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

Elizabeth Bradley, Kenneth A. Anderson, Thomas M. Marchitto, Laura Rassbach de Vesine, James W. C. White, and David M. Anderson

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

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 getting errors
- Works with any kind of core

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...

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

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

What this means for you:
- 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

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...
- 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:
- 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

Ice-core age models, too:
- 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

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 getting errors
- Works with any kind of core

Scalable computing environment
- • Parallelizable
- • “Plug-in” architecture
- • Designed to run on modern cyberinfrastructure
- • …including the cloud

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

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 getting errors
- Works with any kind of core

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

In the works:
- 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
- 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 getting errors
- Works with any kind of core