DOI: 10.18727/0722-6691/5383

Distributed Peer Review at ESO: Demonstrating Success and Evolving Through Period 115

Tereza Jerabkova[1,2]
Ferdinando Patat[1]
Dario Dorigo[1]
Fabio Sogni[1]
Francesca Primas[1]
Annalisa De Cia[1]
Elisabeth Renate Hoppe[1]

[1] ESO
[2] Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Brno, Czech Republic

ESO's Distributed Peer Review (DPR) has transformed proposal evaluations by fostering efficiency and community involvement, making it an essential tool for handling the large volume of proposals traditional panels cannot review alone. A key strength of DPR is its inclusion of the entire community, engaging researchers at all career levels and identifying expert reviewers. This article summarises updated findings up to Period 115, focusing on expertise assignment, DPR comment usefulness and user satisfaction. DPR's success supports its planned expansion into fast-track channels and yearly cycles. ESO is also exploring further innovations to optimise this process.

Introduction

The Distributed Peer Review (DPR) paradigm has emerged as a promising alternative to traditional expert panel reviews, driven by the rapidly growing number of proposals submitted to large astronomical facilities. By actively involving Principal Investigators (PIs) and Co-Investigators (Co-Is) in reviewing one another's proposals, DPR seeks to distribute the workload more evenly while maintaining the quality of evaluations. After initial deployments at Gemini Observatory, ESO conducted an early pilot project (Patat et al., 2019), and the Atacama Large Millimeter/submillimeter Array (ALMA) introduced DPR for its Cycle 8 (although based on rankings, not grades as for ESO; Carpenter et al., 2022). The first comprehensive assessment of ESO's DPR outcomes was presented by Jerabkova et al. (2023), laying a solid foundation for broader adoption. In this article, we update and expand upon these insights through Period 115.

Background and objectives

The DPR process aligns with ESO's commitment to fairness and efficiency in proposal evaluations, introduced as a necessary tool to manage the growing number of proposals that made it challenging for traditional panels to maintain high-quality reviews. At present, ESO assigns proposals requesting less than 16 hours of observing time — around 50% of all proposals but only 20% of the allocated time — to DPR, while panels still oversee most of the time allocation. Potentially sensitive cases, such as joint programs with ALMA, XMM-Newton and other exceptions, are reviewed by panels irrespective of the time request, as outlined in ESO's DPR guidelines[1]. The choice to maintain both DPR and traditional panels offers ESO valuable flexibility, allowing it to address specific proposal types that may require panel oversight in the future. This hybrid system, with DPR alleviating the workload on panels, ensures that panels can maintain a high standard of review quality.
The DPR system leverages ESO's User Portal, where each user is required to provide two to five keywords representing their scientific expertise. Similarly, proposals include selected keywords from the same pool, which are used to calculate expertise-match scores between reviewers and proposals. These scores are central to the proposal distribution process. Each submitting team, represented by a PI, must nominate one reviewer who is responsible for evaluating DPR-assigned proposals from their peers.

