Integration of AI in Global Model Development

Ron McTaggart-Cowan[1], Jean-François Caron[1], Syed Husain[1], Linus Magnusson[2] and Leo Separovic[1]

[1] Environment and Climate Change Canada
[2] European Centre for Medium-Range Weather Forecasts

Outline

Part I: A Hybrid Global Modelling System at ECCC

Introduction to component models:

- GDPS: physically based GEM model
- GEML: retrained Graphcast derivative

Implementation of global spectral nudging

Benefits of a hybrid modelling approach

Part II: The Weather Prediction Model Intercomparison Project

Project objectives:

- Forecast data generation and archiving
- Evaluation and diagnostic techniques

Protocol extension via subprojects

Global contributions to WP-MIP

Component Models

GEML

0.25° graph neural network (GraphCast L13)

Retrained using ERA5 and operational analyses

Highly skillful medium-range forecasts

Very fast inference (forecast generation)

Rapid pace of development and improvement

Small code base uses community-developed tools

GEM

0.14° L84 global deterministic prediction system (GDPS)

FISL dynamics and full physics

Full set of forecast fields

Physical and dynamical consistency

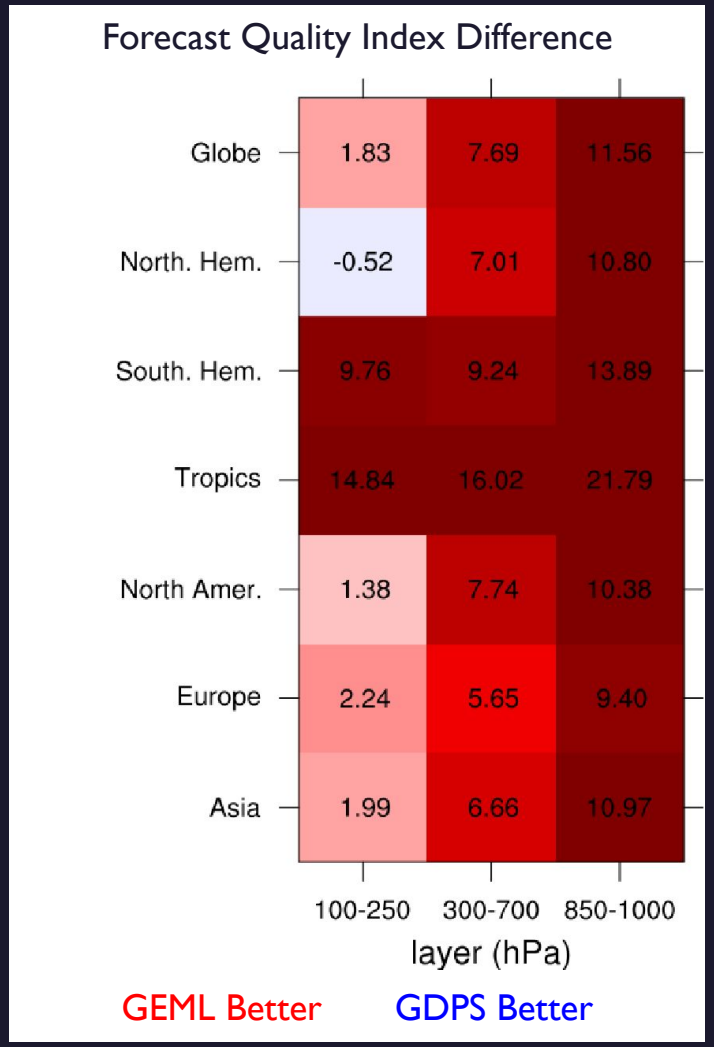
Represents climatological variability and HIW

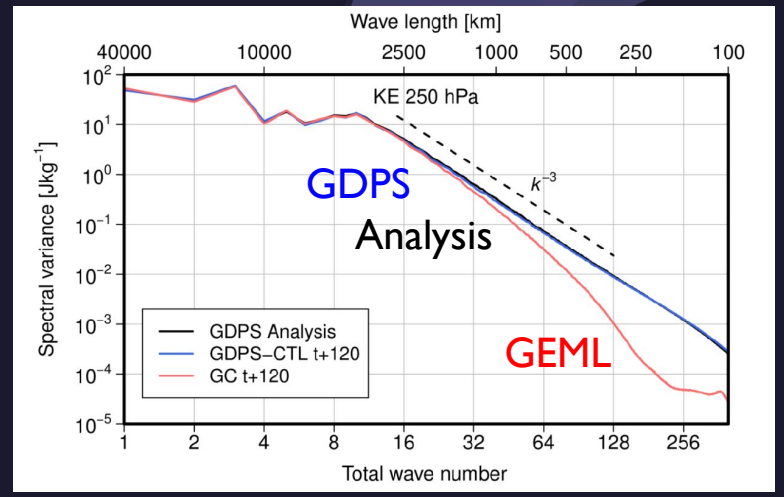
Explainability based on physical principles

Relative Strengths and Weaknesses

Most forecast fields from GEML are significantly (~10%) improved over their GDPS equivalents in RMSE-based scores, particularly in the tropics.

