ECMWF

Advisory Board Meeting #6 Summary of results — Online (12th January 2026)

Board members present: David Wallerstein, Christopher Frans (US Bureau of reclamation), Andrea Manrique Sunen (EDP Renewables), Saeid Vaghefi (WMO), Sebastian Lerch (KIT)

Board members who provided comments by email ahead of the meeting: Gabriela Aznar Siguan (Federal Office of Meteorology and Climatology MeteoSwiss), Jonathan Weyn (Microsoft)

Please note: these comments have been included in the present summary of results.

ECMWF staff: Frederic Vitart (Team Lead Extended-range Forecasts Team), Olga Loegel (User Engagement and Outreach Associate), Joshua Talib (Scientist on Extended-Range Forecasts), Jörn Hoffmann (Team Lead Applications and Innovation Partnerships)

Board meeting #6 Agenda:

1. Competition status
2. Open access to first period outputs
3. New results analysis tools
4. Sustaining team motivation & long-term engagement
5. Edition 2 concept

Discussion points

Board Members commented on different aspects of the proposed approach and suggested amendments.

Item #1: Competition status

- One member expressed strong satisfaction with the format of the SON Awards webinar and noted that it was engaging and effective. In particular, it was appreciated that the presenting teams relied on distinct and contrasting methodological approaches.
- Regarding future Awards webinars, the Board agreed that the current format, spotlighting a limited number of teams, works well.
- The Board noted that many teams appear to be actively iterating on and improving their models over time. Hearing from top-performing teams about how methodological refinements translated into improved scores was considered scientifically interesting, but Board members cautioned that this approach could become repetitive if the same teams are featured across multiple award instances.

- To address this, it was suggested that future webinars introduce specific thematic or targeted focuses, such as highlighting best performance in a particular geographical region, or strong skill during specific extreme event types, in order to ensure that a wider diversity of teams and contributions are represented.
- The Board recommended providing presenters with a light, guiding presentation template, for example drawing on selected questions from the model methodology questionnaire. This template should serve as a common framing tool only, without requiring disclosure of sensitive or proprietary information, recognising that teams may wish to keep certain aspects of their approaches confidential.

Item #2: Open access to first period outputs

- Board members expressed interest in understanding how extensively SON-period forecast data and the publication of methodology summaries are being used by the community, in particular the released forecast datasets and the published methodology summaries.
- It was clarified that, due to technical limitations, usage metrics cannot currently be tracked for the methodology questionnaires. However, the organisers will explore possibilities to monitor access to the submitted forecast datasets.

Item #3: New results analysis tools

- It was suggested that additional explanatory guidance be provided to help prevent potential misinterpretation of leaderboard results and ensure a balanced interpretation of AI and dynamical model performance, notably by non-expert audiences. In particular, the Board recommended adding a clear disclaimer:
    - Stating that the overall performance of a forecasting model cannot be fully characterised by a single verification metric.
    - Highlight the complementary strengths of physical models, such as physical consistency and process-based coherence.
- It was noted that this information could be most appropriately hosted on the Confluence documentation pages linked from the leaderboards.

Item #4: Sustaining team motivation & long-term engagement

- As a potential recognition mechanism, a participant proposal suggested highlighting event-based performance, linking real-world events (e.g. extremes) to model performance. It was noted that such analyses could be showcased during Awards webinars; however, the team clarified that there is currently no capacity to systematically publish event-based evaluations online.
- One Board member suggested exploring the possibility of a journal special collection showcasing top-performing teams and methodological approaches, while another proposed the alternative of a joint publication. It was noted that either option would require a sufficient number of interested teams, willingness from top-performing teams to contribute, and significant effort from contributors. It was agreed that ECMWF will publish a post–Edition 1 manuscript and will give careful consideration to potential co-authorship from top-performing teams.

- The Board emphasised that awards and recognitions should be framed in a way that is authoritative and usable in professional contexts, for example as credible entries on CVs or résumés, explicitly endorsed by ECMWF. ECMWF will thus issue certifications of participation, highlighting the rankings of top-performing teams.
- Board members stressed that awards should be accompanied by clear and visible public recognition, including communication through appropriate ECMWF channels. In response, the organisers noted that options such as an ECMWF press release and a dedicated website article are currently under discussion.
- It was also emphasised that the Quest should articulate, in plain and accessible language, what is being learned through the competition in order to better communicate its value to the general public. Board members underlined that the Quest should fulfil an educational role, beyond the research community. The organisers confirmed that this need has been identified and that the ECMWF communications team has been asked to prepare suitable public-facing material addressing this objective.
- Regarding collaboration pathways, a participant proposal suggested exploring collaboration with top-performing teams towards operational deployment. It was clarified that operational deployment through ECMWF is currently unlikely outside of AIFS-related developments, as present efforts focus on extending AIFS to sub-seasonal timescales and do not include post-processing techniques. Nevertheless, the Board noted that ECMWF should continue reflecting internally on how collaboration with top teams could be explored, even if not directly linked to operational implementation.
- The Advisory Board discussed the importance of maintaining a degree of flexibility for teams that miss forecasts for technical reasons. It was recalled that current practice allows teams to submit missing forecasts by email. Reference was made to similar competitions where one missed submission per period is tolerated. The organisers noted that, while such a rule could not be implemented during the first edition, it represents a possible option to reflect upon for the second edition.
- Finally, it was noted that financial support options are being investigated to enable representatives of top-performing teams to participate in the S2S/S2D event in September 2026, as an additional incentive and engagement opportunity. It was clarified that similar financial support will not be provided for EGU, as the scope of that event extends well beyond AI and sub-seasonal forecasting.

