutorial](#) on how to use the NOMAD API.

[AI TOOLKIT TUTORIALS](#)

There is a new version of NOMAD (1.0) that we currently provide as a beta version. This installation contains most of NOMAD's data and you can already use it to upload and publish more data. Eventually all data will be migrated to this version. It will become the official NOMAD after a short beta phase. We also provide an empty test version of NOMAD. You can use this to try the upload and publish process without any consequences. We will routinely void the test data.

[NOMAD 1.0 BETA](#) [NOMAD 1.0 TEST](#)

# Querying the NOMAD Archive

PUBLISH ▾ EXPLORE ▾ ANALYZE ▾ ABOUT ▾

Entries search ↻

2,149,963 RESULTS

FILTERS C X < > :

Material >

Elements / Formula >

Elements 0 

Symmetry >

Method >

Simulation >

DFT >

GW >

Experiment >

EELS >

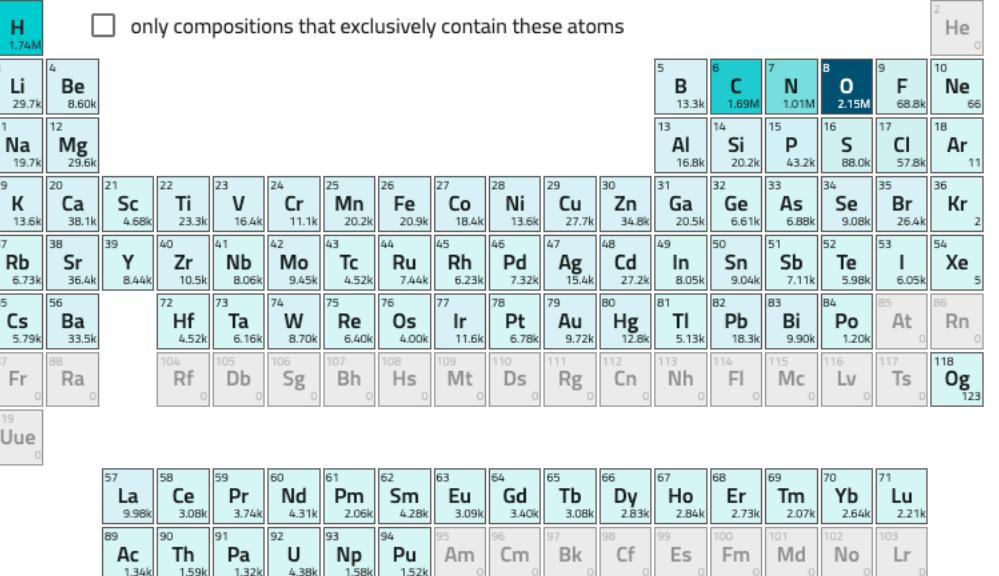
Properties >

Elements / Formula

Elements

only compositions that exclusively contain these atoms

linear ▾ +



# Querying the NOMAD Archive



The screenshot shows the NOMAD software interface. At the top, there are navigation tabs: PUBLISH, EXPLORE, ANALYZE, and ABOUT. Below these is a search bar labeled "Entries search" with a question mark icon. A red arrow points to the search bar, and a red circle highlights the search icon (a magnifying glass) next to it. To the right of the search bar is a back arrow and a help icon.

The main area has a header "2,149,963 RESULTS" and a section titled "ELEMENTS / FORMULA". On the left, there is a sidebar with filter categories: Material, Elements / Formula, Elements, Symmetry, Method, Simulation, DFT, GW, Experiment, EELS, and Properties. The "Elements / Formula" filter is currently selected, indicated by a blue border. A red arrow points to the "Elements / Formula" filter icon. The "Elements" filter is also highlighted with a blue border. The "Elements" section displays a periodic table where each element cell contains its symbol, name, and atomic mass. An unchecked checkbox labeled "only compositions that exclusively contain these atoms" is positioned above the table. The table is presented in a grid format with 18 columns and 10 rows. The last column and row are partially cut off. The entire interface has a light blue background with white text and icons.