However, GEML fields are severely smoothed for wavenumbers well into the synoptic scales (~2000 km)





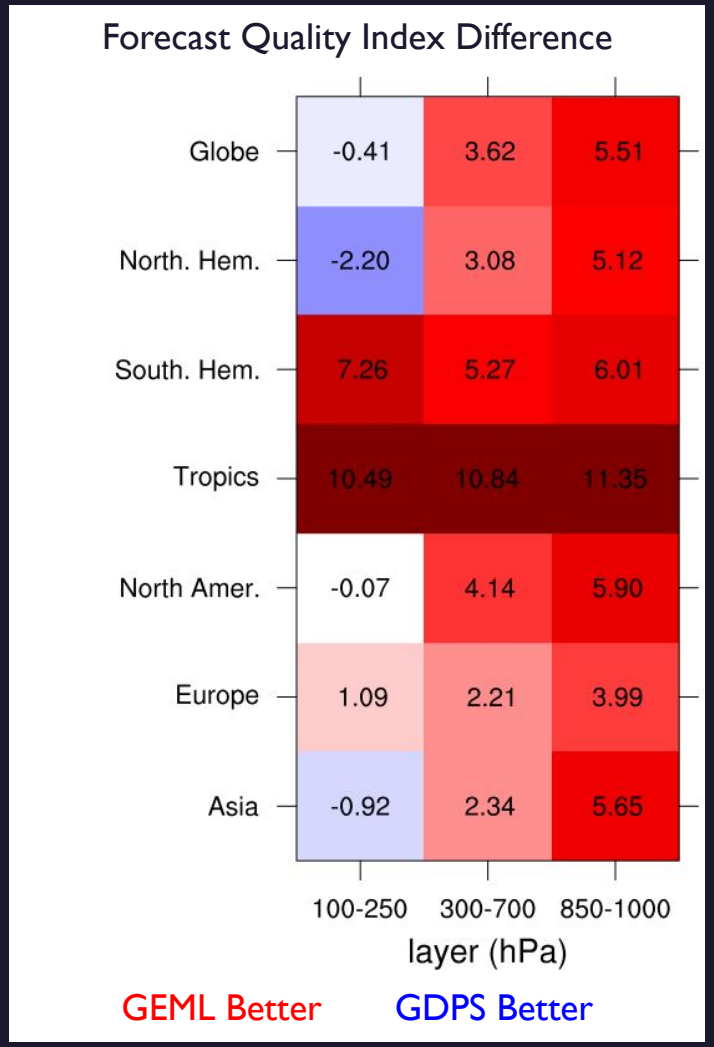
Kinetic energy spectrum at 250 hPa after 120 h for GDPS (blue) and GEML (red), with operational analyses in black for reference. Power spectra differences for other fields and levels are similar.

Aggregate RMSE-based forecast quality index for Boreal winter 2022 test period.

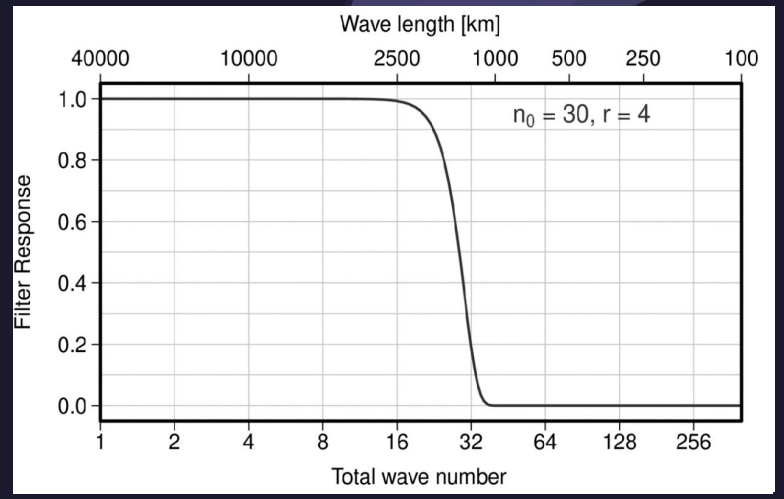
Relative Strengths and Weaknesses

GEML has "learned" not to try to forecast poorly predicted features from the L2 loss function, avoiding the double-penalty of misplacements.