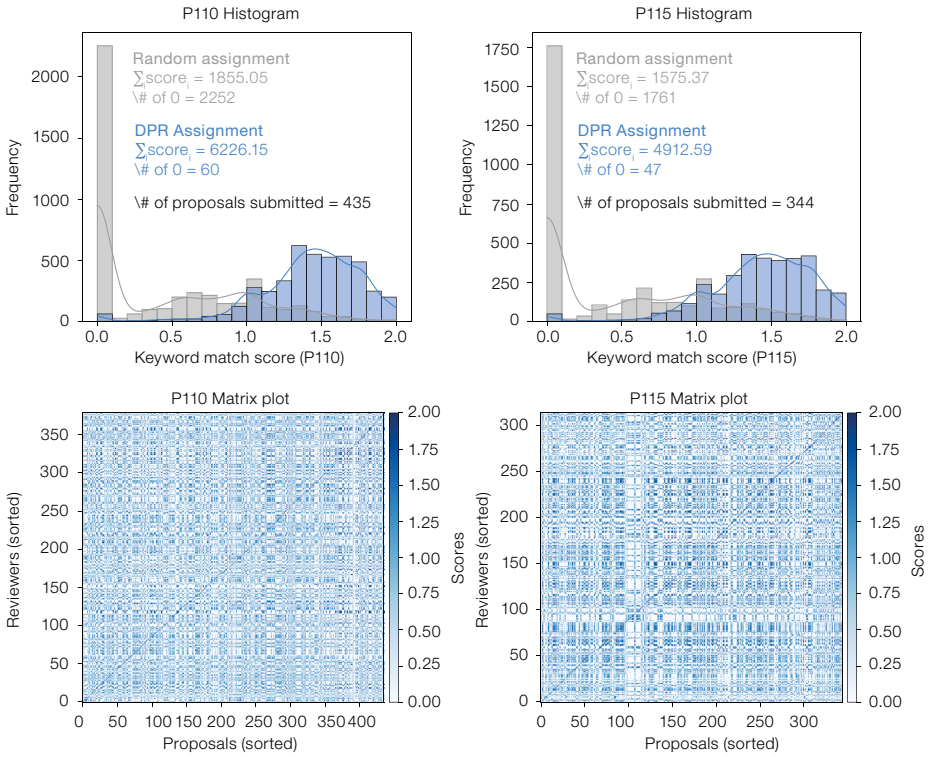

Figure 1. Expert reviewer assignments in the DPR for Period 110 (left) and Period 115 (right). The bottom panels present the *assignment matrix*, in which DPR proposals are on the horizontal axis and DPR reviewers on the vertical axis. Although the matrix is not square, it is sorted so that the main diagonal corresponds to each proposal's delegated reviewer — often the PI or a co-I. The colourmap shows keyword-based expertise scores: a high score along the diagonal verifies that PIs or co-Is indeed have strong expertise for their own proposals, serving as a self-consistency check.

In the top panels, the grey lines indicate how a purely random assignment would appear. Because most keyword-matched scores are zero, the random assignment fails to align expertise with proposals. By contrast, the final assignment (shown in blue) largely avoids zero-score pairings, confirming the necessity of a more sophisticated reviewer assignment process.

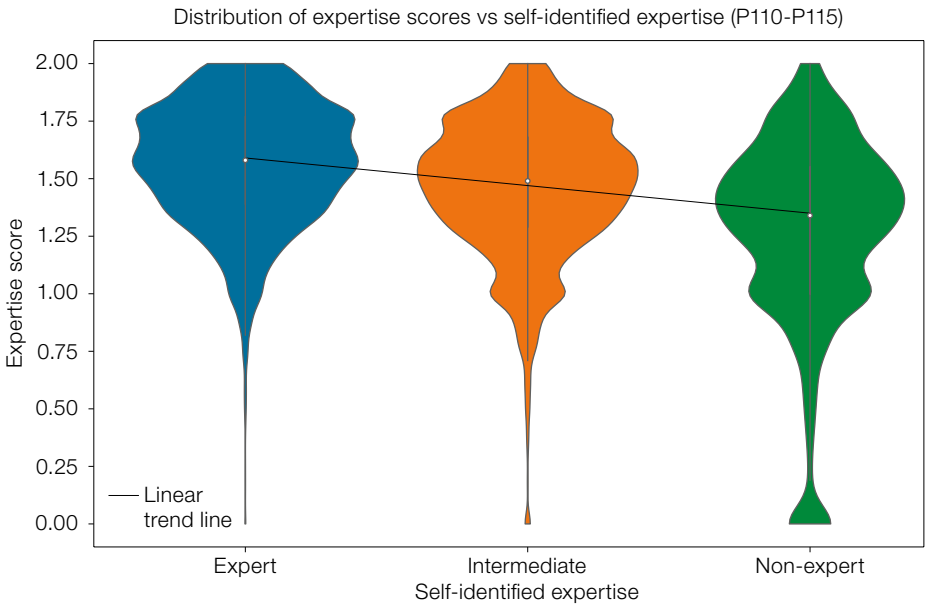

Figure 2. Comparison between the reviewer-proposal match scores — calculated from specified keywords — and reviewers' self-assessed expertise. This Figure demonstrates how closely the keyword-based matching aligns with the reviewers' own perception of their expertise, serving as a validation for the automated assignment approach.

