

EGU21-2401, updated on 16 Oct 2021

<https://doi.org/10.5194/egusphere-egu21-2401>

EGU General Assembly 2021

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



## The SWAG solution for probabilistic predictions with a single neural network

Yann Haddad<sup>1,2</sup>, Michaël Defferrard<sup>2</sup>, and Gionata Ghiggi<sup>1</sup>

<sup>1</sup>Environmental Remote Sensing Laboratory (LTE), EPFL, Lausanne, Switzerland (yann.haddad@epfl.ch)

<sup>2</sup>Signal Processing Laboratory (LTS2), EPFL, Lausanne, Switzerland

Ensemble predictions are essential to characterize the forecast uncertainty and the likelihood of an event to occur. Stochasticity in predictions comes from data and model uncertainty. In deep learning (DL), data uncertainty can be approached by training an ensemble of DL models on data subsets or by performing data augmentations (e.g., random or singular value decomposition (SVD) perturbations). Model uncertainty is typically addressed by training a DL model multiple times from different weight initializations (DeepEnsemble) or by training sub-networks by dropping weights (Dropout). Dropout is cheap but less effective, while DeepEnsemble is computationally expensive.

We propose instead to tackle model uncertainty with SWAG (Maddox et al., 2019), a method to learn stochastic weights—the sampling of which allows to draw hundreds of forecast realizations at a fraction of the cost required by DeepEnsemble. In the context of data-driven weather forecasting, we demonstrate that the SWAG ensemble has i) better deterministic skills than a single DL model trained in the usual way, and ii) approaches deterministic and probabilistic skills of DeepEnsemble at a fraction of the cost. Finally, multiSWAG (SWAG applied on top of DeepEnsemble models) provides a trade-off between computational cost, model diversity, and performance.

We believe that the method we present will become a common tool to generate large ensembles at a fraction of the current cost. Additionally, the possibility of sampling DL models allows the design of data-driven/emulated stochastic model components and sub-grid parameterizations.

### Reference

Maddox W.J, Garipov T., Izmailov P., Vetrov D., Wilson A. G., 2019: A Simple Baseline for Bayesian Uncertainty in Deep Learning. arXiv:1902.02476