When compared at the same (low) effective resolution, GEML gains are halved; however, it clearly improves large-scale predictions.

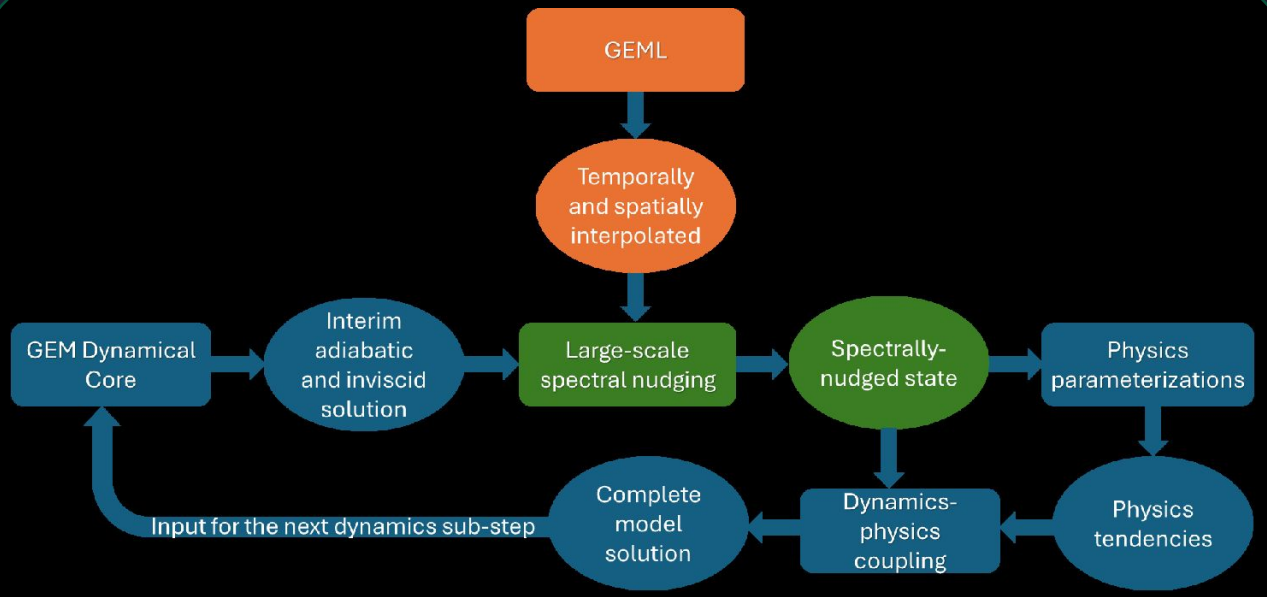


Aggregate RMSE-based forecast quality index for Boreal winter 2022 test period for filtered forecasts.



Response function of filter used to create forecast fields of comparable effective resolution for comparison.

Hybrid Model Design



The GDPS large-scale (LS > 2500 km) state is nudged towards GEML predictions at each time step.

Global spectral nudging is used to leverage the abilities of each of the component systems.

$$F = F_{GDPS} + \frac{\beta(\zeta)\chi(t)\Delta t}{\tau(t)}(F_{GEML} - F_{GDPS})_{LS}$$

Nudging is applied to winds and temperature in the free troposphere.

