Joint Working Group on Forecast Verification Research – Recent activities and VerAlops workshop summary

Barbara Casati (ECCC, Canada)

WGNE-40 November 2025

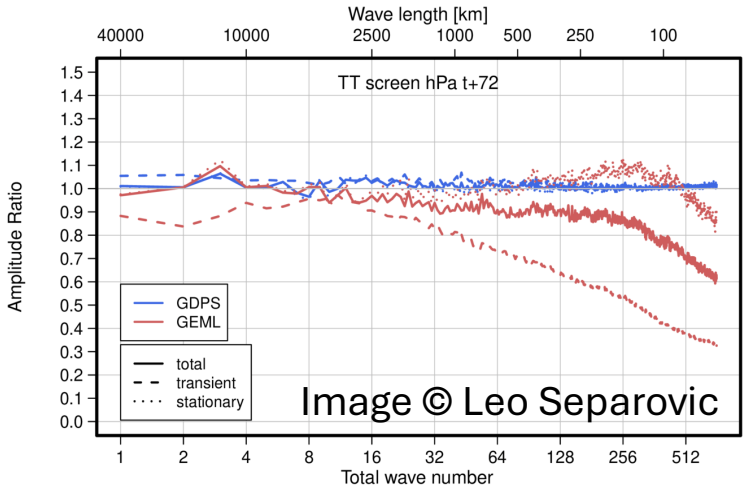
Co-Chair: Barbara CASATI, Environment and Climate Change Canada (ECCC), **Canada**
Co-Chair: Caio COELHO, Instituto Nacional de Pesquisas Espaciais INPE), **Brazil**
~~Eric GILLELAND, National Centre for Atmospheric Research (NCAR), United States of America~~
James BENNETT, Commonwealth Scientific and Industrial Research Organisation (CSIRO), **Australia**
Angie PENDERGRASS, Cornell University, **United States of America**
Ramon de ELIA, Servicio Meteorológico Nacional (SMN), **Argentina**
Javier GARCIA-SERRANO, University of Barcelona, **Spain**
Alfred Lawrence KONDOWE, Tanzanian Meteorological Authority (TMA), **Tanzania**
Zied BEN BOUALLEGUE, European Centre for Medium-Range Weather Forecasts (ECMWF), **United Kingdom**
Anumeha DUBE, National Centre for Medium Range Weather Forecasting (NCMRWF), India
Sabrina WAHL, Deutscher Wetterdienst (DWD), **Germany**
Jochen BROECKER, University of Reading, **United Kingdom**
Nicholas LOVEDAY, Bureau of Meteorology, **Australia**

Nov 2024 – 2025 main contributions

1: South America verification Pilot Project

ECMWF and SMN Argentina (R.deElia), monthly telecons since Sept. 2024
summary statistics + in-depth HIW case studies (heat wave, zonda, precipitation)
Outcomes: ECMWF tech memo + 2 reports + peer review publication on zonda

2: Met.Apps. Special Issue in Forecast Verification Research.



Guest Editors: B.Casati, C.**Coelho**, B.**Ebert,** M. **Dorninger.** ~20 papers (several already accepted, a few still in review). Editorial expected by the end of 2025

3: Advance verification methods for AI models

- VerAI workshop June 2025
- VerAIops workshop October 2025
- WGNE+JWGFVR WP-MIP

4: Contribution to renewal of WIPPS verification standards

includes power-spectra diagnostics to compare ML-based and traditional prediction systems

WORLD METEOROLOGICAL ORGANIZATION

WWRP WORLD WEATHER RESEARCH PROGRAMME

WCRP World Climate Research Programme

WGNE Working Group on Numerical Experimentation

JWGFVR

Environment and Climate Change Canada
Environnement et Changement climatique Canada

Workshop on

Verification of AI models in Operational Centers

Canadian Meteorological Center, 22-23 October 2025

https://hpfx.collab.science.gc.ca/~bca851/VerAIops2025/index.html

Background: Robust, fair, and transparent verification methods are essential to understand the characteristics and capabilities of the new ML prediction systems in comparison to traditional NWP, to guide their effective use, and to develop optimal hybrid approaches to weather forecasting.

Aims: Advance our understanding; plan WP-MIP verif activities; feed into WIPPS operational standards

How to:

1. adapting established (operational) verification practices to the specific features of AI-based forecasts
2. testing existing (spatial) metrics and developing new diagnostics for meaningful comparisons

⇒ ensure that WMO Members and forecast users can **trust and interpret** outputs from both paradigms with confidence.

Participation: more than 200 scientists from MRD+MSC (ECCC) and WGNE+JWGFVR international research network

Leveraging on the outcomes of the preceding VerAI workshops

Jochen Broecker, University of Reading, UK

- VerAI 2024 workshop: https://www.met.reading.ac.uk/~pt904209/VerAI2024/index.html
- VerAI 2025 workshop: https://www.met.reading.ac.uk/~pt904209/VerAI2025/index.html

Three main themes were identified :

1. Statistical properties of the AI-based models and how to evaluate them;
2. Meaningful benchmarking and reliable reference datasets;
3. physical realism and explainability.

The latter relates to **the legal and ethical requirements for "explainable AI"**

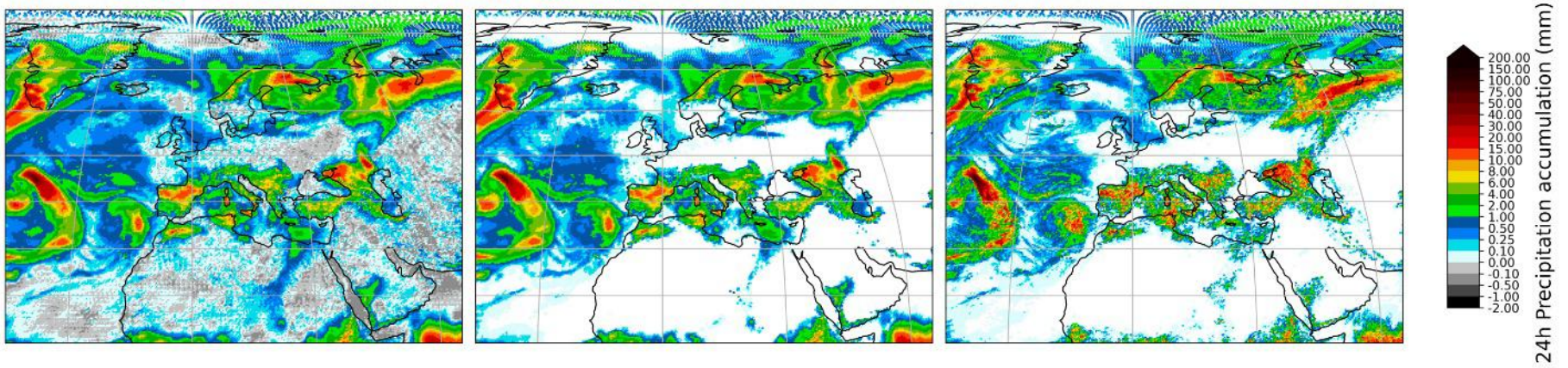
- See recommendations from the "Hiroshima Process International Code of Conduct for Advanced AI systems" and "G7 Leaders' Statement" (https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems and https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-ai-process)
- See also the EU AI Act, recital 27 https://artificialintelligenceact.eu/recital/27/

new tools that enable **transparency** with respect to AI-generated forecasts, allowing appropriate **traceability** and **explainability** of AI-based weather prediction, are needed!

