

EGU21-9254

<https://doi.org/10.5194/egusphere-egu21-9254>

EGU General Assembly 2021

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Machine learning-based uncertainty quantification for data assimilation: a simple model experiment

Juan Ruiz¹, **Maximiliano Sacco**², Yicun Zhen³, Pierre Tandeo³, and Manuel Pulido⁴

¹University of Buenos Aires, CIMA/DCAO (UBA-CONICET), Buenos Aires, Argentina (jrui@cima.fcen.uba.ar)

²National Meteorological Service of Argentina, Buenos Aires, Argentina

³IMT-Atlantique, Brest, France

⁴Universidad Nacional del Nordeste - CONICET, Corrientes, Argentina

Quantifying forecast uncertainty is a key aspect of state-of-the-art data assimilation systems which has a large impact on the quality of the analysis and then the following forecast. In recent years, most operational data assimilation systems incorporate state-dependent uncertainty quantification approaches based on 4-dimensional variational approaches, ensemble-based approaches, or their combination. However, these quantifications of state-dependent uncertainties have a large computational cost. Machine learning techniques consist of trainable statistical models that can represent complex functional dependencies among different groups of variables. In this work, we use a fully connected two hidden layer neural network for the state-dependent quantification of forecast uncertainty in the context of data assimilation. The input to the network is a set of three consecutive forecasted states centered at the desired lead time and the network's output is a corrected forecasted state and an estimation of its uncertainty. We train the network using a loss function based on the observation likelihood and a large database of forecasts and their corresponding analysis. We perform observing system simulation experiments using the Lorenz 96 model as a proof-of-concept and for an evaluation of the technique in comparison with classic ensemble-based approaches.

Results show that our approach can produce state-dependent estimations of the forecast uncertainty without the need for an ensemble of states (at a much lower computational cost), particularly in the presence of model errors. This opens opportunities for the development of a new type of hybrid data assimilation system combining the capabilities of machine learning and ensembles.