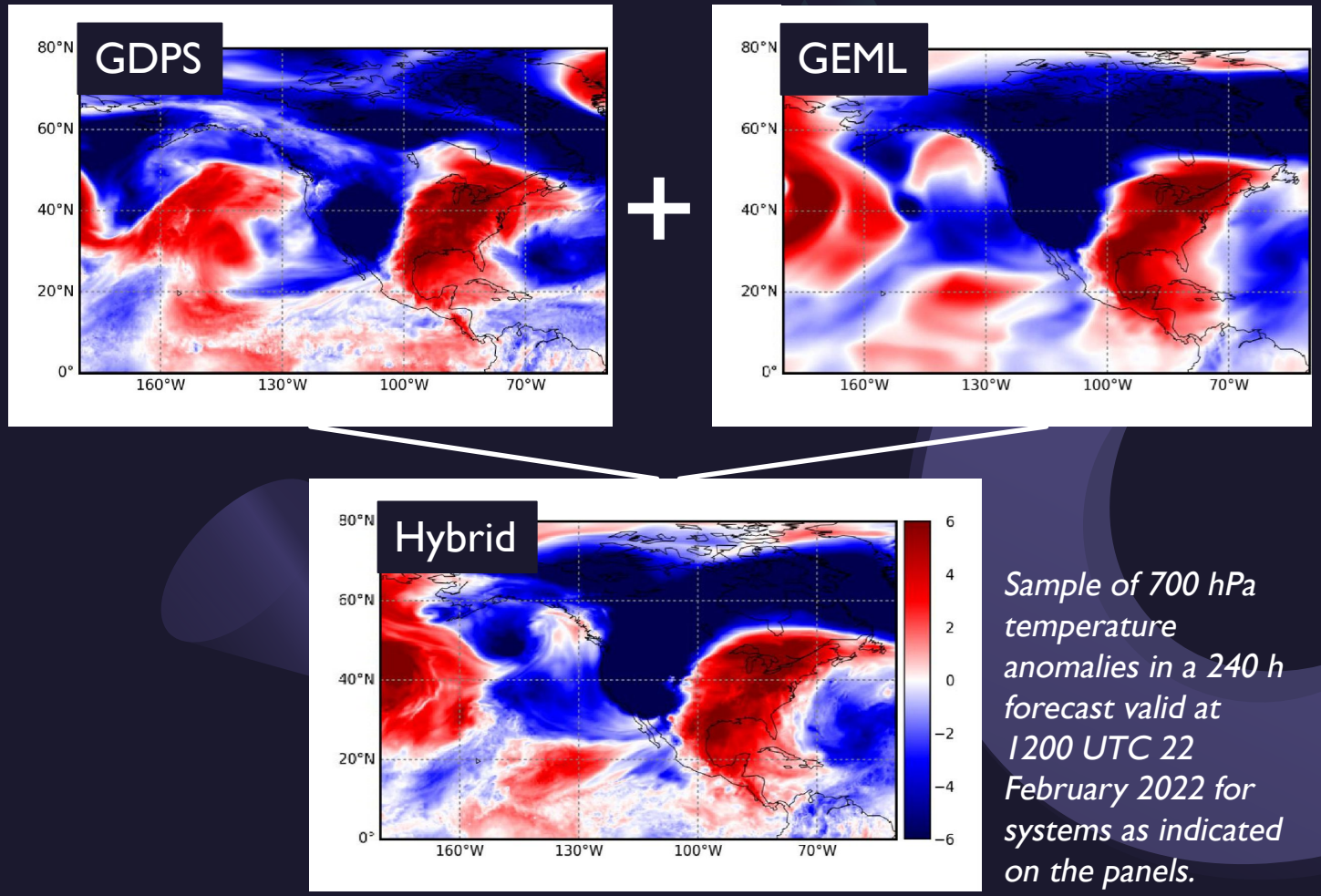
Relaxation time-scale τ=12 h keeps solutions close while allowing some freedom for the physical model.

Spectral Nudging in Action

The large-scale state of the hybrid model remains close to the GEML "guide".

The forward cascade combines with local forcings to ensure that GEM maintains a physically coherent small-scale state.

Results are similar for moisture despite the lack of nudging increment (not shown).



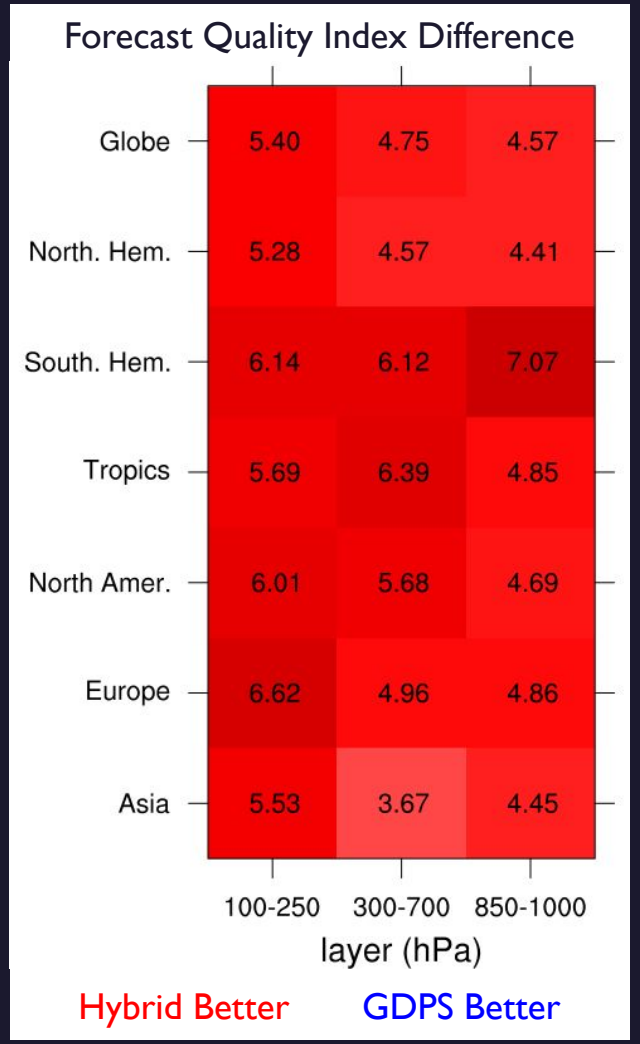
Sample of 700 hPa temperature anomalies in a 240 h forecast valid at 1200 UTC 22 February 2022 for systems as indicated on the panels.