# Querying the NOMAD Archive

**API Code**

URL to this query on the repository API:

```
http://nomad-lab.eu/prod/rae/api/repo?dft.xc_functional=GGA&dft.compound_type=ternary&dft.crystal_system=cubic&dft.code_name=VASP&atoms=>0
```

Access the archive as JSON via curl:

```
curl "http://nomad-lab.eu/prod/rae/api/archive/download?page=1&per_page=10&order_by=upload_time&order=-1&domain=dft&query=pub1&dft.xc_f
```

Access the archive in python:

```
import requests
response = requests.post('http://nomad-lab.eu/prod/rae/api/archive/query', json={
    'query': {
        'domain': 'dft',
        'dft.xc_functional': 'GGA',
        'dft.compound_type': 'ternary',
        'dft.crystal_system': 'cubic',
        'dft.code_name': 'VASP',
        'atoms': ['O']
    }
})
data = response.json()
```

Access the archive with the *NOMAD client library*:

```
from nomad import client, config
config.client.url = 'http://nomad-lab.eu/prod/rae/api'
results = client.query_archive(query={
    'domain': 'dft',
    'dft.xc_functional': 'GGA',
    'dft.compound_type': 'ternary',
    'dft.crystal_system': 'cubic',
    'dft.code_name': 'VASP',
    'atoms': ['O']})
print(results)
```

# Querying the NOMAD Archive

## Powerful Artificial-Intelligence Tools for Materials Science

Find New Patterns and Information in Materials-Science Big Data



[Query the archive →](#)

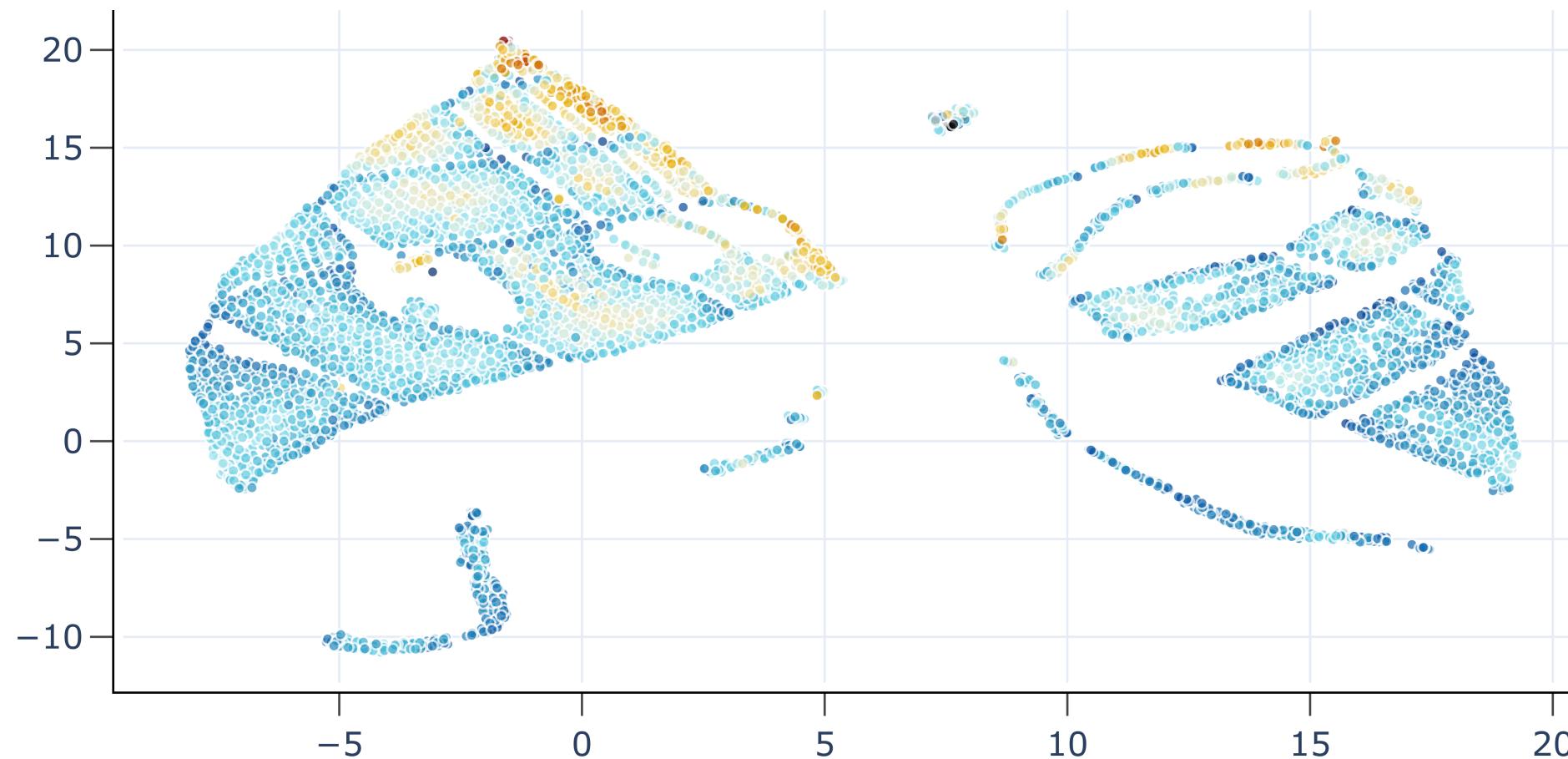
[View tutorials →](#)

[Reproduce results →](#)

[Get to work →](#)

(i)

# Querying the NOMAD Archive



2-dimensional embedding, found with umap. Different colors means different atomic concentration values.

