Contribution ID: **53**

Type: **Oral presentation**

Significance-tested and physically constrained interpretation of a deep-learning model for tornadoes

Wednesday, 7 October 2020 11:00 (30 minutes)

We have developed a convolutional neural network (CNN) to predict tornadoes at lead times up to one hour in a storm-centered framework. We trained the CNN with data similar to those used in operations –namely, a radar image of the storm and a sounding of the near-storm environment. However, CNNs and other ML methods are often distrusted by users, who view them as opaque "black boxes"whose decisions cannot be explained. To address this problem, the field of interpretable ML has emerged, providing methods for understanding what an ML model has learned. However, interpretation methods can be misleading, often producing artifacts ("noise") rather than illuminating the true physical relationships in the data. To address both of these problems (opaque models and noisy interpretation results), we have applied several interpretation methods to the CNN for tornado prediction, each augmented with either a statistical-significance test or physical constraints.

Specifically, we use four interpretation methods: the permutation importance test, saliency maps, class-activation maps, and backward optimization. For the permutation test, we use four different versions of the test and apply significance-testing to each, which allows us to identify where the ranking of predictor importance is robust. For saliency and class-activation maps, we use the "sanity checks" proposed by Adebayo et al. (2018; http://papers.nips.cc/paper/8160-sanity-checks-for-saliency-maps), augmented with formal significance tests. These tests ensure that interpretation results cannot be reproduced by a trivial method like an untrained edge-detection filter. For backward optimization, which produces synthetic storms that minimize or maximize tornado probability (prototypical non-tornadic and tornadic storms, respectively), we use physical constraints that force the synthetic storms to be more realistic.

To our knowledge, this work is one of the few applications of ML interpretation, and the only one with significance-tested or physically constrained ML interpretation, in the geosciences. As ML becomes more integrated into geoscience research and everyday applications, such work will be crucial in building ML systems that are trusted and understood by humans.

Thematic area

1. Machine Learning for Product development - Including NWP Post-processing, Non-linear Ensemble Averaging, Development of new NWP Products

Primary authors: Ms LAGERQUIST, Ryan (Cooperative Institute for Research in the Atmosphere (CIRA)); Dr MCGOVERN, Amy (University of Oklahoma); GAGNE, David (National Center for Atmospheric Research); Dr HOMEYER, Cameron (University of Oklahoma)

Presenter: Ms LAGERQUIST, Ryan (Cooperative Institute for Research in the Atmosphere (CIRA))

Session Classification: Session 4 (cont.) and Session 5: ML for Product Development and ML for Model Identification and development

Track Classification: ECMWF-ESA Workshop on Machine Learning for Earth System Observation and Prediction