

To ease the reading, our responses to the reviewers will be written in royal blue, and the actions taken will be in *italic*.

## 1 Answers to Referee #1

First of all, thanks for the thorough review of the article.

At the beginning of Sec. 3.4, authors define the results of the workflows as “solid”. However, e.g. in the case of MER, 88% of the VASP relaxations just work without any need of error management, and the workflow only deals with 0.1% more. In many cases (the remaining  $\approx 23\%$ ) still there is the need of manual preparation. Therefore, this strikes me as suprising, since one of the focuses of the paper is to describe how the described platform can remove human intervention. In this case, human intervention is still needed, and it is not even significantly reduced from when VASP itself would require it (only 0.1% of the total cases, with 23% of cases where manual preparation is anyway needed). I’d like to stress that I understand that humans are helped in creating the input cells; still, the results don’t strike me as “solid”, and also the authors acknowledge in the abstract that performance is only “good” or even just “decent”. Therefore, I don’t believe that, for instance, the sentence in the conclusion “Our framework has proven to successfully automate two different...” is accurately describing the advantage of the framework over VASP itself

**Author:** We have softened the “solid” statement by a “reasonable”, considering that many of the calculations correspond to transition states and the difficulties and amount of time that typically it is employed in located these important points, we do consider that the time reduction is significant, although we can agree that more improvements are needed. This is also the main reason why the code is not yet available (only upon request) as we would like to optimize it further as the Reviewer indicates. In addition, while we still need some degree of human intervention to deal with a considerable number of structures, automation of the whole process brings three main benefits:

- Reduction of error prone calculation set up.
- Inheriting the structure usually halves the computing time.

- Automatic error checking procedure allow to implement a filter without need to check the structures one by one.

This is crucial for massive reaction networks involving more than one metal and when different materials (alloys) are investigated. *We have softened our claim and now indicate that this is a good starting point for automatization of complex reactions (end Section 3.4 and Conclusions).*

In addition to the point above, one way they mention they used to achieve convergence is to replace the metal slab with another one. But isn't this a different system? What if I really want to simulate that specific material?

**Author:** *It seems we were not sufficiently clear in this point. We employ a seed calculated for a particular metal and take this one as reference for the rest of the calculations of the same intermediate.*

Authors mention that putting data on ioChem-BD makes their research FAIR. However, I could only find less than 10 systems in the database, while in Fig. 3 they report over 300 runs. Do they intend to make this data public to make the paper really FAIR? Otherwise, this is just a proof-of-concept demonstration but not really a FAIR paper.

**Author:** *In our previous version only the crucial calculations were upload (those corresponding to the seed). Now, the embargoed link contains all the data corresponding to the calculations presented. All the calculation are now uploaded to ioChem-BD accessible through an embargoed link.*

It is not clear to me what amount of reproducibility the ioChem-BD guarantees. Can the authors describe advantages and limitations of the database? For instance:

1. what is available to allow an external researcher to reproduce the simulation, and what is not?

**Author:** *All the inputs used are available as well as the VASP and pseudopotential version used for every calculation.*

2. Are input files of VASP available (I think some of them are available, but some only in parsed form like the initial coordinates?);

**Author:** *Yes, the raw inputs are available for all the calculations.*

3. Are output files of VASP available (only a parsed .cml is provided? are raw outputs available? Is it possible to add a link to the CML specifications/schema? Is it possible to provide information on the code and version that performed the parsing?)

**Author:** No, the output is always parsed to a custom .cml format that contains all the pertinent data. Keeping the full output is precisely what ioChem-BD tries to avoid.

4. Is it possible to get inputs and outputs also for the other computational steps (I think only the final ones of VASP are provided?)

**Author:** To save space, the electronic and ionic steps except for the converge structure are not uploaded to ioChem-BD. Parsing subroutine is a part of ioChem-BD.

5. Is it possible to retrieve information on how the inputs of the simulations were obtained? (e.g. if the input of VASP was obtained by a relaxation, or the simulation was the restart of another one, is this specified somewhere?)

**Author:** We did not do it but it can be easily implemented in the script that we used to upload the data to the ioChem-BD platform.

**Author:** *We have added additional information at the end of Section 3.3 on the structure and properties of ioChem -BD that clarifies all these particular aspects.*

Also, at the end of Sec. 3.4 they speak about NEB calculations - is it possible to inspect them and see the results?

**Author:** No, at this moment the visualization of NEBs is not implemented in the platform. Developers are working on these features.

I believe that the paper requires an overall revision for what concerns the use of the English language.

- There are quite a few grammar mistakes (e.g. “Our framework show” instead of “Our framework shows” in the abstract, “that can be search” instead of “that can be searched” at the end of Sec. 1, 2 “a Gamma-centered mesh have” instead of “a Gamma-centered mesh has” in Sec. 2, “Niquel” instead of “Nickel” in one of the captions of Fig. 1, etc. (there are quite a few more occurrences later).

