

Deep Learning Approaches for Efficiently Emulating Noah Land Surface Model Soil Hydrology

Master's Thesis Progress Report
Mitchell Dodson
August 29, 2024

1 Hypothesis

Deep learning methods can reduce evaluation time of Noah-LSM by emulating the nonlinearities of the data at a coarser time step, and by representing the relationship between initial states, static parameters, atmospheric forcings, and output states in terms of a computationally simple composition of matrix transformations.

2 Work Performed

- Acquired 10 years' worth of $1/8^\circ$ NLDAS-2 atmospheric forcings and Noah-3.6 land surface data as hourly grib files.
- Extensively researched the history, implementation, and hydrologic theory underpinning the Noah model.
- Developed data pipeline for converting data into regional HDF5 files that are spatially and temporally chunked so that sporadic data can be accessed in a fast and parallelizable manner.
- Validated the integrity of and collected a variety of bulk statistics for the full dataset.
- Created a series of efficient and flexible data generators that index, extract, format, and interleave data from regional HDF5 files for tasks including model training and evaluation and data visualization.
- Implemented a custom neural network training framework based on JSON configurations, and used it to fit LSTM models with a variety of datasets, inductive biases, and parameters.
- Experimented with a wide range of sampling methods, model architectures, training strategies, and objective function modifications, resulting in several discoveries that considerably improved models' inference ability.
- Evaluated bulk statistics of model results, as well as weekly error statistics on a spatial grid.

3 Future Plans

- Solve banding issue in true (Noah) residual histogram bins.
- Investigate individual cases where model errors are high, and determine the meteorological and hydrological conditions that contribute to model error.
- Develop error statistics with respect to rainfall types and amounts, and land surface initial states.
- Identify drought and flood case studies that demonstrate the effectiveness of the model and outline more detailed information about the meteorological context of the events.

- Integrate the models into NASA SPoRT's production environment and directly compare their effectiveness and efficiency to SPoRT's implementation of the Noah model using CPU and memory profiling strategies.
- Test models' physical integrity using synthetic atmospheric forcings.
- Write and publish thesis as a journal article focused on the background and use cases of the Noah model, the context in which a machine learning approach is advantageous, the model construction and training strategies I used, and the performance of my models in terms of bulk statistics and in the context of specific case studies.

4 Timeline

As of Spring 2024, I completed all the classwork needed to graduate. I will finish the technical effort of my research during the Fall 2024, as well as the majority of the writing, with the goal of initiating the publishing process over the winter. I plan to present my results at the American Meteorological Society conference in January 2025, and will defend in February or March of the Spring 2025 semester.