The Journal of Open Source Software

umami: A Python package for Earth surface dynamics objective function construction

Katherine R. Barnhart[1, 2], **Eric Hutton**[3, 4], **and Gregory E. Tucker**[1, 2, 3]

1 University of Colorado at Boulder, Department of Geological Sciences **2** University of Colorado at Boulder, Cooperative Institute for Research in Environmental Sciences **3** University of Colorado at Boulder, Community Surface Dynamics Modeling System Integration Facility **4** University of Colorado at Boulder, Institute for Arctic and Alpine Research

DOI: 10.21105/joss.01776

Software

- Review
- Repository
- Archive

Submitted: 13 September 2019
Published: 29 October 2019

Summary

Models of Earth's surface dynamics are typically designed to simulate timescales that range from years to geologic epochs (10^6+ years). They represent and evolve a primary state variable, Earth's surface. Some applications may use other state variables (e.g., soil thickness). A diverse array of physical and chemical processes may be present. For example, the making and moving of water; the creation of soil and sediment from bedrock; the transport of mobile material due to hillslope processes, river erosion, and landslides; and the deposition of that material into the geologic archive. These models are used for applications as diverse as understanding limits on the height of mountain ranges and predicting the erosion, transport, and fate of contaminated material on human timescales.

A diversity of data is used to compare models with observations. These data serve as "simulated equivalents", quantities with which to assess model performance. A common observable compared with model output is topography. High resolution topography, occasionally at two points in time, provides a dataset rich in the spatial dimension and sparse in the temporal. However, because model initial condition topography is not often known, we do not often expect modeled and observed topography to line up exactly. Thus, statistical derivatives of topography are often used. Other data sources (e.g., cosmogenic radionuclide derived erosion rates) provide time and space integrated measures.

There is, however, no clear consensus regarding which set of simulated equivalents is most appropriate or most successful for assessing the performance of Earth surface dynamics models. This presents a challenge when this type of model is used in formal model analysis efforts. For example, calibration requires a formal objective function. Under these circumstances, there is a need for generic tools that facilitate exploration and usage of many plausibly successful simulated equivalents. This is the need that the umami package was designed to address.

Description of umami

Umami is a package for calculating objective functions or objective function components for Earth surface dynamics modeling. It was designed to work well with models built with the Landlab Toolkit (Hobley et al., 2017) or with the `terrainbento` multi-model analysis (Barnhart, Glade, Shobe, & Tucker, 2019). Umami is complementary to existing topographic analysis tools such as LSDTopoTools (Clubb et al., 2017; Mudd, Attal, Milodowski, Grieve, & Valters, 2014), TopoToolbox (Schwanghart & Kuhn, 2010; Schwanghart & Scherler, 2014) and the Topographic Analysis Kit (Forte & Whipple, 2019). Rather than performing topographic analysis, umami is used to distill model output into a form usable by model analysis methods such as sensitivity analysis, calibration, and validation.

Umami offers two primary classes: a `Residual`, which represents the difference between model and data, and a `Metric`, which is a calculated value on either model or data. The set of currently supported calculations are found in the `umami.calculations` submodule. Both the `Metric` and `Residual` classes are designed to be fully specified through a YAML-style input-file or python Dictionary interface. Many different calculations can be accomplished through parameter specification. This supports reproducible analysis and systematic variation in metric construction. For example, when used with `terrainbento`, one input file can describe the model run, and one input file can describe the model assessment or model-data comparison. This streamlines model analysis applications by making driver files more re-usable and by placing the code that accomplished calculations in the umami package rather than within the driver file. Umami also provides multiple output formats (YAML and Dakota), the latter of which is designed to interface with Sandia National Laboratory's Dakota package (Adams et al., 2019).

The novel contribution of the umami package is not primarily found in the specific calculations accomplished (e.g., some of them are as straightforward as the mean of a state variable). Instead it is the flexible and extensible nature of the input file format and the `Metric` and `Residual` classes. Additionally, the package can be extended through the addition of new calculation methods.