Verification, Diagnostics and Falsification

From Zied Ben-Bouallègue (ECMWF): while traditional evaluation methods focus on forecast *verification* and model *diagnostics*, the emergence of ML-based systems introduces the need for *falsification* of the underlying data-driven rules.

Image from Moldovan et al (2025)



(a) AIFS previous (b) AIFS revised (c) IFS

This new conceptual framework calls for the development of tools (e.g. see M.Höver, S.Wahl talks) for testing for physical inconsistencies (e.g. negative precipitation, featuring in some data-driven models) and compliance with physical laws (such as energy and dynamical balance, e.g. Bonavita 2024 GRL), which are needed to provide confidence in AIWP and enhance explainability by aligning with physical reasoning.

Verification measures for AI training: inherent responsibilities and new meta-verification tools

Environment and Climate Change Canada
Environnement et Changement climatique Canada

JWGFVR

From B.Casati presentation:

- With the advent of ML, Verification metrics have assumed a new role, from assessing the performance after the forecast is issued, to be cost functions for training – and hence shaping – the weather forecast.
- This comes with inherent responsibilities, since the metric used to train defines the characteristics of the produced data-driven forecast (e.g. RMSE → smooth forecast).
- Subproduct: AI training can be used as a meta-verification tool, since reveals the forecast characteristics which can be used for hedging, and artificially improve, the score.
- Examples shown at the workshop are:
  1. the very well-known smoothing if training with RMSE (cf. C.Subic presentation),
  2. precipitation over-forecasting if training with CSI (cf. D.Brunet presentation).

Benchmarking reference datasets and incestuousness

Meaningful benchmarking and reference datasets are a concern (see S.Lerch, B.Casati, J.Broeker talks), as fair comparisons are challenging:

- AIWP models are optimized based on the RMSE as a loss function, which might put them at an unfair advantage in RMSE-based comparisons – this consists in **metric incestuousness!**
- What is the ground truth (ERA5 vs. IFS analysis vs. own analysis vs. obs)? What is the effect of fine-tuning AIWP models towards a specific observational dataset? Incestuousness between ERA5 as verification reference and ERA5-trained AI-models is concerning – this consists in **dataset incestuousness!**
- **Benchmarking**: AIWP models are usually compared to raw NWP forecasts, whereas a (more) practically relevant and fair comparison might be against post-processed NWP forecasts - S.Lerch introduces the potential CRPS: deterministic model dressed+debiased, enables the use of CRPS = proper scoring rule

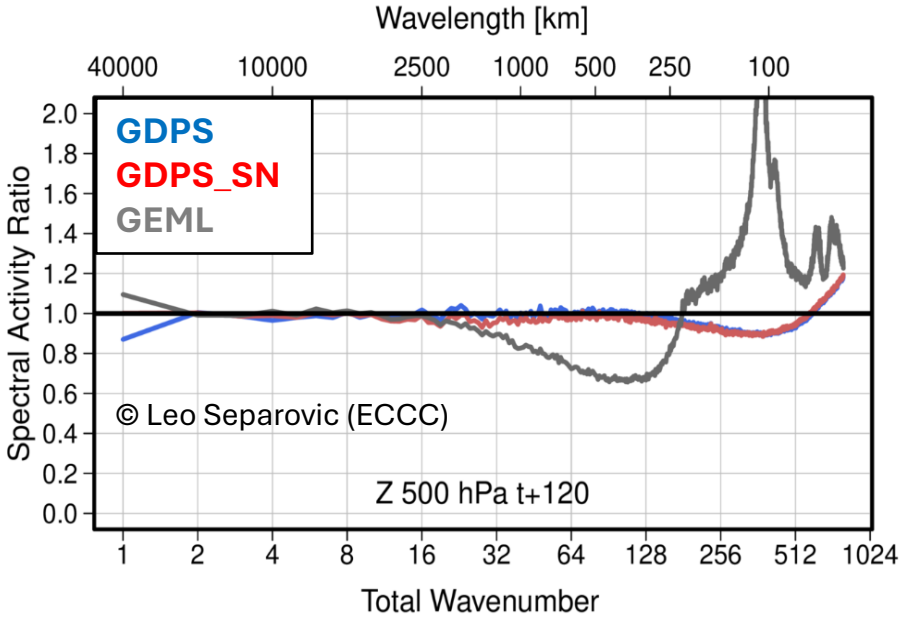
➡ **WP-MIP experimental design** – verification of SIC and OIC, against IFS and own analysis + obs, stationary vs transient weather, portfolio of different metrics, …

Enhanced spatial verification diagnostics

Environment and Climate Change Canada
Environnement et Changement climatique Canada

Traditional summary statistics are limited. Complementing them with spatial verification diagnostics can provide more in-depth knowledge of the forecast characteristics. Examples:

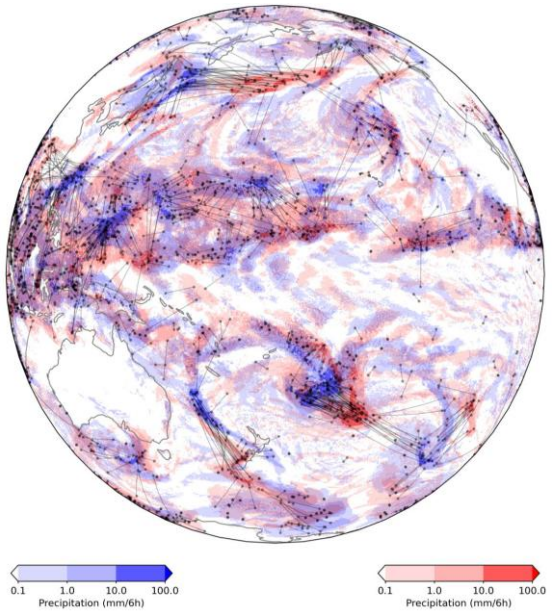
spectral decomposition diagnoses the forecast effective resolution and enables identifying unphysical compensating behaviors that are hidden when using integral metrics (cf. Leo Separovic talk).



Also other spatial metrics, such as the FSS and its modified versions, the wavelet-based SBE, and the SSMI were proposed as additional diagnostics (cf. B.Antonio talk).