Key findings and updates

Algorithm performance

The assignment algorithm remains a critical component of DPR's success. As illustrated in Figure 1, the final matching of proposals to reviewers outperforms a purely random approach, resulting in an optimal distribution of expertise across the submitted proposals. This robust method ensures the integrity of the review process, even as the number of proposals and reviewers continues to grow. In our previous work (Jerabkova et al., 2023), we focused on Period 110 and partially on Period 111, establishing the first statistical analysis of the DPR's performance. Building on those findings, the current results reinforce how crucial it is to match each proposal with a suitably qualified reviewer.

Advantages and challenges of keyword-based assignments

Keywords continue to play a dual role in facilitating expertise matching. On the positive side, they are intuitive to set and interpret, enabling the community to self-regulate how expertise is represented. However, some challenges persist:

1. Static Nature of Keywords: because keywords do not automatically evolve alongside rapidly advancing scientific fields, mismatches can arise over time.
2. User Inconsistencies: Some users fail to update their keywords regularly, or they assign them improperly, both of which can adversely affect the accuracy of reviewer–proposal matches.

Figure 2 demonstrates the link between keyword-based match scores and reviewers' self-identified expertise. While these results validate the general reliability of using keyword vectors to define reviewer expertise, they also underscore potential pitfalls arising from outdated or misapplied keywords. Ongoing efforts by Amado et al. (in preparation) aim to refine the keyword framework, making it more flexible and adaptive to scientific evolution. In parallel, ESO is communicating closely with ALMA, which has begun adopting a machine learning approach for reviewer assignments (Carpenter, Corvillón & Shah, 2024).

By addressing the limitations of keyword-based assignments and leveraging emerging technologies, DPR can continue to evolve as a sustainable and fair peer-review method that adapts to an ever-increasing volume of astronomical proposals.

User satisfaction

Jerabkova et al. (2023) first examined user satisfaction data from Period 110, finding that feedback under DPR was generally better received than traditional panel comments — especially for rejected proposals, where constructive input is critical. Subsequent user surveys are systematically run each Period. They are built into the DPR evaluation system and receive responses from typically 50% of the PIs. The outcomes indicate that since the implementation of DPR PIs with accepted proposals consistently rate DPR feedback as valuable, whereas rejected proposals attract more mixed responses. Despite these variations, a large fraction of DPR users now report that comments

Table 1. Summary for each period for both DPR and panels.

Period	Proposals	Feedback	Accepted	Rejected	%
DPR P110	435	1358	349	1009	31.22%
Panel P110	429	124	50	74	28.90%
DPR P111	419	2708	1138	1570	64.63%
Panel P111	401	247	121	126	61.60%
DPR P112	451	1911	732	1179	42.37%
Panel P112	442	222	113	109	50.23%
DPR P113	402	2555	1239	1316	63.56%
Panel P113	378	206	112	94	54.50%
DPR P114	403	2413	1003	1410	59.91%
Panel P114	448	245	134	111	54.69%
DPR P115	344	1481	698	783	43.05%
Panel P115	416	123	73	50	29.57%