Item #5: Edition 2 concept options

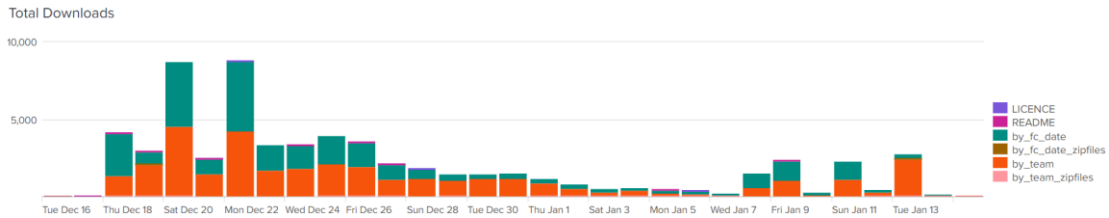
- The Advisory Board generally agreed that retaining the current core framework of the AI Weather Quest represents a natural and coherent continuation into a second edition.
- One Board member noted that, from a longer-term scientific perspective, seasonal and longer-range predictions hold strong scientific and societal relevance. However, it was emphasised that while foundation models are increasingly gravitating towards these timescales, their current ability to deliver robust and competitive forecasts at seasonal horizons remains limited. It would thus be relevant to maintain the Quest's focus on subseasonal timescales for now, while gradually preparing the ground for potential seasonal benchmarks as the underlying technology matures.

- The Board agreed that tropical cyclone-related metrics and indices such as the Madden–Julian Oscillation (MJO) are scientifically interesting additional variables. One Board member noted that the inclusion of the MJO would require additional explanation and contextualisation for participants.
- Board members expressed support for the geographical focus (e.g. regionally targeted evaluation).
- One Board member proposed an additional potential focus on physical consistency, including examples such as:
    - Precipitation conditioned on pressure regimes,
    - Temperature persistence conditioned on large-scale circulation patterns,
    - Penalisation of rapid pressure-pattern decorrelation,
    - Regime classification skill (e.g. NAO, blocking).

  While acknowledged as scientifically relevant, it was noted that such diagnostics would be challenging to implement within the current Quest framework, both methodologically and operationally.

Next steps

- Refine the Awards webinar format by:
    - Introducing targeted thematic focuses (e.g. regional performance or event-based skill) to broaden representation. **JT** - For the DJF webinar, the focus will be on large-scale temperature anomalies in the Northern Hemisphere extra-tropics. A case study covering American, European, and Eurasian temperature anomalies is envisaged. For MAM, a thematic focus on East African rainfall is of interest, with potential links to the ACACIA project.
    - Developing a light, guiding presentation template for presenters, without requiring disclosure of sensitive information. **JT** – Two different guidelines could be set-up:
        - For first time presenters (≈5 minutes presentation, complemented by Q&A), it could feature the team overview, a description of training datasets and computational architecture associated with the best-performing model, a brief overview of alternative model configurations submitted to the Quest, as well as key challenges encountered and planned future developments.
        - For recurring presenters (≈3 minutes presentation, complemented by Q&A), it could feature the team overview and a summary of model developments since the previous competitive season.
- Explore possibility to monitor usage of released forecast datasets, in order to better assess how open-access outputs are being utilised by the community. **JT** – Download tracking has been accessed and indicates very high community interest in the released forecast datasets:



- Prepare a post–Edition 1 scientific manuscript, with careful consideration of potential co-authorship from top-performing teams, and assess, based on interest and feasibility, whether complementary collective publication formats (e.g. joint paper or special collection) could be pursued at a later stage.
  JT – Submission of a high-impact scientific manuscript is planned by the end of 2026. An internal organisation meeting is scheduled in March to discuss the manuscript outline. Top-performing teams will be contacted later in the year regarding potential contributions.
- Develop additional explanatory guidance on leaderboard interpretation, including a disclaimer clarifying that model performance cannot be assessed through a single metric, contextual information on the complementary strengths of physical and AI-based models, and links to relevant verification documentation. **JT** – Draft text prepared for inclusion on the Detailed guide for leaderboards page on Confluence: "*Disclaimer: Caution should be taken when interpreting model performance based on a single evaluation metric. For simplicity and consistency, the AI Weather Quest leaderboards compare sub-seasonal forecasting systems using globally-averaged RPSSs. While the globally-averaged RPSS provides a useful summary of probabilistic forecast skill, it does not capture all aspects of forecast quality, nor does it reflect performance across different regions or use cases. Additionally, ML-based systems may not explicitly enforce physical consistency across variables or conservation laws that are inherent to dynamical models."*
- Strengthen recognition by issuing ECMWF-endorsed certificates of participation and achievement. **OL** – Prepare and send to all SON period participants.
- Increase public visibility mechanisms by continuing discussions on an ECMWF press release and a dedicated website article accompanying major awards, and supporting the preparation of plain-language communication materials explaining what the Quest is learning and why it matters. **OL** – Coordinate with ECMWF Communications.
- Continue internal reflection on collaboration pathways with top-performing teams, recognising that direct operational deployment is currently limited, but that other forms of engagement or collaboration may be explored.
  OL – Follow up during internal progress meetings.
- Consider increased flexibility rules for missed submissions as an input to the design of a potential second edition, drawing on practices from similar competitions.
  OL – Integrate into Edition 2 conceptual reflections.
- Continue investigating financial support options to enable representatives of top-performing teams to participate in the S2S/S2D event in September 2026.
  OL – Follow up on implementation.