UNIVERSITY OF OXFORD

Comparing AI-based models to traditional NWP models with distance metrics, such as the PAD (cf. Skok talk), is of high interest, given the positive results AI models exhibit on TC track position (cf. G.Chen talk) and correct large scale positioning / phase error (cf. C.Subic and Leo Separovic talks).

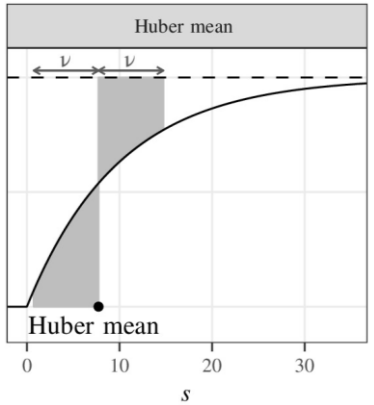


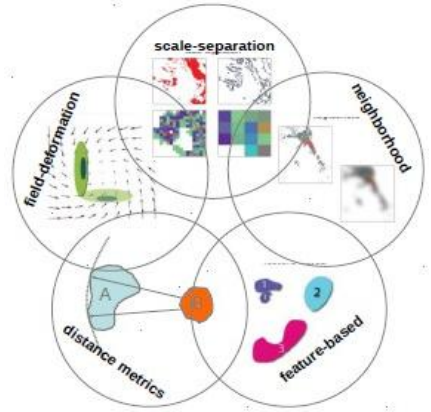
- Skok (2023): Precipitation Attribution Distance (PAD) Atm. Research
- Skok and Lledó (2025), *QJRMS*.

UNIVERZA V LJUBLJANI

Fair and Proper Scoring Rules for Ensembles

A large portfolio of **fair and proper scoring rules** has been developed in the past decade, and should be imported **into operational verification practices**, to fairly compare traditional NWP-based ensembles and the data-driven ensembles, which are advantaged by the significantly larger number of members (cf. J. Broeker, A.Dirkson talks).





Romain Pic: proper scoring rules have however some **limitation in interpretability** - wish to render them more interpretable, while maintaining their statistical rigor!

Aggregation and transformation principles can be leveraged for **bridging the gap** between **consistent and proper scoring rules** and **spatial verification methods**. Examples mentioned during the workshop:

- Verification in AI model lower dimensional "latent space" (cf. J. Broecker talk);
- Verification of deterministic forecasts in spectral space (cf. C. Subic and L. Separovic talks)
- Application of proper scoring rules to spectral components was also suggested for ensembles (Q&A following A.Dirkson talk)

Extreme Events

- **Benchmark extreme-event datasets** are being developed, for comparing AI versus NWP models by displaying verification results on common case studies:
  - ❑ Extreme Weather Bench (cf. A.McGovern talk)
  - ❑ ECMWF extreme event catalogue (L.Magnusson)
  - ❑ EW4All "iconic" case studies

Brightband

ECMWF

- The issue of gathering a sufficiently large sample for robust analysis of extreme events, without overlapping with AIWP training periods, was discussed.
  - In the previous VerAI workshop it was advanced the idea of artificially inserting into the training dataset more extreme events (e.g. storms or hurricanes), to help the data-driven models learn their dynamics; however, this would distort their climatological frequency.
  - In the VerAIops workshop, "AI Retraining Without Iconic Events (AIRWIE)" was a proposed solution to establish a leave-one-out cross-validation protocol for selected iconic events (cf. S.Nath talk).

UNIVERSITY OF OXFORD

- Existing extremal dependence indices (e.g. Ferro & Stephenson, 2011, W&F) and consistent scoring functions (e.g. Taggart, 2021, QJRMS) should be adopted in operational verification practices, to avoid the forecast dilemma (Lerch et al, 2017, Stat Sci) leading to hedging (cf. D.Brunet talk).

Workshop outcomes

https://hpfx.collab.science.gc.ca/~bca851/VerAlops2025/index.html

Plan for verification activities within the WGNE Weather Prediction Model Intercomparison Project (WP-MIP), which ultimately aims to guide the WIPPS standardized verification. Several researchers from international institutions have agreed on performing a coordinated verification exercise on the WP-MIP dataset

- summary verification statistics, for experiences with own and same initial condition, against observation and against own and IFS analysis, to address the incestuousness problem for climatology and anomalies, to separate stationary versus transient weather skill
- explainability: apply diagnostic techniques investigating error sources; develop diagnostics and a validation framework addressing physical consistency and falsification
- Spatial verification methods: apply spectral-based verification, neighbourhood, morphing, distance metrics. Asses the discriminatory power and ability of these spatial techniques to reproduce human judgement by comparing to a blind-subjective WP-MIP assessment.
- Apply extreme dependence indices and consistent scoring functions, which discourage hedging, for forecast of extreme events
- tropical cyclone verification and regional studies, e.g. over Africa or South America, leveraging on local knowledge and observations are also planned.

Coordination performed on a **github platform** – open community, anyone can propose contribution!
The group plans to have the results of these studies be published in an **AMS special collection** (still to be proposed, with submission start-date in June 2026).
The possibility of contributing the developed methods to a **centralized verification tool** (METplus vs Scores) and to display verification results on a **centralized webpage** have been also discussed … (**TBD**)

Extras

Goals for 2026

1. **Contribute to the WP-MIP project in collaboration with WGNE: identify and/or develop novel evaluation techniques, appropriate for the comparison of data-driven and physically-based models.**

2. **Support, guide and contribute to the new spatial verification project for ensembles (initiated by Romain Pic, Geneva University), which aims to bridge the gap between the spatial verification community and the proper scoring rule community**.

3. Organize AI-verif sessions at EGU and EMS, and the third edition of the Ver-AI workshop on Verification of AI-Based Meteorological Forecasts, following the success of the previous two editions (see links below)

    1. https://www.met.reading.ac.uk/~pt904209/VerAI2024/index.html
    2. https://www.met.reading.ac.uk/~pt904209/VerAI2025/index.html

4. Start planning for the next International Verification Method Workshop and associated tutorial, likely to happen in 2027 (TBC).

5. Web resources: the JWGFVR has started migrating and renewing the Forecast Verification FAQ (new url to come)

Joint Working Group in Forecast Verification Research

https://community.wmo.int/activity-areas/wwrp/wwrp-working-groups/wwrp-forecast-verification-research

Mission: The JWGFVR aims to advance the development and application of improved diagnostics and verification methods to assess the quality and enable improvement of weather and environmental predictions, for time scales encompassing from weather forecasts to sub-seasonal and seasonal predictions, to decadal and climate projections.