- Moreover, I never found (or could find) the use of the term “avoidhuman”, that also sounds to be as having a negative connotation, and I would suggest therefore to replace with some other term (“automation”?)
- In addition, sometimes the use of wording is unusual or incorrect, and in some cases I feel that it make it hard to understand the actual meaning of some sentences. I report here some examples: “infinite xyz coordinate listing” in Sec. 1 (I guess they mean “very long” rather than infinite); the mention in the abstract that the framework performs an “experimental” procedure is very confusing (I understood much later that it is instead a computational paper, and it is not describing an experimental protocol); some sentences are long and not clear, like in Sec. 1 “As the applications grow and the access to massive computers and robust codes extends worldwide structural data, spectroscopic fingerprints, general properties can be generated as databases for molecules, nanostructures and materials.” or in Sec. 2 “All the intermediates belong to the same reaction network, being the transition states all the possible elemental steps involving the intermediates.”.

**Author:** *The writing / English has been revised thoroughly. The grammar has been revised. As for the avoid-human, we agree that automatization is a conceptually finest term. However, as we have been the pioneers in the use of the Authorea platform the tag was almost converted in a private joke. Therefore, we prefer to ask the Editor if he finds this use offensive.*

At the end of Sec. 3.1, authors say “After a few tests, further improvements were integrated to the transfer algorithm.” but it is not clear in detail which improvements were integrated, and the technical details of these (i.e., it is not possible with the information provided to try to reproduce their results).

**Author:** *We performed a bond analysis using the structures generated for the MER workflows. This bond analysis lead to a correction of the bond distance between different atoms and the metal surface. We applied this correction to the TS workflows to improve the convergence. The end of Section Workflow Design (3.2) has been modified to clarify this improvement.*

Is the code described in the paper available somewhere? In order to have a really “FAIR” and reusable dataset, it would be important to be able to

rerun the same simulations/workflow.

**Author:** Our code is still in beta version and only works with our particular setup and machines. Implementations are needed for instance for its transfer to the shared supercomputer facilities. We fully agree with the Reviewer that the code needs to be made public. However, at present we still consider that documentation is suboptimal, some of the routines need to be improved and the adaptability to different platforms does not allow us to be fully confident. Meanwhile, the code will be available only upon request.

In Fig. 2, it would be nice to be able to compare at the same time the three structures coming out of the ‘transfer’ step in the TS branch (now one has to do two/three clicks to switch, and cannot show them at the same time, and it is therefore hard to compare them in detail)

**Author:** We consider that this will only add complexity to the figure. Instead, we suggest the Reviewer to open the ioChem-BD structures in different tabs for a proper comparison.

In Fig. 3, the use of an interactive graph is not providing any additional information or advantage. Authors might want to improve this by, e.g., providing additional information in an interactive way (e.g. by providing the materials belonging to a given part of the graph on hover, at least for the orange and red bars, or in some other way).

**Author:** Figure 3 conveys the information we aimed at. Of course extra implementations could be added but we think the interactive nature is already justified.

Can authors describe how much the results of Fig. 3 (blue bars) are actually the results of simple VASP simulations, and how much instead possible VASP crashes are (maybe?) already partially automatically corrected thanks to the use of fireworks?

**Author:** "Additional Steps Required" refers to the crashes solved with error checking, while "Good" refers to flawless VASP calculation. *The caption of Fig. 3 is now changed to clarify this information.*

In Fig. 2, the caption mentions squares and circles. Actually authors might want to refer to rectangles and rectangles with rounded corners, or change the shapes.

**Author:** *The caption of Fig. 2 now refers to color instead of shape.*

Links to “source calculation” in the pop-ups of Fig. 2 do not work (at least for me).

**Author:** Authorea blocks external pop-ups, we suggest the Reviewer to open the link in a new tab.

It is not clear that, in order to access the ioChemDB, one first needs to click on the “Data link” link at the beginning of 3.2, and then access to the embargoed resources is possible. It took quite a while to me to figure it out (I don’t know if this is a choice of the authors, or a technical feature of Authorea).

**Author:** The article will contain public links. Reviewers need to unlock the embargoed dataset with a link to access the data only temporarily.

In the paper, the ‘transition state search’ is sometimes referred to as “TS”, some other times as “TSS”. Please make uniform.

**Author:** *Now only "TS" label is used.*

In caption of Fig. 1, authors speak about “the first 5 steps”, but it is not clear which ones they are - numbers from 1 to 5, including subletters? the top 5 boxes (numbers from 1 to 4)? Something else? Can authors describe in more detail how (or with which method) adsorption states are identified in step (4) of Fig. 1?

**Author:** *Caption of Fig. 1 is modified clarify the steps.*

It is interesting that it is possible to publish data on ioChem-BD with an embargo, and access for reviewers. Can authors explain what is the procedure to make the data public, at which state of publication this will happen, who is in charge of this (authors, the ioChem-BD platform maintainers, ...) and who can guarantee that this will actually happen before publication? E.g., should the Authorea staff not make the paper public (after acceptance) until data is not made public?