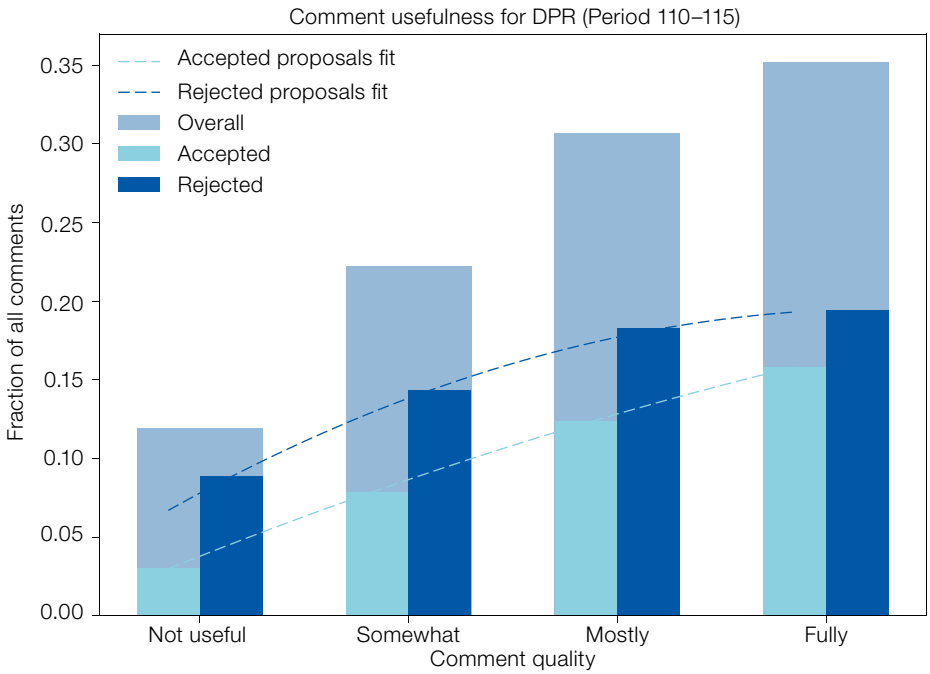

Figure 3. Histogram of user ratings for DPR feedback spanning Periods 110–115. The four categories — *not useful, somewhat useful, mostly useful, fully useful* — reflect PIs' perceptions of review comments they received. The left panel (or first subplot) shows ratings for accepted proposals, and the right panel (or second subplot) shows ratings for rejected proposals. A notable uptick in *fully useful* feedback is seen over time for both categories, suggesting that DPR has steadily improved in delivering constructive and actionable comments.