Promote good verification practices :

- Verification web-page
- Verification tutorials
- Verification software
- WMO recommendation reports and verification standards for operational centers -> INFCOM

Advance verification research:

- Spatial verification method intercomparisons
- International verification methods workshops
- Verification challenges
- Special issues & publications

Support verification activities in **WWRP and WGNE/WCRP**

- South Am. PP
- AvRDP2
- HIW
- SAGE
- PCAPS
- InPRHA
- TC-PFP
- URBAN
- PEOPLE
- ADVANCE
- WGNE WP-MIP
- WGSIP
- WGCM

Verification Workshops and Tutorials

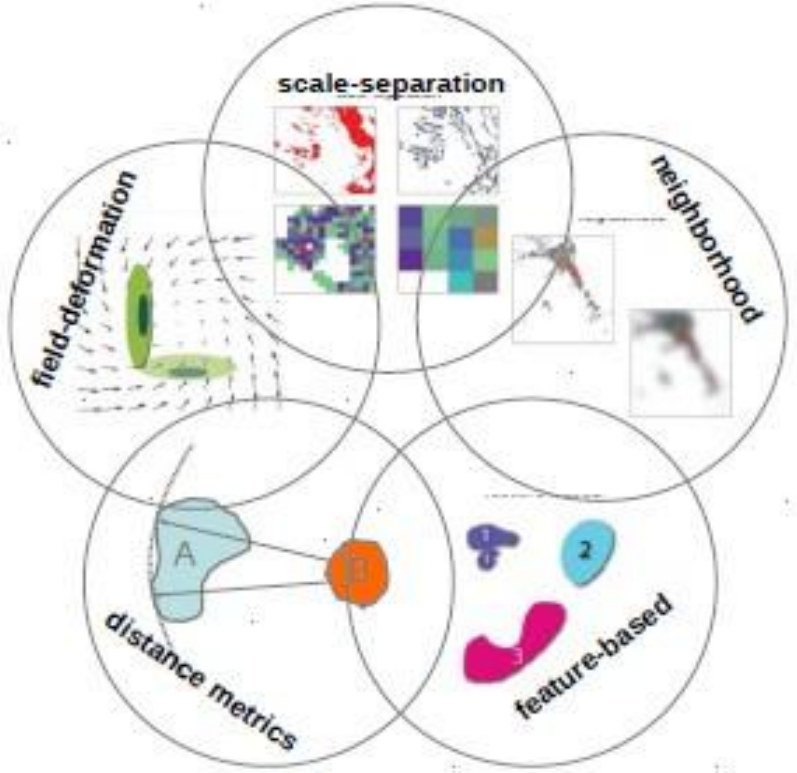
- 30 July – 1 Aug **2002, Boulder**: "Making Verification More Meaningful" (Barb Brown).
- 15-17 Sept **2004, Montreal**: 2nd International Verification Workshop (Laurie Wilson)
- 31 Jan – 2 Feb **2007, Reading**: 3rd International Verification Workshop & Tutorials (Anna Ghelli)
  https://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/2007-international-verification-methods
  Ebert and Ghelli (2008) ed. Met Apps Special Issue; Casati et al (2008) review article
- 4 -10 June **2009, Helsinki**: 4th International Verification Workshop & Tutorials (Pertti Nurmi)
  https://space.fmi.fi/Verification2009/
- 1-7 Dec **2011, Melbourne**: 5th International Verification Workshop & Tutorials (Beth Ebert)
  Ebert et al (2013) review article
- 13-19 March **2014, New Delhi**: 6th International Verification Workshop & Tutorials (Raghu Ashrit)
- 3-11 May **2017, Berlin**: 7th International Verification Workshop & Tutorials (Martin Goeber)
  https://www.7thverificationworkshop.de
  Dorninger et al (2018) ed. Met Zet special issue; Dorninger et al (2020) ed. Met Apps special issue.
- 9-20 November **2020, online,** 8th International Verification Method Workshop (Barbara Casati & Manfred Dorninger)
  https://jwgfvr.univie.ac.at Casati et al (2021) BAMS workshop summary
- 21-25 June **2021, online:** MPE-CDT + JWGFVR verification summer school
  https://mpecdt.ac.uk/mpe-cdt-jwgfvr-forecast-verification-summer-school
- 15-22 May 2024, Cape Town: 9th International Verification Method Workshop (B.Casati) and Tutorial (C.Marsigli), SAWS host (S.Landman): https://ivmw2024.weathersa.co.za MetApps Special Issues (B.Casati C.Coelho)

SPATIAL VERIFICATION METHODS

- Account for **coherent spatial structure** and the presence of **features**
- Aim to provide information on **error in physical terms (meaningful verification): e.g.** assess **scale structure** and **displacement error** (separately from **intensity error**)
- Account for **small time-space uncertainties** (avoid **double-penalty** issue)

- Spatial Verification Inter-Comparison Project (ICP) Gilleland et al (2010), BAMS
- Mesoscale Verification Intercomparison in complex Terrain (MesoVICT) Dorninger et al (2018), BAMS

Description, case-studies and references at:
https://projects.ral.ucar.edu/icp/



Open source community verification tools:

- SpatialVx R verification package (E.Gilleland)
- MET and METplus (NCAR): https://dtcenter.org/community-code/metplus

WILEY

Main contribution 2:

Special Issue in Meteorological Applications
Title: Recent Advancements in Forecast Verification Research

Chief Editors: Dino Zardi (universita di Trento, Italy) **Cristina Charlton Perez** (UK Met Office)
Guest Editors:
Dr. Barbara Casati, Environment and Climate Change Canada (ECCC), Canada
Dr. Caio Coelho, Instituto Nacional de Pesquisas Espaciais (INPE), Brazil
Dr. Manfred Dorninger, University of Vienna, Austria
Dr. Beth Ebert, Bureau of Meteorology, Australia

Submission Deadline: 31st May 2025

Coordinated by the World Meteorological Organization (WMO) Joint Working Group on Forecast Verification Research (JWGFVR) https://community.wmo.int/en/activity-areas/wwrp/wwrp-working-groups/wwrp-forecast-verification-research, **a joint group of the World Weather Research Programme (WWRP) and the Working Group on Numerical Experimentation (WGNE) of the World Climate Research Programme (WCRP)**

Will contain papers of work presented at the **9th International Verification Methods Workshop (IVMW)**, May 2024, **ECMWF Workshop on Diagnostics for Global Weather Prediction**, September 2024, **and additional contributions** on topics including:

- physical process diag
- error tracking method
- spatial verification me
- verification of high im
- unconventional observations (data from citizen science, social media etc.) for verification;
- verification of probabilistic and ensemble forecasts;
- inference and properties of verification methods (meta-verification studies);
- representativeness and observation uncertainty in verification practices;
- user-oriented verification and estimation of forecast value (assessment of the whole forecast-quality to user-value chain);
- verification tools and software;
- verification practices in operational environments;
- verification of AI-ML-based forecasts.