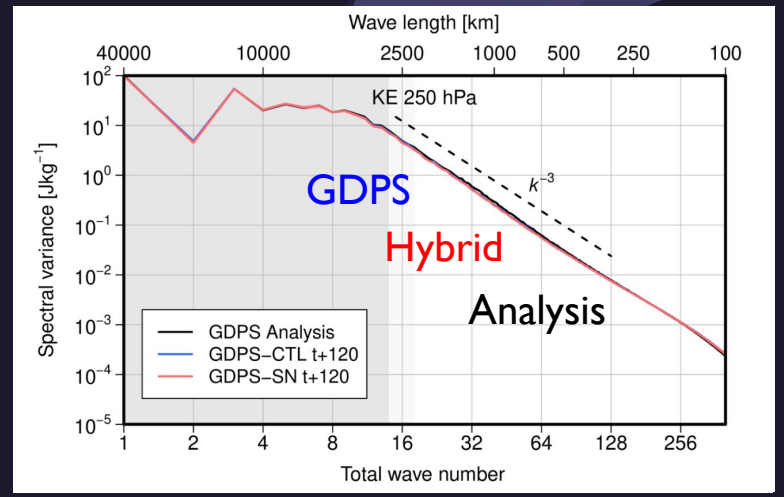
Hybrid Model Forecast Skill

Forecasts from the hybrid model are significantly (~5%) more skillful than the operational GDPS.

This improvement comes **with** a full energy spectrum and high effective resolution.

A complete set of physically consistent forecast fields is produced.