# Querying the NOMAD Archive

## Powerful Artificial-Intelligence Tools for Materials Science

Find New Patterns and Information in Materials-Science Big Data



[Query the archive →](#)

[View tutorials →](#)

[Reproduce results →](#)

[Get to work →](#)

ⓘ

# AI tutorials for materials science



## Learn from tutorials

We develop and implement methods that identify correlations and structure in big data of materials. This will enable scientists and engineers to decide which materials are useful for specific applications or which new materials should be the focus of future studies. The following BEGINNER and INTERMEDIATE LEVEL tutorials are designed to get started with the AI Toolkit.



### Filter Tutorials

### BEGINNER LEVEL

**Symbolic regression via compressed sensing: a tutorial**

Authors: Emre Ahmetcik | Angelo Ziletti | Runhai Ouyang | Luigi Sbailò | Matthias Scheffler | Luca M. Ghiringhelli

**Introduction to decision-trees methods**

Authors: Daniel Speckhard | Andreas Leitherer | Luca M. Ghiringhelli

**Introduction to clustering**

Authors: Luigi Sbailò | Luca M. Ghiringhelli

- Tutorials span the most popular artificial-intelligence methods, such as clustering, dimension reduction, symbolic regression, decision trees, random forest, kernel ridge regression and deep neural networks.
- Applications to materials science are shown in a preparatory manner.
- Hands on exercises are presented throughout tutorials.
- Introductory videos, prepared in occasion of workshops and lectures held at the NOMAD Laboratory, are available.

# Querying the NOMAD Archive

## Powerful Artificial-Intelligence Tools for Materials Science

Find New Patterns and Information  
in Materials-Science Big Data



[Query the archive →](#)



[View tutorials →](#)



[Reproduce results →](#)



[Get to work →](#)

ⓘ

# Contribute

The *Artificial-Intelligence Toolkit* hosts tutorials for hands-on learning of the most popular AI tools.

The *NOMAD Archive* can be queried from the *Artificial-Intelligence Toolkit*, and the data retrieved can be analysed on the fly with the desired AI tools.

In the *NOMAD Artificial-Intelligence Toolkit*, we maintain several notebooks demonstrating the latest applications of AI to materials science data.

## Contact us!

[ghiringhelli@fhi.mpg.de](mailto:ghiringhelli@fhi.mpg.de) [scheffler@fhi.mpg.de](mailto:scheffler@fhi.mpg.de)

# FAIRMAT HANDS-ON TUTORIAL SERIES

[www.fair-di.eu/fairmat-tutorials-3](http://www.fair-di.eu/fairmat-tutorials-3)



FAIR Data Infrastructure  
for Physics, Chemistry,  
Materials Science,  
and Astronomy e.V.

About

Pillars

Events

## FAIRMAT HANDS-ON TUTORIAL SERIES

[Overview](#) [Registration](#) [Tutorial 1](#) [Tutorial 2](#) [Tutorial 3](#) [Tutorial 4](#) [Tutorial 5](#)

### TUTORIAL 3: INTRODUCTION TO THE ARTIFICIAL-INTELLIGENCE TOOLKIT

#### Overview Talk

Apr 6, 15:00 CEST

*Sergei V. Kalinin* (The University of Tennessee, Knoxville): Bayesian Optimization, structured Gaussian processes, and hypothesis learning for materials and physical discovery

#### Hands-on Tutorial

Apr 6, 15:45 CEST

*Luca Ghiringhelli* (FAIRmat): Overview of the artificial-intelligence toolkit

Apr 6, 16:30 CEST

*Luigi Sbailò* (FAIRmat): Querying the Archive, performing artificial-intelligence analysis, and introducing the AI-toolkit local app

Apr 6, 17:15 CEST

*Luca Ghiringhelli* (FAIRmat): Learning artificial-intelligence tools, reproducing publication, overview of selected tutorials and exercises assignment

# Acknowledgments



Luca Ghiringhelli



Matthias Scheffler

Claudia Draxl, Christian Carbogno, Markus Scheidgen, Markus Rapp, Gian-Marco Rignanese, James Kermode, Gábor Csányi, Ádám Fekete, Thomas A. R. Purcell, Mohammad-Yasin Arif, Lauri Himanen, Alvin Noe Lavines, Aakash Ashok Naik, Emre Ahmetcik, Andreas Leitherer, Daniel Speckhard, Marcel F. Langer, Benjamin Regler, Fawzi Mohamed, Angelo Ziletti