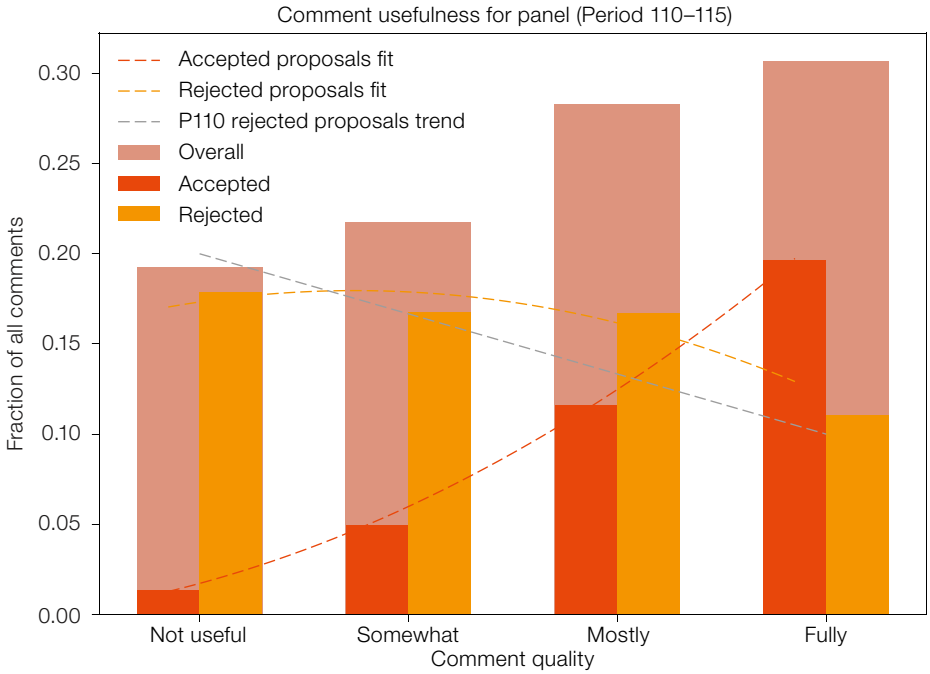

Figure 4. The same as Figure 3, but illustrating user ratings of panel reviews over Periods 110–115. While user satisfaction with panel comments remains lower overall compared to DPR, an upward trend in *mostly* and *fully useful* ratings is observed — particularly for accepted proposals. This improvement may stem from reduced reviewer workloads, thanks to the adoption of DPR, which allows panels to dedicate more time to each proposal. However, panel comments for rejected proposals still show less favourable ratings, indicating the need for continued attention to feedback quality in these cases.

are mostly or fully useful, underscoring the value of this more distributed approach to peer review.

In parallel, panel-based evaluations have shown some improvement over the same periods, likely thanks to the 50% reduction in workload made possible by DPR. Historically, users have viewed panel feedback — especially for rejected proposals — as less beneficial, a sentiment strongly reflected in the Period 110 survey results. However, current data (spanning Periods 110 to 115), as shown in Figures 3 and 4, reveal a modest but encouraging upswing in the proportion of panel comments deemed mostly or fully useful, at least for accepted proposals. While this positive trend has not yet matched DPR levels, it offers promise that, with continued effort and streamlined workloads, panel comments can further improve — particularly where it matters most, i.e. for proposals that ultimately receive a rejection. Table 1 shows a number summary for each period for both DPR and panels, showing that the analysis is based on robust numbers.

Career stage and reviewer bias

Analyses of DPR grading trends reveal no statistically significant bias in the scores assigned by reviewers at different career stages. Although there is a slight tendency for senior scientists to award poorer grades, students continue to be a pivotal part of the process, often providing some of the most constructive and well-received comments. These observations are consistent with earlier results (Jerabkova et al., 2023), reinforcing the conclusion that career-stage differences are not a major driver of bias in the DPR framework.

An additional point worth highlighting is the inclusivity of the DPR model, which allows students and junior researchers to serve as reviewers — an option not typically available under traditional panels. Interestingly, the feedback they supply is frequently rated as most useful by proposal authors (see Figure 5), validating the merit of incorporating perspectives from earlier-career scientists. When students or junior researchers feel uncertain about their evaluations, they may request permission to consult their supervisors, who must abide by the same confidentiality