~ 20 papers received (several accepted, a few in review) Editorial (lead by the JWGFVR co-chairs) expected by the end of 2025

Meteorological Applications Special Issue Call for Papers, please refer to:
https://rmets.onlinelibrary.wiley.com/hub/journal/14698080/call-for-papers

Main contribution 3: Promote research to advance AI forecast verification, including the assessment of physical-coherence and process-diagnostics

1. Several JWGFVR members (Zied Ben Bouallegue, Barbara Casati, Caio Coelho, Sabrina Whal, Nicholas Loveday, Jochen Broecker, Anumeha Dube) are involved in **the Weather Prediction Model Intercomparison Project (WP-MIP, https://www.wcrp-esmo.org/activities/wp-mip)** initiated by WGNE**. The project aims to create a testbed of physical, hybrid and AI weather prediction models**, to enable intercomparison and hence accelerate the development of the systems themselves**. The JWGFVR will play a central role in the WP-MIP project, which is identifying, developing and implementing novel evaluation techniques, appropriate for comparing data-driven and physically-based models**. These include scale-separation (spectral) methods, optimal transport techniques, process-coherence diagnostics, extreme proper scoring rules. Initial JWGFVR contribution has been to the white paper, which lay the basis of the WP-MIP.
2. Two JWGFVR members (Jochen Bröecker and Zied Ben Bouallegue**) organized the Ver-AI workshop** (https://www.met.reading.ac.uk/~pt904209/VerAI2025/index.html) on Verification of AI-Based Meteorological Forecasts, 23-24 June 2025, at Reading University, which is envisaged to be the official platform for WP-MIP verification research exchanges on regular basis.
3. **The JWGFVR is planning a face-to-face meeting in October 2025 in Montreal, with a two-day workshop on AI verification in operational practices, connected to WP-MIP. This workshop aims also to respond to the WIPPS requirements of including AI and hybrid models into the standardized verification exchange** (currently under revision). https://hpfx.collab.science.gc.ca/~bca851/VerAIops2025/

Verification of AI models in Operational Centers

Montreal and online, 22-23 October 2025

Contact and Organizer: *barbara.casati@ec.gc.ca*

https://hpfx.collab.science.gc.ca/~bca851/VerAIops2025/

Workshop background information

The recent rapid improvement of AI-based weather prediction has prompted national meteorological centres to begin operationalizing pure-AI and hybrid models, integrating AI systems with the traditional physically-based Numerical Weather Prediction (NWP) systems. Robust, fair, and transparent verification methods are essential to understand the characteristics and capabilities of these new systems in comparison to traditional NWP, to guide their effective use, and to pave the way toward optimal hybrid approaches to weather forecasting.

This workshop brings together international scientists and researchers from the WWRP Joint Working Group on Forecast Verification Research (JWGFVR), from the WCRP Working Group on Numerical Experiment (WGNE), and from the Meteorological Service of Canada (MSC) and the Meteorological Research Division (MRD) of Environment and Climate Change Canada (ECCC), to address emerging challenges in the operational verification of AI models alongside traditional NWP systems. We aim to examine how established verification practices can be adapted to account for the distinct characteristics of AI-based forecasts, explore metrics and novel diagnostics for meaningful comparisons, and promote best practices to ensure that WMO Members and forecast users can trust and interpret outputs from both paradigms with confidence. Building on discussions from the two preceding (2024, 2025) VerAI workshops hosted at the University of Reading (UK), this event also aims to establish a baseline for verification activities within the WGNE Weather Predition Model Intercomparison Projet (WP-MIP), and will inform the ongoing revision of the WIPPS standardized verification.

Contributions and Registration

Contributions on verification approaches for comparing AI, hybrid and traditional NWP models are welcome. Topics of interest include (but are not limited to) the following verification research topics:

- Developing best practices for fair comparison between AI, hybrid and physical NWP models.
- Verification observation reference and incestuousness of verification against own analysis.
- Spatial methods: neighborhood, filtering and scale-separation approaches; field morphing and displacement methods.
- Process diagnostics and physical coherence assessment.
- Verification of extreme and rare events.

Main contribution 4 - Facilitate operational forecast verification score exchange, including modernization of the WIPPS verification standards

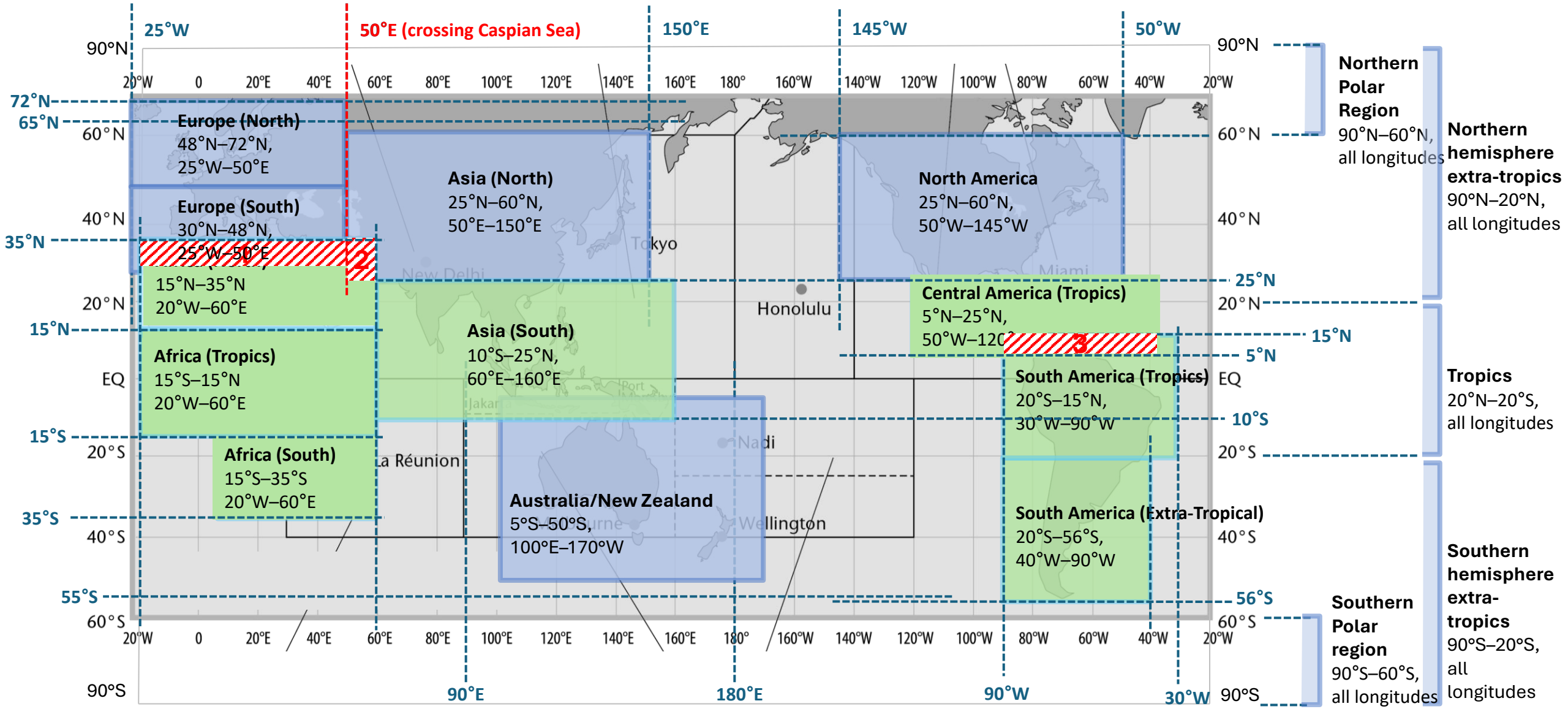
led by Thomas Haiden (ex JWGFVR), include contributions from Barbara Casati and other previous JWGFVR members (Beth Ebert, Chiara Marsigli).

