UncertainData.jl: a Julia package for working with measurements and datasets with uncertainties.

Kristian Agasøster Haaga[1, 2, 3]

1 Department of Earth Science, University of Bergen, Bergen, Norway **2** K. G. Jebsen Centre for Deep Sea Research, Bergen, Norway **3** Bjerknes Centre for Climate Research, Bergen, Norway

Summary

`UncertainData.jl` provides an interface to represent data with associated uncertainties for the Julia programming language (Bezanson, Edelman, Karpinski, & Shah, 2017). Unlike `Measurements.jl` (Giordano, 2016), which deals with exact error propagation of normally distributed values, `UncertainData.jl` uses a resampling approach to deal with uncertainties in calculations. This allows working with and combining any type of uncertain value for which a resampling method can be defined. Examples of currently supported uncertain values are: theoretical distributions, e.g., those supported by Distributions.jl (Besançon et al., 2019; Lin et al., 2019); values whose states are represented by a finite set of values with weighted probabilities; values represented by empirical distributions; and more.

The package simplifies resampling from uncertain datasets whose data points potentially have different kinds of uncertainties, both in data values and potential index values (e.g., time or space). The user may resample using a set of pre-defined constraints, truncating the supports of the distributions furnishing the uncertain datasets, combined with interpolation on pre-defined grids. Methods for sequential resampling of ordered datasets that have indices with uncertainties are also provided.

Using Julia's multiple dispatch, `UncertainData.jl` extends most elementary mathematical operations, hypothesis tests from HypothesisTests.jl, and various methods from the Stats-Base.jl package for uncertain values and uncertain datasets. Additional statistical algorithms in other packages are trivially adapted to handle uncertain values and datasets from `UncertainData.jl` by using multiple dispatch and the provided resampling framework.

`UncertainData.jl` was originally designed to form the backbone of the uncertainty handling in the CausalityTools.jl package, with the aim of quantifying the sensitivity of statistical time series causality detection algorithms. Recently, the package has also been used in paleoclimate research (Vasskog et al., 2019).

References

Besançon, M., Anthoff, D., Arslan, A., Byrne, S., Lin, D., Papamarkou, T., & Pearson, J. (2019). Distributions.jl: Definition and modeling of probability distributions in the juliastats ecosystem. *arXiv e-prints*, arXiv:1907.08611. Retrieved from http://arxiv.org/abs/1907.08611

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, *59*(1), 65–98. doi:10.1137/141000671

Giordano, M. (2016). Uncertainty propagation with functionally correlated quantities. *arXiv preprint arXiv:1610.08716*.

Lin, D., White, J. M., Byrne, S., Bates, D., Noack, A., Pearson, J., Arslan, A., et al. (2019). JuliaStats/Distributions.jl: a Julia package for probability distributions and associated functions. doi:10.5281/zenodo.2647458

Vasskog, K., Svendsen, J.-I., Mangerud, J., Agasøster Haaga, K., Svean, A., & Lunnan, E. M. (2019). Evidence of early deglaciation (18 000 cal a bp) and a postglacial relative sea-level curve from southern Karmøy, south-west Norway. *Journal of Quaternary Science*.