ABSTRACT
The North Atlantic Ocean is key to climate through its role in heat transport and storage, but the response of the circulation’s drivers to a changing climate is poorly constrained. Climate models show that the circulation is weakening, and here the cause is revealed addressing the fundamental question regarding the existence of dynamical coherent regions and the associated mechanisms. Using the transparent machine learning method Tracking global Heating with Ocean Regimes (THOR), dynamical regimes are identified and tied to distinct currents and mechanisms such as the formation of deep water masses, and the location of the Gulf Stream and the Trans-Atlantic Current. Drivers of the circulation are identified using the two-dimensional fields depth, dynamic sea level and wind stress. Moving beyond a black box approach, THOR is engineered to reveal its source of predictive skill. A labeled dataset is engineered by an explicitly interpretable equation transform and k-means application to ocean model data, allowing theoretical inference. A multilayer perceptron is then trained, explaining its skill using relevance maps and theory. With abrupt CO2 quadrupling, the circulation weakens due to a shift in deep water formation areas and a northward shift of the Gulf Stream and an eastwards shift in the Trans-Atlantic Current. If CO2 is increased 1% yearly, similar but weaker patterns emerge. THOR is scalable and applicable to a range of models, needing only commonly available fields, and could accelerate model analysis and facilitate process oriented intercomparisons.