The WIPPS Task Team on NWP Standardized Verification (TT-NWP-SV) is revising the WIPPS verification standards for the WMO (manual n. 485) coordinated score exchange between National Weather Services (https://community.wmo.int/en/forecast-verifications), to become effective in September 2026, mandatory in September 2028.

- New guidance includes domains covering all WMO regions (whereas the current guidance omits South America, Africa, and other regions) and recommends exchange for regional models (whereas the current guidance treats only global models).
- The TT is revising the meteorological variables (upper-air and surface, recommended and mandatory) to strengthen alignment between the Deterministic and Ensemble Prediction System standardized verification exchange.
- The TT is including power-spectra diagnostics to compare AI-based and traditional prediction systems, to identify smoothing effects and compensating errors on different scales, as well as to quantify skill sources due to representing transient weather versus climatology.

INFCOM: Task Team NWP Standard Verification (TT-NWPSV)

Proposed revised and new domains – overlapping area



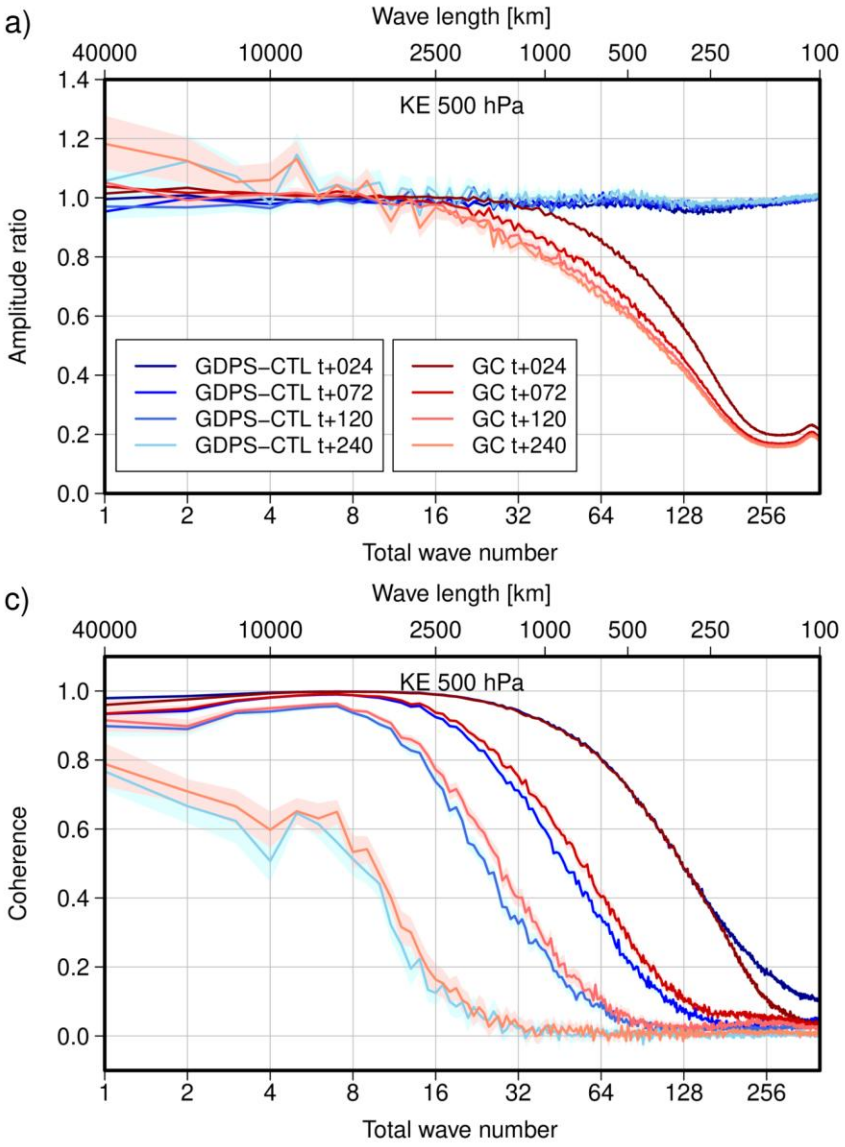
VERIFICATION OF AI-MODELS WITH POWER SPECTRA (1/4)

f and a denote forecast and analysis, and the corresponding anomalies are defined as $f' = f - f_C$ and $a' = a - a_C$

The activity is simply defined as the standard deviation σ of the anomalies. The activity ratio is a measure of bias / amplitude error

The anomaly correlation (= covariance / product f' and a' standard deviations) coefficient provides a complementary measure of accuracy / skill / phase error

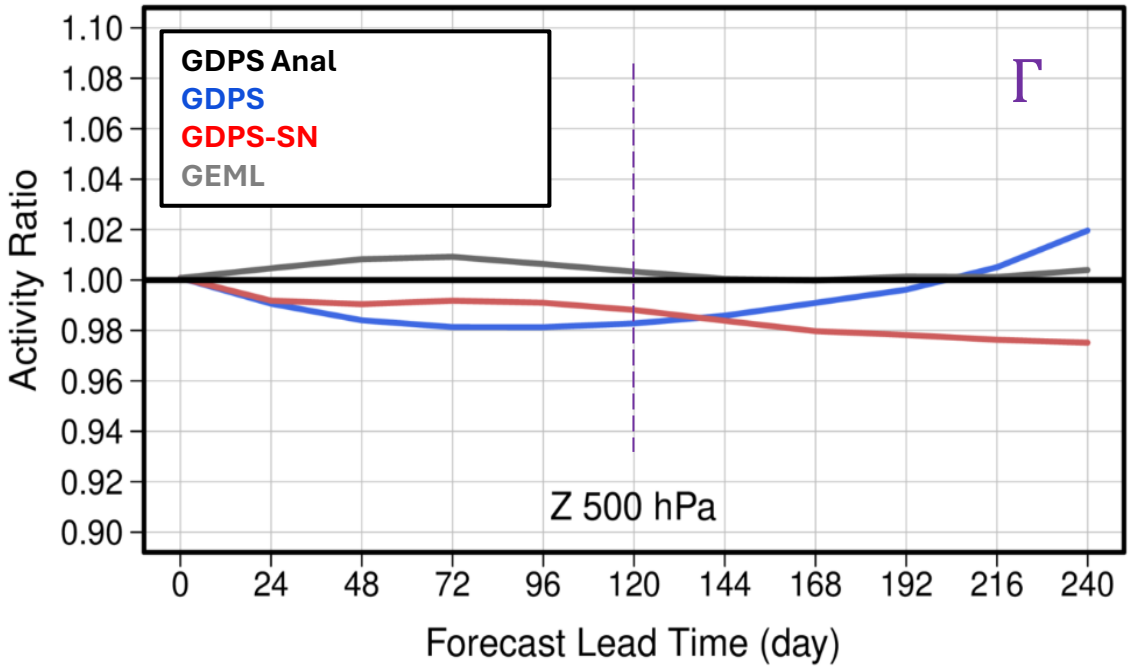
The gridded weather fields can be expressed as truncated spherical harmonic expansions; the variance and covariance can be represented as a sum over power spectra, as a function of total wavenumber n. Activity ratio and anomaly correlation can be calculated as function of spherical harmonic wavenumber n, which is associated to a physical scale L = 2πR/n (in km)