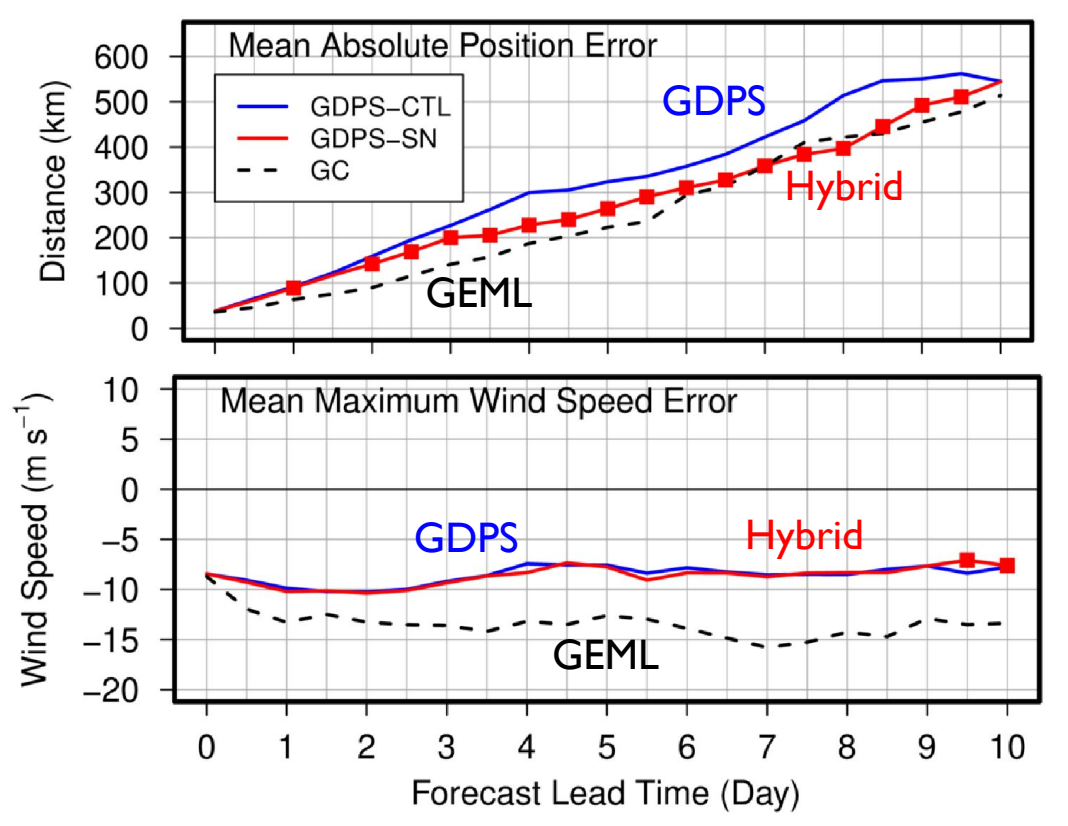




Kinetic energy spectrum at 250 hPa after 120 h for GDPS (blue) and hybrid (GDPS-SN; red), with operational analyses in black for reference. Power spectra differences for other fields and levels are similar.

Aggregate RMSE-based forecast quality index for Boreal winter 2022 test period for filtered forecasts.

High Impact Weather: Tropical Cyclones



Global tropical cyclone evaluation for JFM 2022. Statistically significant differences ($p<0.05$) between the hybrid (GDPS-SN) and GDPS are shown with filled squares.

Tropical cyclone tracks are significantly improved through spectral nudging because the large-scale state provides a better steering flow.

The well-documented weak-intensity bias in AIWP models does not affect the hybrid system because finer-scale structures (including tropical cyclones) are not directly impacted by nudging.

Part II

The Weather Prediction Model Intercomparison Project

WP-MIP

Coordinators: Ron McTaggart-Cowan (ECCC) and Linus Magnusson (ECMWF)

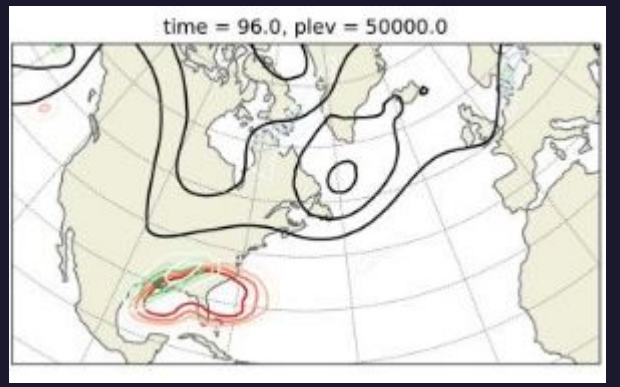
WP-MIP Objectives

> “ A centralized archive for distributed development and diagnostics

Create a testbed to accelerate development of physical, hybrid and AI weather prediction models

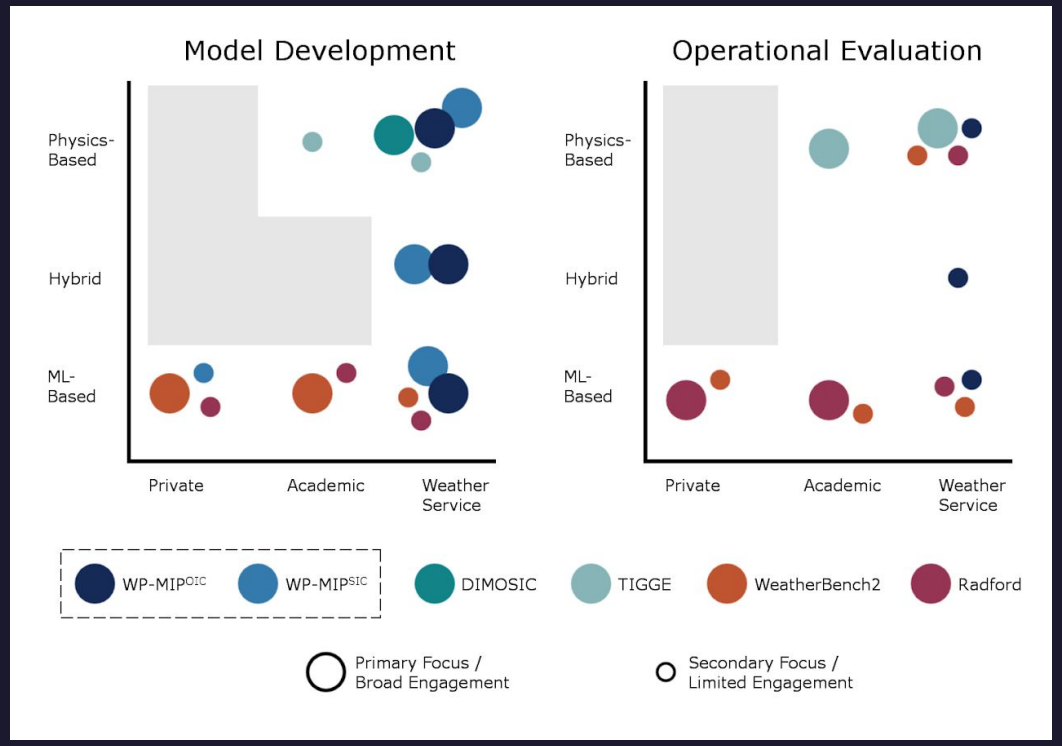
Connect model developers and evaluation experts to identify forecast strengths and weaknesses, ultimately providing guidance for optimal evaluation methods

