

EGU23-953, updated on 23 Apr 2024

<https://doi.org/10.5194/egusphere-egu23-953>

EGU General Assembly 2023

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



s2spy, a package to boost (sub) seasonal forecasting with artificial intelligence

Yang Liu¹, **Bart Schilperoort**¹, Jannes van Ingen², Sem Vijverberg², Peter Kalverla¹, and Dim Coumou²

¹Environment and Sustainability, Netherlands eScience Center, Amsterdam, the Netherlands

²Water and Climate Risk, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

Reliable (sub) seasonal (S2S) forecasts remain a huge scientific challenge. The lead-time is too long to benefit from the atmosphere's inertial memory, but too short for the atmosphere's boundary conditions to be felt strongly. Only for specific "windows of predictability" (i.e. specific regions, timescales and climatic background states), skillful forecasts are possible, in an otherwise largely unpredictable future. Due to a number of successes in S2S forecasting, the interest in machine learning (ML) is growing fast. However, we argue there is a need for more standardization, consensus on best practices, higher efficiency, and higher reproducibility. Typical S2S ML use-cases, such as (1) pure statistical forecasting based on observations, (2) transfer learning, and (3) post-processing of dynamical model ensembles, require a large coding and preprocessing effort. Such experiments are not trivial to set up, and without sufficient experience and expertise there is a large risk of improper cross-validation and/or improper and non-standard verification.

Within a 3-year project, we are developing a high-level Python package called **s2spy**. Our aim is to make ML workflows more transparent and easier to build, and to facilitate standardization and collaboration across the S2S community. **s2spy** also contributes to a higher reproducibility and works towards a wider acceptance of standards and best practices. We will present our vision and the capabilities of our package, show-casing that we can build a model from raw climate data up to verification and explanation in only a few lines of code.