

pyPDAF: a Python version of the Parallel Data Assimilation Framework (PDAF)

Yumeng Chen

University of Reading

Abstract:

pyPDAF is a Python version/interface to the widely used community open-source software framework for data assimilation, PDAF. This Python package provides all latest functionalities of PDAF including ensemble forecasting, DA methods, diagnostics and observation handling approaches. pyPDAF is particularly suited for online data assimilation, i.e. direct coupling into a model where no file exchanges on disk are necessary, with Python-based models, e.g., machine learning based models. For offline data assimilation, where DA is performed on restart files, pyPDAF can be used with any

models. As Python provides flexible syntax without compilation, a rich library ecosystem and convenient package management systems, pyPDAF can also facilitate the DA system implementation and prototyping for PDAF-based DA systems. We demonstrate that, even though a DA system based on pyPDAF could have an overhead compared to directly

using PDAF, it still shows acceptable performance in relatively high-dimensional setups. pyPDAF is available from the Conda package manager and can be compiled from its sources with limited dependencies. These features are attractive for research-oriented DA applications. One of the future plans for pyPDAF is to accommodate the free threading feature introduced by Python 3.13 to support the OpenMP parallelization in PDAF.