Provide the foundation for research into the next generation of weather prediction systems



Warm conveyor belt diagnostics for DIMOSIC models (Julian Quinting, KIT).

WP-MIP in a Global Context



WP-MIP is unique in its broad inclusion of NWP, hybrid and AIWP models

Focus on national meteorological centres to accelerate the deployment of the next-generation of operational models

Direct comparison between state-of-the-art systems will help us to develop metrics to assess relative strengths and weaknesses

Project Overview

Scope

Global deterministic 10-day predictions for the first iteration of the project

Include all forms of weather prediction models: AI-based (AIWP), hybrid and traditional NWP

Focus on operational forecasting centres

Objectives

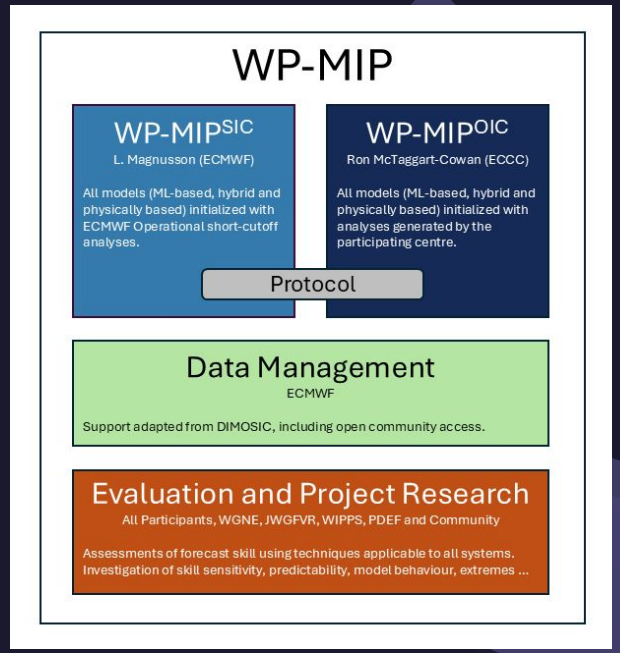
Compare AIWP inferences to hybrid and pure-NWP predictions for physical consistency, balance, effective resolution, extremes…

Document differences in sensitivity to initial conditions between AIWP and NWP models

Build a set of evaluation tools that can be used to compare skill across the full spectrum of model architectures

Protocol

Forecasts generated at 3-day intervals for 2024 to avoid AIWP fine-tuning periods.



Verification Strategy

Traditional verification (subset of WIPPS standards):

- Against observations (upperair+surface)
- Against own and independent analysis (ECMWF or … ?)

Innovative approaches:

- Processes / physical coherence
- Scale separation, filtering and aggregating planetary+synop+meso
- Address differences in resolutions, identify scale-related processes
- Phase vs amplitude, displacement vs intensity = Optimal transport / field morphing
- Address double penalty, measure displacements

Reference datasets:

- Gridded (satellites) products as references?
- Share / use common obs reference datasets …
- Analysis incestuousness and obs error to be factored in!
- Multi-center analysis (TIGGE archive)

