

# The Babelizer: language interoperability for model coupling in the geosciences

Eric W.H. Hutton<sup>1</sup>, Mark D. Piper<sup>1</sup>, and Gregory E. Tucker<sup>1, 2, 3</sup>

<sup>1</sup> Community Surface Dynamics Modeling System, University of Colorado Boulder <sup>2</sup> Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder <sup>3</sup> Department of Geological Sciences, University of Colorado Boulder

DOI: [10.21105/joss.03344](https://doi.org/10.21105/joss.03344)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Kristen Thyng](#) ↗

## Reviewers:

- [@PythonCHB](#)
- [@lheagy](#)

Submitted: 09 April 2021

Published: 22 June 2021

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

## Summary

The `babelizer` is a Python utility that generates code to import libraries from other languages into Python. Target libraries must expose a Basic Model Interface (BMI) ([Hutton et al., 2020](#); [Peckham et al., 2013](#)) and be written in C, C++, or Fortran, although the `babelizer` is extendable, so other languages can be added in the future. The `babelizer` provides a streamlined mechanism for bringing scientific models into a common language where they can communicate with one another as components of an integrated model.

## Statement of need

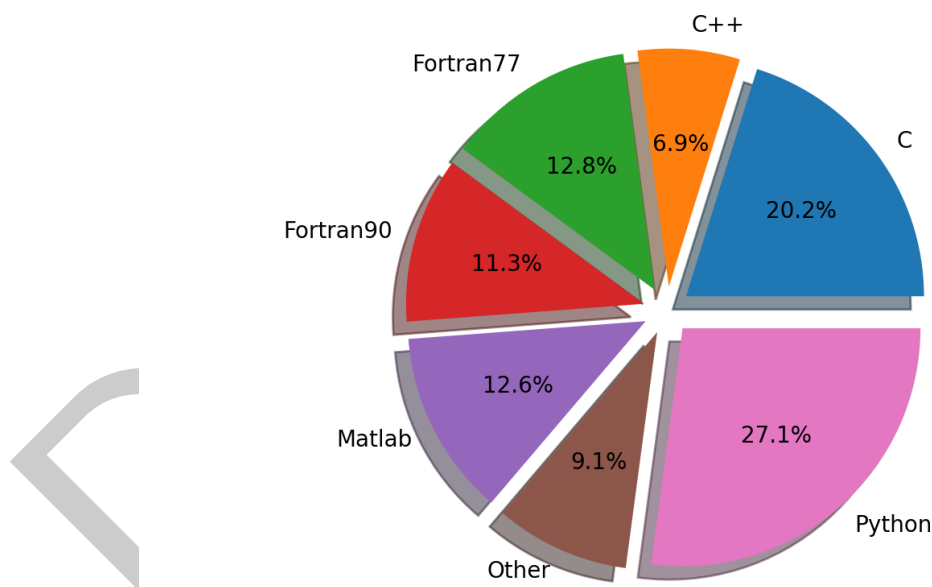
With an integrated multicomponent approach to modeling, scientific modelers—not just software developers—connect components to form integrated models, where plug-and-play components can easily be added or removed ([Collins et al., 2005](#); [David et al., 2013](#); [Gregersen et al., 2007](#); [Tucker et al., 2021](#)). This is in contrast to older methods, where a single modeling group would construct a monolithic model built on the tight integration of software written within an isolated framework. A single person or group would control model development. Outside contributors would go through a gatekeeper to ensure compatibility. The software elements that made up the model would be tied to the larger model and, generally, not used outside of the framework.

Component modeling democratizes model building by empowering the larger scientific community to develop model components. This allows for more innovation and experimentation driven from the bottom up by a community. It reduces redundancy—rather than reinventing software, scientists can find and use models that suit their needs—and it allows scientists to focus on new, unsolved, problems.

There are disadvantages, however. Without a single group to guide model development, there is a greater risk that community-developed models will become incompatible with one another. With hundreds of scientists developing models in isolation, there is a greater likelihood models will be written with idiosyncratic designs, incompatible grids, incompatible time steps, and even in different programming languages. The Earth-system modeling community has developed tools to help solve some of these problems. For example, the Basic Model Interface standardizes model interactions. The Earth System Modeling Framework (ESMF) ([Collins et al., 2005](#)) grid mappers are able to map quantities from one grid to another. The Python Modeling Toolkit `pymt` ([Hutton & Piper, 2020](#)) performs time interpolation, grid mapping, and unit conversion. In this paper, we present a solution to the language incompatibility problem.

## 39 Overcoming the language incompatibility problem

40 To get an idea of the range of programming languages used in Earth-system modeling, we can  
41 look to the Community Surface Dynamics Modeling System (CSDMS) model repository. As  
42 of June 2020, the repository holds over 370 open source models and tools submitted by the  
43 community. These contributions span a range of languages, with Python, C, and Fortran being  
44 the most popular (Figure 1). The mix of languages raises an interesting challenge in creating  
45 an interoperable modeling framework. Our solution is to use a hub-and-spoke approach, where  
46 Python is the hub language that connects to other languages. We chose Python as the hub  
47 because of its popularity in the scientific community, its extensive collection of third-party  
48 libraries (including model coupling frameworks such as the `pymt`), and its existing ability to  
49 communicate with other programming languages. We have built the `babelizer` to generate  
50 the spokes that connect Python to other languages. Using the CSDMS model repository as a  
51 guide, if we build translators for the open source languages C, C++, Fortran, and Python, we  
52 will cover 80 percent of the contributed models. A drawback of using Python is that it can  
53 be relatively slow compared to compiled languages like C and Fortran; however, the models  
54 being wrapped are compiled and run in their native language, which is where the bulk of the  
55 computation takes place, with the `babelizer` providing only a thin wrapper layer.