Two model-data comparison metrics in umami are novel. First, the `joint_density_misfit` provides a misfit metric of the joint density of two state variables. This comparison measure was inspired by the use of channel χ index value (Perron & Royden, 2013) and topographic elevation to assess river channel long profiles (χ-z plots). While it is not clear how to best quantify the difference between a modeled and observed χ-z plot, it is straightforward to calculate the sum of squared residuals between the joint density of χ and topographic elevation for a modeled and observed landscape. In umami, this comparison is generalized to the joint density of two state variables.

Second, the `discretized_misfit` calculation seeks to reduce the dimension of a state variable misfit. In some applications it is appropriate to make a direct comparison between a measured and modeled state variable (e.g., difference measured and modeled topography). However, if an application uses a model domain with tens of thousands of grid cells, a user is faced with a challenging choice: either treat each residual as an individual observation or reduce all residuals to a single value through a sum of squared residuals (or other norm). The `discretized_misfit` calculation permits the identification of consistent "categories" of grid cells based on other grid fields. A sum of squared residuals is then calculated within each of these categories.

Acknowledgements

Support for this work was provided by an NSF EAR Postdoctoral Fellowship to Barnhart (NSF Award Number 1725774). Funding for Landlab was provided by NSF (NSF Award Numbers 1147454 and 1450409). Landlab is additionally supported by the Community Surface Dynamics Modeling System (NSF Award Numbers 1226297 and 1831623).

References

Adams, B. M., Eldred, S., Geraci, G., Hooper, R. W., Jakeman, J. D., Maupin, K. A., Monschk, J. A., et al. (2019). Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.10 User's Manual. *Sandia National Laboratories, Tech. Rep. SAND2014-4633*.

Barnhart, K. R., Glade, R. C., Shobe, C. M., & Tucker, G. E. (2019). Terrainbento 1.0: a Python package for multi-model analysis in long-term drainage basin evolution. *Geoscientific Model Development*, *12*(4), 1267–1297. doi:10.5194/gmd-12-1267-2019

Clubb, F. J., Mudd, S. M., Milodowski, D. T., Valters, D. A., Slater, L. J., Hurst, M. D., & Limaye, A. B. (2017). Geomorphometric delineation of floodplains and terraces from objectively defined topographic thresholds. *Earth Surface Dynamics*, *5*(3), 369–385. doi:10.5194/esurf-5-369-2017

Forte, A. M., & Whipple, K. X. (2019). Short communication: The Topographic Analysis Kit (TAK) for TopoToolbox. *Earth Surface Dynamics*, *7*(1), 87–95. doi:10.5194/esurf-7-87-2019

Hobley, D. E. J., Adams, J. M., Nudurupati, S. S., Hutton, E. W. H., Gasparini, N. M., Istanbulluoglu, E., & Tucker, G. E. (2017). Creative computing with Landlab: an open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics. *Earth Surface Dynamics*, *5*(1), 21–46. doi:10.5194/esurf-5-21-2017

Mudd, S. M., Attal, M., Milodowski, D. T., Grieve, S. W. D., & Valters, D. A. (2014). A statistical framework to quantify spatial variation in channel gradients using the integral method of channel profile analysis. *Journal of Geophysical Research: Earth Surface*, *119*(2), 138–152. doi:10.1002/2013JF002981

Perron, J. T., & Royden, L. (2013). An integral approach to bedrock river profile analysis. *Earth Surface Processes and Landforms*, *38*(6), 570–576. doi:10.1002/esp.3302

Schwanghart, W., & Kuhn, N. J. (2010). TopoToolbox: A set of Matlab functions for topographic analysis. *Environmental Modelling & Software*, *25*(6), 770–781. doi:10.1016/j.envsoft.2009.12.002

Schwanghart, W., & Scherler, D. (2014). Short Communication: TopoToolbox 2 - a MATLAB-based software for topographic analysis and modeling in Earth surface sciences. *Earth Surface Dynamics*, *2*(1), 1–7. doi:10.5194/esurf-2-1-2014