Interactive comment on "Technical note: A prototype transparent-middle-layer data management and analysis infrastructure for cosmogenic-nuclide exposure dating" *by* Greg Balco

Richard Selwyn Jones (Referee)

richard.s.jones@monash.edu

Received and published: 26 May 2020

Summary and recommendation

The technical note describes ICE-D – a concept and computational infrastructure for managing cosmogenic nuclide exposure data. This prototype infrastructure is divided into three layers: data/observations, "middle-layer" age calculation, and analysis/visualisations. The new concept of this database is that the observational data is stored without the derived ages/rates, which are instead calculated based on up-to-

date methods when the user accesses the online database. To optimise data management and analysis, ICE-D has been divided into three region-specific databases: Antarctica, Greenland and Alpine. The author justifies the concept of this project, and describes the different components of the infrastructure.

First, I must congratulate the author for the time- and money-consuming development of ICE-D, which has already become a well-used platform for the cosmogenic nuclide community. More than just a database, the concept of observation-only storage with, what the author describes as, a "transparent-middle-layer" infrastructure has the potential to advance data management beyond the immediate research field. The topic is well within the scope of GChron, and the paper is well-structured, clearly and concisely written, with useful illustrations of the concept and infrastructure.

I recommend publication after consideration, and perhaps discussion, of points below.

Detailed comments

The middle-layer calculations

The use of the ICE-D database is dependent on the Balco et al. (2008) age calculator, which needs to be more explicitly addressed. There is nothing necessarily wrong with using this age calculator – in fact, the calculator is arguably as accurate as any other, and more efficient – but there are differences in calculation methods between the various online calculators, and even between the published version 1 (Balco et al., 2008) and the current version 3, which will lead to different resulting exposure ages. I appreciate that the author does not want to get into the details of the calculator, especially as some of this information will become outdated as the calculator updates to use new production rate information, scaling models, etc. However, I think that this paper – as well as the ICE-D website – should more clearly outline what methods/values are specific to this particular age calculator, and to what degree the ages may differ to those generated from other calculators.

Section 4.2 could be expanded to 1) outline in more detail the changes from version 1 to version 3, and 2) discuss the differences between this age calculator and the other available online calculators – for example, why might ages calculated with a different calculator but using the same production rate scaling method, get different exposure ages to that calculated within ICE-D?

There are also slight limitations to using this age calculator that should be acknowledged. For example, the user cannot get ages from ICE-D that are: calculated for a user-specified (local, or alternative global) production rate, as with CREp (Martin et al., 2017); or, calculated with the full range of observational data uncertainty, as with CRONUSCalc (Marrero et al., 2016) – correct me if I am wrong.

Access to data

In order for this database project to suit the needs of the community, the data will need to be accessible for use beyond the analysis and visualisation provided by ICE-D. The ICE-D online interface provides all of the available observational data, as well as dynamically-calculated exposure ages, in a fairly clear format and navigable structure. But this information could be made more accessible.

The information is not all stored on the same webpage, and it is assumed that the user should just copy the required information from the various tables. There are some limitations to this current set up. The tables are not always easy to copy-and-paste from the webpage to a local file (e.g. spreadsheet), particularly with the nuclide boxes. Not all sample information is available in a single table, meaning that the user has to move around the various locations to extract all of the data. All of the necessary sample information for age calculation is only available together in the format required for the Balco et al. (2008) version 3 calculator, and otherwise has to be found on each sample-specific webpage. What if a user wants to gather all sample information (observations necessary for independent age calculation and perhaps also ages provided by ICE-D) for a particular region, site or publication?

To overcome these limitations may require a substantial restructuring of the ICE-D online platform, and I leave it to the author to decide whether they want to make such changes. I have two suggestions:

1. Provide a table with all available sample information (including ICE-D dynamically-calculated ages) for data when grouped by region, site or publication. This could either be within the current page structure, as a link to a separate page with a table generated for that specific dataset, or as a downloadable file (e.g. tab-delimited .txt file) for that specific dataset.

2. Provide an option to extract user-specified sample information from ICE-D. This could be a link to a code package, or command line that could be entered on a local computing system. Either of these could be added to the data pages, again, when grouped by region, site or publication.

Additional comment

I also have slight concerns about the long-term maintenance of ICE-D and sustainability of the project – which is an issue for any such database project – but this has largely been addressed by the other reviewer.