**Figure 1:** The distribution of programming languages used in the models in the CSDMS model repository. (Data from [https://csdms.colorado.edu/wiki/CSDMS\\_models\\_by\\_numbers](https://csdms.colorado.edu/wiki/CSDMS_models_by_numbers).)

## 56 Design of the babelizer

57 The `babelizer` is a command-line utility that generates the glue code to bring a model  
58 exposing a BMI from another language into Python. Because the BMI is a well-defined  
59 standard, the `babelizer` requires only a small amount of metadata to generate the glue code.  
60 The metadata depends somewhat on the language being wrapped, but includes the name of

61 the library providing the BMI, the name of an entry point into the library, the language the  
62 library was written in, and any necessary compiler flags. With this metadata, the babelizer  
63 creates a new git repository, a Python package containing the Python interface to the model,  
64 documentation, and sets up continuous integrations and a test suite for the model's BMI. The  
65 model can then be imported and run through Python.

66 The user provides metadata describing their model through a *toml*-formatted file (see [Figure 2](#)  
67 for an example). The babelizer uses the metadata to fill a set of *jinja*-formatted template  
68 files to construct the new repository (or update an existing repository). The entire repository  
69 is almost completely auto-generated, which means it can easily be regenerated. The only files  
70 a user need edit are the main configuration file, `babel.toml`, and any optional model data  
71 files, which are installed along with the new component.

```
[library]
[library.PRMSurface]
language = "fortran"
library = "bmiprmsurface"
header = ""
entry_point = "bmi_prms_surface"

[build]
undef_macros = []
define_macros = []
libraries = []
library_dirs = []
include_dirs = []
extra_compile_args = []

[package]
name = "pymt_prms_surface"
requirements = ["prms", "prms_surface"]

[info]
github_username = "pymt-lab"
package_author = "Community Surface Dynamics Modeling System"
package_author_email = "csdms@colorado.edu"
package_license = "MIT"
summary = "PRMS6 surface water process component"

[ci]
python_version = ["3.9"]
os = ["linux", "mac", "windows"]
```

**Figure 2:** The babelizer configuration file (`babel.toml`) for the Precipitation-Runoff Modeling System v6 surface water component, PRMSSurface ([Piper et al., 2020](#)). Running the babelizer with this file produces most of the repository [https://github.com/pymt-lab/pymt\\_prms\\_surface](https://github.com/pymt-lab/pymt_prms_surface).

72 Data files provided to a babelized component are intended to be used either by a user of  
73 the new component or by a separate framework that imports the component. There is little  
74 restriction on the contents of the files, but typically they are sample input files that a user  
75 of the component can use to run the model. Another use-case is where the component will  
76 be used within a separate modeling framework and that framework may require additional  
77 metadata. As an example, the `pymt` is a modeling framework able to work with generic BMI  
78 models. In addition to the BMI, the `pymt` requires descriptive information about the model

(e.g. authors, license, references, summary of what it does, etc.) as well as *jinja*-formatted sample input files. The *pymt* uses the template input files as part of a utility for a user to programmatically generate model input files without having to know anything about the idiosyncratic details of those model input files. Within such a framework, therefore, a user is given model components with a standardized way to create input files, as well as a common Python interface to run and interact with the model.

## Acknowledgements

This work is supported by the National Science Foundation under Grant No. 1831623, *Community Facility Support: The Community Surface Dynamics Modeling System (CSDMS)*.

## References

- Collins, N., Theurich, G., Deluca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C., & Da Silva, A. (2005). Design and implementation of components in the earth system modeling framework. *The International Journal of High Performance Computing Applications*, 19(3), 341–350.
- David, O., Ascough, J. C., Lloyd, W., Green, T. R., Rojas, K. W., Leavesley, G. H., & Ahuja, L. R. (2013). A software engineering perspective on environmental modeling framework design: The object modeling system. *Environmental Modelling & Software*, 39, 201–213. <https://doi.org/10.1016/j.envsoft.2012.03.006>
- Gregersen, J., Gijsbers, P., & Westen, S. (2007). OpenMI: Open modelling interface. *Journal of Hydroinformatics*, 9(3), 175–191. <https://doi.org/10.2166/hydro.2007.023>
- Hutton, E. W. H., & Piper, M. D. (2020). *The python modeling toolkit* (Version v1.0.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.3644240>
- Hutton, E. W. H., Piper, M. D., & Tucker, G. E. (2020). The basic model interface 2.0: A standard interface for coupling numerical models in the geosciences. *Journal of Open Source Software*, 5(51), 2317. <https://doi.org/10.21105/joss.02317>
- Peckham, S. D., Hutton, E. W., & Norris, B. (2013). A component-based approach to integrated modeling in the geosciences: The design of CSDMS. *Computers & Geosciences*, 53, 3–12. <https://doi.org/10.1016/j.cageo.2012.04.002>
- Piper, M., McDonald, R., Hutton, E., Markstrom, S., Parker, N., & Tucker, G. (2020). Coupling hydrologic models with data services in an interoperable modeling framework. *Earth and Space Science Open Archive ESSOAr*. <https://doi.org/10.1002/essoar.10504855.1>
- Tucker, G. E., Hutton, E., Piper, M., Campforts, B., Gan, T., Barnhart, K. R., Kettner, A., Overeem, I., Peckham, S. D., McCready, L., & Syvitski, J. (2021). Numerical modeling of earth's dynamic surface: A community approach. *EarthArXiv*. <https://doi.org/10.31223/X51615>