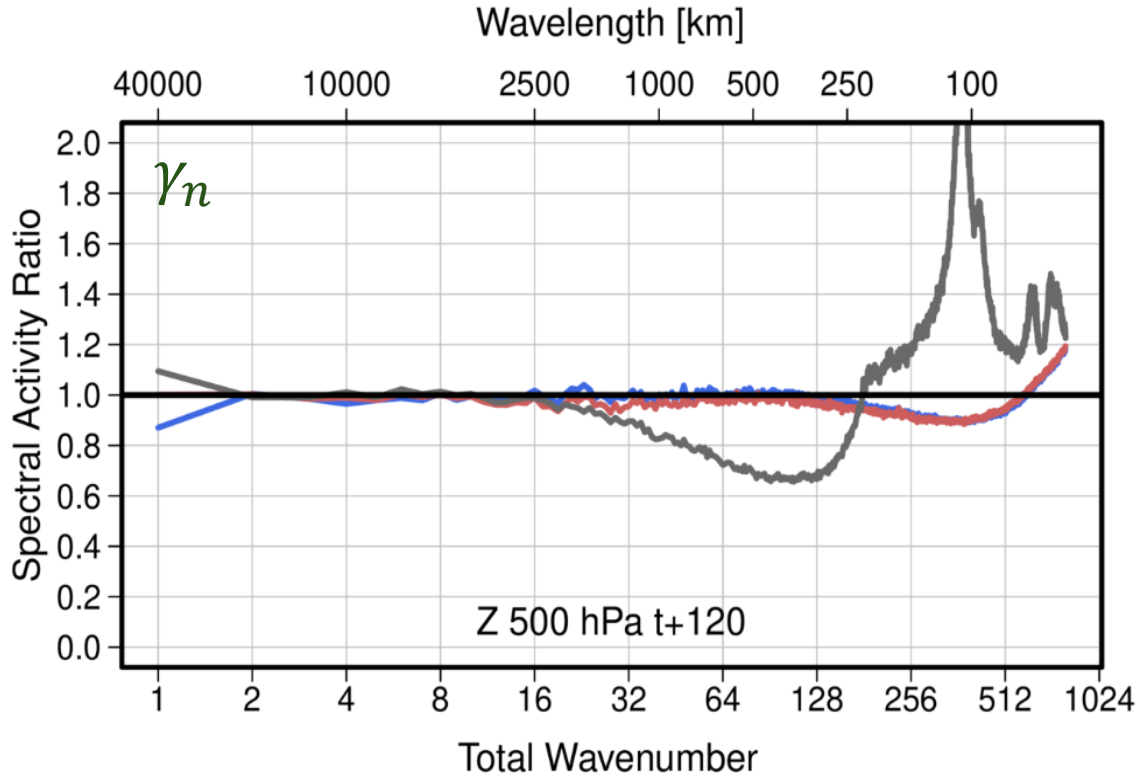


All material is © from Leo Separovic (ECCC) and Husain et al (2024) https://arxiv.org/abs/2407.06100

VERIFICATION OF AI-MODELS WITH POWER SPECTRA (2/4)

- Example from activity and its spectral decomposition for 500-hPa GZ.
- GEML activity appears to be the closest to the analysis (grey curve).
- Only by using spectral metrics one can diagnose that GEML has unphysical compensating effects on synoptic to meso- and micro- scales