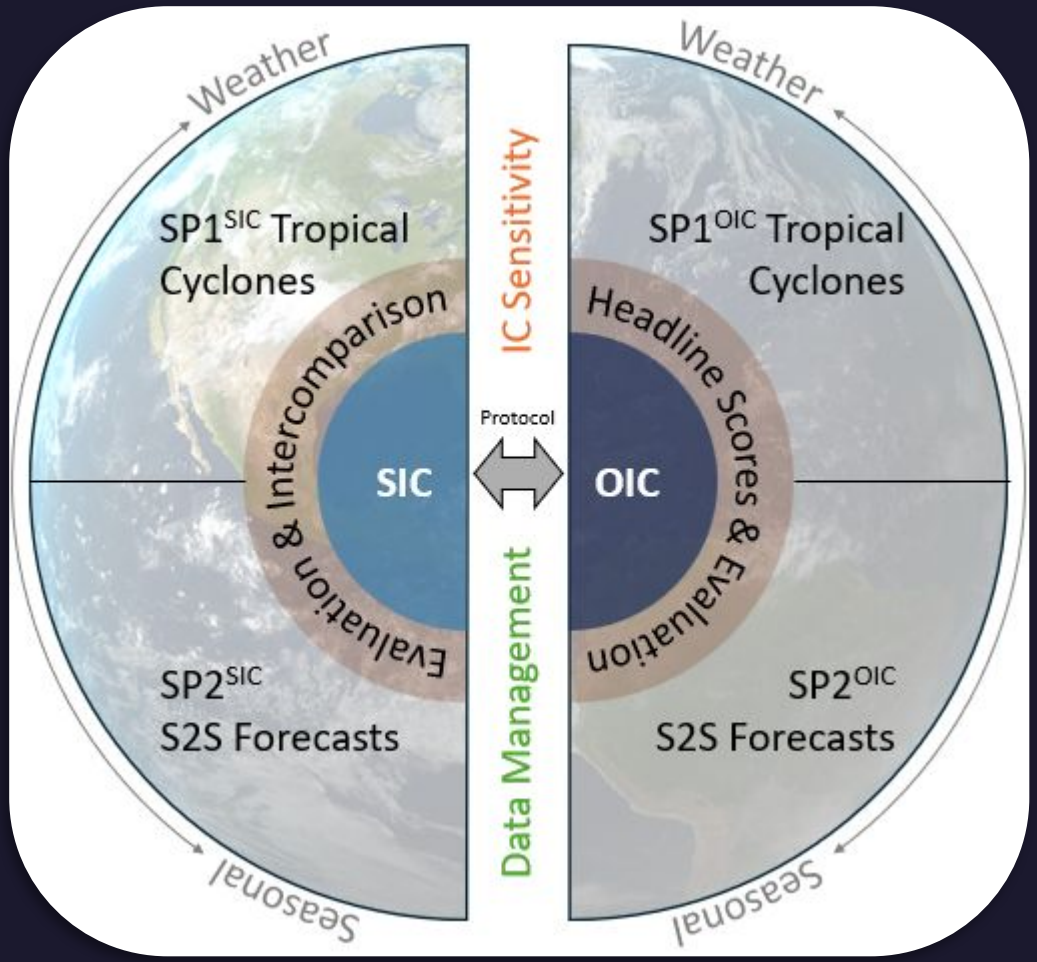
WP-MIP Extensions

The "core" WP-MIP protocols (SIC and OIC) will generate datasets for evaluation and IC sensitivity studies.

Sub-Projects (SP) can be proposed to augment the protocol for specific diagnostic needs:

- SP1: Tropical cyclone evaluation requires initializations every day
- SP2: Preliminary S2S-relevant diagnostics need runs extended to 15 days

Ensembles could be introduced like this in a future WP-MIP iteration



Global Community

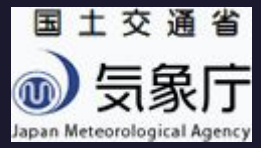
WMO Integrated Processing and Prediction System (WIPPS)

JWGFVR

Working Group on Predictability, Dynamics and Ensemble Forecasting (PDEF)

Working Group on Subseasonal to Interdecadal Prediction (WGSIP)

ECMWF



METEO FRANCE

KIAPS
KOREA INSTITUTE OF ATMOSPHERIC PREDICTION SYSTEMS

Discussion

A hybrid system that leverages the strengths of NWP and AIWP paradigms using spectral nudging significantly improves operational forecasts

- The GEML inferences serve as a "guide" for the large-scale state
- The GDPS fills the energy spectrum with physically consistent predictions for the full set of forecast fields

The strengths and weaknesses of each class of models (NWP / AIWP / hybrid) will be diagnosed and evaluated as part of WP-MIP

- A centralized database of forecasts facilitates distributed analysis
- New evaluation techniques will permit meaningful comparisons
- Evaluation recommendations for model developers and operational centres