**Author:** This question has two different sides. From the authors point of view, and in our particular case, we are fully committed to open science and open data and we have already made public more than 150 datasets for the activities of our group since 2015. From the Editors point of view, our recommendation would be that they shall be the responsables to ensure that the data is public under the Authorea platform.

## 2 Answers to Referee #2

Thank you for your revision, in response to your questions:

There are many cases of automated computational frameworks - indeed this uses Fireworks, which has been used by the Materials Project. There does not seem to be much in terms of comparing this specific effort to others. For example, a quick search turns up work by (Montoya & Persson, 2017)

**Author:** We would like to thank the Reviewer for providing us with the reference. We have surveyed the literature and included a paragraph in the Introduction where the previous initiatives regarding automatization for adsorption have been carried out. *Additional bibliography has been added to the introduction to give a vision about the state of automation in heterogeneous catalysis*

The authors note automated upload to the ioChem archive and the use of embargo/review processing. These are great, although it's not clear whether the archive itself is used in the algorithm outlined in the work. Does the algorithm "learn" from successful relaxations or transition state searches?

**Author:** In its present version, our algorithm do not contain a statical learning subroutine, we are already working in this kind of implementation.

The authors mention comparisons to graph theory / SMILES - but it's not clear that the work here generates a graph or descriptor. If I use either of these methods, I can generate a particular species or even geometry. Does this work generate a unique descriptor / key like SMILES for accessing the data later?

**Author:** Graph theory is used to detect possible bond-breakages at the end of a calculation, during the checking step but still in the input layer. We convert the adsorbate into a graph to check if there are more than one not connected molecules. The SMILES label is generated by ioChem-BD and used internally as fingerprint and substructure search when the dataset becomes public. *Additional information has been added to Sections 3.2 and 3.3 to clarify these points.*

Do you need to know the specific database record number?

**Author:** Every dataset is associated with a DOI and every item into the dataset has a URL identifier. You can access to the dataset with this DOI

or using the search tools integrated in ioChem-BD (Author, date, SMILES, chemical formula, etc.). *Additional information has been added to section 3.3.*

How do you get that from ioChem?

**Author:** Dataset main page contains the DOI identifier and item main page contains the URL identifier at the bottom of the page. *We have expanded the storage data Section (3.3) with additional information of the ioChem-BD features.*

There are many cases in the text where specific details are omitted. Several of these are listed below - but as it stands, I cannot learn much from the article because the details required for reproducibility are lacking. What are these improvements?

**Author:** We performed a bond analysis using the structures generated for the MER workflows. This bond analysis lead to a correction of the bond distance between different atoms and the metal surface. We applied this correction to the TS workflows to improve the convergence. *The end of Section Workflow Design (3.2) has been modified to clarify this improvement.*

What are the specifics of the bond analysis mentioned?

**Author:** We analyzed the differences between the adsorption distances of different metals/elements of the MER workflows, and then applied the correction to the TS workflows. *The end of the Section Workflow Design (3.2) has been modified to clarify this improvement.*

What are the specific corrections to the bonding geometry between the molecule and the slab?

**Author:** These corrections were calculated as the difference between the average adsorption distance of the element of the lowest atom of the molecule and the surface for the host and guest slabs.

The authors note efforts to automate preparation, transfer and checking, etc. It would be useful to detail some of the effort involved in these and link to the actual scripts/code, which would likely be useful to a wide number of readers, especially those also using VASP.

**Author:** Our code is still in beta version and only works with our particular setup and machines. Implementations are needed for instance for its



transfer to the shared supercomputer facilities. We fully agree with the Reviewer that the code needs to be made public. However, at present we still consider that documentation is suboptimal, some of the routines need to be improved and the adaptability to different platforms does not allow us to be fully confident. Meanwhile, the code will be available only upon request.

The authors also note general success with the automated work, but very little about the specifics. For example, while more optimization steps are noted as required in some cases, the authors don't mention how many were in the initial processing, nor how many additional steps were required later. In all cases, the descriptions in this section are merely qualitative commentary, not a discussion of how many cases fit each pattern (e.g., broken bond, migration to a different adsorption site, etc.)

**Author:** These are the most common errors during the workflows:

- Maximum number of ionic steps reached. The calculations needed additional steps to converge. With an addition of 10% of ionic steps, the geometries converged correctly.
- Bond breakage of highly reactive species. e.g: We were unable to obtain  $CH_2Br$  over Fe.
- Migration. This one was very common for monoatomic and diatomic species over top positions that migrate to hcp or fcc positions. e.g: Br over Cu.

For the MER the most common error was type 1 and for TS was type 2. *We added additional information about the errors to Test and Results Section (3.4).*