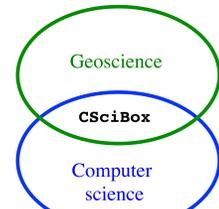


CSciBox: A Software System for Age-Model Construction and Evaluation

PP41D-1425



Elizabeth Bradley¹, Kenneth A. Anderson¹, Thomas M. Marchitto², Laura Rassbach de Vesine¹, James W. C. White², and David M. Anderson^{2,3}

¹ Department of Computer Science, University of Colorado, Boulder, Colorado, USA

² Institute for Alpine and Arctic Research (INSTAAR), University of Colorado, Boulder, Colorado, USA

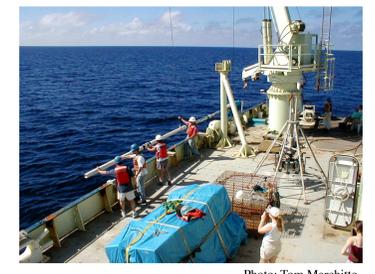
³ NOAA National Climatic Data Center, Boulder, Colorado, USA

lizb@colorado.edu

The goal:

Revolutionize climate scientists' ability to work with age models by providing:

- A powerful set of analysis and design tools that work on multiple types of cores
- A flexible, extensible framework that handles many different data sets and analysis workflows
- A single way to access the best community tools, in a unified way, with support for using them in informed and appropriate ways
- In the form of open-source code



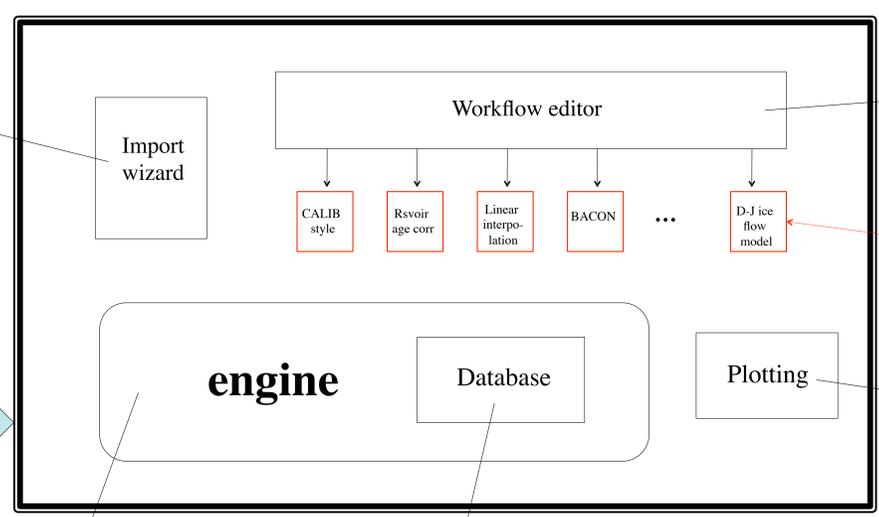
What this means for you:

- Minimum effort to start work on data that you already have
- Flexible enough to adapt to data that you may gather in the future
- Helps deal with different naming conventions (and spelling errors)
- Works with any kind of core

Import wizard

Guesses the mapping between the column headers in the spreadsheet and the terms that CSciBox knows about

Offers the user customized dropdowns in each field to modify these guesses



Workflow editor

- "Workflow" — a set of modular, interchangeable computation elements (e.g., calibration, interpolation) that are assembled together to solve a larger problem
- CSciBox's computation plan browser lets you see what steps are in an existing workflow
- ...or create a new one:

Below: a workflow that performs a CALIB-style calibration with the IntCal 2013 calibration curve, then calculates missing ages using linear interpolation:

What this means for you:

- Create new analysis programs without doing any coding
- Easy to define, run, and evaluate variants of a particular analysis
- Can plug in your own code, which then inherits CSciBox's GUI, etc.
- Workflows stored with the data → analyses are transparent, repeatable, and easy to update

What this means for you:

- A powerful analysis and design environment for working with age models...
- ...that doesn't require you to sling data, write code, or configure plotters
- Wide variety of browsers and editors to view and process core data and supporting data sets
- Support for a wide range of graphs for visual comparisons of multiple data streams

Graphical User Interface

- Manipulate programs and data using graphical menus, icons
- View and work with your data in the same place, simplifying your workflow
- No need to learn or memorize esoteric commands to process your data...

These dropdowns let you plot any variable(s) against any other

Full distributions are available:

Ice-core age models, too:

Repository

- Stored in MongoDB, a scalable NoSQL database designed for fast queries
- hbase version also available for really large datasets
- As well as a version that stores data in the cloud

Scaleable computing environment

- Parallelizable
- "Plug-in" architecture
- Designed to run on modern cyberinfrastructure
- ...including the cloud

What this means for you:

- Focused queries provide fast, tailored access to data
- Share data with your workgroup (students, etc.) without having to email files around
- CSciBox can work with multiple repositories, stored locally or in the cloud; for very large data sets, repositories can be distributed across a cluster

CSciBox is open source!

- Source code (python) available on github
- But you don't have to know python to run it; we also have one-click installers
- GNU public license; free to modify/extend/use as you see fit
- Code is designed to make it easy to "plug in" your favorite analysis tool, thereby adding it to the workflow editor's arsenal

your tool here

github.com/ldevesine/Calvin github.com/ldevesine/Calvin/releases

What this means for you:

- You can run CSciBox on your laptop in standalone mode with a local repository and no internet connection (e.g., on an airplane)
- Or on a remote compute cluster, with your laptop as the front end (e.g., when the problem demands that level of effort or storage)
- With the same interface, functionality, etc. (though the performance will of course be different!)

In the works:

- Lots more tools for different kinds of cores
- The plug-in architecture: making it easy to integrate your code into CSciBox...
- ...and handle any associated data-conversion issues (viz., data dictionaries)
- Artificial intelligence and age models
 - intelligent choice of parameter values for calls to BACON, Mai Winstrup's HMM-based layer counter — as well as intelligent interpretation and use of their results
 - automated exploration of the space of possible models

your tool here

We are trying to build — and support — a user community. Please join us!

www.cs.colorado.edu/~lizb/cscience.html



This material is based upon work sponsored by the National Science Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.