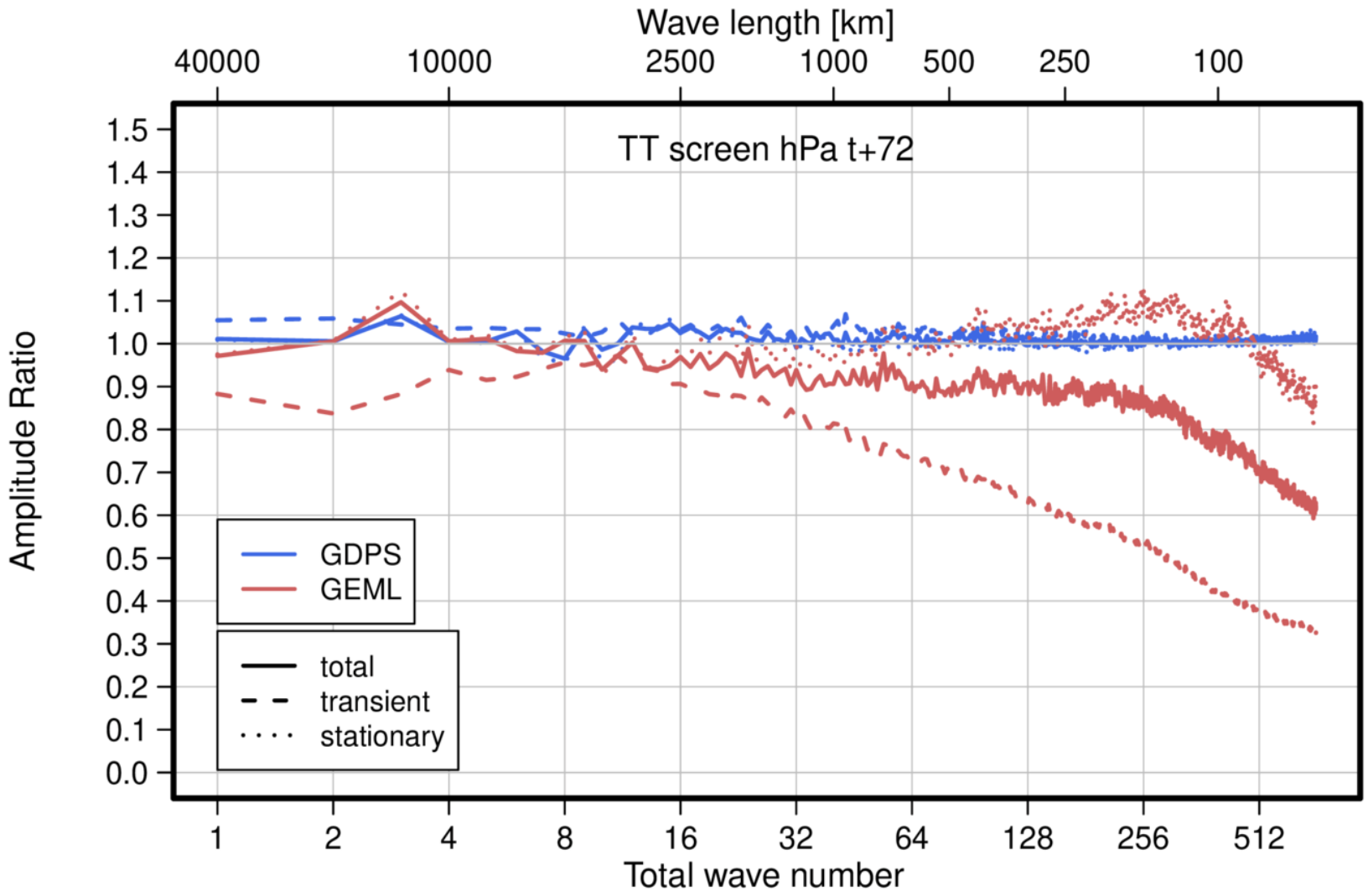


VERIFICATION OF AI-MODELS WITH POWER SPECTRA (3/4)

Separate signal into **climatology and anomalies**:

$$f = \overline{f} + f'$$
$$a = \overline{a} + a'$$

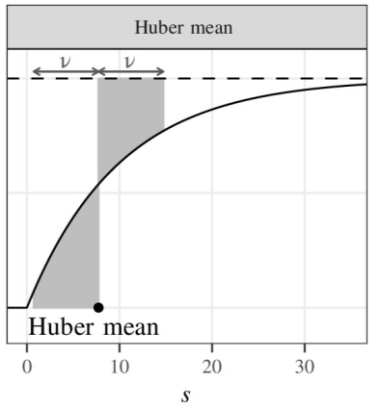
- Total: $\gamma_n(f, a)$
- Stationary, climatology: $\gamma_n(\overline{f}, \overline{a})$
- Transient, anomalies: $\gamma_n(f', a')$



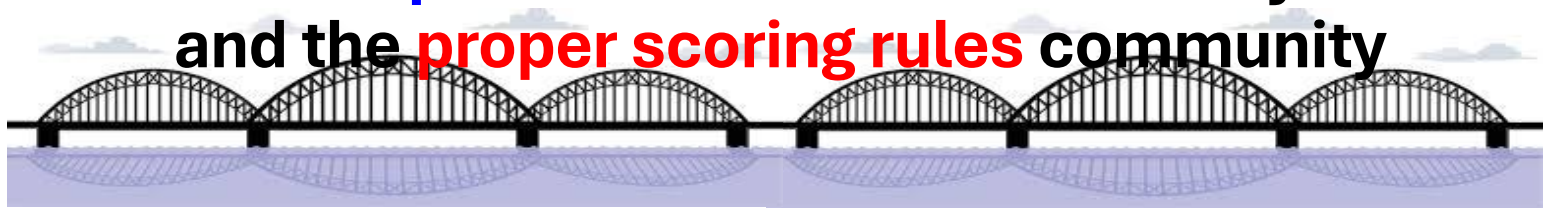
VERIFICATION OF AI-MODELS WITH POWER SPECTRA (4/4)

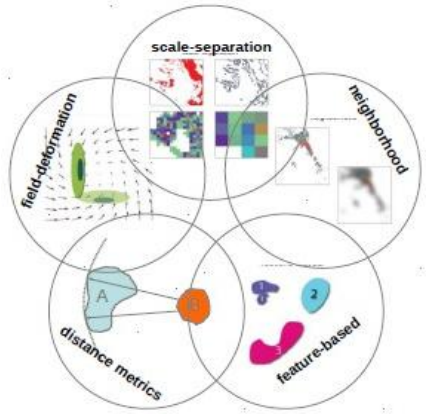
Recommendations for good verification practices:

1. To gain more insight into the effective resolution at sub-synoptic and mesoscale ranges, scores (e.g. activity) should be scale-separated using, metrics based on spectral decomposition.
   - There are compensating errors on the different scales, which need to be detected
2. Spectral metrics should be calculated and inspected separately for the stationary components (e.g. sample or long-term climatology) and transient anomalies.
   - AI models are trained on the climatology, separating it from transient weather reveals the performance aside reproducing expected weather



Bridging the gap between the spatial verification community and the proper scoring rules community





Led by **Romain Pic** (ECR, Geneva University, Switzerland) with support from all **JWGFVR** members

Participants: ~50 international verif. researchers, across 12 meteorological centers and 17 academic institutions.

Goals: bridge the gap between the **spatial verification community** and the **proper scoring rule community**, via the initiation of a **third spatial verification inter-comparison project with focus on ensembles**.

1. Organized scientific exchanges between the two communities
2. Understand and explicit the links as well as the gaps between the two approaches
3. Develop new proper spatial verification methods for ensemble forecasts
4. Focus on specific applications: physical realism, NWP vs AI, …
5. Facilitate meta-verification studies

Previous Spatial VxICP had dedicated testbed dataset: create one or leverage on existing datasets?

Modus Operandi:

Current: online meetings (~ every 2 months) featuring invited expert presentation followed by discussion – enhance dialogue between the two communities and identify the challenges to be addressed to bridge the gap.

Forthcoming:

- session at a conference (e.g., EGU) and hybrid workshop
- review of spatial verification and identification of the challenges
- white paper to clearly and publicly define the project …

Software development and training

Scores, developed by Nicholas Loveday (BoM), is an open-source python package (xarray based) for forecast verification https://scores.readthedocs.io/

Nicholas recently helped add the following methods:

- Block bootstrapping (with circular option)
- Spearman's Rank Correlation Coefficient
- SEEPS

```
>>> import numpy as np
>>> import xarray as xr
>>> from scores.processing import block_bootstrap
>>> obs = xr.DataArray(np.random.rand(100, 100), dims=["time", "space"])
>>> fcst = xr.DataArray(np.random.rand(100, 100), dims=["time", "space"])
>>> bootstrapped_obs, bootstrapped_fcst = block_bootstrap(
...     [obs, fcst],
...     blocks={"time": 10, "space": 10},
...     n_iteration=1000,
... )
```

Simple example of using block-bootstrapping in *scores* with some synthetic data

Each new method added into *scores* has an associated tutorial. For example, see https://scores.readthedocs.io/en/latest/tutorials/SEEPS.html for the SEEPS tutorial.

In the tutorial, an AIWP model (Graphcast) and a physical NWP model (ECWMF IFS HRES) are evaluated against ERA5.