



## Deriving meaning from metrics – a new approach for machine learning nowcasting verification

Jakub Lewandowski<sup>1,2</sup>, Leif Denby<sup>3</sup>, and Andrew Ross<sup>2</sup>

<sup>1</sup>Institute of Meteorology and Water Management - National Research Institute, National Meteorological Protection Center, Poland (jakub.lewandowski@imgw.pl)

<sup>2</sup>School of Earth and Environment, University of Leeds, Leeds, UK

<sup>3</sup>Danish Meteorological Institute

Nowcasting - the prediction of weather conditions over the next few hours - is critical for mitigating the impacts of severe convective storms. Machine learning offers new opportunities for improving nowcasting, particularly for convective precipitation, where traditional numerical models struggle. Yet, despite rapid progress in model development, evaluating these models remains a major challenge. Current verification practices typically rely on a narrow set of standard metrics that often fail to capture the complexity of atmospheric phenomena and cannot distinguish between different types of errors, providing limited insight into the specific weaknesses of the models.

This research introduces a comprehensive verification framework that combines carefully crafted datasets with sensitivity analyses, aiming to transform metric-based evaluation into a more informative process. Synthetic datasets are generated using ArtPrecip, a novel tool that randomly generates radar-like precipitation fields while allowing full control over properties such as motion, initiation, and evolution. Observational radar data are classified based on synoptic setting and observed precipitation properties, using different dimension-reduction methods. Sensitivity analyses examine how existing metrics respond to various error patterns, providing guidance on interpreting benchmark results.

The resulting system provides a well-defined and well-described set of benchmarks and enables reproducible, objective, and meaningful comparison of models. By addressing gaps in evaluation methodology, this work contributes to a more robust assessment of machine learning nowcasting skill and its applicability to severe weather forecasting.