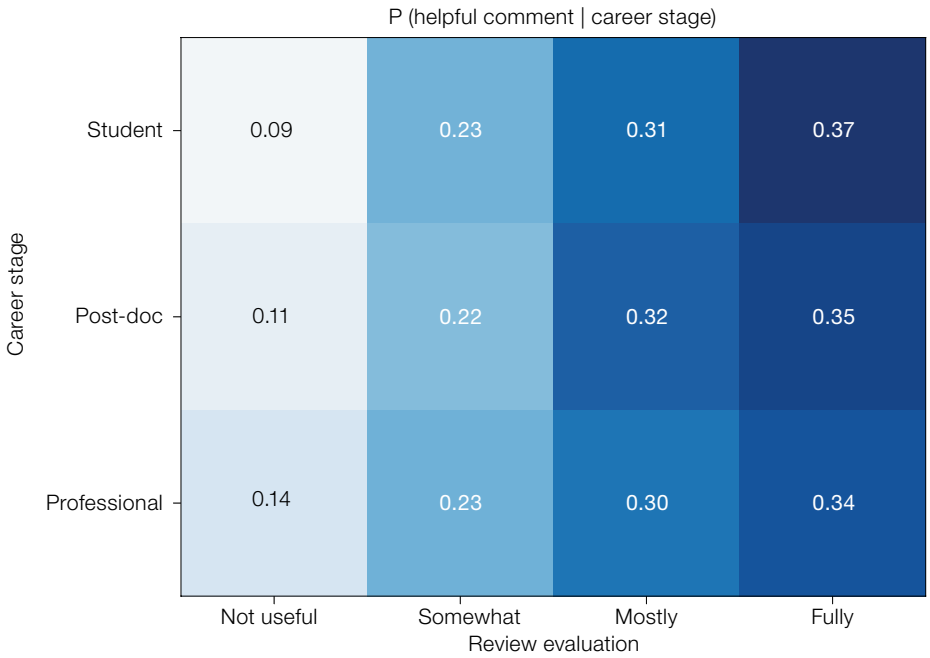

Figure 5. Color-coded conditional probability map showing how perceived comment usefulness varies with reviewer career stage in the DPR. Each bin displays the probability that a reviewer at a given career stage receives a specific usefulness rating from the PIs, with values normalized across the usefulness axis (i.e. each row sums to one). The data confirm that comments from students are consistently rated as highly useful, while senior researchers — though often more stringent in grading — also provide valuable, high-quality feedback. These results align with findings from Period 110, reinforcing the important contribution of early-career scientists to the peer-review process.

agreements. This process has been invoked in several instances each semester, helping to ensure that reviewers at all career stages can participate confidently and responsibly.

Reviewer demographics and diversity

One notable aspect of DPR is that the reviewer pool reflects the broader community, with a slight overrepresentation of postdoctoral researchers. Unlike traditional panels, ESO does not control the composition of DPR reviewers in terms of gender, seniority or nationality. In contrast, panel composition is carefully curated to ensure diversity across these dimensions (Primas et al., 2024; Primas et al., in preparation). While this organic representation of the community is valuable, it may be interesting in the future to explore algorithms that ensure diversity among the reviewers assigned to each proposal. Achieving this would require a significantly larger pool of reviewers to balance these considerations effectively.

Implications and future directions

The successful integration of DPR into ESO's evaluation process demonstrates its viability as a scalable alternative to traditional panel reviews. The reduction in panel workload and the high satisfaction rates among users reinforce the value of DPR. Nonetheless, there is room for improvement:

– Adaptive Algorithms: ESO is exploring more sophisticated assignment algorithms, including those leveraging machine learning, to further enhance the match between proposals and reviewers.
– Keyword Refinement: Efforts are underway to make the keyword system more agile and reflective of contemporary science.
– Increased Feedback Participation: Strategies to improve the response rate for feedback surveys are being developed to expand the statistical basis for future analyses.

ESO plans to use DPR within its yearly proposal cycle, likely covering an even larger fraction of submissions to keep the expert panel load at the current level. While ALMA has implemented the DPR model for 100% of their submitted proposals, ESO plans to retain its panels for specific cases, such as joint ALMA–ESO proposals and those with significant time requests, ensuring a balanced approach that combines DPR's efficiency with the expertise of traditional committees[a]. Studies by Carpenter et al. (2022) on ALMA's implementation of DPR provide valuable insights that support these developments.

Expansion to fast-track channel

ESO is confident in deploying DPR for its upcoming fast-track channel, as highlighted by Patat et al. (2024). This channel is expected to streamline the evaluation on short timescales while maintaining the high standards of review quality established by DPR.

Conclusion

With DPR now operating across six periods, its benefits for the ESO community are clear. The process has reduced panel reviewer workloads, maintained fairness in proposal evaluations, and provided helpful feedback to PIs. The ongoing refinement of algorithms and reviewer profiling ensures that DPR will continue to serve as a model for peer review innovation in astronomy and beyond.

References

Carpenter, J. M. et al. 2022, PASP, 134, 045001
Carpenter, J. M., Corvillón, A. & Shah, N. B. 2024, arXiv:2410.10009
Jerabkova, T. et al. 2023, The Messenger, 190, 63
Patat, F. et al. 2019, The Messenger, 177, 3
Patat, F. et al. 2024, The Messenger, 193, 45
Primas, F. et al. 2024, Proc. SPIE, 13098, 130980H

Links

[1] ESO's DPR guidelines: https://www.eso.org/sci/observing/phase1/distributed-peer-review.html

Notes

[a] In the current implementation the proposals are split 50/50 in number, while the time share is 80/20 (Panels/DPR).