Centre for Mathematics of Planet Earth

Ver-AI 2025 Workshop

on Verification of AI-Based Meteorological Forecasts

23th-24th of June, Department of Meteorology, Brian Hoskins Building.

Organisers: Jochen Bröcker (Reading), Zied Ben Bouallègue (ECMWF)
Administration: Shirls Smits (Reading)

23rd of June

Time	Speaker	Title
12.00		*Welcome Lunch*
13.00	Massimo Bonavita	*On the verification of Machine Learning Weather Prediction Models*
13.40	Anna-Luise Ellis	*ML Model Architectures' Power Spectra Characteristics and their Relationships to ML Model Artefacts*
14.10		*Coffee*
14.30	Leo Separovic (virtual)	*A Spectral Framework for Analyzing Strengths and Weaknesses of Machine Learning-Based Weather Prediction*
15.00	Britta Seegebrecht	*On the relation between activity and the power spectrum*
15.30		*Coffee*
15.50	Helen Dacre	*Midlatitude Cyclone Intensity Biases in Machine Learning Weather Prediction Models*
16.20	Linus Magnusson	*The Weather Prediction Model Intercomparison Project (WPMIP)*
17.00	Stéphane Vannitsem	*Skill and predictability of AI models: Comparison of GraphCast and Pangu-weather*
17.30		*Closing, 19.00 Dinner at Zerodegrees*

24th of June

Time	Speaker	Title
9.00	Jose Rodriguez	*Systematic errors in global circulation models and machine-learning models for NWP*
9.30	Dieter Van den Bleeken	*Forecast skill of regional AI weather models: a comparison of stretched grid and limited area designs.*
10.00	*Coffee and Discussion first round*	
11.00	*Discussion second round*	
12.00	*Lunch*	
13.00	Kaustubh Mittal	*Forecast bust characteristics in Europe: Variability across NWP and AI models*
13.30	Romain Pic	*On the verification of weather forecasts for extremes: a statistical review*
14.10	David Landry (virtual)	*Revisiting multi-dimensional rank histograms to evaluate generative weather forecasting models*
14.40	Maxime Taillardat	*On "proper" human verification: how to make subjective evaluation objective*
15.10	*Coffee and Panel Discussion*	
16.10	*Closing*	

Discussion themes

Theme 1	*Physical realism, artifacts and predictability.* Do outputs of AI models exhibit some form of realism? Are there common artifacts in AI models and do tools to detect them exist? Are these artefacts affecting error growth and predictability properties of AI models?
Theme 2	*Benchmarking, intercomparison, and systematic errors.* Benchmarks provide a playing field for intercomparison exercises (eg WP-MIP, a benchmark built under the WMO umbrella). What have we learned so far about systematic errors in AI forecasts? When do AI forecasts generally outperform/ underperform compared with NWP forecasts?
Theme 3	*Scoring, multivariate aspects and spatial evaluation.* How good are AI forecasts in representing multivariate aspects of the weather? Which scores and diagnostic tools should be used to assess these aspects? Would spatial verification tools be useful in this context?