Can we Design a New NWP Data Assimilation System Based Entirely on AI Techniques? Advantages & Challenges

S.-A. Boukabara[1], E. Maddy[1]

[1]National Environmental Satellite Data and In formation Service, NOAA

Global Numerical Weather Prediction (NWP) has been a great success story over the past many years and has contributed immensely to societies across the globe. This success has benefited from the large and diverse set of observations that resulted from an efficient coordination at the international level between all providers. These observations have indeed led, through data assimilation mechanisms, to an increase in accuracy and quality of the initial conditions that are key for accurate global predictions. However, the NWP community currently faces a number of challenges in the way traditional approaches of assimilating observations in NWP are implemented. Chief among these challenges is the inefficiency to easily handle significant increases in data volume from new and emerging observing systems including from satellites. This issue is expected to get exacerbated with the upcoming smallsats and cubesats that are planned for the near future as well as from the multiplication of new sources of observations from ground-based and above-ground platforms (water-based and air-based unmanned vehicles, Internet of Things IoT, etc). Other challenges in NWP include the handling of complex phenomenologies such as precipitation, presence of ice, high-contrast scenes at the coasts, along fronts, etc. that often lead to NWP's data assimilation assumptions being invalid. These assumptions include local linearity of cost functions, Gaussian distribution of errors, etc. We argue that new approaches based on modern, proven techniques should be considered in order to tackle these challenges. A new approach to perform large-volume data fusion and assimilation, based entirely on Artificial Intelligence (AI) modern techniques including machine learning and computer vision techniques, is proposed and presented in this study. We will assess its efficiency and the quality of its outcomes with a focus on satisfying physical constraints. This approach to data assimilation is applied to real environmental data measured from both satellites-based and surface-based observing systems to reproduce traditional Numerical Weather Prediction (NWP) data assimilation performances from the U.S. National Oceanic and Atmospheric Administration (NOAA). Of note, these AI techniques have already been successfully tested in other fields to merge large varieties of data, and this effort aims at assessing the feasibility of leveraging them for the purpose of NWP data assimilation and

Earth System Modeling (ESM) in general. We will explore both advantages and challenges. The preliminary results confirm that significant efficiency could be achieved in environmental data assimilation using AI. This efficiency allows us to address one of the long-vexing issues of handling the Big data challenge due to the exponential increase in the volume of satellite and ground based data that exists and is expected to increase in the future. We assess this efficiency gain by assessing the reduction in the amount of time required to perform the AI-based assimilation, compared to traditional approaches. In this demonstration of the feasibility of an AI-based system, we focus on performing a multi-variable data fusion/assimilation focusing on a representative but limited set of variables at different levels in the atmosphere. Namely atmospheric temperature, moisture, wind and cloud. The resulting analysis from the data assimilation or fusion modes, i.e. with and without the use of an NWP forecast model as background respectively, is generated at global scales and at varying spatial resolutions. This study shows that physical constraints in the analysis, an important aspect of any data assimilation, could be satisfactorily accounted for to a certain degree, as part of the AI-based training approach. These results are extremely encouraging but are considered only a first initial step toward an entirely AI-based environmental data fusion/assimilation system. This approach, if expanded and matured further, has the potential to allow us to easily widen the applicability to a holistic Earth System Model environment, open the door to new nontraditional sources of environmental data, and perhaps more importantly, allow us to handle situations where assumptions of traditional data assimilation techniques are not necessarily valid. For example in cases where cost functions are highly non-linear or even discontinuous, such as in rainy conditions or in cases where observations or geophysical variables have non-Gaussian error characteristics. In this AI-based approach, these assumptions are indeed not necessary since the data assimilation is done using a highly non-linear approach and does not make simplification assumptions in order to parametrize the minimization cost function. We focus in this study on the assessment of the quality of the AI-based NWP analysis by assessing its accuracy when compared to an independent reference, and by assessing its characteristics including spatial variability, inter-parameters correlations, kinetic energy conservation, hydrostatic and geostrophic balances conservation, and other assessment metrics to ensure physics constraints are respected and analyses are interpretable and trustworthy.